UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Academic Unit of Clinical and Experimental Sciences

Characterisation of the Molecular Mechanisms Underpinning Idiopathic Pulmonary Fibrosis (IPF) Using Quantitative Proteomics

by

Leanne Amanda Wickens

Thesis for the degree of Doctor of Philosophy

September 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

Clinical and Experimental Sciences

Doctor of Philosophy

CHARACTERISATION OF THE MOLECULAR MECHANISMS UNDERPINNING IDIOPATHIC PULMONARY FIBROSIS (IPF) USING QUANTITATIVE PROTEOMICS

By Leanne Amanda Wickens

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial lung disease. Pathogenesis involves the aberrant accumulation of myofibroblasts and excessive deposition of matrix components. Its aetiology is unknown and treatment options are limited. The histone deacetylase inhibitor FK228 has been shown to have anti-proliferative effects on fibroblasts, and therefore may be a candidate for IPF therapy.

This thesis describes the in-depth proteomic characterisation of fibroblasts from patients with IPF, and their response to FK228. Comparison of protein expression between IPF and non-IPF fibroblasts revealed over 120 aberrantly expressed proteins in this disease, including those involved in transcriptional regulation, which have the potential for use as diagnostic or monitoring biomarkers. Extensive validation is now required for their translation to the clinic. IPF biomarkers are greatly sought after as there are currently none in clinical use. FK228 treatment caused a decrease in IPF fibroblast metabolic activity and proliferation through cell cycle arrest, and significantly affected the expression of hundreds of proteins. The proteomic responses of IPF fibroblasts grown in 2D and 3D culture were shown to differ substantially; there was little difference in the number of proteins quantified between culture systems, but treatment affected the expression of different proteins. This highlights the importance of culture system selection for studying IPF and for testing therapeutics effectively *in vitro*.

The method for profiling cultured fibroblasts using proteomics was firstly developed in the MRC-5 cell line, to optimise cell lysis, protein isolation and digestion, and mass spectrometry analysis for global protein identification. Intracellular and extracellular proteomic analysis of these cells following optimisation, with and without transforming growth factor beta-1 (TGF-β_1) stimulation is also discussed in this thesis. This dataset provides to the author's knowledge the most comprehensive characterisation of the fibroblastic response to this profibrotic cytokine to date, with more than 5000 proteins quantified. Data analyses highlighted biological processes that may be dysregulated in fibrotic disease, including cell cycle regulation and cellular organisation.

Finally, the first steps in the development of a method using immunoprecipitation enrichment to study the IPF fibroblast acetylome in response to FK228 are described, for the identification of off-target effects. Fewer than 100 acetylated peptides were identified by mass spectrometry analysis, a considerably lower number than found in previously published reports, highlighting the inefficiency of the technique for global enrichment of acetylated peptides in this study.

Contents

List of Figures

List of Tables

Academic Thesis: Declaration of Authorship

I, Leanne Amanda Wickens, 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.

Characterisation of the Molecular Mechanisms Underpinning Idiopathic Pulmonary Fibrosis (IPF) Using Quantitative Proteomics

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. None of this work has been published before submission.

Signed: .....

Date:

Contributors

All experiments and data analyses were performed by the author in either the Brooke Laboratory, Clinical and Experimental Sciences, Faculty of Medicine, Southampton University Hospital, University of Southampton, or in the Centre of Proteomic Research, Institute for Life Sciences, University of Southampton with the following exceptions:

Chapter 3

Heat maps and protein interaction networks were generated by Akul Singhania, Clinical and Experimental Sciences, Faculty of Medicine, Southampton University Hospital, University of Southampton, assisted by the author in terms of what analyses were required.

Chapters 4 and 5

Lung biopsies were obtained from non-IPF and IPF patient donors by clinicians at Southampton General hospital and the Royal Brompton hospital. Primary fibroblasts were isolated from biopsies by members of the Brooke Laboratory including Dr Franco Conforti and Dr Mark Jones.

Heat map analysis was performed by the author, using a template R script originally created by Dr Alistair Bailey.

Chapter 4

FACS analysis was carried out by the author, under the supervision of Dr Franco Conforti.

Chapter 6

All experiments using the Choudhary method for analysis of acetylated peptides were performed by both the author and Dr Caterina Folisi, Dipartimento di Biomedicina Clinica e Molecolare Sezione Malattie Respiratorie Università di Catania, Italia. Processing and analysis of data was performed by the author.

Acknowledgements

Over the course of my PhD studies, there have been many people that have helped and encouraged me, and without their support the completion of this thesis would not have been possible. I cannot thank all of these people here, but there are several people I would like to give special thanks to.

Firstly, I would like to thank both the NIHR Southampton Respiratory Biomedical Research Unit (RBRU) and the Centre for Proteomic Research at the University of Southampton for funding my work, as without their support this research would not have been possible. I am especially grateful to the RBRU for extending my funding for an extra three months as I approached the end of my laboratory work, and to both funders for enabling me to travel to America to present my research, which was an amazing experience.

I would like to thank all current and former members of the Centre for Proteomic Research that have helped me over the course of my PhD; I am very grateful to Dr Erika Parkinson for her guidance with laboratory work and data analysis, and for the pep talks and coffee; my thanks also go to Dr Dominic Burg, for always being happy to answer my hundreds of questions regarding proteomics in the first couple of years, and to Dr Harry Whitwell for his support. Thanks also to Dr Alistair Bailey, particularly for teaching me the basics of R programming, and to Dr Caterina Folisi for her help in the laboratory with the acetylation study and for keeping me entertained with her stories.

Within the University of Southampton, Faculty of Medicine, I would like to thank all members of the Brooke Laboratory for their help and assistance during my studies. Special thanks go to: Dr Franco Conforti, for his guidance with laboratory techniques such as primary cell culture and FACS analysis, as well as for the many interesting discussions we shared concerning IPF pathogenesis; Dr Mark Jones and Dr Francesco Cinetto, for their help and assistance with regards to the IPF 3D model study; and to Dr David Smart, for teaching me the basics at the very beginning, and for his knowledge and help throughout my studies. I would also like to thank Akul Singhania and Topher Woelk for their help with the analysis of my first big proteomic dataset.

I wish to give a big thank you to my supervisors for all their advice and guidance over the last four years. I am incredibly grateful to Dr Paul Skipp, for his help day to day and for his support throughout the process; due to his encouragement, I always left his office after meetings feeling positive, even when times felt particularly difficult. I am also very thankful to Professor Donna Davies for sharing her advice, knowledge and support, especially when I

was new to my PhD studies. Finally, I would also like to thank Dr Katharine O'Reilly for her advice on my work and for giving me insights into IPF from a clinical perspective. The guidance of all my supervisors enabled me to develop greatly as a researcher during the course of my studies.

Many people in academia have supported me throughout my PhD experience in a non-academic capacity. I am especially grateful to Lyndsy Ambler, Richard Latouche, Emily Bowler, Natalie Day, Alison Hill, Kate Packwood and Oliver Hutton; their friendship and support as we all went through this experience together made it much easier than it would have been otherwise. Lyndsy deserves additional thanks for babysitting my cells during the times I was working at Highfield campus and for listening to all of my PhD woes.

Finally, I would like to thank all of my friends and family at home for being there for me throughout my PhD, and for taking an interest regardless of whether they understood what I was talking about. Special thanks to Eleanor Hopkinson for regularly checking in on me, and to the rest of the Yeovil group, and the Exeter group too. Thank you to Julie Bullock for always being happy to listen. Thanks also to my first year housemates for making my move to Southampton the easiest and most fun it could be.

Most of all, I would like to thank Ben Bullock, my mum and my dad for their love and support throughout my academic career so far, for constantly encouraging me and for picking me up whenever I needed it. I could definitely not have done this without you.

Abbreviations

AEC	Alveolar epithelial cell
AMRT	Accurate mass retention time
AQ	Absolute quantification
AT1	Alveolar type 1
AT2	Alveolar type 2
BAL	Bronchoalveolar lavage
BALF	Bronchoalveolar lavage fluid
BSA	Bovine serum albumin
CAPS	3-[cyclohexylamino]-1 propane sulfonic acid
cDNA	Complementary DNA
CID	Collision-induced dissociation
CPFE	Combined pulmonary fibrosis and emphysema
Ct	Cycling threshold
CTCL	Cutaneous T-cell lymphoma
CTGF	Connective tissue growth factor
DDA	Data-dependent acquisition
DEP	Differentially expressed proteins
DIA	Data-independent acquisition
DM	Diabetes mellitus
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMT	Epithelial-mesenchymal transition
ESI	Electrospray ionisation
FACS	Fluorescence-activated cell sorting

FBLIM-1	Filamin binding LIM protein 1
FBS	Foetal bovine serum
FDA	Food and Drug Administration
FDR	False discovery rate
FGF	Fibroblast growth factor
FIP	Familial interstitial pneumonia
fNSIP	Fibrotic non-specific interstitial pneumonia
FVC	Forced vital capacity
GER	Gastroesophageal reflux
GO	Gene ontology
HAT	Histone acetyltransferase
HBSS	Hanks' Balanced Salt solution
HDAC	Histone deacetylase
HDACI	Histone deacetylase inhibitor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLA	Human leukocyte antigen
HRCT	High resolution computed tomography
IAA	Iodoacetamide
IAP	Immunoprecipitation buffer
IEF	Isoelectric focusing
IIP	Idiopathic interstitial pneumonia
IMS	Ion mobility separation
IP	Immunoprecipitation
IPF	Idiopathic pulmonary fibrosis
IPG	Immobilized pH gradient
iTRAQ	Isobaric tags for relative and absolute quantification
LAP	Latency-associated peptide
LC	Liquid chromatography
LTBP-1	Latent TGF-β-binding protein
MALDI	Matrix-associated laser desorption/ionisation

MAPK	Mitogen-activated protein kinase
MHC	Major histocompatibility complex
MMP	Matrix metalloproteinase
MRM	Multiple reaction monitoring
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS^E	Elevated energy mode
MTS	(3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium)
NAC	N-acetylcysteine
NADPH	Nicotinamide adenine dinucleotide phosphate
NQO1	NAD(P)H dehydrogenase [quinone] 1
NSIP	Non-specific interstitial pneumonia
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PEG	Polyethylene glycol
PI	Propidium iodide
PI3K	Phosphoinositide 3-kinase
PLGS	Protein Lynx Global Server
PLOD1/2	Procollagen-lysine 1, 2-oxoglutarate 5-dioxygenase 1/2
PMF	Peptide mass fingerprinting
PMS	Phenazine methosulfate
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
qPCR	Quantitative polymerase chain reaction
QQQ	Triple quadrupole
QTOF	Quadrupole-time-of-flight
RNA	Ribonucleic acid

ROS	Reactive oxygen species
RT	Room temperature
RT(-qPCR)	Reverse transcription
SAHA	Suberoylanilide hydroxamic acid
SDS	Sodium dodecyl sulfate
SILAC	Stable isotope labelling with amino acids in cell culture
SiRNA	Small interfering RNA
SRM	Single reaction monitoring
TBS	Tris-buffered saline
TEAB	Triethyl ammonium bicarbonate
TFA	Trifluoroacetic acid
TGF-β_1	Transforming growth factor beta-1
TMT	Tandem mass tag
TNF-α	Tumour necrosis factor alpha
TOF	Time-of-flight
TSA	Trichostatin A
UIP	Usual interstitial pneumonia
UPLC	Ultra-performance liquid chromatography
UPR	Unfolded protein response
VEGF	Vascular endothelial growth factor
α-SMA	Alpha-smooth muscle actin

1. Introduction

1.1. Structure and function of the lung

The respiratory system in humans begins at the nose and mouth and leads down into a pair of lungs, which are situated in the thorax. The lungs are organs specialized for gas exchange, during which inhaled oxygen passes into the blood in exchange for waste carbon dioxide. The lungs themselves are made up of a mass of tubes, starting with the trachea (or windpipe) that leads from the throat and splits into two tubes, termed the bronchi. Each bronchus extends into the left or right lung. The nasal passage, trachea and bronchi together make up the upper airways. The upper airways are lined by a layer of mucus, which is secreted by goblet cells interspersed in the epithelial cell lining and protects the lower airways by trapping potentially harmful inhaled particles such as bacteria. The cuboidal epithelial cells that predominantly line the upper airways have cilia, which are hair-like projections that beat in a synchronised motion for mucociliary clearance out of the respiratory system (Knight and Holgate, 2003).

The lower airways are comprised of bronchioles, networks of tubes that extend from the bronchi and lead down into the lowest section of the lungs, the alveoli. These microscopic air sacs are found at bronchiole termini and are surrounded by capillaries from the vascular system for efficient diffusion of gases. Two types of epithelial cells line the alveoli, alveolar type 1 (AT1) and alveolar type 2 (AT2) cells. AT1 cells are large, extremely thin squamous cells which cover approximately 95% of the alveolus (Mason, 2006) and are specialised for efficient gas exchange to and from the capillaries. AT2 cells are cuboidal and rather than having a role in gas exchange, they secrete surfactant to provide low surface tension and prevent alveolar collapse, and to protect against inhaled particles (Mason and Williams, 1977) and upon injury these cells can proliferate and transdifferentiate into AT1 cells to replace those that have been damaged (Mason, 2006). Underlying these alveolar epithelial cells (AECs) is a basement membrane, a thin layer of specialised extracellular matrix (ECM) which supports these cells and also acts as a barrier between the epithelium and the mesenchymal compartments (Paulsson, 1992).

The mesenchymal connective tissue comprises many kinds of stromal cells, including fibroblasts and smooth muscle cells. Fibroblasts are structural cells that support tissues by producing an ECM scaffold. Due to this function they play a large role in the wound healing response. AECs that become injured secrete profibrotic molecules that drive fibroblast

differentiation into myofibroblasts, which take on an "activated" phenotype (Thannickal et al., 2003) and migrate to sites of injury. They then secrete ECM components such as fibrillar collagen, elastin and fibronectin into the alveolar interstitium (Fernandez and Eickelberg, 2012) to support re-epithelialisation of the wound. Once the wound is healed the myofibroblasts undergo apoptosis and homeostasis is restored (Wynn and Ramalingam, 2012).

The wound healing response is a well organised system, however, it can become dysregulated in disease and can result in excessive scarring and organ damage. Dysregulated interactions between injured AECs and fibroblasts is the hypothesized pathogenic mechanism in the interstitial lung disease idiopathic pulmonary fibrosis, which is the research area of this thesis, and is discussed below.

1.2. Background to idiopathic pulmonary fibrosis

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive respiratory disease with no known cause, involving the formation of abnormal scar tissue in the lungs. It is the most common, and most deadly, of all idiopathic interstitial pneumonias (IIPs) (van den Blink et al., 2010; Raghu et al., 2011; Kim et al., 2006), a group of idiopathic heterogeneous inflammatory/fibrotic diseases affecting the lung parenchyma (Korfei et al., 2013). The prognosis can be worse for IPF than for many types of cancer (Maher et al., 2007). At diagnosis, patients are usually 50-70 years of age, primarily male, with mortality commonly occurring within 3-5 years (Raghu et al., 2011; King et al., 2011; Fernandez and Eickelberg, 2012).

Incidence of IPF in the population is at a rate of between 4.6 and 16.3 per 100,000 people, and evidence suggests that the number of cases have been increasing over recent years (Raghu et al., 2011; Gribbin et al., 2006). In 2006, Gribbin *et al*. estimated the annual increase in incidences of IPF in the United Kingdom to be 11%, with incidence doubling between 1990 and 2003, and suggested that IPF is diagnosed in more than 4000 patients each year. This rate increase was not thought to be attributed to an aging UK population, or an increased ascertainment of mild cases of IPF (Gribbin et al., 2006). In 2011, Navaratnam *et al*. investigated UK death certificate data over a 40-year period and found that the number of recorded deaths from IPF (bearing in mind that some may be misdiagnoses due to the difficulty in distinguishing the different IIPs) had risen dramatically, from just under 500 patients to over 3000, suggesting a yearly mortality rate increase of 5% (**Figure 1.1**). Based

on this and data from primary care records between 2000 and 2008, they estimated that each year in the UK 5000 people will be diagnosed with IPF and 5000 people will die from this disease, more than from ovarian cancer, lymphoma or leukaemia (Navaratnam et al., 2011). A more recent study by Hutchinson *et al*. (2015) assessed all available studies of incidence of IPF on the Medline and Embase databases. They identified 34 studies in 21 countries between 1968 and 2012, and the majority of them showed an increase in incidence over time, indicting a worldwide increase. This increase in incidence has prompted IPF to become a major research area over the last few decades (Richeldi, 2015).

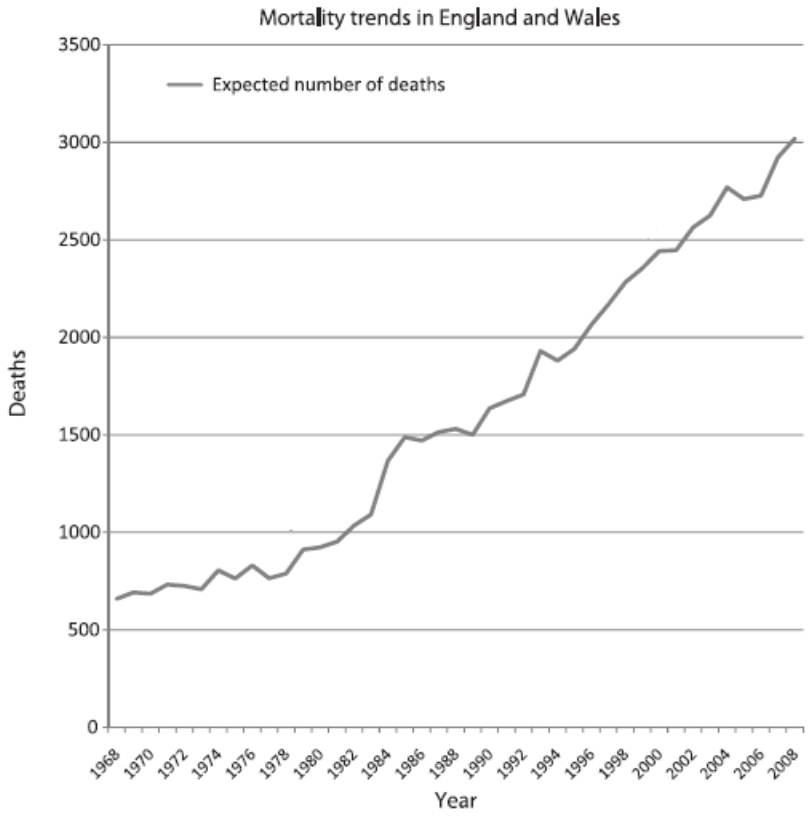


Figure 1.1 The estimated number of deaths from IPF, standardised to the England and Wales population of 2008. Adapted by permission from BMJ Publishing Group Limited from Navaratnam *et al.* (2011).

Progression of the condition varies between patients, with the majority experiencing a gradual decline in health while others experience periods of acute respiratory deterioration, known as acute exacerbations; the reason for this is unknown (Raghu et al., 2011). Ten to fifteen percent of patients experience a rapid decline overall, with a matter of months between their first symptoms and death (Maher, 2013). Several studies have been carried out

to try to predict disease course; an example of this is a retrospective study by Schmidt *et al*., who investigated whether it could be predicted by a patient's earlier trend in pulmonary function tests (Schmidt et al., 2014). They concluded that early mortality could be predicted using this data, as patients with a stable forced vital capacity (FVC) after 1 year were more likely to survive a second year than those who had experienced a decline in FVC. However, they could not predict the manner of future disease progression.

An acute exacerbation describes the acute worsening of a disease condition due to an unknown cause. IPF is the fibrotic lung disease in which acute exacerbations are most prevalent (Antoniou and Wells, 2013). It can only be confirmed when other potential causes, such as infection, have been excluded (Molyneaux and Maher, 2013), for example acute respiratory decline following viral infection would be termed "viral exacerbation" rather than "acute exacerbation". Despite this, patients suffering acute exacerbations are often treated with antimicrobials or steroids, although no benefit of steroids has been proven and their use is linked to a high mortality rate (Papiris et al., 2014).

The incidences of acute exacerbations are highly variable between patients and within a patient's disease course. In a retrospective study of acute exacerbation featuring 461 IPF patients, 35% of patients required hospitalisation during a mean of 24 months' follow-up period due to rapid respiratory decline, and acute exacerbation was the most common cause. 25% of these patients had multiple episodes of exacerbation, and 50% of those admitted to hospital died. This study indicates that acute exacerbations can be a predictor of poor survival (Song et al., 2011).

1.3. Diagnosis of IPF

Typical symptoms displayed by IPF patients can include cough, unexplained chronic exertional dyspnoea (shortness of breath), a crackling sound during breathing clinically defined as "bibasilar fine inspiratory crackles" (Raghu et al., 2011; van den Blink et al., 2010), finger clubbing, or comorbid conditions such as emphysema (Raghu et al., 2011). Diagnosis can be difficult as these symptoms are not confined to this disease alone; histologically IPF is similar in presentation to non-specific interstitial pneumonia (NSIP), another IIP with a better prognosis (Maher et al., 2007; Latsi et al., 2003). To be classed as truly idiopathic, known causes of other interstitial lung diseases must be excluded for diagnosis (Soc, 2000).

IPF is distinguished by a histopathological pattern referred to as Usual Interstitial Pneumonia (UIP) (King et al., 2011) defined as evidence of structural aberration and heterogeneous fibrosis in the lung parenchyma, with possible areas of honeycombing; and "fibroblastic foci", areas of proliferating fibroblasts and myofibroblasts, may also be observed (**Figure 1.2**). Fibrosis typically progresses from sub-pleural areas. A UIP pattern can be diagnosed using High-Resolution Computed Tomography (HRCT), or by surgical lung biopsy. It is possible to observe multiple HRCT patterns and/or pathologic patterns within the same patient. For example, more than one pattern may be observed in one biopsy, or if a biopsy is taken from more than one area, different patterns may be seen in these areas (Travis et al., 2013), which can increase the difficulty of diagnosis. Diagnosis typically requires a multidisciplinary approach, with clinician, radiologist and pathologist communication (Raghu et al., 2011). This can help to determine whether a biopsy or HRCT scan would be more informative, and thus whether biopsy is necessary (Travis et al., 2013). Biopsy becomes increasingly dangerous as gas transfer levels decline towards 30% of normal, so may not be considered for patients with later stage IPF (Latsi et al., 2003). Thoracic surgical procedures, as well as infection and reflux are also thought to be possible trigger factors for acute exacerbations (Antoniou and Cottin, 2012).

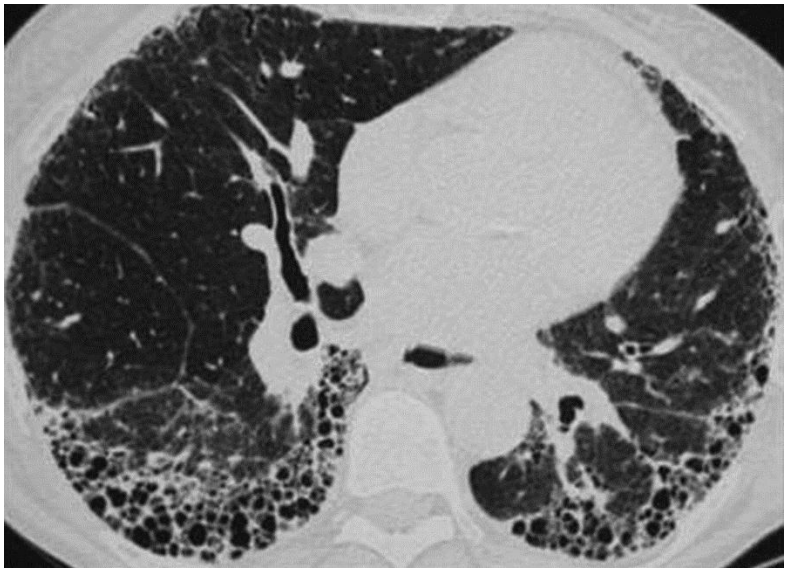


Figure 1.2 A typical usual interstitial pneumonia (UIP) pattern by high-resolution computed tomography (HRCT). Fibrosis is heterogeneous and sub pleural honeycombing is present. Image republished from Valeyre (2011).

In addition to the difficulty of diagnosis, unpredictable disease progression and increasing incidence of this disease, the aetiology of idiopathic pulmonary fibrosis is unknown and the pathogenic mechanisms are relatively poorly understood. Past and current hypotheses concerning the pathogenesis of IPF are discussed below.

1.4. Pathogenesis of IPF

There are multiple theories concerning the disease mechanism of IPF. Leading hypotheses involve inflammatory pathways or epithelial injury/aberrant wound response pathways (Strieter and Mehrad, 2009). Despite the fact that anti-inflammatory drugs such as corticosteroids appear to be ineffective, and that levels of inflammation in the lung do not correlate well to the severity of the disease (Strieter and Mehrad, 2009; Selman et al., 2001), chronic inflammation is still believed by some to be the mechanism of IPF pathogenesis (Todd et al., 2012). One of the more popular hypotheses is that inflammation is a side effect of IPF, and that the disease occurs through the cellular pathway, beginning with alveolar epithelial cell micro-injury, causing dysregulation of homeostasis (Fernandez and Eickelberg, 2012) and aberrant wound healing in aged and genetically susceptible individuals; this is discussed below.

1.4.1. Dysregulated alveolar epithelial-mesenchymal interactions in fibrosis

It is thought that repetitive injury by an unknown stimulus to alveolar epithelial cells causes their exaggerated apoptosis, and their abnormal proliferation and activation in an effort to repair the lung (King et al., 2011). Prostaglandin E2 is an antifibrotic lipid mediator produced by COX2 that has been shown to have a protective role for alveolar epithelial cells against FasL-induced apoptosis. It is found in lower levels in IPF lungs due to defective expression of COX2, making these cells more sensitive to apoptosis; this may be a reason why AEC apoptosis is wide-spread in IPF compared to normal wound healing (Maher et al., 2010). Aberrantly activated AECs release profibrotic mediators (Sakai and Tager, 2013) also produced by immune cells that are subsequently recruited to the wounded area. These profibrotic molecules stimulate local fibroblast differentiation into contractile myofibroblasts. By releasing these mediators in response to injury, injured epithelial cells modify the way that fibroblasts behave in a paracrine manner (Sakai and Tager, 2013). Normally, myofibroblasts create an ECM scaffold for repair and remodelling, contract to close the wound and undergo apoptosis as the tissue is repaired, but in IPF the profibrotic

signal persists and they accumulate in the interstitium, continuing to produce and secrete ECM components excessively (Hinz et al., 2007; Zhang and Phan, 1999). It has been shown that inhibition of COX2 increases the resistance of fibroblasts to FasL-mediated apoptosis and that exogenous prostaglandin E2 restores sensitivity, in opposition to its effects on AECs (Maher et al., 2010). Accumulating activated fibroblasts cause further injury to the alveolar epithelial cells (Sakai and Tager, 2013) suggestive of a feedback loop, resulting in less healthy tissue and less efficient gas exchange in the lungs (**Figure 1.3**) (Davies et al., 2012). Areas of myofibroblast accumulation, termed fibroblastic foci, are typically found near to areas of dense scarring (Davies et al., 2012). It has been indicated that an increased presence of fibroblastic foci is associated with poorer survival (King et al., 2011).

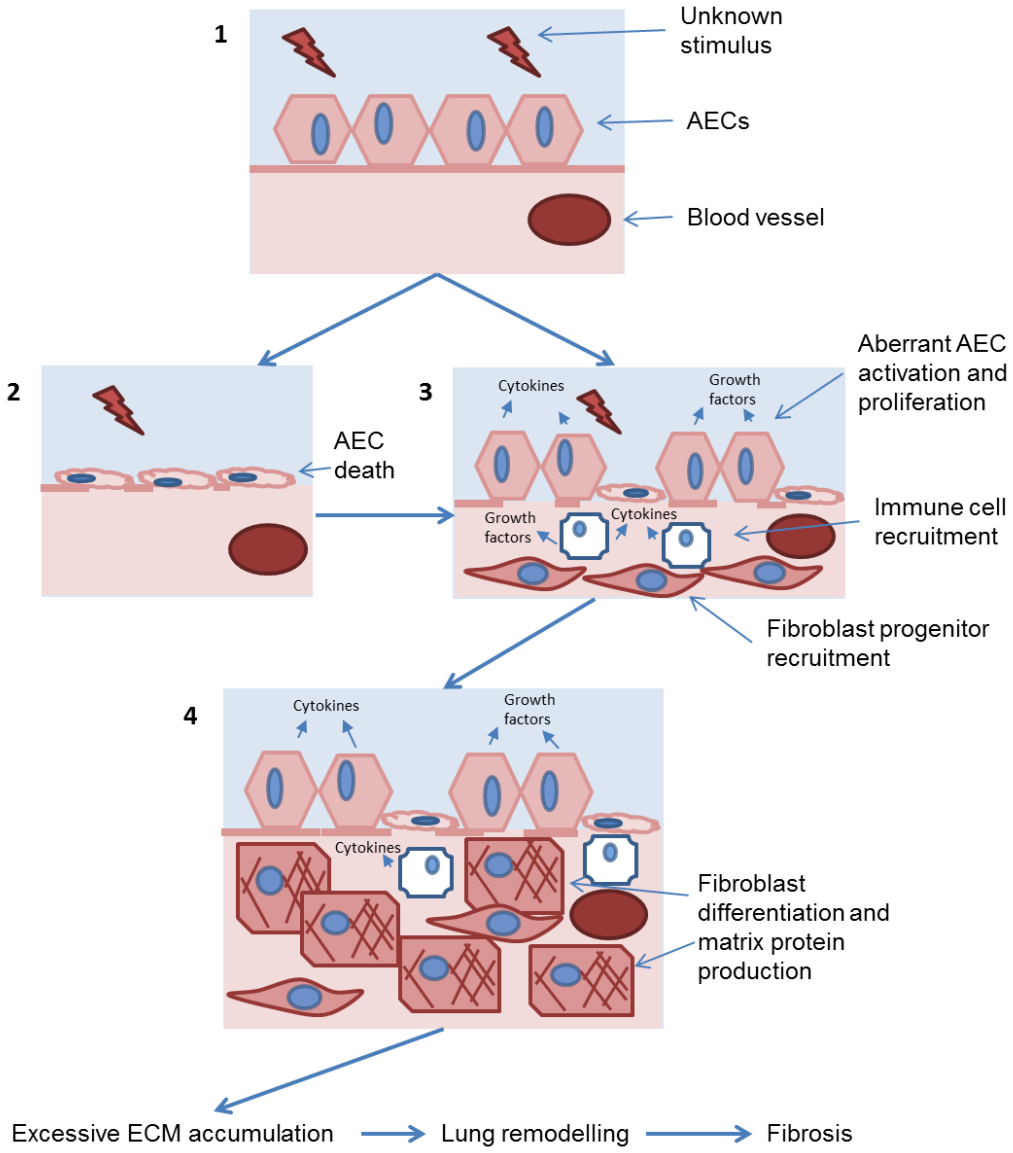


Figure 1.3 The pathogenesis of idiopathic pulmonary fibrosis. Repetitive micro-injuries by unknown stimuli to susceptible lung (**1**) causes alveolar epithelial cell death (**2**) and subsequent aberrant proliferation and activation of AECs to try to repair the wound (**3**). Activation of AECs leads to their release of cytokines and growth factors such as transforming growth factor beta, also produced by recruited immune cells. These signalling molecules stimulate the differentiation of fibroblast progenitors from a number of origins into extracellular matrix-secreting myofibroblasts (**4**), which accumulate to form fibroblastic foci and deposit ECM proteins into the interstitium, resulting in stiffening of the lung tissue and impaired gas exchange.

The extracellular matrix itself is composed of fibrous proteins including collagen, elastin and fibronectin, within a gel-like structure of proteoglycans, hyaluronan and other glycoproteins.

They all interact to provide a structural network for fibroblast attachment, and in part they regulate the tissues by interacting with the cells themselves (Wight and Potter-Perigo, 2011; Malmstrom et al., 2004a). Cells can sense changes in ECM composition and organisation, allowing it to influence factors such as cell adhesion and proliferation, and at the same time the cells remodel the ECM by producing increased or reduced amounts of ECM components, regulating its composition (Malmstrom et al., 2004a). If the tissue becomes injured the ECM structure is destroyed, and the cells must remodel it so that fibroblasts and immune cells can enter and repair the wound and restore homeostasis (Wight and Potter-Perigo, 2011; Malmstrom et al., 2004a). Matrix metalloproteinases (MMPs) have been shown to be actively involved in ECM remodelling, and several including MMP-1 (collagenase-1) and MMP-7, expressed primarily by alveolar epithelial cells, and MMP-2, expressed in both epithelial cells and fibroblasts, have been shown to have strongly upregulated expression in IPF lungs (Zuo et al., 2002; King et al., 2011). Stimulated by profibrotic mediators, fibroblasts synthesize ECM components so that the ECM becomes stabilised and more cells can proliferate, differentiate and contribute to repair (or fibrosis if the stimulant persists) (Wight and Potter-Perigo, 2011). In IPF, excessive ECM secretion into the interstitium results in tissue thickening, stiffening and fibrosis, destroying the architecture of the lung (**Figure 1.4**).

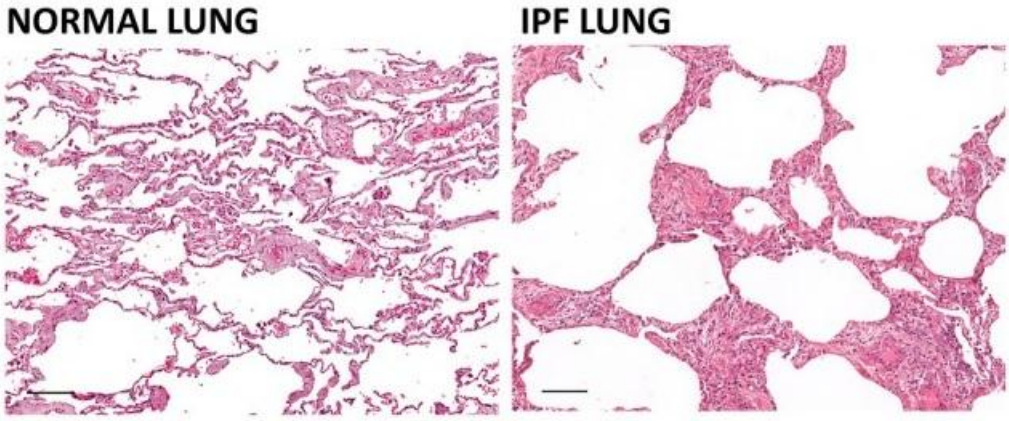


Figure 1.4 Hematoxylin and eosin stain of IPF and normal lung tissue. IPF lungs typically have thick bands of fibrous connective tissue compared to normal; excessive matrix deposition leads to scar formation that destroys lung architecture and compromises gas exchange. Scale bar = 100 μm. Adapted from Gardet *et al*. (2013).

The origin of the myofibroblasts contributing to fibrosis is under speculation. During tissue remodelling, many cells capable of contributing to this process are needed temporarily to

quickly restore homeostasis, therefore it is thought that myofibroblast precursors are recruited from many sources. Locally residing fibroblasts appear to be the most common source (Hinz et al., 2007) recruited through cytokine release by nearby injured AECs and responding immune cells. Other sources include pericytes and smooth muscle cells from the local vasculature (Hinz et al., 2007). The bone marrow-derived progenitor cells, mesenchymal stem cells (MSCs) and circulating fibrocytes, have also been implicated as myofibroblast progenitors (Bucala et al., 1994). Fibrocytes have been shown to express α-smooth muscle actin (α-SMA) and have contractile properties *in vitro* when differentiated (Abe et al., 2001). α-SMA is the most commonly used marker of differentiated myofibroblasts. It is found in undifferentiated fibroblasts, but in myofibroblasts it is incorporated into stress fibres, which provide contractility for wound closure during the healing response (Hinz et al., 2007; Malmstrom et al., 2004a). It can be difficult to distinguish myofibroblasts from smooth muscle cells using α-SMA since they share features of both fibroblasts and smooth muscle cells (Hinz, 2007). Fibrocytes have also been shown to migrate to the lungs and sites of injury in a bleomycin-mouse model of IPF via chemotactic signals released by the injured tissue (Phillips et al., 2004). However, much of the evidence for extrapulmonary fibroblast progenitors comes from animal models of fibrosis (Lama and Phan, 2006), and the extent of their involvement in IPF pathogenesis is also unknown.

There is current debate as to whether an epithelial-to-mesenchymal transition (EMT) process, in which differentiated epithelial cells undergo morphological changes to assume a mesenchymal phenotype, occurs in the lung (Willis et al., 2005) as a possible source of myofibroblast precursors. There is evidence suggesting it does occur *in vitro* in alveolar epithelial cells grown on plastic and *in vivo* in mouse models (Willis et al., 2005; Kim et al., 2006). In both cases a reduced expression of epithelial markers and an increase in expression of mesenchymal markers such as α-SMA could be observed within AECs in response to the profibrotic mediator transforming growth factor beta-1 (TGF-β_1). Willis *et al.* also demonstrated by immunohistochemical staining co-expression of both epithelial and mesenchymal markers within AECs in IPF lung biopsy tissue (Willis et al., 2005), suggestive of EMT. Kim *et al*. found co-expressing cells in IPF lungs and injured mouse lungs, and showed that mouse AECs cultured on fibronectin (a prominent component of the ECM) dramatically change phenotype resembling a mesenchymal cell, suggesting that AECs may be stimulated to trans-differentiate upon wound repair and ECM remodelling (Kim et al., 2006). However, it is yet to be proven that fibroblasts derived from epithelial cells do contribute to fibrosis in IPF. If it does occur, this marker expression pattern could be used as an indicator of disease progression (Lomas et al., 2012).

1.4.2. Transforming growth factor beta 1 (TGF-β_1) in fibrosis

Transforming growth factor beta is a profibrotic cytokine present as 3 isoforms in mammals, each with their own biological functions (Sakai and Tager, 2013). TGF-β signalling is important in embryonic development for tissue growth and for maintaining tissue homeostasis from there on (Sakai and Tager, 2013). TGF-β_1 is most predominantly expressed (Khalil et al., 2001) and is the most well characterised profibrotic molecule (Lepparanta et al., 2012), causing the differentiation of fibroblasts into myofibroblasts. It has been suggested that it may have opposing effects on AEC and fibroblast apoptosis compared to prostaglandin E2 in IPF (Maher et al., 2010). Upon release by injured epithelial cells and immune cells along with other growth factors such as connective tissue growth factor (CTGF), and tumour necrosis factor alpha (TNF-α), TGF-β_1 ligand binds to fibroblast TGF-β_1 tyrosine kinase receptors, which activates the canonical Smad pathway, and the signal is propagated to the nucleus via various Smad mediators to alter expression of particular genes (**Figure 1.5**). TGF-β_1 can also signal via Smad-independent signalling pathways to regulate gene expression, such as via various MAP kinase pathways (Zhang, 2009). TGF-β_1 is held in a latent complex upon secretion by associating with latency-associated peptide (LAP) to form a small latent complex, to which latent TGF-β-binding protein (LTBP-1) binds. LTBP-1 targets the complex to ECM structures, and has been shown to be significantly upregulated in the lungs of IPF patients, especially in areas of myofibroblast accumulation (Lepparanta et al., 2012).

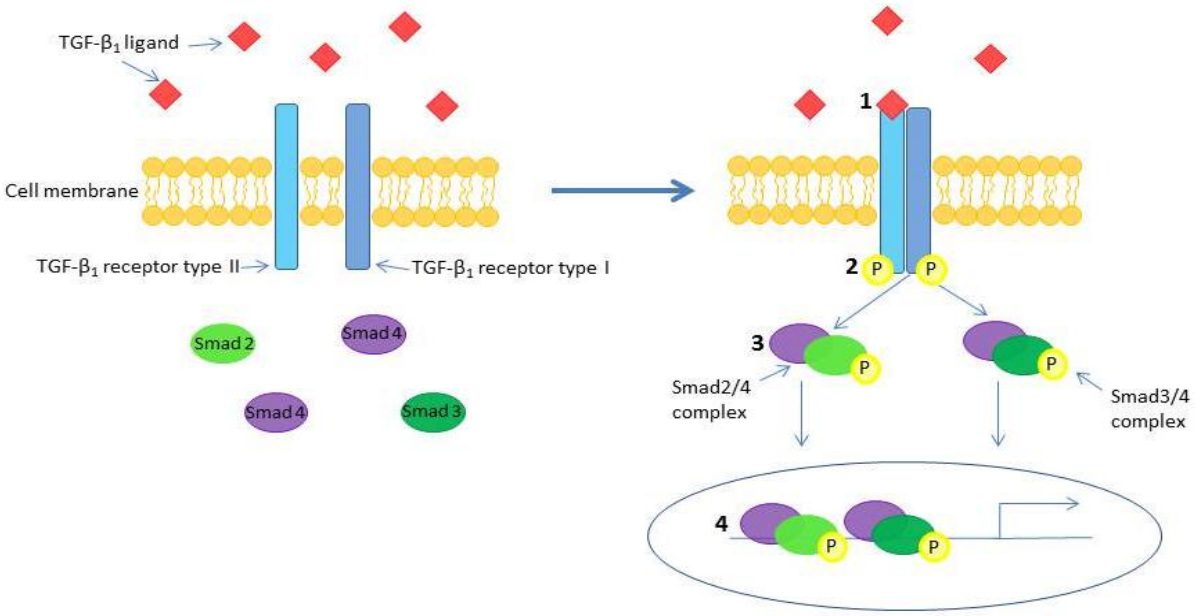


Figure 1.5 The canonical TGF-β$_1$ signalling pathway. TGF-β$_1$ ligand is secreted by epithelial cells and immune cells surrounding the area of injury. It binds to autophosphorylated TGF-β$_1$ receptor type II which recruits, heterodimerizes with and then phosphorylates TGF-β$_1$ receptor type I. TGF-β$_1$ receptor type I initiates signalling by phosphorylating effector Smad proteins, Smad2 and Smad3, which in turn heterodimerize with Smad4 to form Smad2/4 and Smad3/4 complexes. These complexes translocate to the nucleus to modify expression of genes such as profibrotic type 1 collagen, α-smooth muscle actin and other ECM components.

Integrins, cell surface (transmembrane) molecules that mediate cell-cell and cell-ECM adhesion, have also been described as having a role in fibroblast differentiation by activating latent TGF-β$_1$ (Todd et al., 2012; Amara et al., 2010), particularly αvβ6 integrin whose expression on epithelial cells has been shown to be increased in the lungs of UIP pattern pulmonary fibrosis patients (Sakai and Tager, 2013). Alveolar macrophages from IPF patients have also been shown to secrete active TGF-β$_1$ (Khalil et al., 2001). However, the mechanism of TGF-β$_1$ activation is little understood (Lepparanta et al., 2012).

1.4.3. Chronic inflammation in fibrosis

Another common observation in the IPF lung is inflammation of the parenchyma (Todd et al., 2012) although it is no longer the prevailing hypothesis for the cause of IPF.

Inflammation can be assessed by bronchoalveolar lavage (BAL) for the presence of immune cells, during which the lungs are washed with saline, the fluid is collected and the contents analysed. Neutrophilia is observed in BAL fluid of 70-90% of patients, and eosinophilia in 40-60% (Kliment and Oury, 2010; Soc, 2000). These conditions are associated with a worse prognosis since inflammatory cells release reactive oxygen species (ROS), proteases and peroxidases which can increase oxidative stress, damage the lung tissue and have an adverse effect on ECM remodelling. ROS generation by NADPH (nicotinamide adenine dinucleotide phosphate) oxidases has been implicated in pulmonary fibrosis due to these effects (Kliment and Oury, 2010; Amara et al., 2010). Oxidative stress could be a stimulus that causes micro-injuries to epithelial cells and maintains growth factor signalling. However, inflammation is often mild (Raghu et al., 2011).

The pathogenesis of IPF, how the disease manifests itself once triggered, is still not entirely understood, and the cause remains unknown. There also are many factors associated with IPF that have been shown to increase the risk of having the disease; these are discussed below.

1.5. Potential risk factors for IPF

Idiopathic pulmonary fibrosis is of unknown aetiology, but there are many associated risk factors. A risk factor is a term used to describe a variable that is associated with the increased risk of a disease. It is likely that IPF is induced by a complex mix of both environmental and genetic factors.

1.5.1. Environmental factors

Cigarette smoking is considered to be a main risk factor of IPF (Raghu et al., 2011; King et al., 2011). Approximately 60-70% of IPF patients are current or former smokers, according to epidemiologic studies (Schwartz et al., 1991) and many studies have reported a prevalence of smoking in IPF (Baumgartner et al., 1997; Taskar and Coultas, 2006; Enomoto et al., 2003). However, the relationship and potential causal mechanism is relatively undefined, though it is possible that it may be due to increased oxidative stress (Oh et al., 2012).

Metal and wood dusts also feature as major environmental and occupational risk factors for IPF (King et al., 2011; Raghu et al., 2011). There have been several studies in which an occupational exposure to dusts, past and/or present, has been strongly associated with an

increased risk of IPF. The first retrospective questionnaire case-control studies (Taskar and Coultas, 2006) by Scott (1990) and Hubbard (1996) reported that a history of working with metal/wood dusts or livestock was more prevalent among IPF patients than age/sex matched controls. Further studies by Garcia-Sancho *et al.* (2011) and Baumgartner *et al*. (2000) supported this finding. Other risk factors observed by Garcia-Sancho *et al.* included a family history of pulmonary fibrosis (the most significant risk factor in this study), cigarette smoking and past exposures to smokes, gases or chemicals. Further occupational risk factors identified by Baumgartner *et al.* included farming, raising birds, vegetable/animal dust, stone cutting/polishing and hairdressing. This group found that the risk of IPF generally increased with time of exposure to these risk factors. Retrospective studies such as these can be prone to recall bias and exposure misclassification (Taskar and Coultas, 2006), but these results suggest that IPF could be caused, at least in part, by environmental factors.

1.5.2. Comorbidities

Often in current and/or former smokers with IPF, an emphysematous pattern can be seen as well as UIP on HRCT. This has been termed "combined pulmonary fibrosis and emphysema" (CPFE) (Cottin et al., 2007) although according to the 2011 ATS/ERS/JRS/ALAT consensus statement regarding IPF diagnosis and management, it has not been confirmed as a distinct entity (Raghu et al., 2011). Emphysema and IPF have opposing effects on lung volume, elasticity, airflow and parenchymal structure (Schwartz et al., 1991). A comparison of survival rates between IPF patients with emphysema and IPF patients without emphysema demonstrated a higher mortality with CPFE than with IPF alone (median 25 vs. 34 months, respectively) (Mejia et al., 2009) which may be associated with pulmonary arterial hypertension. Of a cohort of 110 IPF patients, 28% also had emphysema, which was significantly associated with males and with smoking. It is not known whether each condition progresses independently or whether there is interplay between the two, resulting in a worse prognosis (Mejia et al., 2009); patients may require treatment for both conditions at once (Raghu et al., 2011).

It is hypothesised that aspirated reflux could be a trigger of an aberrant wound healing response and accumulation of extracellular matrix in IPF. Abnormal gastroesophageal reflux (GER) is common in IPF patients and it is considered to be a risk factor (Raghu et al., 2011). Evidence of microaspiration can be gained by detecting bile salts or pepsin enzymes that are found in non-acid or weak-acid reflux in the BAL fluid of IPF patients. It is now more widely appreciated that GER is associated with respiratory disease. Two studies, one pilot study by Tobin *et al*. (1998) and another larger study by Raghu *et al*. (2006) have investigated a possible relationship between GER and IPF. Both studies found that there was

a significantly higher prevalence of reflux in the IPF group (mostly asymptomatic) of 25% and 47% respectively, suggesting an association between them. However, because GER is common amongst the general population (Fahim et al., 2011b), it is difficult to evaluate whether there is a strong link between GER and IPF. GER may instead be a secondary event, occurring because of decreased lung compliance that fibrosis has caused (Fahim et al., 2011a).

Gribbin *et al.* (2009) found a possible association between GER and IPF and also diabetes mellitus (DM) and IPF in their study of 920 cases of IPF with 4 matched controls per case. They used prospective data from the THIN primary care database (The Health Improvement Network) in which general practitioners enter diagnostic and prescribing data as part of routine clinical care, and found that there were increased risks linked to use of ulcer drugs, and also use of insulin. Enomoto *et al*. (2003) also reported a prevalence of DM in a small cohort of 52 IPF patients, as well as smoking. However, a conflicting study by Miyake *et al.* (2005) reported no significant association between DM and IPF. DM is mentioned only very briefly in the 2011 ATS/ERS/JRS/ALAT consensus statement (Raghu et al., 2011) when mentioning the Gribbin study; it is unclear whether there is a strong association between DM and IPF, and if there is, why diabetes mellitus would increase the risk of developing IPF.

1.5.3. Family history/genetic predisposition

More than 95% of cases of IPF are sporadic, however it can also be familial (Raghu et al., 2011), with 1 in 5 patients reporting a family history of the disease (Alder et al., 2008). Familial interstitial pneumonia, or FIP, is a common term used in the literature as IIPs are frequently grouped together when discussing familial pulmonary fibrosis, with IPF as the most common subtype (Steele and Schwartz, 2013). FIP appears to present an autosomal dominant manner of inheritance, with anticipation (Lawson et al., 2011; Ravaglia et al., 2014), but it is quite indistinguishable from sporadic IPF. In a retrospective study by Ravaglia *et al.* (2014) of 30 cases of FIP and 127 patients with sporadic IPF, they could find no significant differences in symptoms, or overall outcomes, except that the FIP patients were more likely to be female and had a lower age of onset in younger generations than the non-FIP patients.

Increased frequencies of many gene polymorphisms have been observed in IPF, particularly in familial cases, including genes encoding cytokines, enzymes and antibodies (Raghu et al., 2011; Fahim et al., 2011a; Strieter and Mehrad, 2009), suggesting a genetic susceptibility of individuals to this disease.

Gene mutations in surfactant proteins have been linked to pulmonary fibrosis, particularly *SFTPC* and *SFTPA2*, as they result in disturbed homeostasis and cellular stress, particularly to the alveolar epithelia that secrete surfactant (Steele and Schwartz, 2013). Misprocessed or misfolded proteins can trigger an unfolded protein response (UPR), which if activated for a sustained time period can lead to alveolar epithelial cell injury and death, and so may be involved in IPF pathogenesis (Lawson et al., 2011). In addition, a variant in the promoter of mucin 5B (*MUC5B*) expressed by bronchial epithelial cells has been associated with IPF in both sporadic and familial cases and may predispose to both types of IPF (Seibold et al., 2011; Steele and Schwartz, 2013). A study by Seibold *et al.* (2011) found that the minor allele of this single nucleotide polymorphism was present in 34% of 83 familial IP patients studied, 38% of 492 IPF patients, and only 9% of the 322 controls. *MUC5B* expression was 14.1 times higher in IPF patients than controls. This association has since been confirmed in another study of 341 IPF patients and 802 controls (Zhang et al., 2011). It is possible that upregulated expression of mucin 5B may be involved in IPF pathogenesis.

The polymerase telomerase that functions to elongate chromosome telomeres, the tandem TTAGGG repeats that protect chromosome shortening during DNA replication, has been shown to have germ line mutations in components *hTERT* and *hTR* in families with pulmonary fibrosis. According to a report by Armanios *et al.* (2007), severe telomere shortening in peripheral blood leukocytes was observed in several families with FIP, indicating reduced telomerase activity. Lawson *et al.* (2011) reported that 10% of this FIP cohort had either of these mutations. It was further noted by Alder *et al.* (2008) that pulmonary fibrosis patients had short telomeres in both their leukocytes and alveolar epithelial cells compared with control subjects in their cohort, not only in familial IIP but also in sporadic IPF. When telomeres become too short a DNA damage response is activated, which can lead to cell death. Telomere shortening occurs with aging, so this finding may link this with onset of IPF, as aging is also a risk factor for IPF (Thannickal, 2010).

Genetic alterations, apoptosis resistance, uncontrolled cell proliferation and changes in cell signalling are pathogenic hallmarks of both IPF and cancer. Two tumour suppressor genes, *p53* and *FHIT*, that are susceptible to mutation and are heavily implicated in uncontrolled tumour proliferation, have been demonstrated as also being mutated in IPF, and therefore may contribute to excess fibroblast proliferation and extracellular matrix deposition and may also be linked to higher incidence of lung cancer with fibrosis (Vancheri et al., 2010).

In a study by Steele *et al.*, cigarette smoking was also found to be a risk factor for familial IP, suggesting that an interaction between both genetic and environmental factors may be involved in the pathogenesis of FIP (Steele et al., 2005).

Despite many gene polymorphisms and mutations being associated with predisposition to IPF, none of these are consistently associated; therefore, genetic testing for diagnosis is not recommended (Raghu et al., 2011). Further studies are needed.

As there are so many factors involved concerning the cause of IPF, its risk factors and how it is aggravated, there are many avenues to explore in terms of treatment. It is becoming increasingly likely that IPF is caused by a multitude of factors, making the search for therapeutic options more challenging.

1.6. Treatments for IPF

1.6.1. Current therapies

The main currently recommended treatment for IPF is surgery, especially when patients do not respond to any other medical therapy and progress to advanced disease (Brown et al., 2015). Thabut *et al*. (2003) showed in a study of 46 IPF patients referred for surgery that lung transplantation reduced the risk of death by 75%. However, this option is limited to patients that fit very specific criteria, and if lung transplantation is recommended and performed, 5-year survival rates are approximately 50% (Raghu et al., 2011). Prior to 2015, if a patient was prescribed drugs they were often immunosuppressants and corticosteroids to reduce inflammation (Seymour, 2010), although as discussed previously, inflammation can occur but more likely as a side effect rather than a cause (Todd et al., 2012). Corticosteroids can also cause unpleasant side effects, which is not ideal if the patient is already seriously ill.

Many novel drugs for IPF have been synthesized and brought to clinical trial, but unfortunately most have shown little or no benefit (King et al., 2011). In the 2011 official ATS/ERS/JRS/ALAT statement concerning IPF (Raghu et al., 2011), many drug options were explored, such as the antioxidant *N*-acetylcysteine (NAC), Interferon-γ 1b (an immunoregulatory cytokine that limits fibroblast proliferation and collagen synthesis (Rafii et al., 2013)) and anticoagulants, yet it was recommended that the majority of patients should not be treated with any of them due to lack of evidence of their efficacy, or because their effectiveness would not outweigh the risk to the patient. It was also observed in the PANTHER-IPF study of combination treatment with NAC, prednisone (a corticosteroid) and azathioprine (an immunosuppressant), that this widely-used therapy actually results in increased risk of death when compared to placebo (Idiopathic Pulmonary Fibrosis Clinical Research et al., 2012). Instead the ATS/ERS/JRS/ALAT statement recommended long-term oxygen therapy and palliative care. However, the most recent update of this statement,

released in 2015, now conditionally recommends two drug treatments, Pirfenidone (trade name Esbriet) and Nintedanib (trade names Ofev and Vargatef) (Raghu et al., 2015).

Out of the many phase III clinical trials over the past decade, only Pirfenidone and Nintedanib have shown enough clinical benefit to be recommended as first-choice therapeutic options for IPF patients. Again, these are only used for patients who fit specific criteria. Results of the clinical trials showed that in patients with mild-to-moderate IPF, both drugs reduced decline in FVC by approximately 50% in 1 year, and had acceptable safety profiles, leading to their subsequent approval (Bonella et al., 2015).

Pirfenidone has been described as "a new molecular entity" with an unknown mechanism of action (Seymour, 2010); however, there is evidence to show that it causes downregulation of collagen I in the presence of TGF-β_1 in a dose-dependent manner *in vitro* (Nakayama et al., 2008) and that it reduces fibroblast proliferation and α-SMA expression *in vitro* (Conte et al., 2014). It has been shown to have anti-fibrotic (Schaefer et al., 2011) and anti-inflammatory properties in experimental models, and may have a role in inhibiting EMT (Hisatomi et al., 2012). In clinical trials, Pirfenidone demonstrated beneficial effects such as slower decline in both percentage of FVC and IPF-related mortality, plus improvements in survival with no progression, and was approved in the European Union in 2011 (Xaubet et al., 2014) and by the US Food and Drug Administration (FDA) in 2014 (Bonella et al., 2015). Adverse effects include gastrointestinal symptoms, nausea, photosensitivity, and fatigue, but these are often mild and can be lowered by reducing the dose of Pirfenidone (Takeda et al., 2014). It is currently recommended that this drug can be used in patients willing to tolerate adverse effects for small benefits (Raghu et al., 2011). As previously mentioned, Pirfenidone cannot be used to treat advanced IPF, the stage at which sufferers are in most need of treatment.

Nintedanib has very recently come to the forefront of treatment options for IPF. This drug is a potent, small molecule, triple angiokinase inhibitor. Formally known as BIBF 1120, Nintedanib specifically inhibits vascular endothelial growth factor (VEGF) receptor, platelet-derived growth factor (PDGF) receptor and fibroblast growth factor (FGF) receptor tyrosine kinases (Roth et al., 2009), whose corresponding ligands are profibrotic mediators. It was originally developed as a treatment for cancer due to its ability to inhibit VEGF-R and therefore reduce angiogenesis, a process required for tumour growth and metastasis. Along with data from experiments carried out in animal models of fibrosis using an analogue, BIBF 1000, that suggested its efficacy in inhibiting fibrosis, this data led to its clinical development (Woodcock et al., 2013).

Findings from a 12 month, randomized phase II IPF trial with Nintedanib in 2011 (the TOMORROW trial) reported by Richeldi *et al.* indicated that it may possibly reduce the decline in lung function of IPF patients (annual rate of decline in FVC was used as the primary endpoint). There were also fewer incidences of acute exacerbation in the test group compared to placebo, and patients reported a better quality of life in the test group (Richeldi et al., 2011). These promising findings resulted in its continuation into phase III trials. The setup of this study was two replicate 12 month, randomized, double-blind trials (named INPULSIS™-1 and -2) with over 1000 participating IPF patients, to further evaluate the efficacy of Nintedanib at the higest dose given in the phase II trials in comparison with placebo (Richeldi et al., 2014a). In both trials Nintedanib significantly reduced the rate of lung function decline over the period of the study. There was an inconsistent association with the drug and the risk of acute exacerbation as seen in the phase II trial, potentially due to the difficulty in diagnosing and characterising an exacerbation (Richeldi et al., 2014b). In both phase II and phase III trials, the most common adverse events were gastrointestinal symptoms, particularly diarrhoea, but treatment discontinuation due to these effects was minimal.

Upon consideration of results including this phase III trial, the FDA granted breakthrough therapy designation to Nintedanib in July 2014 for the treatment of IPF, to help accelerate the process of its approval due to the seriousness of this disease and the promising results it had provided, given the lack of treatment available for IPF (Calandra, 2014). It was approved in October 2014 (Fala, 2015). There are also clinical trials still ongoing as the safety and efficacy of Nintedanib is not yet fully confirmed; there is also a trial using combination treatment with both Nintedanib and Pirfenidone (Woodcock et al., 2013) (NCT01417156).

Despite the presence of Pirfenidone and Nintedanib, there is no cure for IPF and the outlook for the majority of patients is still bleak, and therefore it is imperative that the search for effective treatments continues.

1.6.2. Histone deacetylase inhibitors as a therapeutic option for IPF

The pathogenesis of IPF shares some similarities with that of cancer. IPF has been described as a neoproliferative disorder (Vancheri, 2013) with pathogenesis involving uncontrolled cell proliferation, dysregulated cell signalling and delayed apoptosis. Nintedanib was originally developed as a cancer treatment but was found to also have beneficial effects in IPF (Woodcock et al., 2013). It is therefore possible that cancer therapeutics may hold the key to IPF treatment (Vancheri et al., 2010).

Histone deacetylases (HDACs) are considered drug targets for cancer as well as other diseases, and are the focus of many clinical trials (Choudhary et al., 2009). Acetylation is a type of post-translational modification, whereby an acetyl group is added to a lysine residue of a protein, often histones, by histone acetyl-transferases (HATs); along with deacetylation by HDACs these enzymes regulate the expression of specific genes, such as those involved in cell cycle regulation and cell proliferation and apoptosis. This regulation occurs through influencing the accessibility of DNA to transcription elements, by causing disruption of the association between histones and DNA (Emanuele et al., 2008) (**Figure 1.6**). Approximately 2-5% of genes are regulated by acetylation of histones (Emanuele et al., 2008). An imbalance in the antagonistic activity of HATs and HDACs has been associated with disease, particularly tumour proliferation; there is evidence that HDACs are often overexpressed in cancer, leading to their aberrant recruitment to gene promoters and transcriptional repression of tumour suppressor genes (Ropero and Esteller, 2007).

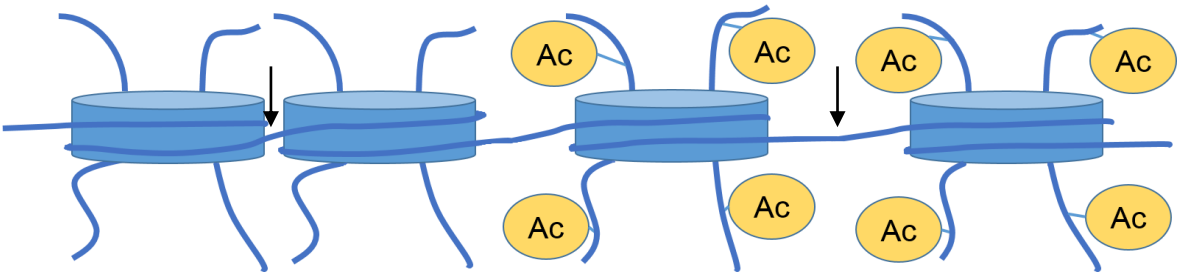


Figure 1.6 Histone acetylation increases the accessibility of transcriptional machinery to DNA. The addition of an acetyl group to histone tails within chromatin removes a positive charge, reducing the strength of the interaction between histones and negatively-charged DNA, causing local areas of chromatin to decondense, rendering it more accessible for transcription (Struhl, 1998).

Preventing IPF progression upon diagnosis would require inhibition of IPF fibroblast and myofibroblast proliferation and differentiation, processes involving genes regulated by acetylation state. HDAC inhibitors (HDACIs) interfere with deacetylation, thereby themselves regulating gene expression. It is possible that HDACIs could be used as treatments for IPF. They have been shown to have anticancer activity, with effects resulting in growth arrest and apoptosis of tumour cells (Balakin et al., 2007), whilst being shown to have low toxicity for healthy cells (Armeanu et al., 2005) and three have been approved as anti-cancer therapeutics (Glass and Viale, 2013).

Investigations have already been carried out into the use of HDAC inhibitors in fibrosis in general, including polycystic kidney disease (Pugacheva et al., 2007; Cao et al., 2009) and renal fibrosis (Pang et al., 2009). More recently, Zhang *et al.* used suberoylanilide hydroxamic acid (SAHA), a broad-spectrum HDACI already FDA-approved for cancer treatment, to show that the increased expression of collagen III-α1 mRNA and protein expression in IPF fibroblasts is reduced upon HDACI treatment. The same outcome was observed in a murine model (Zhang et al., 2013b). In addition, Davies *et al.* (2012) observed that the Class I HDACI Spiruchostatin A inhibited the proliferation of both primary IPF and control fibroblasts in a time- and concentration-dependent manner with little cytotoxicity, and suppressed myofibroblast differentiation and the collagen gene expression of TGF-β_1-stimulated fibroblasts.

FK228, or Romidepsin (trade name Istodax) is a HDAC inhibitor structurally similar to Spiruchostatin A (Davies et al., 2012), and is also already approved by the FDA as an anti-cancer therapeutic; it can be used for treatment of cutaneous T-cell lymphoma (CTCL) and peripheral T-cell lymphoma (Glass and Viale, 2013). FK228 is a natural, bicyclic compound, and acts as a pro-drug as it is activated upon entering a cell. It can act on both histone and non-histone targets (Konstantinopoulos et al., 2006). Blocking the action of HDACs results in an increase in the acetylation of histones, increasing the accessibility of transcriptional machinery to DNA, and thus increasing the expression of genes such as tumour suppressors and cell cycle inhibitors to inhibit tumour progression. For example, the cell cycle inhibitor p21 has been demonstrated to be inactive in some cancers through hypoacetylation of the promoter, and HDACI treatment causes an increase in promoter acetylation and gene expression, and inhibition of tumour growth (Gui et al., 2004; Ropero and Esteller, 2007).

FK228 has been demonstrated to be more cytotoxic to tumour cells than to normal cells (Ueda et al., 1994), and it has an acceptable safety profile. However, the precise mechanism of action against cancerous T cells in T cell lymphoma is unknown (Valdez et al., 2015). Adverse events according to clinical trials are generally mild and the most common side-effects include fatigue, nausea, vomiting and anorexia (Piekarz et al., 2009; Whittaker et al., 2010). FK228 is administered to patients intravenously over a 4-hour period, at a recommended dose of 14 mg/m^2, over a 28 day cycle in which it is given on days 1, 8, 15 and 21 (Foss et al., 2016). Cycles are repeated provided that it continues to be tolerated and show benefits to the patient.

Following the Spiruchostatin A study, Davies *et al.* have shown that FK228 inhibits fibroblast proliferation and differentiation in the same manner, even more potently than Spiruchostatin A (Davies, 2011). They have also shown that it suppresses the expression of

α-SMA, collagen III and HDAC4, with little toxicity to AECs, and that it also suppresses bleomycin-induced fibrosis in an animal model (Davies et al., unpublished). These results suggest that FK228 could be a suitable therapeutic option for the treatment of IPF. This is the concept that is to be investigated further in this thesis.

In order for FK228 to be considered as a potential IPF treatment, it would need to be taken to clinical trial. A crucial part of designing a clinical trial for evaluating a potential therapeutic agent is choosing the most appropriate primary endpoint, so that a significant difference can be measured (Nathan and Meyer, 2014). There is no standard defined clinical primary endpoint for IPF treatment; using mortality as an endpoint is quite extreme, whilst an alternative endpoint, the 6-minute walk test (measuring a patient's heart rate and how far they can walk) lacks standardization (Raghu et al., 2011). The most commonly used endpoint is change in FVC (Saketkoo et al., 2014). This has been associated with survival, but whether it should be used as a surrogate endpoint for survival is disputed (Raghu et al., 2011).

One potentially effective method of analysing whether a drug is having the desired effect is to analyse the modulation in expression of one or more identified protein biomarkers of a disease in response to treatment. There are currently no validated protein biomarkers of IPF that could be used in the clinic (Vij and Noth, 2012), either as a diagnostic tool or an indicator of treatment efficacy. The aim of this project is to identify proteins with differential expression in IPF fibroblasts compared to non-IPF fibroblasts using a proteomic approach, and to analyse changes in expression after treatment with FK228. Following validation these differentially expressed proteins would have potential as candidate biomarkers, and could subsequently be monitored in clinical trials.

1.7. Protein biomarkers and proteomics

A biomarker has been defined as being "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" by the FDA Biomarkers Definitions Working Group (Atkinson et al., 2001). These can include vital signs, physiologic testing, radiographic studies and biochemical analyses (Flynn et al., 2015). This study focuses on biomarkers with respect to biological molecules within tissues or fluids, that can act as molecular signatures of a condition, and can be used as indicators of change between different states e.g. before and after drug treatment. Biomarkers can be used in clinical trials

to determine whether there is a response to treatment; they can also be used for early diagnosis and predicting disease progression, which, if a treatment becomes validated for IPF, could greatly improve the prognosis of patients. Currently, there are no single biomarkers that can accurately indicate whether a patient has an idiopathic interstitial pneumonia (van den Blink et al., 2010). An ideal biomarker would be one that can be easily and accurately measured, reproducibly, and an extra benefit would be if it could be used for serial monitoring to help doctors keep track of disease progression (Vij and Noth, 2012). Additional important factors when considering biomarkers for clinical testing include the safety and ease of obtaining the sample to be assessed, and availability and cost of the test itself (Flynn et al., 2015).

Biomarker discovery is typically a very long process, and many candidate biomarkers fall at any of the several hurdles during this time. The National Cancer Institute has proposed the concept of five phases of biomarker discovery and development (Pepe et al., 2001) - these were developed with cancer in mind but the structure is applicable to other diseases such as IPF. The studies described in this thesis are part of Phase 1, the "preclinical exploratory" phase - looking for characteristics unique to IPF or to FK228 treatment, such as differentially expressed proteins, that could be candidate biomarkers of disease or treatment. Any potential candidates identified in these studies would need to pass subsequent phases of biomarker development before clinical use. Validation and development of an assay begins in Phase 2.

There are several reasons for the current lack of biomarkers for IPF (amongst many other diseases) – one is the complexity of the disease itself. IPF is heterogeneous, with multiple complex and interactive mechanisms of pathogenesis (Ley et al., 2014). This makes it difficult to identify biomarkers with disease association as there are many potential biological processes and signalling pathways to explore. There are also a large number of overlapping diseases such as other idiopathic interstitial pneumonias that also have unknown aetiology. Biomarkers for IPF need to be specific and able to distinguish between the different IIPs.

Another limitation of biomarker development is the clinical trial phase – IPF is a relatively rare disease, therefore it can be difficult to recruit enough patients that match the eligibility criteria to gain a big enough sample population. Even then, the aggressiveness of some patients' progression may result in their death early in a study. Biomarker studies often lack replication in addition to size – these kinds of studies require biomarker validation across independent cohorts with a standardised method. Relevant controls including other interstitial lung diseases are required to test for specificity, as well as longitudinal follow up

(Zhang and Kaminski, 2012). Validation is also needed across ages, sexes and ethnicities to ensure generalisation (Ley et al., 2014). To meet all of these requirements, the studies would need to be incredibly large and would be complex to set up; it would also be very expensive.

Given the challenges outlined above relating to biomarker discovery and development, it is not difficult to understand why there is a lack of validated biomarkers for IPF – it is a long and complex process with no guarantee that a candidate biomarker will translate to clinical use. However, the idea of using one or more biomarkers of IPF as diagnostic, monitoring and/or prognostic tools remains an attractive one, due to the aforementioned difficulties in diagnosis and the current lack of methods to predict disease course; the ability to do this is incredibly important for both care of patients and for research (Vij and Noth, 2012). The complex pathogenesis of IPF and the lack of effective therapies makes the search for predictors of survival and response to potential treatments critical, in spite of the challenges.

There are two approaches to identifying biomarkers - selecting individual candidates based on existing evidence, and using omics technologies to screen for candidates which may be related to the disease (Ley et al., 2014). The former approach has the benefit of pathological reasoning, yet is slow and inefficient, whilst the latter is much faster as it gives the opportunity for identifying a large number of candidates, although this method generates a lot of irrelevant data and each candidate requires further validation and preferably disease association. The studies within this thesis use the omics screening technique for biomarker identification for its speed and efficiency.

Protein biomarkers can be identified and analysed using proteomic approaches. The proteome encompasses all of the expressed proteins in a cell, tissue or organism, at a particular time and under certain conditions. Proteomics is the study of the proteome, and aims to identify, quantify and characterise these proteins in order to understand their cellular functions (Graham et al., 2005; Lane, 2005). This includes studying protein isoforms, modifications and interactions, which altogether contribute to the complexity of the proteome. On top of this, the proteome is constantly changing in response to stimuli, both internal and external, making it very dynamic; a single proteomic experiment can only demonstrate a snapshot of what is occurring in a particular cell or tissue at the time of protein extraction. A lot of studies focus on sub-proteomes, groups of proteins with a common property (e.g. cellular localization) instead of the entire proteome in an effort to simplify the very complex mixtures (Graham et al., 2005; de Godoy et al., 2006). Many experiments are carried out in a high throughput manner using the “shotgun” method – in which a whole proteome is subjected to analysis – rather than using a targeted approach

(Patel et al., 2009). This is a type of bottom-up proteomic study that enables global protein identification and quantitation within a sample, and was used in this study.

Proteomic research has developed very fast, from the use of two-dimensional gel electrophoresis in early experiments (Graham et al., 2005; O'Farrell, 1975) to immunoassays, to the latest highly advanced approaches using high resolution mass spectrometers. Since it is such a broad field, the following section covers only relevant topics to this investigation – mass spectrometry and how it can be used to achieve relative and absolute quantification of proteins.

1.7.1. Background to mass spectrometry

Mass spectrometry (MS) is an analytical technique used to identify and quantify molecules. The beginnings of MS date back to the early 1900's (Lane, 2005) and only as recently as the mid-1990's emerged as a routine method for protein identification (Patel et al., 2009; Pandey and Mann, 2000). MS-based proteomics can yield information about cellular composition and specific organelles, as well as protein interactions and post-translational modifications (PTMs), such as acetylation, phosphorylation and carbonylation (Walther and Mann, 2010). Protein samples are typcially digested into peptides, usually by the serine protease trypsin, for analysis by MS as whole proteins yield limited sequence data and peptides can be more easily sequenced for protein identification. Liquid chromatography (LC) is often coupled to MS for separation of peptides, to simplify the sample for MS analysis. LC makes use of the chemical properties of peptides by separating them using reversible binding to a resin on a column; for example, reverse-phase LC separates peptides by hydrophobicity by gradually increasing the percentage of organic solvent. The LC system used in this study is nanoflow-UPLC (ultra-performance LC); the use of nanoscale operations increases sensitivity, and requires less sample for analysis, a distinct advantage when only a small amount of sample is available, as is often the case with clinical samples. UPLC uses very high pressure, which is necessary to drive the sample through a column with an internal diameter in the μm range for greater peptide separation (Aebersold and Mann, 2003).

MS analysis involves the travel of charged molecules in electromagnetic fields within a vacuum, so peptides must be ionized, usually positively, and in gaseous phase before they can be analysed (Walther and Mann, 2010). Ions are separated according to their mass-to-charge ratios (*m/z*) using one or more mass analysers. The resulting mass spectra can be pre-processed using specialised software and searched against a database containing protein sequences to allow protein identification.

1.7.2. Methods of ionization

The two main methods used for ionization of peptides are electrospray ionization (ESI) and matrix-associated laser desorption/ionisation (MALDI). These are considered "soft ionisation" techniques as they generate ions in the gas phase without breaking bonds. Which ionisation technique to use is dependent on the sample to be analysed, for example the complexity and the phase it is in.

Electrospray ionisation

ESI is the most commonly used technique of the two, and is the chosen method of peptide ionisation for this study, due to its efficiency in ionising complex samples. ESI forms gaseous ions from solution phase samples, making it highly compatible with liquid chromatography. A high voltage with positive potential is applied to the liquid sample as it travels through a conductive capillary tube, and it is drawn out forming an elliptic shape. When voltage is increased further it forms a conical shape, known as a "Taylor cone". Droplets form and start to spray charged droplets, moving towards a negative counter electrode due to their positive charge. During this process the solvent begins to evaporate, and the droplets may pass through a heated capillary or a curtain of heated nitrogen that causes further evaporation of solvent, generating smaller, even more highly charged droplets. These charged droplets eventually lose their solvent sheath, and the ions are accelerated into the mass spectrometer (**Figure 1.7**) (Lane, 2005; Wilm, 2011). The complete mechanism of the ESI process is not fully understood (Lane, 2005). Coupled with liquid chromatography, ESI can be used for the analysis of complex samples. Depending on the mass analyser used there may also be no limitation on mass, meaning that even very large protein complexes can be studied, a previous limiting factor in mass spectrometry. It is very efficient in ionization, meaning that very low abundant proteins can be identified, making the system high-throughput (Wilm, 2011).

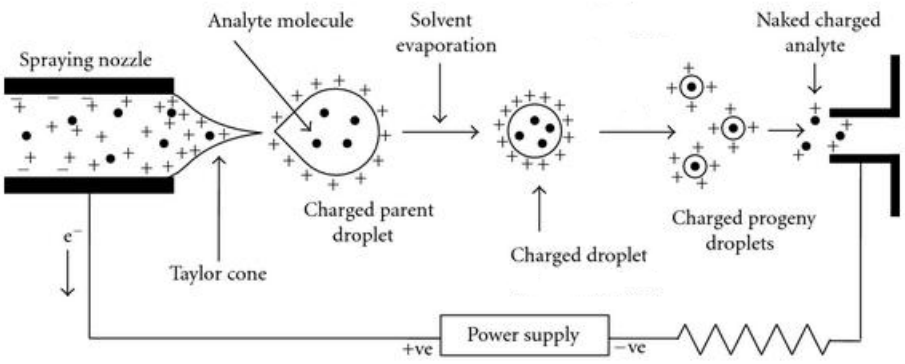


Figure 1.7 A representation of the electrospray ionisation process. Adapted from Banerjee and Mazumdar (2012).

Matrix-associated laser desorption/ionisation (MALDI)

The MALDI technique uses a laser to ablate and ionize peptide samples that have been co-crystallised with matrix material. The matrix is a low molecular mass UV absorbing chemical that absorbs some wavelength of radiation from the laser, and expands into gaseous phase due to rapid heating, taking analyte molecules along with it into the gas phase (Lane, 2005). Ionization occurs via transfer of protons from the matrix, but again this mechanism is not fully understood (Wilm, 2009; Lane, 2005). Efficient energy transfer provides high accuracy and sensitivity (Lewis JK WJ, 2000). MALDI is typically used for analysis of simpler peptide mixtures rather than highly complex samples.

1.7.3. Mass analysers

Peptide ions are separated and detected according to their *m/z* ratio by the mass analyser of the mass spectrometer. There are four kinds used in proteomics: the quadrupole, the time-of-flight (TOF) analyser, the ion trap and the fourier-transform analyser. The first two are relevant to this study and will be discussed here.

Quadrupoles

A quadrupole mass analyser is made up of four parallel rods equidistant from a central axis, along which ions introduced from ESI or MALDI travel. One set of opposing rods has a positive voltage, and the other set has a negative voltage. These rod pairs create mass filters that overlap, meaning that ions must have *m/z* values above a critical *m/z* threshold or below a critical *m/z* to pass through. Others with an *m/z* outside of these criteria are discharged on the rods. In this way the quadrupole can be used to select ions with a particular *m/z* (Walther and Mann, 2010). The ratio is proportional to the voltage, so if the voltage is increased the

selected *m/z* will be higher. By scanning the quadrupole voltage, a full mass spectrum can be created (Lane, 2005).

Two or more mass analysers are typically coupled together to perform tandem mass spectrometry, or MS/MS. Triple quadrupoles are often used for MS/MS, where the second quadrupole acts as a collision cell to produce fragment ions, and the quadrupoles either end are used for scanning (Lane, 2005). A quadrupole can be also be used in conjunction with a time-of-flight (TOF) analyser (QTOF), to identify molecules by their arrival time at the detector (Walther and Mann, 2010).

Time-of-Flight Analysers (TOFs)

The premise of TOF is that ions with different masses, when accelerated through a vacuum, will travel the same distance in different amounts of time (i.e. an ion of larger mass will travel slower than an ion of smaller mass). The analyser measures the time taken for ions to move from the ion source to a detector to determine their *m/z*. The additional retention time information is an advantage as it acts as confirmation of detection (Bataineh et al., 2006). Ions are pushed down into the TOF analyser and are forced to travel in a "V" or "W" shape by a "reflectron" or ion mirror situated at the bottom of the analyser; the ions with more energy penetrate deeper before being reflected upwards, so they have to travel a slightly longer distance, arriving at the detector at the same time as the ions with lower energy (Lane, 2005). This is shown in **Figure 1.8**, a schematic of the Synapt G2-S mass spectrometer from Waters Corporation that will be used in this study**.** This mass spectrometer is a QTOF system, which as mentioned previously, implements both a quadrupole mass filter and a time-of-flight mass analyser, and has a collision cell in-between. Combining the quadrupole with the TOF analyser results in increased mass accuracy, along with higher sensitivity and resolution (Lane, 2005).

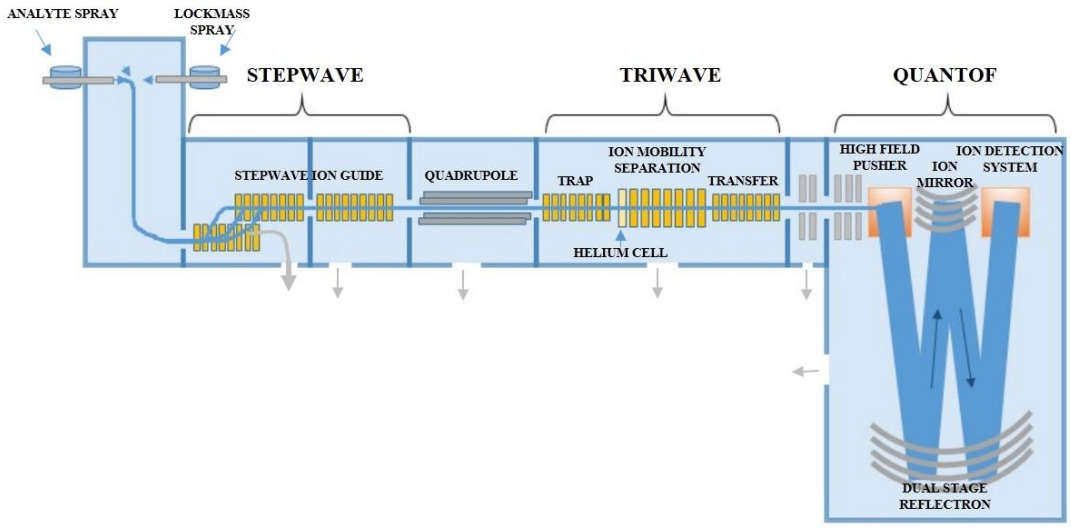


Figure 1.8 A schematic of the Waters Synapt G2-S HD mass spectrometer. The G2-S includes StepWave technology to filter neutral ions, an ion mobility cell which separates ions according to shape (implemented before fragmentation in the transfer cell) and quadrupole/time-of-flight mass analysers.

1.7.4. Modes of data acquisition

Mass spectrometry can be performed using a data-dependent mode of data acquisition (DDA) or a data-independent mode of acquisition (DIA). MS/MS is traditionally performed using DDA. Precursor peptide ions are scanned upon entry into the mass spectrometer and precursor ion peaks are selected (i.e. peptides) in the quadrupole according to defined criteria such as ion intensity and charge state. Ions are selected in turn and are subjected to fragmentation, and the *m/z* of the fragment ions is measured in the second mass analyser. Once analysed, ions of a specific *m/z* can be dynamically excluded from following cycles, making way for peptide ions of lower abundance to be fragmented (Peng and Gygi, 2001). Fragment ion data along with the precursor *m/z* is used to identify the peptide sequence of the precursor, as the difference in mass between consecutive fragment ions corresponds to the exact mass of one amino acid (Walther and Mann, 2010); thus fragmentation yields more information about the peptide ion for more confident protein identifications. MS/MS is a very powerful tool for its ability to isolate and sequence peptides within very complex mixtures (Peng and Gygi, 2001). However, some low abundant peptides can be missed due to non-selection during DDA (de Godoy et al., 2006).

During LC-MS/MS, precursor ions are identified for analysis in MS mode, and the instrument switches from survey scanning in MS mode to MS/MS mode, in order to select

each precursor for fragmentation and measure it. Whilst the instrument is in MS/MS mode, it is not performing MS survey scans and therefore precursor ions not identified for fragmentation and analysis are missed, since peptide ions are continually sprayed into the mass spectrometer. These are usually ions of lower abundance, as the switch to MS/MS mode is typically triggered by the intensity of ions being analysed (Blackburn et al., 2010), in the form of an eluting peak rising above a threshold defined by the user (Plumb et al., 2006). This data-dependence causes undersampling (Geromanos et al., 2009), and irregular sampling intervals, resulting in lower chromatographic peak quality, and data which may not necessarily be reproducible. Also, the point at which the peak reaches the threshold to switch to MS/MS mode may not be at the apex of the peak where the precursor ion is at its most intense, which would further reduce the quality of the product peak (Blackburn et al., 2010). However, if the frequencies of MS survey scans and the ion intensity threshold are increased, fewer ions are selected for fragmentation, resulting in lower sequence coverage. Therefore, the user must decide how to balance the trade-off (Blackburn et al., 2010).

MS^E is a data-independent mode of acquisition (DIA), in which the mass spectrometer continually alternates between low energy mode (MS) and elevated energy mode (MS^E) without ion selection (Silva et al., 2006). LC-MS^E from Waters Corporation was the first platform developed for data-independent analysis (Blackburn et al., 2010).

An illustration of MS^E is shown in **Figure 1.9**. Peptides separated by LC are introduced into a QTOF mass spectrometer operating in MS^E mode. The quadrupole allows all precursor ions to pass through to the TOF analyser without selection. The collision cell energy continuously cycles between low and elevated energy; thus precursor ions are analysed in low energy mode and are fragmented in elevated energy mode in one analytical run. The fragment ions are matched back to their precursors by their retention time and intensity using an Ion Accounting algorithm (Waters) to create reconstructed fragment ion spectra for each precursor ion, which can then be used for peptide/protein identification. Data-independent analysis will be carried out using a Synapt G2-S Waters mass spectrometer during this study (as shown previously in **Figure 1.8**).

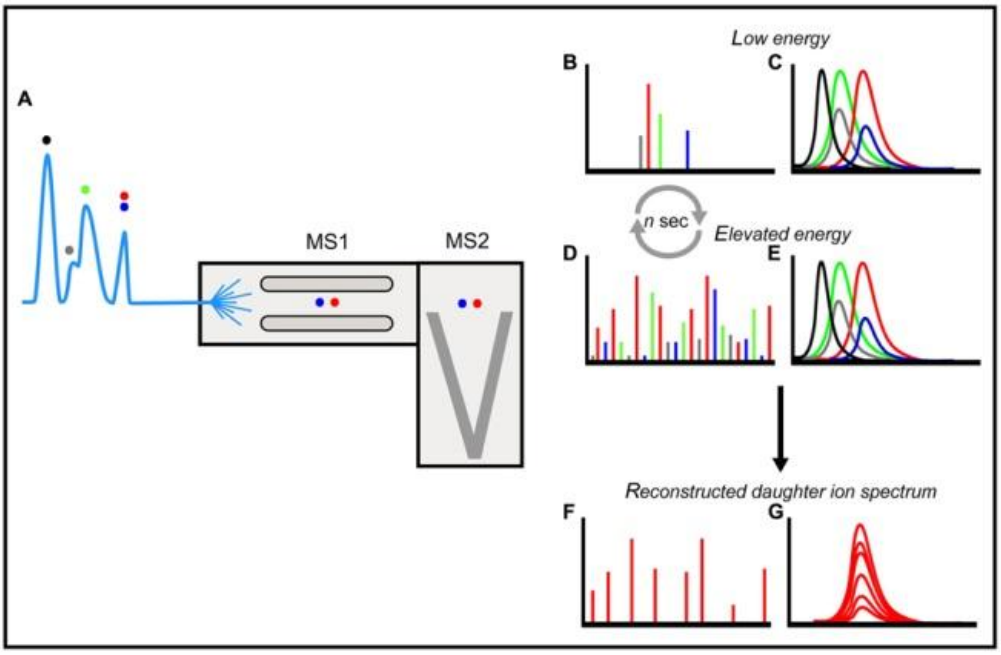


Figure 1.9 An overview of LC-MSE data-independent mode of acquisition. Peptides separated by LC (A) are sprayed into a QTOF mass spectrometer by ESI. The quadrupole (MS1) does not select ions based on intensity as it would with DDA, and all parent/daughter ions are analysed in the TOF analyser (MS2). Through continuous cycles of low and elevated energy in the collision cell, spectra are generated for precursor (B) and fragment ions (D) without bias. An Ion Accounting algorithm (Waters) is used to match parent and daughter ions through comparing retention times and intensity (C and E) to create a reconstructed daughter ion spectrum (F and G) that can be used for each parent ion to identify peptides and proteins. Diagram republished from Kramer *et al.* (2012).

MS^E avoids the issues of undersampling by fragmenting all precursors at once across their whole peak profile, independent of ion intensity, without the need to go "offline", and the switching between MS and MS^E modes of acquisition occurs at regular intervals, meaning that sampling is also not irregular (Blackburn et al., 2010). In this manner, MS^E provides more information than MS/MS as there is no selection or possible discrimination, and all peptide ions are analysed, giving higher coverage. Less sample is needed as the sample does not need to be run several times in order to collect data for all of the ions, as with MS/MS (Kramer et al., 2012). Limitations to MS^E are precursor ion intensity below the limit of detection (Blackburn et al., 2010), and the number of correctly assigned fragment ions to their corresponding precursors by the processing software (Bond et al., 2013). This mode of acquisition also generates very large amounts of data, often several gigabytes per raw file which must then be processed into a format suitable for database searching and protein

identification, and stored. The Waters Synapt G2-S mass spectrometer used in this study also has ion mobility separation (IMS) capability, referred to as HDMS^E (Pringle et al., 2007; Bond et al., 2013). IMS improves data quality through providing further separation in the gas phase, by separating ions according to their cross-sectional area (shape), producing specific drift times for each ion. For example, Shliaha *et al.* have reported more accurate fragment-to-precursor ion assignment when IMS is applied before fragmentation, leading to 60% higher proteome coverage (Shliaha et al., 2013). IMS can be performed before or after fragmentation. Comparison of these different methods of data acquisition indicates that the UPLC-HDMS^E method employing the data-independent acquisition approach would produce the optimal results for the global profiling of cellular proteomes in this study.

1.7.5. Protein identification

To identify proteins, raw data is processed and then searched against protein sequence databases. During the first stage of processing MS^E data, attributes such as the *m/z*, the retention time, the IMS drift time and the peak apex are identified for all MS peaks in the data, before lock mass correction (lock mass is a compound of known *m/z* that is used as an internal mass standard, to counteract drifts in calibration of the mass spectrometer (Cox et al., 2011). This generates a list of all precursor and fragment ions. After deconvolution of isotopes and charge states, a time alignment algorithm is applied to match fragment ions back to their parent ions by parent ion apex retention time. This creates a list of accurate mass retention time (AMRT) precursor/product ion pairs that is used for database searching and protein identification. As mentioned previously, MS^E data are searched using an Ion Accounting search algorithm (Waters) (Blackburn et al., 2010). It is an iterative process and each iteration is more selective, sensitive and specific to achieve confident identifications (Li et al., 2009). False positives are identified using a decoy database that contains both positive and reverse sequences. Peptides only identified in the reversed database are assumed to be false positive results (Graham et al., 2005). Fixed and variable modifications of peptides can be included in the search.

The success of protein identification is dependent on factors such as protein coverage and mass accuracy (Graham et al., 2005). Coverage is dependent on the sensitivity, sequencing speed and dynamic range of the mass spectrometer (de Godoy et al., 2006), as well as sample complexity. Data analysis itself is often the limiting factor in MS.

For a comprehensive comparison of the proteomes of two or more conditions, the data produced by mass spectrometry is much more meaningful if the proteins can be quantified as well as identified. MS is not inherently quantitative as ionisation efficiency can differ

between peptides. However, current technology enables proteins to be quantified relatively within a sample, and more recently in estimates of absolute amounts, which provides the ability to compare protein expression levels across multiple samples.

1.7.6. Relative quantification of proteins

There are a number of methods that provide relative quantification of proteins as well as their identification. They are commonly used in shotgun experiments (Patel et al., 2009). They often involve labelling, although more recently, label-free methods are being employed. Some of the most commonly used techniques for relative quantification will be discussed here: iTRAQ, SILAC and label-free quantification.

iTRAQ (isobaric tags for relative and absolute quantification) involves tagging primary amine groups of peptides with up to 8 mass tags that have the same molecular weight but fragment differently during MS/MS analysis (Graham et al., 2005). Fragmented tags are reporters of different mass. The intensity of each reporter indicates the relative abundance of that peptide/protein in the sample (Walther and Mann, 2010; Ross et al., 2004). The chemical labelling is efficient and data analysis is relatively simple, however this method also requires high fragmentation efficiency and comes with a high cost (Nikolov, 2012).

SILAC (Stable Isotope Labelling with Amino Acids in Cell Culture) is a popular method for differential labelling of proteins. Cells are grown in either normal or SILAC medium which contains "heavy" arginine and lysine amino acids labelled with ^{13}C and/or ^{15}N, which become fully incorporated by the cellular proteins, and the samples are mixed in a 1:1 ratio before LC-MS/MS analysis (de Godoy et al., 2006; Walther and Mann, 2010). Digestion with trypsin, which cleaves proteins at the C-terminus of arginine/lysine residues, will result in SILAC peptides having the heavy amino acid at their C-terminus. The mixture will then contain pairs of peptides that can be recognised by having the exact difference in mass between the normal and heavy amino acid, referred to as SILAC peptide pairs. Comparing peak intensities will indicate the relative abundance of each (Walther and Mann, 2010). SILAC offers efficient labelling, but heavy-labelled SILAC media has a high cost (Nikolov, 2012).

Due to limitations with labelling strategies such as complex sample preparation, cases of incomplete labelling and high cost, a current focus is on label-free methods. These can be based on comparing precursor intensities or spectral counts between samples. One study by Patel *et al.* used an LC-MS^E method of label-free quantification using the addition of a known concentration of an internal standard protein digest to compare different sample runs for comparison with 1-dimensional PAGE and iTRAQ labelling methods. Relative

quantification of proteins analysed using the label-free method was carried out by using the peak area/intensity of the peptides of the internal standard as a normalisation tool to correct the peak area/intensity of the peptides in the sample, and comparing these measurements between sample peptides. They found that the label-free method confidently identified 421 proteins, compared to 178 using iTRAQ and 235 using 1D-PAGE. The label-free method also took the shortest amount of time to prepare and run, and the least amount of sample was needed to run the experiment. The main disadvantage of this method was the size of the data file (18 Gb compared to 1.2 Gb for both other methods) (Patel et al., 2009). A study by Evans *et al*. showed that as of 2011 label-free methods have overtaken labelling techniques such as iTRAQ and SILAC as the method of choice for quantitative proteomic analysis (Evans et al., 2012; Bond et al., 2013).

1.7.7. Absolute quantification of proteins

A limitation to relative quantification is that samples cannot be compared across independent batches. Absolute Quantification (AQ) allows the study of the relationship between proteins by direct comparison in one sample or across multiple samples, which can be simple or complex, and can help understanding of complex molecular networks (Silva et al., 2006).

The gold standard of measuring proteins in absolute amounts is to use a multiple reaction monitoring (MRM) technique, a variation of selected reaction monitoring (SRM). SRM is typically performed on a triple quadrupole (QQQ) instrument, with the first and third mass filters used to select the *m/z* value for a particular precursor ion and a specific fragment ion, respectively, and the second acting as a collision cell. Multiple precursor/fragment ions pairs are monitored, providing a specific retention time and intensity value for each. This non-scanning approach results in high sensitivity, and the narrow mass ranges for selection provide high selectivity (Lange et al., 2008). In an MRM assay, multiple SRM *m/z* pairs can be monitored at once within the same run. For absolute quantification, internal standard protein digests can be spiked into the sample that are identical to endogenous peptides in the sample and are isotopically labelled e.g. chemically, at a known concentration. Comparison of the concentration of these labelled peptides with their identical (apart from a shift in mass) endogenous counterparts in a sample is the most accurate method of quantification (Schmidt and Urlaub, 2012; Kuzyk et al., 2013).

An example of a labelled internal standard technique is Absolute SILAC, which uses the basis of the SILAC technique to accurately quantify proteins within a complex sample. Recombinant protein internal standards that are SILAC-labelled are added before digestion.

Again, these standards must be almost identical to the endogenous proteins under study. Proteins can be quantified in this manner to the attomole level (Hanke et al., 2008).

MRM assays are a targeted approach and require prior knowledge of the identity of analytes in a given sample, thus MRM is an appropriate method for studying a protein of interest. As it requires the design and synthesis of the labelled internal standard, it can be time consuming and costly (Schmidt and Urlaub, 2012).

A simple method to estimate the absolute quantification of proteins in a sample for global "shotgun" experiments is to spike a protein digest internal standard(s) into the sample to be analysed, at a known concentration. During data analysis, the MS signal response of the three most intense tryptic peptides of the internal standard is used to determine a universal signal response factor, which is compared to the MS signal response of the three most intense tryptic peptides of each protein in the sample, thereby quantifying those proteins in the sample (Silva et al., 2006). This is referred to as the Hi3 method of quantification (Doneanu et al., 2012). This internal standard can be unlabelled and does not require to be matched to endogenous proteins in a sample; rather it should be distinguishable from endogenous peptides so that it can be set apart from them during data analysis, such as by originating from a different species. Thus this method of absolute quantification, whilst being less accurate (such that the quantification should be treated as only an estimate of absolute abundance) is less costly and needs no prior knowledge of peptide composition of the sample for its use.

For proteomic experiments in this investigation, the Hi3 method has been used to provide estimates of absolute quantification. It is hoped that a high number of protein identifications can be achieved for each sample, and that both their presence ("presence" as being above levels of detection) and their levels of expression can be compared between conditions. If changes in protein expression in response to a change in a variable such as exposure to FK228 or an exogenous signalling molecule can be observed, these proteins could have potential for use as a biomarker.

1.8. Aims

As discussed above, idiopathic pulmonary fibrosis is a lethal respiratory disease, likely resulting from alveolar epithelial cell injury and aberrant wound healing involving increased myofibroblast formation, in response to unknown stimuli. Furthermore, there are no therapeutic options for this disease that improve survival. *In vitro* and *in vivo* evidence

suggests that the HDAC inhibitor FK228, currently a treatment for T cell lymphoma, could be an effective treatment for IPF, and should be taken forward to clinical trial. Protein biomarkers are a useful tool in clinical trials for evaluating drug efficacy and monitoring for expected effects, therefore the identification of novel biomarkers of IPF and of response to FK228 would accelerate the full evaluation of this drug and may also further the current understanding of IPF pathogenesis.

The hypotheses to be tested in this study are as follows:

- The proteome of the fibroblast is modified during differentiation, as well as the profile of secreted proteins, and the changes in protein expression provide insights into the biological processes and signalling pathways that may become dysregulated in fibrotic disease

- The proteome of primary fibroblasts from IPF patient biopsy cultures can be comprehensively and quantitatively analysed by mass spectrometry, and that the proteomic profile of an IPF fibroblast differs from that of a non-IPF fibroblast, allowing for the identification of differentially expressed proteins and biological processes

- The cellular response of IPF fibroblasts to treatment with the histone deacetylase inhibitor FK228 can be analysed using a proteomic approach, in more than one type of cell culture system, to identify proteins that have modulated expression following treatment

- IPF fibroblasts grown in 3D culture have different proteomic profiles to those grown in 2D monolayer culture, reflective of their responses to differences in their cellular environment between culture systems, such as the stiffness of the culture plastic and the deposition of extracellular matrix

- A method to study the acetylome (acetylated protein profile) of IPF fibroblasts with and without FK228 treatment by mass spectrometry can be developed, which would lead to the identification of off-target effects of this HDAC inhibitor and thus novel drug targets

2. Materials and Methods

2.1. Materials

2.1.1. Reagents

Reagent	Manufacturer	Code (if available)
Acetic acid, glacial	Fisher Scientific (Leicestershire, UK)	A/0360/PB17
Acetonitrile, HPLC gradient grade	Fisher Scientific (Leicestershire, UK)	A/0627/17
Acetyl lysine antibody, agarose	Immunechem (British Columbia, Canada)	ICP0388
Acetyl lysine antibody, rabbit	Immunechem (British Columbia, Canada)	ICP0380
Acetyl-histone H3 (Ac-Lys^{9}), rabbit	Merck Millipore (Hertfordshire, UK)	06-599
Actin α-smooth muscle, mouse	Sigma-Aldrich (Gillingham, UK)	A5228
Adenine	Sigma-Aldrich (Gillingham, UK)	A2786
Ammonium bicarbonate	Sigma-Aldrich (Gillingham, UK)	A6141
Bovine serum albumin	Sigma-Aldrich (Gillingham, UK)	A4503
CAPS (3-[cyclohexylamino]-1 propane sulfonic acid)	Sigma-Aldrich (Gillingham, UK)	C-2632
CellTiter 96® AQueous Non-Radioactive Cell Proliferation Assay	Promega UK (Southampton, UK)	G5421
Chloroform (for chloroform/ethanol precipitation)	Sigma-Aldrich (Gillingham, UK)	Z2432
Chloroform (for methanol/chloroform precipitation)	Fisher Scientific (Leicestershire, UK)	C/4920/17
Dimethyl sulfoxide	Sigma-Aldrich (Gillingham, UK)	472301

(DMSO)		
Disodium tetraborate	Sigma-Aldrich (Gillingham, UK)	S9640
DL-Dithiothreitol	Sigma-Aldrich (Gillingham, UK)	D9779
DNase kit	Life Technologies (Paisley, UK)	AM1906
Dulbecco's Modified Eagle Medium, phenol red-free	Life Technologies (Paisley, UK)	31053028
Dulbecco's Modified Eagle Medium, phenol red	Life Technologies (Paisley, UK)	11960-044
Epidermal Growth Factor (EGF)	Sigma-Aldrich (Gillingham, UK)	E9644
Ethanol	Sigma-Aldrich (Gillingham, UK)	32221
F-12	Life Technologies (Paisley, UK)	21765-037
FBLIM-1 antibody, mouse	Santa-Cruz Biotechnology (Germany)	sc-271417
FK228/Romidepsin	Selleckchem (Suffolk, UK)	S3020
Foetal bovine serum	Life Technologies (Paisley, UK)	10500-064
Formic acid, 0.1% in water	Fisher Scientific (Leicestershire, UK)	LS118-212
Fungizone	Life Technologies (Paisley, UK)	15290026
Glycerol	Sigma-Aldrich (Gillingham, UK)	G6279
Glycine	Thermo Fisher Scientific (Leicestershire, UK)	220910010
Glycogen solution	Roche (Burgess Hill, West Sussex, UK)	10901393001
Goat anti-Mouse IgG (H+L) Cross Adsorbed Secondary Antibody, DyLight 800 conjugate	Thermo Fisher Scientific (Leicestershire, UK)	SA5-10176
Goat anti-Rabbit IgG (H+L) Secondary Antibody, DyLight 680 conjugate	Thermo Fisher Scientific (Leicestershire, UK)	35568
H3 Pan antibody, rabbit	Merck Millipore (Hertfordshire, UK)	07-670
Hanks' Balanced Salt solution without Ca2+ and Mg2+ (HBSS -/-)	Life Technologies (Paisley, UK)	14170-088
HEPES	Sigma-Aldrich (Gillingham, UK)	H3375/H4034
Hi3 *E.coli* standard	Waters (Manchester, UK)	186006012
Hydrochloric acid	Sigma-Aldrich (Gillingham, UK)	7102

Hydrocortisone	Sigma-Aldrich (Gillingham, UK)	H0888
Insulin	Sigma-Aldrich (Gillingham, UK)	I9278
Iodoacetamide	Sigma-Aldrich (Gillingham, UK)	I6125
IPG buffer, pH 3-10	GE Healthcare (Buckinghamshire, UK)	17-6000-87
Isopropanol (propan-2-ol)	Sigma-Aldrich (Gillingham, UK)	I9516
L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate	Sigma-Aldrich (Gillingham, UK)	A8960
L-glutamine	Life Technologies (Paisley, UK)	21051-040
Lys C endoproteinase, MS grade	Thermo Fisher Scientific (Leicestershire, UK)	90051
MASSPREP Enolase digestion standard	Waters (Manchester, UK)	186002325
Methanol, analytical grade	Fisher Scientific (Leicestershire, UK)	M/4000/17
Methylene blue	Sigma-Aldrich (Gillingham, UK)	M9140
MMP-2 antibody, rabbit	Cell Signaling (Hertfordshire, UK)	4022
MOPS	Sigma-Aldrich (Gillingham, UK)	M9381
Non-essential amino acids	Life Technologies (Paisley, UK)	11140035
NuPAGE® LDS sample buffer (4x)	Life Technologies (Paisley, UK)	NP0007
NuPAGE® MOPS SDS running buffer (20x)	Life Technologies (Paisley, UK)	NP0001
Paraformaldehyde	TAAB Laboratories Equipment Ltd (Berkshire, UK)	P001
Penicillin-Streptomycin	Life Technologies (Paisley, UK)	15070-063 5k
Potassium dihydrogen phosphate	Fisher Scientific (Leicestershire, UK)	P/4800/53
Potassium chloride	Fisher Scientific (Leicestershire, UK)	P/4280/53
Precision Plus Protein™ All Blue Standards	Biorad (Hemel Hempstead, UK)	161-0373
Precision reverse transcription kit	PrimerDesign (Southampton, UK)	RT-std
Propidium iodide stain	Sigma-Aldrich (Gillingham, UK)	P4170
qPCR primer ACTA2	PrimerDesign (Southampton, UK)	PP-HU-900-ACTA2

qPCR primers Cy5/FAM	PrimerDesign (Southampton, UK)	HK-PP-HU-900
Quantitative polymerase chain reaction (qPCR) mastermix	PrimerDesign (Southampton, UK)	PrecisionPLUS-iC
Recombinant Human TGF-β1 (HEK293 derived)	Peprotech (London, UK)	100-21
RNase A	Sigma-Aldrich (Gillingham, UK)	R4875
Semaphorin-7a antibody, rabbit	Abcam (Cambridge, UK)	Ab133803
Sequencing grade modified trypsin	Promega UK (Southampton, UK)	V5111
Sodium azide	Sigma-Aldrich (Gillingham, UK)	S2002
Sodium chloride	Sigma-Aldrich (Gillingham, UK)	S7653
Sodium dihydrogen phosphate	BDH Lab supplies (UK)	10245
Sodium dodecyl sulfate	Sigma-Aldrich (Gillingham, UK)	L3771
Sodium hydroxide	Fisher Scientific (Leicestershire, UK)	S/4880/60
Sodium phosphate	Sigma-Aldrich (Gillingham, UK)	342483
Sodium pyruvate	Life Technologies (Paisley, UK)	11360039
Thiourea	Sigma-Aldrich (Gillingham, UK)	T7875
Triethylammonium bicarbonate buffer	Sigma-Aldrich (Gillingham, UK)	T7408
Trifluoroacetic acid	Sigma-Aldrich (Gillingham, UK)	299537
Trizma® (Tris-base)	Sigma-Aldrich (Gillingham, UK)	T1503
TRIzol® lysis reagent	Life Technologies (Paisley, UK)	15596-026
Trypan blue stain	Sigma-Aldrich (Gillingham, UK)	T8154
Trypsin-ethylenediaminetetraacetic acid (EDTA)	Life Technologies (Paisley, UK)	15400054
Trypsin, TPCK treated	Sigma-Aldrich (Gillingham, UK)	T1426
Tween® 20	Sigma-Aldrich (Gillingham, UK)	P1379
Urea	Sigma-Aldrich (Gillingham, UK)	U5128
β-Actin antibody, rabbit	Cell Signaling (Hertfordshire, UK)	4970S
β-catenin antibody, rabbit	Cell Signaling (Hertfordshire, UK)	8480P

2.1.2. Reagent/buffer compositions

Lysis buffer:	0.1 M triethylammonium bicarbonate + 0.1% sodium dodecyl sulfate
Urea lysis buffer:	9 M urea 20 mM HEPES, pH 8.0 with NaOH
PBS:	80.06 g NaCl 2.013 g KCl 11.5 g Na_2HPO_4 2.04 g KH_2PO_4 pH to 7.2-7.4 if necessary
PBS-T:	As above, with the addition of: 0.5% Tween® 20
TBS:	50 mM Tris pH 7.5 with HCl 150 mM NaCl
TBS-T:	As above, with the addition of: 0.5% Tween® 20
Electrophoresis buffer:	20 ml 20x MOPS SDS running buffer 500 ml dH2O
Continuous CAPS (3-[cyclohexylamino]-1 propane sulfonic acid) buffer (10x):	22 g CAPS/1 L dH2O pH to 10.5-11.0 with NaOH
Continuous CAPS buffer (Working):	100 ml CAPS (10x) 800 ml dH2O 100 ml methanol
OFFGEL stock solution (1.25x):	600 µl IPG buffer, pH 3–10 6 ml glycerol Make up to 50 ml with dH2O

Mild stripping buffer (Abcam)	15 g glycine
1L:	1 g SDS
	10 ml Tween® 20
	Adjust pH to 2.2, bring volume to 1 L with dH2O
Protein A	50 mM Tris-HCl, pH 7.5
Immunoprecipitation kit	150 mM NaCl
Wash buffer 1:	1% Nonidet P40; 0.5% sodium deoxycholate
	1 complete protease inhibitor cocktail tablet
Wash buffer 2:	50 mM Tris-HCl, pH 7.5
	500 mM NaCl
	0.1% Nonidet P40; 0.05% sodium deoxycholate
Wash buffer 3:	10 mM Tris-HCl, pH 7.5
	0.1% Nonidet P40; 0.05% sodium deoxycholate
Immune-affinity precipitation	50 mM MOPS
(IAP) buffer:	10 mM Na_3PO_4
	50 mM NaCl
Formal Saline:	4% (v/v) formaldehyde in 0.9% (v/v) saline solution
	9 g NaCl
	100 ml 40% formaldehyde
	900 ml dH2O
10mM Borate Buffer	3.82 g disodium tetraborate
(pH 8.5):	(made up to 800 ml with dH2O)
	Adjust final volume to 1 L with dH2O
1:1 Ethanol – HCl:	200 ml EtOH
	200 ml 0.1 M HCl
1% (w/v) Methylene Blue	5g methylene blue
in 10mM Borate Buffer	Made up to 500 ml with 10 mM borate buffer
(pH 8.5):	Filtered

2.1.3. Equipment

Equipment	Manufacturer
20cc reservoir	BD (Ireland)
BD Cell Quest™	BD Biosciences (Oxford, UK)
BD FACS Calibur™	BD Biosciences (Oxford, UK)
Ultracentrifuge Avanti J-20 XPI	Beckman Coulter (High Wycombe, UK)
C18 SpinTips Sample Prep Kit	Protea (WV, USA)
Concentrator 5301	Eppendorf (Stevenage, UK)
Cryogenic vial, 1ml (Sterile), E3110-6122	STARLAB UK (Milton Keynes, UK)
Cryostorage system, K series	Taylor Wharton (USA)
Cytoscape 2.8	Cytoscape Consortium (San Diego, USA)
DirectDetect™	Merck Millipore (Hertfordshire, UK)
Dri-Block® D3-3A (Techne®)	Bibby Scientific Limited (Staffordshire, UK)
Empore™ SPE 96-well (66875-U)	Sigma-Aldrich (Gillingham, UK)
Filter paper (Thin Blot Paper) (1620118)	Biorad (Hemel Hempstead, UK)
GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool)	Israel
Heraeus™ Fresco™ centrifuge	Thermo Scientific (Leicestershire, UK)
iCycler CFX96 Thermal Cycler	Biorad (Hemel Hempstead, UK)
Immobiline DryStrips, 13cm, pH 3-10 (17-6001-14)	GE Healthcare (Buckinghamshire, UK)
Improved Neubauer bright-line haemocytometer	Marienfeld (Lauda-Königshofen, Germany)
Labsystems Multiskan Ascent plate reader and software	Thermo Fisher Scientific (Leicestershire, UK)
Leica DMI6000B light microscope and software	Leica Microsystems GmbH (Wetzlar, Germany)
MassLynx	Waters (Manchester, UK)
Metacore™	Thomson Reuters (New York, NY USA)
MicroModulyo®-230 freeze dryer	Thermo Scientific (Leicestershire, UK)
Microseal® B Seal seals	Biorad (Hemel Hempstead, UK)
Microsoft Office Excel	Microsoft (USA)
Millicell hanging cell culture insert, 1.0 µm (PIRP12R48)	Merck Millipore (Hertfordshire, UK)

Mini-incubator	Biorad (Hemel Hempstead, UK)
Mini-PROTEAN®II cell	Biorad (Hemel Hempstead, UK)
Multiplate™ PCR plates	Biorad (Hemel Hempstead, UK)
Nalgene® Mr Frosty® Cryo 1°C Freezing Container	Thermo Scientific (Leicestershire, UK)
NanoACQUITY UPLC system	Waters (Manchester, UK)
Nanodrop ND-1000 spectrophotometer	Nanodrop (Wilmington, DE, USA)
Nunc™ Cell-Culture Treated Flasks T175	Thermo Fisher Scientific (Leicestershire, UK)
Nunc™ Cell-Culture Treated Flasks T75	Thermo Fisher Scientific (Leicestershire, UK)
Nunc™ Cell-Culture Treated Multidishes (6 well)	Thermo Fisher Scientific (Leicestershire, UK)
Nunc™ Cell-Culture Treated Multidishes (12 well)	Thermo Fisher Scientific (Leicestershire, UK)
Nunc™ Cell Scrapers (32 cm) (179707)	Thermo Fisher Scientific (Leicestershire, UK)
Nunc™ MicroWell™ 96-Well Microplates	Thermo Fisher Scientific (Leicestershire, UK)
NuPAGE Power Ease 500	Life Technologies (Paisley, UK)
NuPAGE® Novex® 4-12% Bis-Tris Protein Gels, 1.0 mm, 12 well	Life Technologies (Paisley, UK)
Odyssey CLx Infrared imaging system/software	Licor (Cambridge, UK)
OFFGEL fractionator	Agilent Technologies (Cheshire, UK)
PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System	Thomas lab (University of Southern California, CA, USA)
Parafilm "M" roll	Sigma-Aldrich (Gillingham, UK)
Power Pac 300	Biorad (Hemel Hempstead, UK)
Protein A Immunoprecipitation kit (Roche)	Sigma-Aldrich (Gillingham, UK)
Protein Lynx Global Server 3.0	Waters (Manchester, UK)
PVDF membrane	Biorad (Hemel Hempstead, UK)
Reverse phase analytical column (NanoACQUITY UPLC HSS T3, C18, 1.8 µm, 75 µm x 250mm)	Waters (Manchester, UK)

Reverse phase trap column (ACQUITY UPLC Symmetry C18, 5 µm, 180 µm x 20mm)	Waters (Manchester, UK)
REVIGO (Reduce and Visualise Gene Ontology)	Rudjer Boskovic Institute (Croatia)
Savant™ SPD111V SpeedVac™	Thermo Scientific (Leicestershire, UK)
Sep-Pak C18 Plus Long Cartridge (WAT023635)	Waters (Manchester, UK)
Sonicator probe	Misonix Incorporated (Farmingdale, NY, USA)
Sorvall® Legend® RT centrifuge	Thermo Scientific (Leicestershire, UK)
Strataclean resin	Agilent Technologies (Cheshire, UK)
STRING 9.1 (Search Tool for the Retrieval of Interacting Genes/Proteins)	CPR, EMBL, SIB, KU, TUD and UZH
Stuart®rotator SB3	Bibby Scientific Limited (Staffordshire, UK)
Synapt G2-S	Waters (Manchester, UK)
T100 Thermal Cycler	Biorad (Hemel Hempstead, UK)
Thermomixer comfort	Eppendorf (Stevenage, UK)
Threshold Inspector	Waters (Manchester, UK)
Ultrasonic bath S15H heated, 1.75 L (FB15049)	Fisher Scientific (Leicestershire, UK)

2.2. Cell culture

2.2.1. Establishment of fibroblast cultures from cryogenic storage

Complete fibroblast medium (Dulbecco's modified Eagle medium (DMEM) supplemented with 1% penicillin/streptomycin (50 IU/ml, 50 µg/ml), 1% sodium pyruvate (1 mM), 1% non-essential amino acids (1 mM), 1% L-glutamine (2 mM) and 10% foetal bovine serum (FBS) (all Life Technologies)) was pre-warmed to room temperature and 9 ml was added to a 15 ml tube. 1 ml medium from the tube was pipetted up and down into a cryovial containing approximately $1x10^6$ frozen fibroblasts to defrost the cell pellet. During thawing, any thawed cells were removed to the tube and 1 ml fresh medium from the tube was used to further thaw the pellet. This was repeated until all cells were thawed. After centrifugation at 300 xg for 5 min, room temperature (RT), cells were resuspended in 1 ml fresh medium and transferred to a 75 cm^2 culture flask (Thermo Fisher Scientific) containing 10 ml medium, and incubated overnight at 37ºC, 5% CO_2. After 24 h medium was replaced to remove any non-adherent cells.

2.2.2. Passaging and counting of fibroblasts

Fibroblast cultures were incubated in 75 cm^2 flasks in 11 ml complete fibroblast medium, at 37°C with 5% CO_2. Medium was replaced every 2-3 days, and cells were passaged after 4-5 days. All culture reagents were pre-warmed prior to use to avoid cellular stress.

For passaging, medium was removed and cells were washed twice with 6 ml Hank's Balanced Salt Solution (HBSS, Life Technologies) and treated with 1 ml 0.05% trypsin-EDTA (Life Technologies) in HBSS at 37°C, 5% CO_2 for 5-10 min. 10 ml medium was added to neutralise the trypsin and cells were centrifuged at 300 xg for 5 min, RT. The cell pellet was resuspended in 1 ml medium. For counting, 50 µl of the cell suspension was added to 50 µl trypan blue stain (Sigma), and 10 µl was pipetted into a chamber of an Improved Neubauer bright-line hemocytometer (Marienfeld) for counting using a light microscope (Leica). Cells in the four 1 mm^2 corners were counted and the average number was recorded. The number of cells per ml were calculated as follows:

Cells per ml = average cell count x 2 (dilution factor) x 10^4 (chamber volume)

Fibroblasts were then transferred to new 75 cm^2 flasks in 11 ml complete fibroblast medium and incubated at 37°C with 5% CO_2.

2.2.3. Isolation of primary fibroblasts from lung biopsy

Surgical lung biopsies obtained with ethical approval from five patients with a confirmed diagnosis of IPF (which will be referred to as IPF patients 1-5 from hereon) or tissue biopsies obtained with ethical approval from three non-IPF donors were used to establish fibroblast cultures. Lung biopsy samples were placed in a petri dish; any staples present in the sample were removed and the sample was dissected into approximately 2x2 mm sections using autoclaved sterile forceps and a scalpel. Pieces of tissue were then transferred to wells of a 6-well culture plate (Thermo Fisher Scientific) containing complete DMEM, 10% FBS (2 ml/well) and were scratched into the bottom of the wells using a scalpel. Medium was changed after 7 days, then replaced every 2-3 days for approximately 2 weeks. Outgrown fibroblasts were dissociated from the wells and placed into a 75 cm^2 tissue culture flask. At 70-80% confluence, cells were collected, counted and split into several 75 cm^2 culture flasks containing 10 ml complete medium to increase cell number. Once confluent, fibroblasts were collected, counted and put into culture or frozen for cryogenic storage (see sections **2.2.2** and **2.2.4**). Lung biopsies were obtained by clinicians at Southampton General hospital and the Royal Brompton hospital. Fibroblasts were isolated from biopsy by colleagues from the Brooke Laboratory, Faculty of Medicine, Southampton University Hospital, University of Southampton.

2.2.4. Cryogenic storage of fibroblasts

For cryopreservation, fibroblasts were washed in HBSS and dissociated from flasks using trypsin-EDTA in HBSS according to the protocol for cell passaging (section **2.2.2**). After cell detachment, resuspension in medium and cell counting, fibroblasts were centrifuged for 5 min at 300 xg, RT. Medium was removed and cells were resuspended in freezing medium (90% DMEM (10% FBS), 10% DMSO (Sigma)), $1x10^6$ cells/ml. Cells were aliquoted into cryovials (STARLAB UK), 1 ml cell suspension/vial. The vials were stored in a Nalgene cryo-container (Thermo Scientific) containing isopropyl alcohol at -80ºC overnight, and then moved to a liquid nitrogen vessel (Taylor Wharton) after 24 h.

2.3. Methylene blue assay following TGF-β_1 challenge of MRC-5 fibroblasts

100 µl complete medium (10% FBS) was added to 6 x 8 wells of two 96-well culture plates (Thermo Fisher Scientific). Cells were seeded into all 8 wells in the first columns, at $1.2x10^4$

cells per well in 100 µl medium. A 1 in 2 serial dilution was performed across the 6 columns of the plates. 4 rows were designated “control” and 4 rows were designated “treatment” to give 4 rows of each condition.

After 24 h, all cells in one plate were fixed at T0. To fix, medium was removed from each well and cells were washed in 100 µl HBSS. 100 µl 4% (v/v) formaldehyde (TAAB Laboratories Equipment Ltd) in 0.9% (v/v) saline solution (see section **2.1.2**) was added to the cells for 30 min, RT. This was removed and cells were washed in 100 µl PBS. 100 µl 0.05% sodium azide (Sigma) in PBS was added to each well and the plate was wrapped in Parafilm M roll (Sigma) and foil and stored at 4°C. Half of the cells in the second plate were treated with 10 ng/ml TGF-β_1 (Peprotech) in fresh medium; fresh medium was added to control cells.

After 48 h of treatment, cells in the second plate were fixed as previously. After the addition of 100 µl PBS for washing, the T0 plate was removed from storage and both plates were emptied and blotted until dry. Cells in both plates were stained using 100 µl 1% (w/v) methylene blue (Sigma) in 10 mM borate buffer, pH 8.5 (see section **2.1.2**), per well. After staining for 30 min, RT, plates were emptied and washed with water, and blotted until dry. 100 µl 1:1 ethanol/HCl (see section **2.1.2**) was added to each plate and absorbance was measured at 630 nM using a Labsystems Multiskan Ascent plate reader and software.

2.4. MTS assay following FK228 challenge of primary cells

IPF and non-IPF fibroblasts were plated at a density of $1.5x10^3$ cells in 100 µl complete medium (10% FBS) per well, in a 96-well culture plate. At ~60% confluence cells were treated with 100 µl FK228 in DMEM at a range of concentrations, as follows: 0 nM (DMSO vehicle control), 0.1 nM, 1 nM, 5 nM, 10 nM and 50 nM. At 48 h/144 h post-treatment, 20 µl combined MTS/PMS solution (CellTiter 96® AQ_{ueous} Non-Radioactive Cell Proliferation Assay reagents, Promega) was pipetted into each well of the 96-well plate. The plate was incubated for 3 h at 37°C, 5% CO_2. The absorbance was then recorded at 490 nm using a Labsystems Multiskan Ascent plate reader and software.

2.5. Cell cycle analysis by flow cytometry following FK228 challenge

Primary IPF and non-IPF fibroblasts were seeded into a 6-well plate each at a density of $7.5x10^4$ cells in 2 ml complete medium (10% FBS) per well. At ~60% confluence original medium was removed and medium containing 2 nM FK228 was added to 2 wells per plate, 1 nM was added to 2 wells per plate, and the highest equivalent volume of DMSO was added to the control wells. Cells were washed twice with 1 ml HBSS and harvested using 300 µl 0.05% trypsin-EDTA in HBSS 48 h post-treatment; after 5 min incubation at 37°C, 5% CO_2, 2 ml medium was added to neutralise the trypsin and cells were centrifuged at 300 xg for 5 min, RT. The cell pellets were resuspended in 1 ml PBS.

Cells of each condition were pooled together and counted in trypan blue stain using a hemocytometer, as outlined in **2.2.2**. Samples were centrifuged at 300 xg for 5 min at 4°C. Pellets were stored in 1 ml ice-cold 70% ethanol at -20°C until further use.

Samples were centrifuged at 5000 xg for 2 min at 4°C. Pellets were resuspended in 400 µl PBS and transferred to a FACS tube, and centrifuged for 5 min at 2500 xg. A master mix of PBS, 5 mg/ml propidium iodide stain (Sigma) and 4 mg/ml RNase A (Sigma) was added to each sample, and samples were incubated at 37°C, 5% CO_2 for 30 min.

Cells were analysed for their DNA content using a FACS Calibur flow cytometer and CellQuest™ software (BD Biosciences) to determine the phase of the cell cycle they had reached upon harvesting, under the supervision of Dr Franco Conforti (Brooke Laboratory, University of Southampton). As fibroblasts are typically adherent cells, doublet discrimination was used to eliminate aggregates of cells, so that two cells in G1 phase clumped together would not count as one cell in G2 phase.

2.6. Challenge of fibroblasts with transforming growth factor beta-1 or FK228 for mRNA/protein expression analysis

2.6.1. Challenge of MRC-5 cells with transforming growth factor beta-1

MRC-5 fibroblasts (passage 25) were seeded into 6-well plates at $7.5x10^4$ cells/2 ml medium per well. At ~60% confluence, fibroblasts were treated in the presence or absence of 10 ng/ml TGF-β_1 (Peprotech) for 48 h. For the secretome experiment, upon treatment, FBS was omitted from the culture medium. At 48 h medium was collected, centrifuged at 300 xg for 3 min, RT to remove dead cells and debris, and supernatants were stored at -80°C until further use. For protein-based experiments, cells were washed twice with 1 ml HBSS and harvested using 300 µl 0.05% trypsin-EDTA in HBSS. After 5 min incubation at 37°C, 5% CO_2 the trypsin was neutralised with 2 ml culture medium and the cell suspension centrifuged at 300 xg for 5 min, RT. The cell pellet was resuspended in 1 ml phosphate buffered saline (PBS) and the cell suspension was counted in trypan blue as outlined in **2.2.2.** Cells were washed a further 3 times in 1 ml PBS with centrifugation at 300 xg for 5 min, RT between washes, before storage at -80°C. For the gene expression experiment cells were washed 1x in 1 ml HBSS and harvested in 500 µl TRIzol® lysis reagent (Life Technologies), and stored at -80°C.

2.6.2. Challenge of primary cells with FK228

2.6.2.1. IPF/non-IPF fibroblast monolayers

Five primary IPF and three primary non-IPF fibroblast cell cultures (all passage 5) were seeded into 6-well plates at $7.5x10^4$ cells/2ml medium per well. At ~60% confluence, fibroblast medium was removed and replaced with either 2 ml of 2 nM FK228 (Selleckchem) within fresh medium or medium containing the equivalent volume of DMSO. Cells were incubated for 48 h at 37°C, 5% CO_2. Medium was collected and centrifuged at 300 xg for 3 min, RT and supernatants were stored at -80°C. Cells were washed twice with 1 ml HBSS, before harvesting using 300 µl 0.05% trypsin-EDTA in HBSS. After 5 min of incubation at 37°C, 5% CO_2, trypsin was neutralised with 2 ml culture medium and cells were centrifuged at 300 xg for 5 min. Cells were resuspended in 1 ml PBS and counted using trypan blue stain

as outlined in **2.2.2.** Cell pellets were washed 4x total in PBS with centrifugation at 300 xg for 5 min between washes, before storage at -80°C.

2.6.2.2. IPF fibroblast 3D culture

Hanging cell culture inserts (Millipore) were added to all wells of 1x 24-well plate (Thermo Fisher Scientific) and 4 wells of another using sterile forceps. 900 µl DMEM (10% FBS), supplemented as outlined in **2.2.1**, was added to the bottom of each well of the plates containing inserts, and 200 µl added to each insert. At all times during cell culture, medium was removed and replaced in the hanging inserts using a pastette. 1 ml HBSS was added to all wells not containing inserts to prevent evaporation. Plates were incubated for 3 h at 37°C, 5% CO_2 to equilibrate the membranes. Medium was then removed and fibroblasts isolated from lung biopsy of an IPF patient (Patient 1), at passage 5, were seeded into the inserts at $1.5x10^4$ cells/insert in 200 µl medium. 900 µl medium was added to the bottom of each well. Medium was removed after 24 h and replaced with equivalent volumes of DMEM (10% FBS) supplemented with adenine (0.18 mM), HEPES (8 mM), insulin (5 µg/ml), hydrocortisone (0.5 µg/ml), EGF (10 ng/ml), F-12 (1 µl/ml, mixed at a 3:1 ratio) and 10 µg/ml L-ascorbic acid 2-phosphate; this medium mix was used throughout the rest of the culture period. Cells were incubated at 37°C with 5% CO_2. Medium was replaced every 3 days of culture.

After 3 weeks of culture fibroblast medium was removed and replaced with fresh medium containing 2 nM FK228 in 14/28 inserts and fresh medium containing the equivalent volume of DMSO in 14/28 inserts.

After 48 h, half of the untreated cells and FK228-treated cells were harvested. Medium was collected and centrifuged at 300 xg for 3 min, and supernatants were stored at -80°C. Fibroblasts were washed twice with 500 µl HBSS and harvested using 70 µl 0.05% trypsin-EDTA in HBSS in the inserts and 200 µl in the wells of the plate. After an incubation period at 37°C, 5% CO_2, the trypsin was neutralised with 200 µl culture medium in the inserts and 400 µl in the wells of the plate. Cells were centrifuged at 300 xg for 5 min. Cell pellets were washed 4x in 1 ml PBS with centrifugation at 300 xg for 5 min between washes, before storage at -80°C until required.

The remaining fibroblasts were cultured for a further 3 weeks, with chronic treatment every 3 days with fresh FK228-containing medium or vehicle. All cells were then harvested in the same manner as previously, and were stored at -80°C until further use.

2.6.3. Challenge of MRC-5 fibroblasts with FK228 prior to immunoprecipitation

MRC-5 fibroblasts (passage 29) were seeded into 6-well plates at $7.5x10^4$ cells/2 ml medium per well. At ~60% confluence, fibroblasts were treated in the presence or absence of 2 nM FK228 or vehicle control in fresh medium for 48 h. At 48 h cells were washed twice with 1 ml HBSS and harvested using 300 µl 0.05% trypsin-EDTA in HBSS. After an incubation period at 37°C, 5% CO_2, the trypsin was neutralised with 2 ml culture medium and centrifuged at 300 xg for 5 min. Cells were resuspended in 1 ml PBS and the cell suspension was counted in trypan blue as outlined in **2.2.2.** Cells were washed a further 3 times in 1 ml PBS before storage at -80°C.

2.7. Culture and harvesting of MRC-5 fibroblasts using the Li method for the analysis of acetylated peptides

Culture and harvesting of fibroblasts for immunoprecipitation was performed according to Li *et al*. (2013). MRC-5 cells (passage 27) were grown according to section **2.2.2** to produce approximately 10–20 mg of soluble protein (16x 175 cm^2 cell culture flasks total). At 80% confluence, cells were harvested using urea lysis buffer (20 mM HEPES pH 8.0, 9 M urea, Sigma). Cells in all flasks were washed with 5 ml cold PBS. 10 ml of urea lysis buffer was added to one flask. Cells were scraped into the lysis buffer, which was collected and added to the second flask; cells were again scraped into the buffer and collected and transferred to the next flask until cells in all 16 flasks had been scraped and collected into a 50 ml tube. The lysate was stored at -80°C until further use.

2.8. RNA extraction and real-time qPCR following TGF-β_1 challenge

Total RNA was isolated from MRC-5 fibroblasts using 500 µl TRIzol® according to manufacturer's instructions and stored at -80°C until further use. RNA was extracted by chloroform/ethanol precipitation. 1/5 volume of chloroform (Sigma) was added to TRIzol®/RNA and shaken vigorously for 15 seconds. Samples were incubated at room

temperature for 10 min, followed by centrifugation at 13,800 xg for 15 min at 4°C. Aqueous phase was transferred to a new tube and 0.5 µl glycogen (Roche) was added and vortexed. Ice cold isopropanol (Sigma) in equal volume to RNA aqueous fractions was mixed into each sample and samples were stored at -80°C until further use. Samples were centrifuged at 16,200 xg for 30 min, 4°C. Isopropanol was poured off and pellets were washed in 500 µl 75% ethanol (Sigma) and vortexed. Samples were centrifuged at 5,400 xg for 5 min 4°C, and ethanol was removed. Pellets were air-dried before DNAse treatment (20 µl DNAse treatment per sample: 1 µl DNAse, 2 µl 10x DNAse I buffer, 17 µl ultrapure MilliQ water; Life Technologies). Samples were incubated for 60 min at 37°C in a water bath. Neutralisation buffer was added (5 µl/sample), mixed and incubated for 2 min at room temperature. Samples were further mixed and incubated for 2 min. Samples were centrifuged at 16,200 xg and stored at -80°C.

RNA was quantified using a Nanodrop™ ND-1000 spectrophotometer at 260 nm absorbance, and additionally at 280 nm to determine protein contamination. The absorbance ratio at 260nm:280nm of 1.8-2 was defined as there being no significant contamination.

RNA was reverse transcribed into cDNA using a Precision Reverse Transcription kit in a T100 thermal cycler (Biorad) according to manufacturer's instructions, and samples were stored at -20°C. Polymerase chain reaction (PCR) was carried out with an iCycler CFX96 thermal cycler (Biorad) to determine expression levels of RNA for α-smooth muscle actin (ACTA2: Primer sequence sense: 5'-AAG CAC AGA GCA AAA GAG GAA T-3', anti-sense: 5'-ATG TCG TCC CAG TTG GTG AT-3', PrimerDesign). Master mixes were Precision Plus Perfect Probe. Housekeeping genes used were ubiquitin C and phospholipase A2 (UBC/A2) (PrimerDesign). All conditions of cycling were as specified by the manufacturer.

mRNA levels were calculated using the ΔΔCt method (Ct = cycling threshold). Briefly, a successful reaction is measured by fluorescence, and fluorescence increases with the number of cycles. When the fluorescent signal reaches a threshold the number of cycles taken to reach it is calculated for each sample. The geometric mean of the housekeeping genes Ct value of a sample is subtracted from the same treatment sample, resulting in a ΔCt value. The ΔCt value for a control sample is subtracted from the other samples to give the ΔΔCt value for all samples.

2.9. Sample preparation for mass spectrometry analysis

2.9.1. Whole cell lysates

Cell pellets were resuspended in 500 µl lysis buffer (0.1 M triethylammonium bicarbonate (TEAB), 0.1% SDS, both Sigma) and lysed by sonication using a microtip (6 pulses for 25 seconds, with 60-second rests in between, power set between 2-3, Misonix Incorporated) on ice. Cell lysate was centrifuged at 17,000 xg for 20 min at 4°C. Protein concentration of supernatant was measured via a DirectDetect® spectrometer (Millipore) according to manufacturer's instructions. Lysis of the IPF 3D culture model samples included an extra sonication step prior to using the sonicator probe; samples were first put in an ultrasonic bath (Fisher Scientific) for 10 min and then vortexed for 5 min until the pellet was no longer visible. Following protein concentration measurement, samples were pooled from 2 hanging cell culture inserts, approximately 100 µg protein.

Methanol/chloroform extraction was performed on 100 µg of each cell lysate. 4 volumes of methanol (Fisher Scientific), 1 volume of chloroform (Fisher Scientific) and 3 volumes of dH_2O were added sequentially to each sample, with ~2 seconds of vortexing after each addition. Samples were centrifuged at 17,000 xg for 1 minute to focus the proteins between organic and inorganic phases. Both phases were discarded as much as possible without disturbing the protein disc. 4 starting volumes of methanol were added and samples were vortexed for ~2 seconds, before centrifugation at 17,000 xg for 2 minutes. Methanol was removed and the pellet was left to air dry. Precipitated proteins were redissolved in 100 µl 6 M urea, 2 M thiourea, 10 mM HEPES (all Sigma) buffer, pH 7.5 and vortexed to completely solubilise the proteins. Proteins were then reduced with 1 mM Dithiothreitol (DTT, Sigma) for 60 min at RT, alkylated with 5.5 mM iodoacetamide (IAA, Sigma) for 45 min in the dark, and digested with endoproteinase LysC, MS grade (ThermoScientific Pierce, 1/50 (w/w)) for 4 h, RT. Samples were diluted 4 times with 20 mM ammonium bicarbonate (Sigma) and further digested with sequencing grade modified trypsin (Promega, 1/50 (w/w)) overnight at RT.

Waters enolase (*Saccharomyces cerevisiae*) standard and Waters Hi3 *Escherichia coli* standard were added to each sample for estimates of absolute quantification, at a concentration that would result in 150 fmol being injected per sample for mass spectrometry analysis.

Peptides from each sample were separated into 12 fractions according to their isoelectric point in OFFGEL stock solution containing IPG buffer, pH 3–10 (GE Healthcare) and glycerol (Sigma) (see section **2.1.2**), on IPG Strips, 13 cm, pH 3–10 (GE Healthcare) using an Agilent 3100 OFFGEL fractionator, according to manufacturer's instructions. Peptides were focused for 20 kVh.

Fractions were acidified to pH <3.0 using trifluoroacetic acid (TFA, Sigma) and loaded onto an Empore C18, 96-well solid phase extraction plate (Sigma) to remove residual salts, carrier ampholytes and glycerol from fractionation (Hubner *et al*., 2008). 150 µl methanol was added to each well to wet the silica, followed by 200 µl 80% acetonitrile + 0.5% acetic acid (both Fisher Scientific). The plate was centrifuged at 100 xg for 2 min before further equilibration of the silica with 150 µl 0.5% acetic acid, centrifugation at 100 xg for 2 min and sample addition. Samples were centrifuged at 250 xg for 1 min. Samples were washed twice with 200 µl 0.5% acetic acid with centrifugation at 250 xg for 2 min after each wash. Peptides were eluted from the plate with 150 µl 80% acetonitrile + 0.5% acetic acid and centrifuged at 100 xg for 2 min.

Samples were lyophilised at 35°C in a Savant SPD111V SpeedVac and resuspended in 10 µl (MRC-5 samples), 20 µl (IPF/non-IPF samples) or 8 µl (3D IPF culture model samples) of 98% dH_2O/acetonitrile (v/v) + 0.1% formic acid (Fisher Scientific). The samples were vortexed for 1 min and centrifuged at 100 xg for 1 min. 10 µl (MRC-5/IPF/non-IPF samples) or 4 µl (3D culture model samples) of each fraction was subjected to UPLC-$HDMS^E$ analysis.

2.9.2. Secreted proteins

Medium samples were centrifuged for 1 min, 17,000 xg and supernatants were transferred to new tubes. StrataClean resin (Agilent) was used to enrich for secreted proteins in the medium. Resin was vortexed and 50 µl was added to 1 ml medium. Samples were mixed for 1 h at 4°C on a rotator. After centrifugation for 3 min, 17,000 xg, supernatant was removed and stored at -20°C, and resin/protein complexes were resuspended in 300 µl 100 mM ammonium bicarbonate.

Assuming 25 µg protein per sample, proteins were reduced with 0.5 µg DTT for 1 h with mixing using an Eppendorf Thermomixer® set to 60°C at 1000 RPM. Proteins were alkylated with 2.5 µg IAA for 45 min in the dark, and digested with 0.5 µg sequencing grade modified trypsin (1/50 (w/w)) overnight at 37°C.

Samples were centrifuged at 17,000 xg for 3 min and peptide supernatant was transferred to new tubes. Resin was washed in 500 µl 100% acetonitrile, vortexed and centrifuged at

17,000 xg for 3 min. Supernatant was added to the previous supernatant yielding a final concentration of 60% acetonitrile. Samples were lyophilised at 35°C in a SpeedVac.

Peptides were reconstituted in 150 µl 3% acetonitrile + 0.1% formic acid and vortexed. Samples were acidified to pH <3.0 with TFA and loaded onto an Empore C18 96-well solid phase extraction plate to remove residual salts, buffers and contaminants, as outlined in **2.9.1**. Samples were concentrated to completion at 35°C and resuspended in 25 µl of 98% dH_2O/acetonitrile (v/v) + 0.1% formic acid containing 100 fmol enolase standard (Waters), before being vortexed for 1 min and centrifuged at 100 xg for 1 min. 2 µl was injected into the UPLC system for UPLC-HDMS^E analysis.

2.9.3. Acetylated peptides (Choudhary method)

Untreated MRC-5 fibroblasts harvested as described in **section 2.6.3** were used for this experiment. Cell pellets were resuspended in 500 µl lysis buffer (see section **2.1.2**) and lysed by sonication using a microtip (6 pulses for 25 seconds, with 60-second rests in between, power set between 2-3) on ice. Cell lysate was centrifuged at 17000 xg for 20 min at 4°C. Protein concentration of supernatant was measured via a DirectDetect® spectrometer.

Proteins were reduced with DTT (1/25 (w/w)) for 1 h at 60°C and alkylated with iodoacetamide (1/5 (w/w)) for 45 min in the dark, RT. Proteins were digested with sequencing grade modified trypsin (Promega, 1/50 (w/w)) overnight at 37°C.

Immunoprecipitation was performed using anti-acetyl lysine agarose (Immunechem) at a ratio of 50 µl agarose: 2 mg protein, according to manufacturer's instructions. Anti-acetyl lysine agarose was washed 3x with 1 ml PBS-T, 2x with 1 mL of 0.1 M NaH_2PO_4 (BDH Lab Supplies) + 1M NaCl (Sigma) and 1x with 1 ml PBS-T, and centrifuged at 1036 xg for 5 min between washes. Cell lysate was added to the beads and the mixture was incubated overnight at 4ºC on a rotator. After incubation beads were washed 4x in 1 ml PBS-T and eluted with 3x 50 µl of 1% TFA. Eluates were pooled and lyophilised at 35°C in an Eppendorf® Concentrator 5301.

Sample clean-up was performed using C18 SpinTips (Protea). Peptides were reconstituted in 200 µl of Rinse buffer (3% acetonitrile + 0.1% TFA) and gently vortexed. The resin of the SpinTip was wetted with 100 µl Equilibration buffer (50% acetonitrile + 0.1% TFA), and the SpinTip was centrifuged at 1000 xg for 2 min, 20°C. Resin was washed in 150 µl Rinse buffer and centrifuged at 1000 xg for 3 min, 20°C, and this was repeated twice for a total of three washes. The peptide sample was loaded into the SpinTip and centrifuged at 1000 xg for 4 min, 20°C. The flow-through was collected and stored at -20°C. The sample was

washed three times with 150 µl Rinse buffer with centrifugation at 1000 xg for 3 min, 20°C between washes. The sample was then eluted into a new tube with 75 µl Elution buffer (65% acetonitrile + 0.1% TFA) and centrifuged twice at 1000 xg for 3 min, 20°C in order for all of the eluate to pass through the resin. An additional step of adding 25 µl Elution buffer and centrifugation was used to fully elute the peptides. All eluate was pooled and concentrated to completion, and stored at -20°C.

Peptides were reconstituted in 10 µl of 98% dH_2O/acetonitrile (v/v) + 0.1% formic acid containing enolase protein digest standard (Waters) at a final concentration of 100 fmol. The sample was vortexed for 1 min, centrifuged at 100 xg for 1 min to collect sample to the bottom of the tube and transferred to an autosampler vial for mass spectrometry analysis. The whole 10 µl was applied for UPLC-HDMS^E analysis.

2.9.3.1. Acetylated peptides (Choudhary method, modification 1)

Untreated MRC-5 fibroblasts harvested as described in **section 2.6.3** were also used for this experiment. Fibroblasts were lysed by sonication on ice using a microtip in 800 µl lysis buffer (see section **2.1.2**) and proteins were reduced, alkylated and digested as described in section **2.9.3**. Following digestion, immunoprecipitation was carried out using a Protein A Immunoprecipitation Kit (Roche, Sigma). The peptide sample was pre-cleared using 20 µl Protein A-agarose to remove non-specific proteins. Beads were added to the sample and incubated for 3 h, 4°C on a rotator to mix. Samples were centrifuged at 12000 xg for 20 s and supernatant was transferred to a new tube. 1 µl of anti-acetyl-lysine antibody (Immunechem) was added to the supernatant and incubated overnight at 4°C with mixing. 50 µl Protein A-agarose was then added to the sample and incubated overnight at 4°C with mixing.

Beads were collected by centrifugation at 17,000 xg for 1 min and supernatant removed. 1 ml wash buffer 1 was added to the beads and the sample was incubated for 20 min, 4°C with mixing, before centrifugation at 17,000 xg for 1 min. This wash was repeated. Beads were washed twice with 1 ml wash buffer 2 in the same manner, and once with 1 ml wash buffer 3 (see section **2.1.2** for components of wash buffers 1, 2 and 3). After collection, peptides were eluted from the beads with 3x 300 µl 1% TFA, and eluates were pooled together. The sample was stored at 4°C. Peptides were resuspended in 2 ml OFFGEL stock solution (see section **2.1.2**) and separated into 12 fractions according to their isoelectric point using an Agilent 3100 OFFGEL fractionator and IPG Strips, 13 cm, pH 3–10 (GE Healthcare) according to manufacturer's instructions. Peptides were focused for 20 kVh before acidification to pH <3.0 with TFA and purification using an Empore C18 96-well solid

phase extraction plate (Sigma) as outlined in **section 2.9.1**. Fractions were lyophilised at 35°C in a SpeedVac and stored at -80°C. For mass spectrometry analysis, peptide fractions were resuspended in 5 µl of 98% dH2O/acetonitrile (v/v) + 0.1% formic acid before being vortexed for 1 min and centrifuged at 100 xg for 1 min. 4 µl was injected onto the LC column.

2.9.3.2. Acetylated peptides (Choudhary method, modification 2)

Fibroblast cell pellets were lysed and proteins were reduced, alkylated and digested, and immunoprecipitation was carried using the original method described in section **2.9.3**, with the exception of two extra washes of the anti-acetyl lysine agarose with 1 ml PBS-T and an extra elution step with 50 µl of 1% TFA before pooling of eluates. Samples were concentrated to completion at 35°C and stored at -20°C. Peptides from each sample were resuspended in 2 ml OFFGEL stock solution (see section **2.1.2**) and were separated into 12 fractions according to their isoelectric points using an Agilent 3100 OFFGEL fractionator on IPG Strips, 13 cm, pH 3–10 (GE Healthcare) according to manufacturer's instructions. Peptides were focused for 20 kVh. Fractions were acidified to pH <3.0 with TFA and purified using an Empore C18 96-well solid phase extraction plate (Sigma) as outlined in **section 2.9.1**. Fractions were dried to completion in a SpeedVac at 35°C and stored at 4°C, before resuspension in 5 µl of 98% dH2O/acetonitrile (v/v) + 0.1% formic acid. Samples were vortexed for 1 min and centrifuged at 100 xg for 1 min, and 4 µl was injected on column for UPLC-HDMS^{E} analysis.

2.9.3.3. Acetylated peptides (Choudhary method, modification 3)

As described in **section 2.6.3,** MRC-5 fibroblasts were stimulated with 2 nM FK228 or vehicle control for 48 hours before harvesting. Cells were lysed in 500 µl lysis buffer (see section **2.1.2**) on ice by sonication. Protein concentration was measured by DirectDetect™ (Millipore). Proteins were precipitated by methanol/chloroform extraction as described in **section 2.9.1**, and redissolved in 6 M urea, 2 M thiourea, 10 mM HEPES buffer, pH 7.5. Proteins were reduced with DTT (1/25 (w/w)) for 60 min at RT and alkylated with iodoacetamide (1/5 (w/w)) for 45 min in the dark. Samples were diluted 4 times with 20 mM ammonium bicarbonate and further digested with sequencing grade modified trypsin (1/50 (w/w)) overnight at RT.

Peptide samples were concentrated to completion at 35ºC and resuspended in 700 µl immune-affinity precipitation buffer (IAP) (50 mM MOPS pH 7.2 (Sigma), 10 mM sodium phosphate (Fisher Scientific), 50 mM sodium chloride (Sigma)). Peptides were then

immunoprecipitated using anti-acetyl lysine agarose (Immunechem), approximately 50 µl beads:1.5 mg protein. Agarose beads underwent several rounds of washing: 6x in 1 ml PBS-T, 2x in 1 ml 0.1 M NaH_2PO_4 + 1M NaCl and then 1x in 1 ml PBS-T, with centrifugation at 1036 xg for 5 min between washes. Peptides were added to the beads and incubated overnight at 4ºC on a rotator. Peptides were washed 5x in 500 µl IAP buffer, 3x in 500 µl dH_2O and eluted from the beads using 4x 50 µl 0.1% TFA. Fractions were concentrated to completion at 35°C before resuspension in 2 ml OFFGEL stock solution (see section **2.1.2**). Peptides were separated according to isoelectric point by OFFGEL fractionation into 12 peptide fractions according to manufacturer's instructions. Each fraction was purified using a C18 Empore 96-well solid phase extraction plate, as described in **section 2.9.1** and lyophilised at 35°C before resuspension in 5 µl 98% dH2O/acetonitrile (v/v) + 0.1% formic acid for mass spectrometry analysis. Samples were vortexed for 1 min and centrifuged at 100 xg for 1 min. 4 µl was injected.

2.9.4. Acetylated peptides (Li method)

Sample preparation for immunoprecipitation was performed according to Li *et al*. (2013). After harvesting the fibroblasts as outlined in **Chapter 2.7**, lysate was thawed and kept on ice, and sonicated using a microtip (6 pulses for 25 seconds, with 60-second rests in between, power set between 2-3). Lysate was cleared by centrifugation at 20,000 xg for 15 min at 15°C. Supernatant was transferred to a new tube to which 1/278 volume of 1.25 M DTT was added; the sample was mixed and reduced for 30 min, RT. 1/10 volume of iodoacetamide was added and mixed into the sample, which was incubated for 15 min, RT in the dark. The lysate was diluted 3-fold with 20 mM HEPES, pH 8.0 which gave a final lysis buffer concentration of 2 M urea, 20 mM HEPES, pH 8.0. After dilution, 1/100 volume of 1 mg/ml trypsin-TPCK (Sigma) in 1 mM HCl (Sigma) was mixed into the sample and the sample was digested overnight with mixing at RT.

1/20 volume of 20% TFA was added to the digested sample for a final concentration of 1% TFA. pH was ensured to be below 3.0 and the sample was placed on ice for 15 min to allow precipitate to form. Precipitate was removed by centrifugation at 1780 xg for 20 min, RT. The supernatant was purified using a Sep-Pak® C18 Plus Long Cartridge (Waters, larger than specified) connected to a 20 cc reservoir (BD).

The column was wetted with 5 ml 100% acetonitrile, and washed sequentially with 1 ml, 3 ml and 6 ml of 0.1% TFA. Sample was loaded into the reservoir in stages (60 ml total) and flowed through the column. Flow-through was collected into a 50 ml tube and stored at -20°C. The column was washed again sequentially with 1 ml, 3 ml and 6 ml of 0.1% TFA,

and then washed with 5% acetonitrile + 0.1% TFA. Wash flow-through was stored at -20°C. The sample was eluted into a new 50 ml tube with 3x 4 ml of 40% acetonitrile + 0.1% TFA (larger volume than specified due to larger column). The eluate was frozen on dry ice and then stored at -80°C overnight. The frozen eluate was then lyophilised using a MicroModulyo-230 freeze dryer for 48 h and stored at -80°C.

Washing of anti-acetyl lysine agarose beads prior to immunoprecipitation was performed according to the manufacturer's instructions (Immunechem) with a 10 second-vortex and centrifugation at 1036 xg for 3 min in between washes. 50 µl of anti-acetyl lysine agarose was washed 3x in 1 ml PBS-T, followed by 2 washes with 1 ml 0.1 M NaH_2PO_4 + 1 M NaCl and 1 wash with 1 ml PBS-T. Beads were then washed twice with 1 ml IAP buffer (see section **2.1.2**) according to Li *et al*. (2013); as much buffer as possible was removed without disturbing the pellet and 50 µl was re-added to make a 1:1 slurry of anti-acetyl lysine agarose: IAP buffer. This was stored at 4°C.

1.4 ml IAP buffer was added to the sample eluate powder, which was gently resuspended. Sample was centrifuged at 100 xg to collect to the bottom of the tube and solution was transferred to a 1.5 ml microfuge tube, before centrifugation at 10000 xg for 5 min, 4°C to clear, and cooled on ice for 15 min. The peptides were transferred to the tube containing anti-acetyl lysine agarose in IAP buffer and this was incubated overnight at 4°C on a rotator.

After incubation the sample was centrifuged at 2000 xg for 30 s, 4°C and the supernatant was transferred to a new tube and stored at -20°C. 1 ml IAP buffer was added to the agarose, centrifuged at 2000 xg, 30 s, 4°C and supernatant was removed. This was repeated for a total of two washes. The beads were then washed with 1 ml chilled dH_2O three times in the same manner. To elute, 55 µl of 0.15% TFA was added to the beads, and after tapping the tube several times it was left to stand at RT for 10 min, with gentle mixing every 2 min. The sample was centrifuged at 2000 xg for 30 s and the eluate was transferred to a new tube. 50 µl of 0.15% TFA was added to the beads and the centrifugation/elution steps were repeated, and the eluate was pooled into the previous eluate tube. As much eluate as possible was removed from the beads. The eluate was also briefly centrifuged to pellet any remaining beads and the supernatant was transferred to a new 1.5 ml tube, and then concentrated to completion at 35°C.

The peptide sample was purified using C18 SpinTips (Protea) as described in section **2.9.3.1**. The eluate was pooled and lyophilised at 35°C, and stored at -80°C.

Peptides were resuspended in 8 µl of 98% dH2O/acetonitrile (v/v) + 0.1% formic acid, vortexed for 1 min, centrifuged at 100 xg for 1 min to collect sample to the bottom of the

tube and transferred to an autosampler vial for mass spectrometry analysis. 4 µl was injected into the UPLC system for UPLC-HDMS^E analysis.

2.10. Mass spectrometry analysis

Nano-LC separation was performed on a NanoACQUITY UPLC system (Waters) with a C18 reverse-phase column (HSS T3, 1.8 µm, 250mm x 75µm, Waters). Samples were firstly loaded onto a reverse phase trap column (Symmetry C18, 5 µm, 180 µm x 20mm, Waters), at a trapping rate of 5 µL/min and washed for 10 min with buffer A (98% dH_2O/acetonitrile (v/v) + 0.1% formic acid). Eluted peptides were separated via the analytical column at a flow rate of 0.3 µl/min using a gradient of 3-50% organic phase (buffer B, 80% acetonitrile/dH_2O + 0.1% formic acid) over 90 min (with the exception of all fractionated acetylated peptide samples which were run over 60 min) followed by a 5 min rinse with 85% buffer B. The column was re-equilibrated with 1% buffer B until the end of the run.

Mass spectrometry analysis of the peptides was performed using a Waters Synapt G2-S HDMS system with MassLynx software (Waters). The mass spectrometer was operated in resolution mode, using positive mode-ESI. Samples were sprayed directly into the mass spectrometer and data was collected in MS^E mode of acquisition, alternating between low energy (5 V) and elevated energy (15 V- 40 V ramp) scans. Ion mobility separation was implemented prior to fragmentation; the IMS wave height was 40 V and the IMS wave velocity was 650 m/s. Glu-fibrinopeptide (*m/z* = 785.8426, 100 fmol/µl concentration) was used as LockMass and was delivered to the reference sprayer at 0.25 µl/min, and sampled every 60 seconds for mass correction.

2.11. Processing and database searching of raw data

Raw data files were processed using ProteinLynx Global Server (PLGS, Waters) version 3.0 executables to generate mass spectra using optimal processing parameters indicated by Threshold Inspector (Waters). Data from fractionated samples were merged into one spectrum.bin file using PLGS version 3.0.2 executables and data were database-searched using an Ion Accounting algorithm also using PLGS version 3.0.2 executables (except for the MRC-5 proteome and secretome data, which used 3.0 executables for all stages). Database-searching used workflow parameters that included a false discovery rate (FDR) of 4%, 500000 Da maximum protein mass, and amino acid sequence digesters LysC (where

appropriate) cleaving at lysine residues and trypsin cleaving at lysine and arginine residues, allowing for 1 or 2 missed cleavages (where appropriate). Carbamidomethylation of cysteine was set as a fixed modification and methionine oxidation was set as a variable modification. For cell lysate proteomic experimental data, asparagine and glutamine deamidation were set as additional variable modifications. Acetylation of lysine was added as a variable modification for acetylation experiments. The human database .fasta file for searching the MRC-5 cellular proteome data was downloaded from the UniProt website on the 11/12/2012; the .fasta file for searching of subsequent data was downloaded on the 29/11/13. Both include the amino acid sequences for Enolase (*Saccharomyces cerevisiae* accession number P00924) and ClpB (*Escherichia coli* accession number P03815).

2.12. Data analysis

2.12.1. Correction and normalisation of mass spectrometry protein data

Processed and searched raw data outputs were converted to Microsoft Excel .xlsx format. Preliminary data analyses were performed in Microsoft Excel. For each experiment, all sets of protein data were merged into a single file to include protein UniProt accession number, protein description and femtomoles on column. The Hi3 method was used to quantify proteins which, as outlined previously, calculates quantification using the top 3 most intense tryptic peptides of the internal standards and of each protein (Silva et al., 2006); only quantified proteins were included in analysis, therefore all included proteins had been matched to at least 3 unique peptides, a measure of confident identification.

The data was normalised using a "top 100" method (see **Appendix A.1** for examples), to account for differences in LC loading. All proteins that were detected with quantification in 100% of the samples in an experiment were used for correction, as these proteins were considered the most stable and abundant and thus the most reliable across all sample runs. The number of femtomoles of each of these proteins were summed within each individual replicate, and the number of femtomoles of each protein in the replicate was divided by this correction value and multiplied by an appropriate multiple of 10 to yield a simpler value to manipulate. The data was then logged to base 2 to transform.

2.12.2. Statistical analyses of mass spectrometry protein data

The data were analysed in two ways; firstly, by a more stringent Student's T-Test, and secondly by a more data-accommodating fold-change analysis, both performed in Microsoft Excel.

To analyse the data by T-Test, only proteins that were present in 3+ replicates of both conditions were included in the dataset, which excluded part of the dataset that contained missing data for analysis. Each T-Test was unpaired with unequal variance, except for the IPF fibroblast monolayer dataset comparing primary donor cultures with and without FK228 treatment; this was paired by donor culture. The cut-off *p*-value for significance was 0.05.

For fold-change analysis, proteins detected in 2+ replicates of a condition were included in the calculation. The average logged femtomole value across each set of 3+ replicates was used to calculate the fold-change in expression of the proteins identified. In each case the average logged abundance value for control or non-IPF fibroblast samples was subtracted from the average logged abundance value for the treated or IPF fibroblast samples and this result was anti-logged to result in a fold-change value. A 2-fold change in expression was used as the cut-off value for significance.

2.12.3. Identification of potential biomarkers

The list of proteins significant by Student's T T-test, by fold-change or only detected in one condition in the IPF fibroblast vs. non-IPF fibroblast dataset was compared to the corresponding list of significant proteins in the IPF fibroblast ± FK228 dataset to identify proteins that have similar or the opposite expression. The two lists were matched up in Microsoft Excel and sorted according to the direction of change in expression in the two datasets, for example, proteins that had increased expression in IPF fibroblasts compared to non-IPF fibroblasts but were no longer detected following treatment with FK228. This was used to identify potential candidate biomarkers of IPF and of FK228 treatment in IPF.

2.12.4. Gene ontology and protein-interaction analyses of mass spectrometry protein data

Differentially expressed proteins (DEPs) – significant proteins according to Student's T-Test, or proteins only detected in one condition (which must be in at least 3 replicates of one condition and 0 of the other) – were subjected to a series of analyses to determine their role inside/outside of the cell, their potential interactions and which signalling pathways they were involved in. Gene ontology analysis was performed using GOrilla (Gene Ontology

enRIchment anaLysis and visuaLizAtion tool) (Eden et al., 2009) to generate enriched GO terms of the DEP dataset against the entire human proteome, which were summarized into distinct terms using REVIGO (Supek et al., 2011). GO analysis was also performed using PANTHER (Mi et al., 2013).

Heat map clustering and direct protein networks shown in **Chapter 3** only were created by Akul Singhania using Metacore (Thomson Reuters) and Cytoscape 2.8. Heat maps in **Chapters 4** and **5** were created by the author using R packages "ggplots" and "d3heatmap", using a template R script originally created by Dr Alistair Bailey. Session info was as follows:

```
R version 3.2.2 (2015-08-14)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 7 x64 (build 7601) Service Pack 1
```

Through interpreting the data from the direct protein networks, individual pathways that were differentially regulated through differential expression of receptors, effector proteins and transcription factors could be identified.

2.12.5. Analysis of the acetylated peptide mass spectrometry data following immunoprecipitation

Preliminary data analysis was performed in Microsoft Excel. Processed and searched peptide data was filtered for the acetylated lysine modification. The protein accession number, protein name, the number of acetylated peptides per protein, the amino acid sequence and the number of b and y ions was recorded for each identified acetylated protein. **Figure 5.7** was produced in R, session info as above.

2.13. Immunoprecipitation of acetylated proteins

Cell pellets were resuspended in 750 µl lysis buffer (see section **2.1.2**) and lysed by sonication (6 pulses for 25 seconds, with 60-second rests in between, power set between 2-3) on ice. Cell lysates were centrifuged at 17000 xg for 20 min at 4°C. Protein concentration of supernatant was measured using a DirectDetect® spectrometer.

Immunoprecipitation was carried out using anti-acetyl lysine agarose (Immunechem) at a ratio of 50 µl agarose:2 mg protein, according to manufacturer's instructions. Anti-acetyl lysine agarose was washed 3x with 1 ml PBS-T, with centrifugation at 1036 xg for 5 min between washes, followed by 2 washes with 1 mL of 0.1 M NaH_2PO_4 + 1M NaCl. Beads were washed lastly with 1 ml PBS-T. Cell lysate was added to the beads and the mixture was

incubated overnight at 4ºC on a rotator. After incubation beads were washed 4x in 1 ml PBS-T and eluted with 50 µl Laemmli buffer. The eluate was vortexed and centrifuged at 1036 xg for 5 min to remove any contaminating beads. The supernatant was concentrated to 20 µl at 37ºC before western blot analysis. For western blotting, an aliquot was also taken of the lysate before and after incubation with the anti-acetyl lysine agarose.

2.14. Western blotting

The lowest concentration of protein per sample set (µg) was used as the loading concentration for each experiment. Cell lysates were made up to equal volumes, dissolved in sample buffer (8 µl 4x sample buffer (Life Technologies), 2µl 0.5 M DTT, final volume 32 µl) and centrifuged at 1036 xg for 1 min. Proteins were heated for 10 min at 70°C on a Techne Dri-Block and separated on a 4-12% gradient NuPAGE Bis-Tris gel in MOPS electrophoresis buffer (both Life Technologies) for 55 min at 200 V. 5 µl Precision Plus Protein™ All Blue (Biorad) was used as a molecular weight standard.

Separated proteins were then transferred to a PVDF membrane using a wet blotting system (Biorad). Transfer was carried out for 2 h, at 100 V in 1x continuous CAPS transfer buffer (see section **2.1.2**). The blot was rinsed in Tris-buffered saline (TBS, see section **2.1.2**) in dH_2O and blocked for 1 h in 10 ml TBS, 3% bovine serum albumin (BSA, Sigma).

All primary antibodies were added to 3 ml TBS-T + 5% BSA, and incubated with PVDF membranes overnight at 4°C on a Stuart roller mixer SRT9D. Secondary antibodies were added to 3 ml TBS-T + 0.5% BSA, and incubated with the membranes for 1 h, RT on a Stuart roller mixer SRT9 in the dark (see **Table 2.1** for list of antibodies used). Membranes were washed 4x in TBS-T (see section **2.1.2**) before, between and after incubations before being dried in filter paper (Biorad). Blots were visualised using an Odyssey infrared imaging system and software (Licor), scanning using 680 and 800 wavelengths at intensities ranging from 0.1-5.0.

Table 2.1 Antibodies used for western blotting.

Antibody	Manufacturer	Dilution
Primary		
Actin α-smooth muscle, mouse	Sigma Aldrich	1:5000
Acetyl lysine, rabbit	Immunechem	1:500
Acetyl histone H3, rabbit	Millipore	1:20000
FBLIM-1, mouse	Santa-Cruz Biotechnology	1:100/1:75
H3 Pan, rabbit	Millipore	1:5000
MMP-2, rabbit	Cell Signaling	1:1000
Semaphorin-7a, rabbit	Abcam	1:1000/1:100
β-actin, rabbit	Cell Signaling	1:1000
β-catenin, rabbit	Cell Signaling	1:1000
Secondary		
Goat anti-Rabbit IgG (H+L) Secondary Antibody, DyLight 680 conjugate	Thermo Fisher Scientific	1:15000
Goat anti-Mouse IgG (H+L) Cross Adsorbed Secondary Antibody, DyLight 800 conjugate	Thermo Fisher Scientific	1:15000

2.14.1. Stripping of western blot membranes

Stripping was carried out according to a protocol from Abcam. Membranes were incubated with 10 ml mild stripping buffer (see section **2.1.2**) at RT for 10 min. Used buffer was removed and replaced with fresh buffer for 10 min, RT. Membranes were washed with 10 ml PBS for 10 min, RT, and then with 10 ml TBS-T for 5 min, RT for a total of two washes. Membranes were re-blocked with 3% BSA for 1 h, RT with mixing and rinsed with TBS and dH_2O before re-probing.

3. Results - Characterisation of myofibroblast formation *in vitro* using quantitative proteomics

3.1. Abstract

Myofibroblasts are key players in the wound healing response, producing extracellular matrix components to form a scaffold for repair and remodelling. However, in fibrotic disease these cells accumulate and deposit excessive amounts of matrix components, reducing tissue compliance.

In this study, myofibroblast formation was characterised at the proteomic level by generating proteomic expression profiles for fibroblasts with and without stimulation with the profibrotic growth factor transforming growth factor-beta, which causes fibroblast differentiation. Proteins were isolated from whole cell lysates, digested and fractionated before subjection to UPLC-HDMS^E with ion mobility enabled. Almost 5000 proteins were identified with quantification; over 60 of these were significantly differentially regulated (p<0.05) between conditions, and more than 70 were detected only in the myofibroblast model.

Statistical and gene ontology analysis revealed that these differentially expressed proteins were mainly involved in cellular organisation and cell cycle regulation, and revealed the modulation of several non-canonical TGF-β_1 signalling pathways. Furthermore, changes in the expression of secreted proteins with and without TGF-β_1 stimulation were also investigated using UPLC-HDMS^E, thus providing a comprehensive proteomic characterisation of myofibroblast formation. Eight of these proteins changed significantly in abundance (p<0.05) following treatment.

The results of this study suggest that there are many more proteins that change in expression during fibroblast differentiation than previously thought, and it is likely that some of these are aberrantly modulated in fibrosis, thus providing many more potential biomarkers for investigation in fibrotic disease.

3.2. Introduction

The myofibroblast is the contractile, secretory, differentiated form of the fibroblast mesenchymal cell, which plays an important role in the wound healing response. Under normal physiological conditions, activated myofibroblasts migrate to sites of injury and produce extracellular matrix (ECM) components to form a scaffold that supports cellular migration, proliferation and differentiation during tissue repair and re-epithelialisation of the wound (Hinz, 2007), and undergo apoptosis following wound closure (Wynn and Ramalingam, 2012). In fibrotic disease, this response can become dysregulated.

The pleiotropic growth factor transforming growth factor-beta (TGF-β) regulates many key cellular functions such as cell growth, proliferation, differentiation, apoptosis and survival (Zhang, 2009). It is an important signalling molecule both in embryonic development for tissue growth, differentiation and cell proliferation to establish the body plan (Derynck and Zhang, 2003) and subsequently for maintaining homeostasis (Sakai and Tager, 2013). TGF-β signalling plays a major role in wound healing as it is required for myofibroblast formation (Hinz et al., 2007) and is also a potent inducer of ECM production (Coker et al., 1997). TGF-β acts via a family of TGF-β receptors with intrinsic serine/threonine kinase activity (Heldin et al., 1997). A TGF-β_1 family member typically binds to its receptor complex and signals through the canonical Smad pathway, which can alter the expression of profibrotic genes (Leask and Abraham, 2004). It also signals via Smad-independent pathways. These include the phosphatidylinositol-3-kinase/protein kinase B (PI3K/AKT) and mitogen-activated protein kinase (MAPK) pathways (Zhang, 2009; Derynck and Zhang, 2003) such as the Ras/MEK/ERK cascade, which can also modify the activity of Smad proteins amongst others (Leask and Abraham, 2004). Through regulation of genes controlling myofibroblast differentiation and function, TGF-β_1 signalling can stimulate matrix remodelling to restore homeostasis through tissue repair.

In various disease pathologies myofibroblasts increase in number and activity, leading to fibrotic changes that characterise lung diseases such as idiopathic pulmonary fibrosis (King et al., 2011). In IPF, repetitive pulmonary alveolar epithelial injury by an unknown stimulus is hypothesized to cause the release of profibrotic mediators including TGF-β_1, along with others such as connective tissue growth factor (CTGF) and ED-A fibronectin (a splice variant of fibronectin) (Leask and Abraham, 2004). This repetitiveness of cellular injury and prolonged TGF-β_1 signalling triggers an aberrant wound healing response, causing continual differentiation of fibroblasts, which form areas of fibroblastic foci and produce excessive ECM components, causing progressive pulmonary scarring. This results in tissue stiffening

and the destruction of lung architecture, a reduction in lung function and/or impaired gas exchange, which is some cases can be fatal (Hinz et al., 2007; Zhang and Phan, 1999).

Proteomic studies of the myofibroblast are currently limited. Since these cells play such a prominent role in IPF pathogenesis, the aim of this investigation was to comprehensively characterise myofibroblast formation, using a quantitative mass spectrometry-based proteomic approach to profile both intracellular and secreted proteins. To achieve this, a pulmonary fibroblast cell line was treated with TGF-β_1 to stimulate cellular differentiation, and the expression of proteins from both cell lysates and cell-free supernatants was compared with those of untreated fibroblasts to identify proteins and processes that are differentially regulated between the two conditions. More than 5000 proteins were identified in total in this study. Proteins with differential expression following TGF-β_1 stimulation could then be used to identify biological processes and mechanistic pathways important in the progression of diseases involving abnormal wound healing responses. These proteins may also be useful as markers to help to distinguish myofibroblasts from fibroblasts, since α-smooth muscle actin (α-SMA), currently the most commonly used marker of differentiated fibroblasts (Hinz, 2007), is not specific for myofibroblasts as it is also a marker of smooth muscle cells (Skalli et al., 1989).

3.3. Results

In order to study the proteomic effects of TGF-β_1 on lung fibroblasts, MRC-5 fibroblasts cultured in phenol red-free Dulbecco's modified Eagle medium containing 10% foetal bovine serum (FBS) were seeded into 6-well plates at $7.5x10^4$ cells/well. At 60% confluence, cells were treated in the absence or presence of 10 ng/ml TGF-β_1 for 48 hours. For secreted protein analysis, phenol red-free Dulbecco's modified Eagle medium without FBS was used upon TGF-β_1 stimulation. Media were collected and cells were harvested using TRIzol® for mRNA analysis, or harvested using trypsin-EDTA, counted using trypan blue stain, washed 4 times in phosphate buffered saline and lysed in 0.1% sodium dodecyl sulfate using a sonication probe for protein analysis. All experiments were performed in triplicate. Please refer to sections **2.2** and **2.6** for full methods.

3.3.1. The mRNA expression of α-smooth muscle actin significantly increases after stimulation with TGF-β_1

The gene expression of α-SMA in fibroblasts has been shown to increase in response to TGF-β_1 (Davies et al., 2012; Desmouliere et al., 1993) and this increased expression has been used as a biomarker of the myofibroblast (Hinz et al., 2007). Therefore, its expression was measured from three biological replicates of MRC-5 fibroblast cultures by RT-qPCR after challenge with 10 ng/ml TGF-β_1 for 48 hours, to validate that the treatment had had the intended effect on the fibroblasts, prior to proteomic analyses. RNA was extracted from TRIzol®-harvested fibroblasts and was DNase-treated before cDNA synthesis. mRNA expression of α-smooth muscle actin was analysed by quantitative reverse transcription PCR and normalized to the housekeeping genes ubiquitin C and phospholipase A2 (UBC/A2) using the ΔΔCt method (see section **2.8**).

Figure 3.1 shows that the mRNA expression of α-SMA significantly increased, approximately 4-fold in response to TGF-β_1 treatment (p=0.0253).

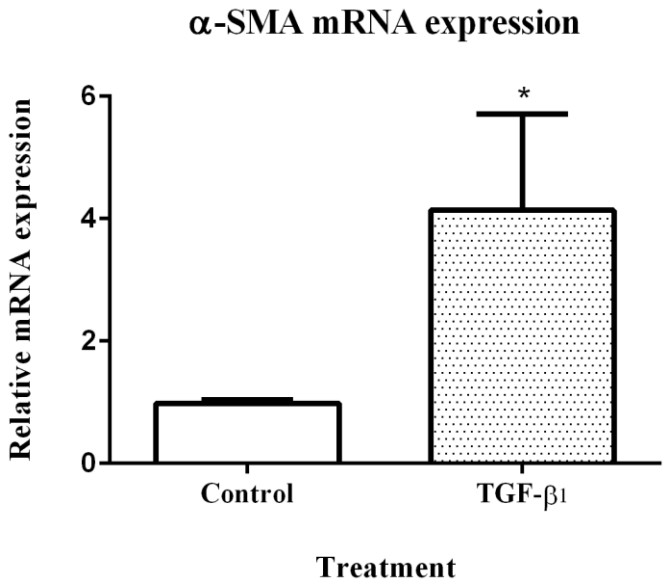


Figure 3.1 mRNA expression of α-smooth muscle actin in MRC-5 fibroblasts with and without treatment with TGF-β_1. α-SMA mRNA expression was measured by quantitative RT-PCR after 48 hours of MRC-5 fibroblast treatment with 10 ng/ml TGF-β_1. Expression was calculated relative to the geometric mean of the housekeeping genes UBC/A2 using the ΔΔCt method. The data are expressed as fold-change in expression relative to one untreated control sample. Bars indicate the mean ± SD.
* = p< 0.05 compared to the non-treated control.

3.3.2. α-smooth muscle actin protein expression increases after stimulation with TGF-β$_1$

The protein expression of α-SMA in the cell lysate following fibroblast stimulation with 10 ng/ml TGF-β$_1$ for 48 hours was also analysed, by SDS-polyacrylamide gel electrophoresis and western blotting. Membranes were stripped and re-probed with histone H3 pan as a loading control (see section **2.14**).

In accordance with previous reports (Davies et al., 2012; Honda et al., 2010), the protein expression slightly increased in MRC-5 fibroblasts after treatment with TGF-β$_1$ in all three biological replicates, indicating that the fibroblasts responded to treatment in the expected manner (a representative image is shown in **Figure 3.2**).

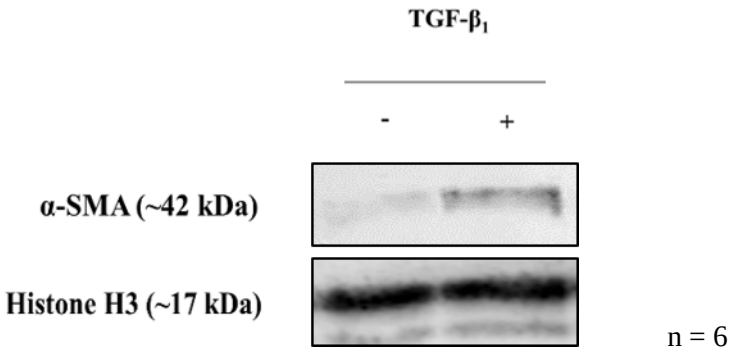


Figure 3.2 Western blot of α-smooth muscle actin expression in MRC-5 fibroblasts with and without treatment with TGF-β1. α-SMA expression was analysed by western blotting after 48 hours of treatment with TGF-β$_1$. Histone H3 pan was used as a loading control.

3.3.3. Comprehensive proteomic analysis of MRC-5 fibroblasts reveals significant changes in protein expression following TGF-β$_1$ stimulation

A quantitative, mass spectrometry-based proteomic approach was used for the comparison of fibroblast and myofibroblast protein expression profiles, to study protein expression changes during the transition to a myofibroblast. The inclusion of an internal peptide digest standard of known concentration to each sample enabled estimates of absolute amounts of each identified protein to be calculated using the Hi3 quantification method, for comparison of expression values across all samples.

As described in **Figure 3.3A,** for mass spectrometry analysis of fibroblast cell lysates, 100 µg of protein from fibroblast cell lysates was precipitated by methanol/chloroform extraction. Proteins were reduced with 1 mM DTT for 1 hour at 60°C, alkylated at room temperature for

45 minutes in the dark with 5.5 mM iodoacetamide, and digested into peptides, firstly with 2 µg endoproteinase Lys-C for 4 hours at 37°C and then with 2 µg sequencing grade modified trypsin at 37°C overnight. Enolase protein digest internal standard was spiked into each cellular peptide sample at a concentration of 150 fmol before isoelectric focusing by OFFGEL fractionation into 12 peptide fractions according to manufacturer's instructions. Each fraction was subjected to purification using an Empore™ C18 solid phase extraction plate to remove residual salts, buffers and contaminants before evaporation to completion and resuspension in loading buffer (3% acetonitrile + 0.01% formic acid) for mass spectrometry analysis (see section **2.9**).

For secreted protein analysis**,** as illustrated in **Figure 3.3B,** fibroblast medium samples were enriched for secreted proteins using StrataClean resin. Proteins were reduced with 0.5 µg DTT for 1 hour at 60°C, alkylated at room temperature for 45 minutes in the dark with 2.5 µg iodoacetamide and digested off the resin overnight with 0.5 µg sequencing grade modified trypsin at 37°C. Following digestion, peptides were also purified using an Empore™ C18 solid phase extraction plate before evaporation of solvent and resuspension in loading buffer A (3% acetonitrile + 0.01% formic acid) containing 100 fmol enolase protein digest internal standard for mass spectrometry analysis (see section **2.9**).

Both the cell lysate samples and secreted protein samples were analysed by UPLC-HDMS^{E}. Peptides were separated by liquid chromatography using a NanoACQUITY ultra-performance liquid chromatography (UPLC) system with a C18 reverse-phase column, at a flow rate of 300 nL/minute, over a 90-minute gradient of 3-50% buffer B (80% acetonitrile/dH_2O + 0.1% formic acid). Peptides were sprayed directly into a Waters Synapt G2-S system operating in positive ion mode, with ion mobility enabled prior to fragmentation. Data was collected in MS^{E} mode of acquisition, alternating between low energy (5V) and high energy (15V - 45V ramp) scans. Glu-fibrinopeptide (m/z = 785.8426, 100 fmol/µl) was used as LockMass and was sampled every 60 seconds for calibration (see section **2.10**).

Raw data files were processed using Protein Lynx Global Server (PLGS) version 3.0. Data were searched against the human UniProt database using an Ion Accounting algorithm also in PLGS 3.0 (see section **2.11**). Estimates of absolute quantification were calculated using the Hi3 method of quantification (Silva et al., 2006).

In total, 5142 unique proteins were identified from the cell lysates. 4996 of these were associated with three or more unique peptides and were quantified using the Hi3 method. In the secretome 215 unique proteins were identified, and of these 173 were quantified.

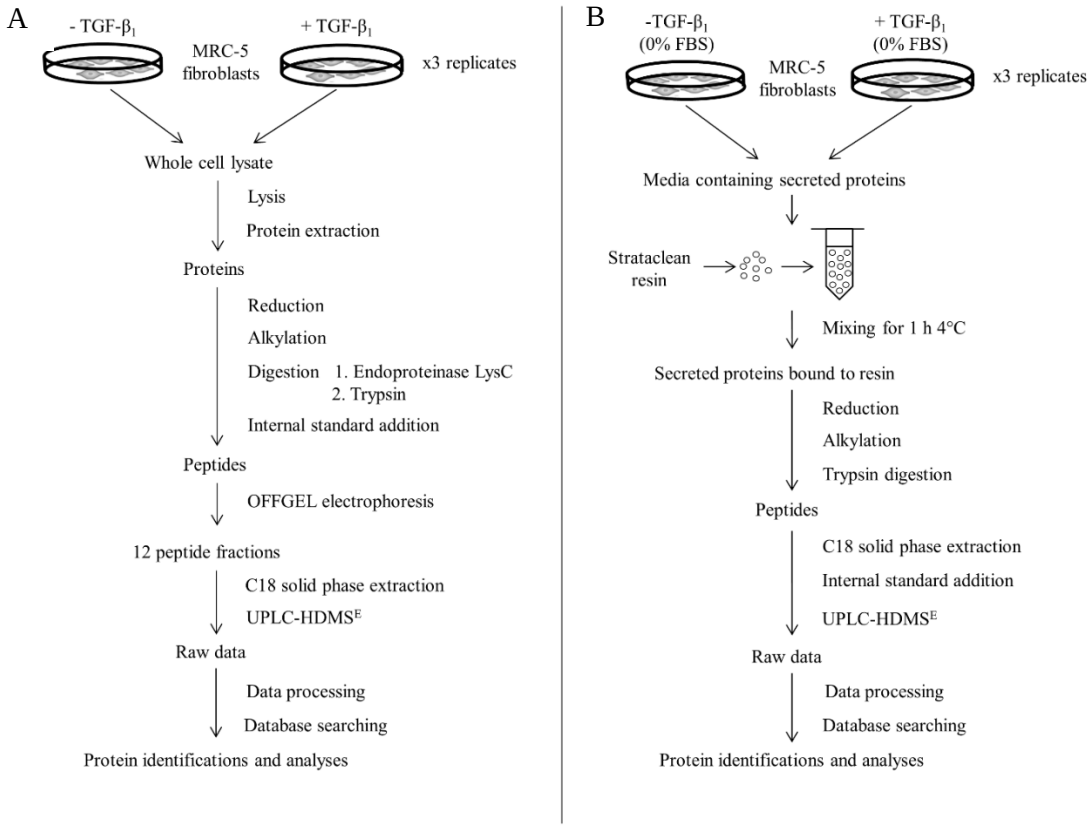


Figure 3.3 A representation of the method workflow used to identify and quantify the intracellular proteome and secretome of TGF-β1-stimulated and unstimulated MRC-5 fibroblasts. MRC-5 fibroblasts were treated with 10 ng/ml TGF-β_1 for 48 h and compared against untreated fibroblasts as controls. **A.)** Protein lysates were reduced, alkylated and digested into peptides using endoproteinase Lys-C and trypsin. Enolase internal reference standard (Waters) was spiked into the digests for Hi3 quantification. Peptides were fractionated by OFFGEL electrophoresis (Agilent). **B.)** Secreted proteins in the media were bound to StrataClean resin (Agilent) and were reduced, alkylated and digested off the resin into peptides using trypsin. Enolase internal protein reference standard (Waters) was spiked into the digests for Hi3 quantification. For both experiments, samples were separated by nano-UPLC and eluted peptides were sprayed directly into a Synapt G2-S mass spectrometer (Waters) operating in MS^E mode with ion mobility enabled. Data were processed and searched using PLGS 3.0 (Waters), prior to statistical analyses and pathway analyses. Each experiment was performed in triplicate.

Protein concentration was normalised (see section **2.12** and **Appendix A.1**) and then two methods of analysis were applied to these proteomic datasets. There are no standard statistical methods that are always applied to proteomic data, so each case must be judged individually as to how it should be analysed. Parametric or non-parametric statistical tests

may be applied, however in this case with n = 3 per condition, it is difficult to judge whether the dataset is normally distributed, therefore data was logged to base 2 prior to statistical analysis.

Firstly, proteins quantified across all samples were analysed for differential expression using a two-tailed unpaired Student's T-Test. *p*-values <0.05 were considered significant. Proteins identified in all three replicates of one condition and in none of the other condition were also considered significant.

Assuming that the data is normally distributed, there is also the issue of missing data, an inherent problem with mass spectrometry data. Stringent tests such as the Student's T-Test cannot handle missing values effectively; therefore, only proteins identified in all replicates were analysed using this test. Reducing the number of proteins that could be analysed to just those identified only in all samples meant that many proteins were not investigated. For a more exploratory analysis of the datasets, proteins quantified in at least two replicates of a condition were then analysed for differential expression according to fold-change, with a 2-fold change in expression used as the cut-off value for significance. Proteins identified in two replicates of one condition and in none of the other condition were also considered significant. This method also has limitations since it lacks statistical confidence, and is still affected by a low number of replicates, but it allows the investigation of proteins that could not otherwise be studied by a more rigorous test. The methods of analyses are illustrated in **Figure 3.4**.

For secretomic analysis, a medium-only sample was analysed by mass spectrometry, and any proteins identified in this sample were not included in T-Test and fold-change analyses to eliminate background interference.

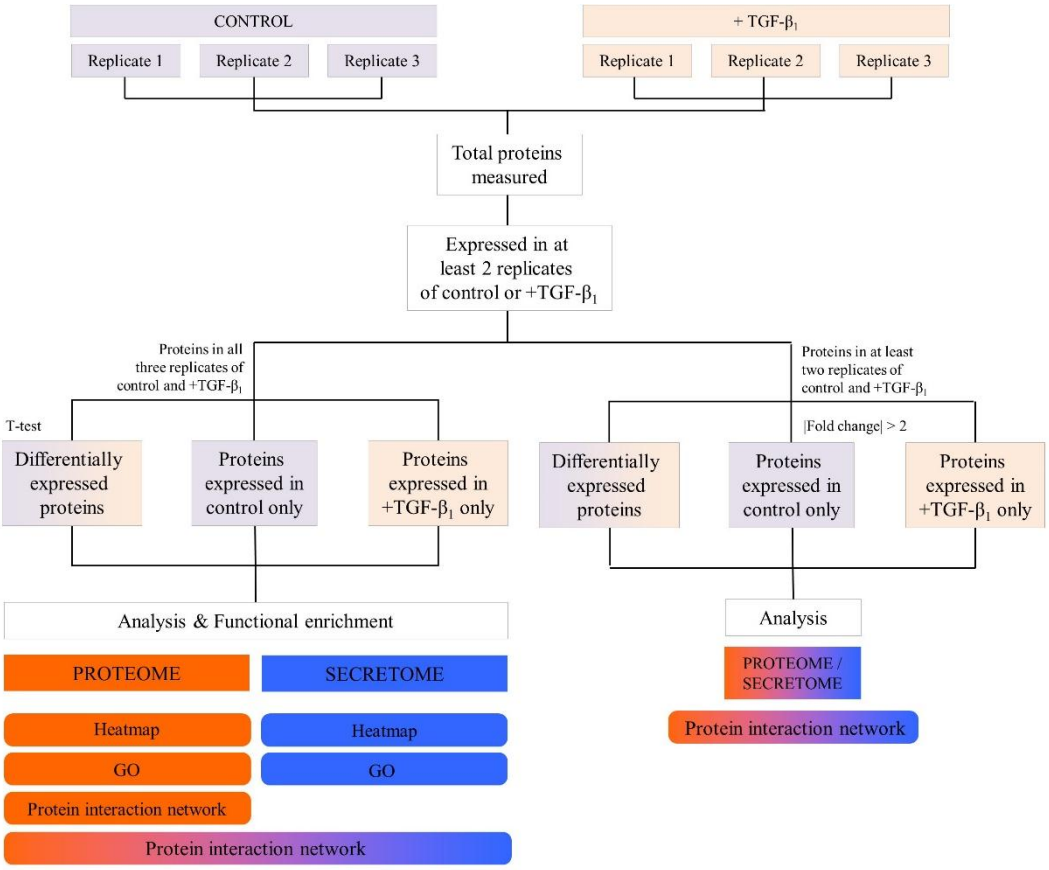


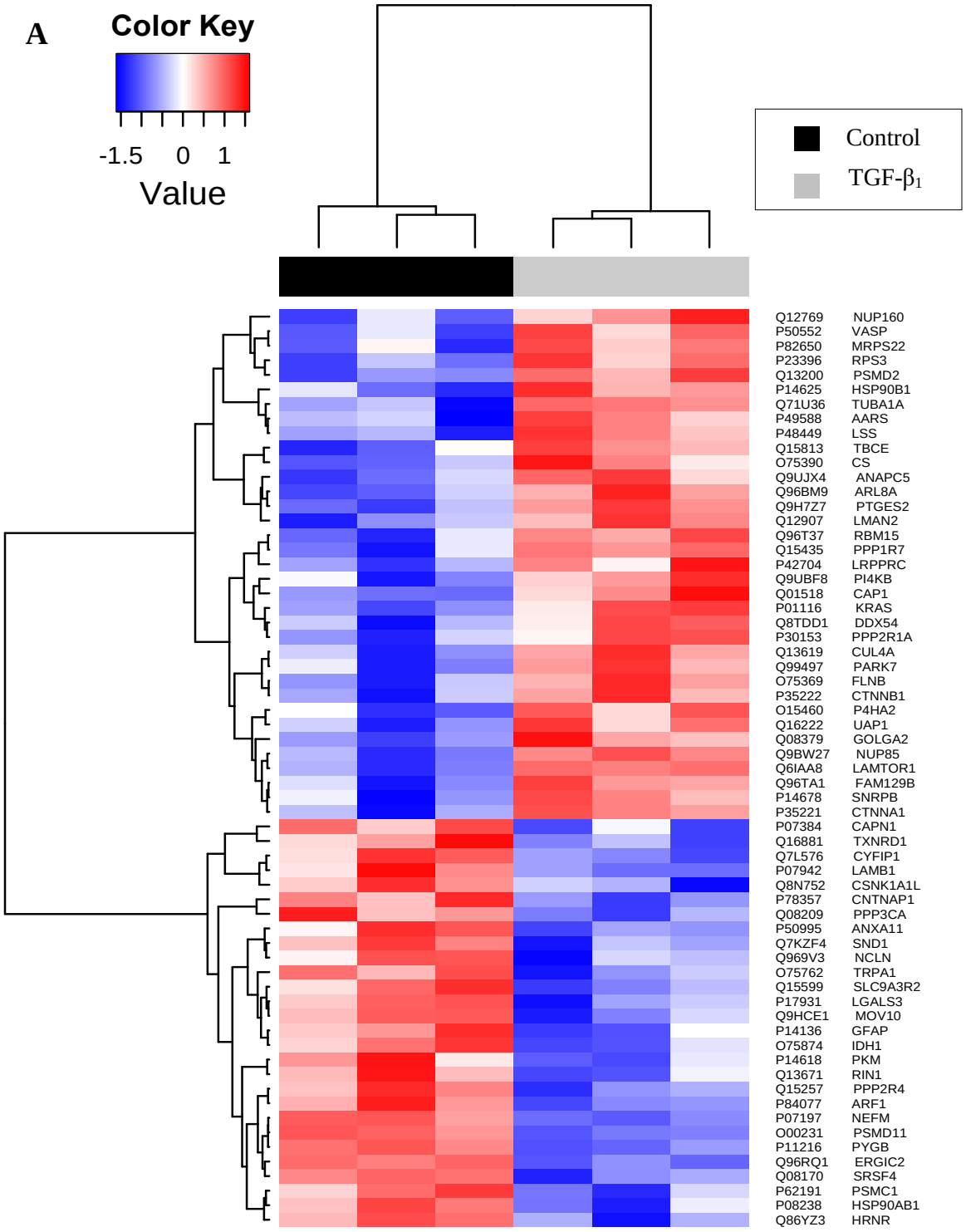
Figure 3.4 Flow chart representing the data analysis pipeline for the cellular proteome and secretome analysis. Datasets were analysed both by Student's T-Test and by fold-change analyses. Gene ontology analyses and network analyses were performed on the resulting significant proteins.

As previously mentioned, for analysis by Student's T-Test only proteins quantified in all three replicates of both conditions were considered. 1589 cellular proteins and 24 secreted proteins were analysed by T-test. Proteins needed to be quantified in all three biological replicates of each condition and not detected in any replicates of the other to be considered to be identified in only one condition; using this criterion 73 proteins were present only in TGF-β_1-treated fibroblast samples and 7 proteins were present only in control samples across the two datasets.

3.3.4. Fibroblast intracellular proteomic data revealed over one hundred proteins with a significant difference in expression following TGF-β_1-treatment or only detected either pre- or post-treatment

For the intracellular proteomic data, 63 proteins had a statistically significant difference in expression between control and TGF-β_1-treated fibroblasts by Student's T-Test, with a *p*-value of less than 0.05. Thirty-five proteins were demonstrated to have significantly upregulated expression, and 28 had significantly downregulated expression in response to treatment. Seven proteins were detected only in control samples; 67 proteins were only detected following TGF-β_1 treatment. The lists of proteins and their corresponding *p*-values can be found in **Appendix B.1-B.2.**

These data are reflected in a heat map in **Figure 3.5A,** in which protein expression values are shown to cluster to condition. **Figure 3.5B** includes proteins that were only quantified in one condition according to the more stringent criteria (137 total proteins). Heat maps were created using R. These lists of significant proteins were taken forward for further analyses.


A
Color Key
-1.5
0
1
Value
Control
TGF-β1
Q12769 NUP160
P50552 VASP
P82650 MRPS22
P23396 RPS3
Q13200 PSMD2
P14625 HSP90B1
Q71U36 TUBA1A
P49588 AARS
P48449 LSS
Q15813 TBCE
O75390 CS
Q9UJX4 ANAPC5
Q96BM9 ARL8A
Q9H7Z7 PTGES2
Q12907 LMAN2
Q96T37 RBM15
Q15435 PPP1R7
P42704 LRPPRC
Q9UBF8 PI4KB
Q01518 CAP1
P01116 KRAS
Q8TDD1 DDX54
P30153 PPP2R1A
Q13619 CUL4A
Q99497 PARK7
O75369 FLNB
P35222 CTNNB1
O15460 P4HA2
Q16222 UAP1
Q08379 GOLGA2
Q9BW27 NUP85
Q6IAA8 LAMTOR1
Q96TA1 FAM129B
P14678 SNRPB
P35221 CTNNA1
P07384 CAPN1
Q16881 TXNRD1
Q7L576 CYFIP1
P07942 LAMB1
Q8N752 CSNK1A1L
P78357 CNTNAP1
Q08209 PPP3CA
P50995 ANXA11
Q7KZF4 SND1
Q969V3 NCLN
O75762 TRPA1
Q15599 SLC9A3R2
P17931 LGALS3
Q9HCE1 MOV10
P14136 GFAP
O75874 IDH1
P14618 PKM
Q13671 RIN1
Q15257 PPP2R4
P84077 ARF1
P07197 NEFM
O00231 PSMD11
P11216 PYGB
Q96RQ1 ERGIC2
Q08170 SRSF4
P62191 PSMC1
P08238 HSP90AB1
Q86YZ3 HRNR

B

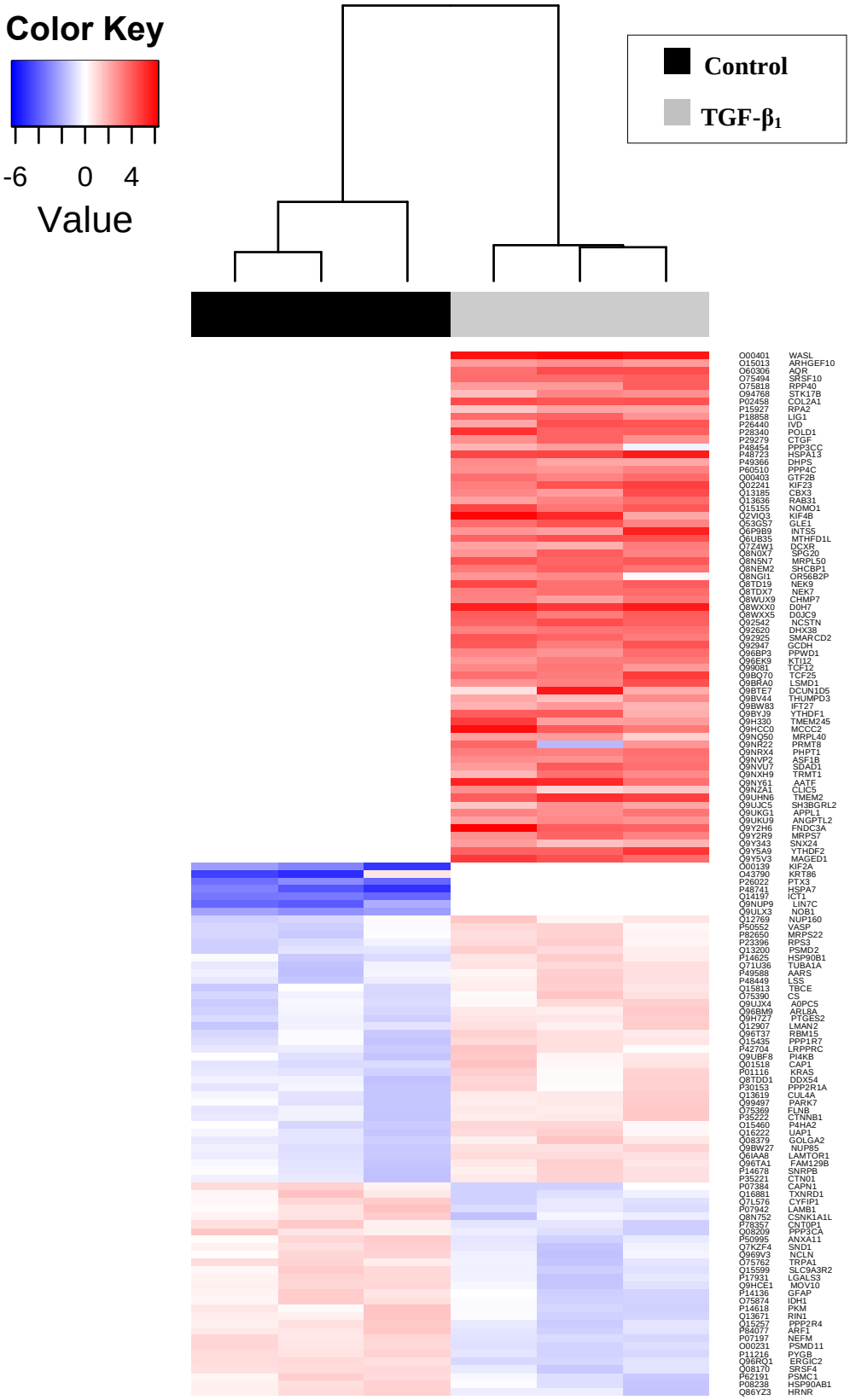

Color Key
-6
0
4
Value
Control
TGF-β1
O00401 WASL
O15013 ARHGEF10
O60306 AQR
O75494 SRSF10
O75818 RPP40
O94768 STK17B
P02458 COL2A1
P15927 RPA2
P18858 LIG1
P26440 IVD
P28340 POLD1
P29279 CTGF
P48454 PPP3CC
P48723 HSPA13
P49366 DHPS
P60510 PPP4C
Q00403 GTF2B
Q02241 KIF23
Q13185 CBX3
Q13636 RAB31
Q15155 NOMO1
Q2VIQ3 KIF4B
Q53GS7 GLE1
Q6P9B9 INTS5
Q6UB35 MTHFD1L
Q7Z4W1 DCXR
Q8NDX7 SPG20
Q8N5N7 MRPL50
Q8NEM2 SHCBP1
Q8NGI1 OR56B2P
Q8TD19 NEK9
Q8TDX7 NEK7
Q8WUX9 CHMP7
Q8WXX0 D0H7
Q8WXX5 D0JC9
Q92542 NCSTN
Q92620 DHX38
Q92925 SMARCD2
Q92947 GCDH
Q96BP3 PPWD1
Q96EK9 KTI12
Q99081 TCF12
Q9BQ70 TCF25
Q9BRA0 LSMD1
Q9BTE7 DCUN1D5
Q9BV44 THUMPD3
Q9BW83 IFT27
Q9BY39 YTHDF1
Q9H330 TMEM245
Q9HCC0 MCCC2
Q9NQ50 MRPL40
Q9NR22 PRMT8
Q9NRX4 PHPT1
Q9NVP2 ASF1B
Q9NVU7 SDAD1
Q9NXH9 TRMT1
Q9NY61 AATF
Q9NZA1 CLIC5
Q9UHN6 TMEM2
Q9UJC5 SH3BGRL2
Q9UKG1 APPL1
Q9UKU9 ANGPTL2
Q9Y2H6 FNDC3A
Q9Y2R9 MRPS7
Q9Y343 SNX24
Q9Y5A9 YTHDF2
Q9Y5V3 MAGED1
O00139 KIF2A
O43790 KRT86
P26022 PTX3
P48741 HSPA7
Q14197 ICT1
Q9NUP9 LIN7C
Q9ULX3 NOB1
Q12769 NUP160
P50552 VASP
P82650 MRPS22
P23396 RPS3
Q13200 PSMD2
P14625 HSP90B1
Q71U36 TUBA1A
P49588 AARS
P48449 LSS
Q15813 TBCE
O75390 CS
Q9UJX4 A0PC5
Q96BM9 ARL8A
Q9H7Z7 PTGES2
Q12907 LMAN2
Q96T37 RBM15
Q15435 PPP1R7
P42704 LRPPRC
Q9UBF8 PI4KB
Q01518 CAP1
P01116 KRAS
Q8TDD1 DDX54
P30153 PPP2R1A
Q13619 CUL4A
Q99497 PARK7
O75369 FLNB
P35222 CTNNB1
O15460 P4HA2
Q16222 UAP1
Q08379 GOLGA2
Q9BW27 NUP85
Q6IAA8 LAMTOR1
Q96TA1 FAM129B
P14678 SNRPB
P35221 CTN01
P07384 CAPN1
Q16881 TXNRD1
Q7L576 CYFIP1
P07942 LAMB1
Q8N752 CSNK1A1L
P78357 CNTOP1
Q08209 PPP3CA
P50995 ANXA11
Q7KZF4 SND1
Q969V3 NCLN
O75762 TRPA1
Q15599 SLC9A3R2
P17931 LGALS3
Q9HCE1 MOV10
P14136 GFAP
O75874 IDH1
P14618 PKM
Q13671 RIN1
Q15257 PPP2R4
P84077 ARF1
P07197 NEFM
O00231 PSMD11
P11216 PYGB
Q96RQ1 ERGIC2
Q08170 SRSF4
P62191 PSMC1
P08238 HSP90AB1
Q86Y23 HRNR

Figure 3.5 Heat maps illustrating the A) differentially expressed proteins and B) differentially expressed proteins as well as proteins expressed only in TGF-β_1-treated fibroblasts and only in control from the cellular proteome analysis. Hierarchical clustering was performed using Pearson correlation metric to calculate distances and clustered using complete linkage. The expression values of proteins were averaged and scaled across every row to indicate the number of standard deviations above (red) or below the mean (blue), denoted as the row Z-score.

3.3.5. Gene ontology analysis revealed a variety of biological processes that are modulated following TGF-β_1 stimulation

Proteins with significantly different expression between untreated and TGF-β_1-treated fibroblasts were subjected to gene ontology (GO) classification using GOrilla (Eden et al., 2009) to identify GO terms significantly overrepresented for the differentially expressed proteins as well as proteins only in TGF-β_1-treated fibroblasts and only in control fibroblasts (Chen et al., 2009). GO terms with a *p*-value of <0.05 were considered significant. The resulting GO terms were collapsed into categories of related terms using REVIGO (Supek et al., 2011).

Analysis showed that these proteins were primarily involved in regulation of biological quality, cellular organisation and the cell cycle. Many of the proteins in the dataset were associated with more than one GO term, contributing to several segments of the pie chart in **Figure 3.6.** Regulation of biological quality is a general term that pertains to "any process that modulates a qualitative or quantitative trait of a biological quality. A biological quality is a measurable attribute of an organism or part of an organism, such as size, mass, shape, colour, etc." In this case, the analysis suggests that proteins that maintain characteristics of the fibroblast changed in expression in response to TGF-β_1, possibly transitioning towards a myofibroblastic phenotype. Forty proteins were associated with this GO term, many of them involved with actin polymerization and actin-based transport, which are important in contraction, motility, cell division and rearrangement of the cytoskeleton, which are consistent with changes in cell morphology and differentiation.

Sixty-six proteins were associated with the related GO term, cellular component organisation. This term collapses down further into organelle/protein complex subunit/macromolecular complex/cytoskeleton/spindle organisation. It also includes the localisation of these components within the cell and includes cell cycle processes in general,

indicating rearrangement of the cell and all of its components taking place in response to TGF-β_1 in order to form a myofibroblast.

Twenty-three proteins within the dataset were linked specifically to cell cycle-related processes, particularly kinesins and replication proteins. Antigen processing and presentation was also a significant GO term due to the presence of kinesins and proteasomes within the dataset. Enrichment of microtubule–based process was due to the kinesins present, as well as dynein heavy chain 7 and tubulin-α-1a.

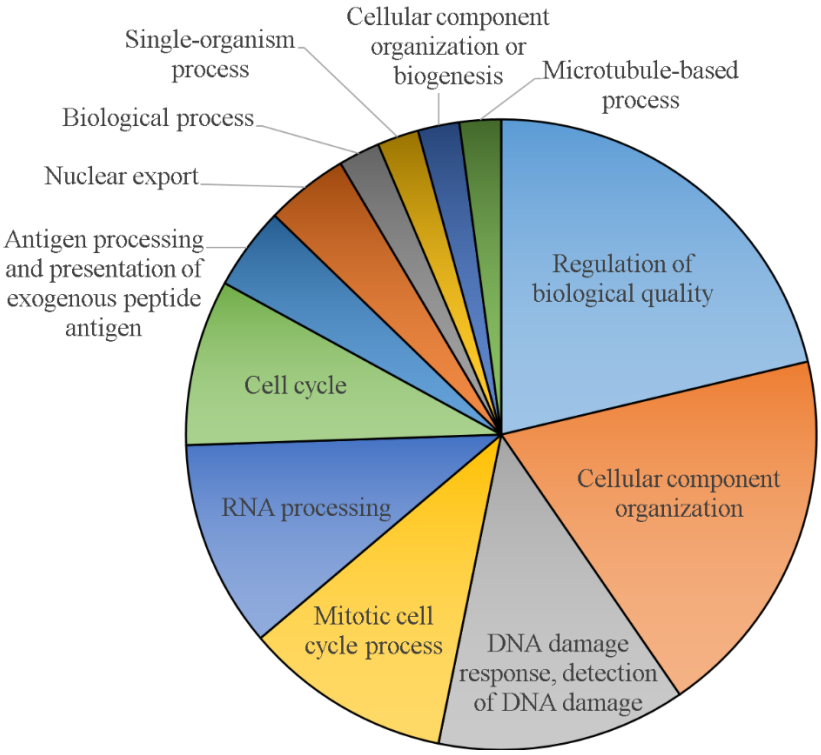


Figure 3.6 Gene ontology enrichment analysis of differentially expressed proteins and proteins expressed only in TGF-β1-treated fibroblasts or only in control from the cellular proteome analysis. Individual GO terms were combined into related terms using REVIGO and are reflected in the pie chart. Size of the terms is adjusted (log10*p*-value) to reflect the *p*-values.

RNA processing and DNA damage response processes were identified as being associated with this dataset. Just 4 proteins were related to DNA damage, all being ribosomal or replication proteins. Proteins related to RNA processing included alanyl-tRNA synthetase and tRNA methyltransferase 1, as well as RNA-binding proteins and nucleoporins. These proteins either were increased in expression or were only detected after TGF-β_1 treatment,

suggesting an increase in translation and synthesis of proteins, potentially required for fibroblast differentiation.

In total, the significantly differentially expressed proteins in this experiment were associated with 74 GO terms, summarized into 12 biological processes, demonstrating that TGF-β_1 affects many proteins that have the same or similar functions and are involved in multiple processes, ensuring the initiation of a large and robust cellular response to this growth factor.

A direct protein interaction network was generated for these differentially expressed proteins or those expressed only in one condition. This was performed using Metacore™ (GeneGo, St. Joseph, MI, USA) and visualized in Cytoscape (version 3.2.1) (Shannon et al., 2003) by Akul Singhania (Faculty of Medicine, University of Southampton NHS Foundation Trust). This analysis highlights protein interactions in the dataset, and the direction of change in abundance was included in the analysis. As shown in **Figure 3.7**, the largest hub in the network was β-catenin, which was significantly upregulated in TGF-β_1-treated fibroblasts and had interactions with nine other proteins in the dataset. β-catenin is a key transcriptional regulator of the Wnt signalling pathway, which plays a major role in embryonic development. Most of its interactions were through binding or transcriptional regulation, including regulation of an important extracellular matrix protein, connective tissue growth factor (CTGF), identified in TGF-β_1-treated fibroblasts only. The second largest hub was calpain-1, a calcium-sensitive protease with roles in cytoskeletal remodelling, apoptosis and the cell cycle (Janossy et al., 2004) that was significantly downregulated in TGF-β_1-treated fibroblasts.

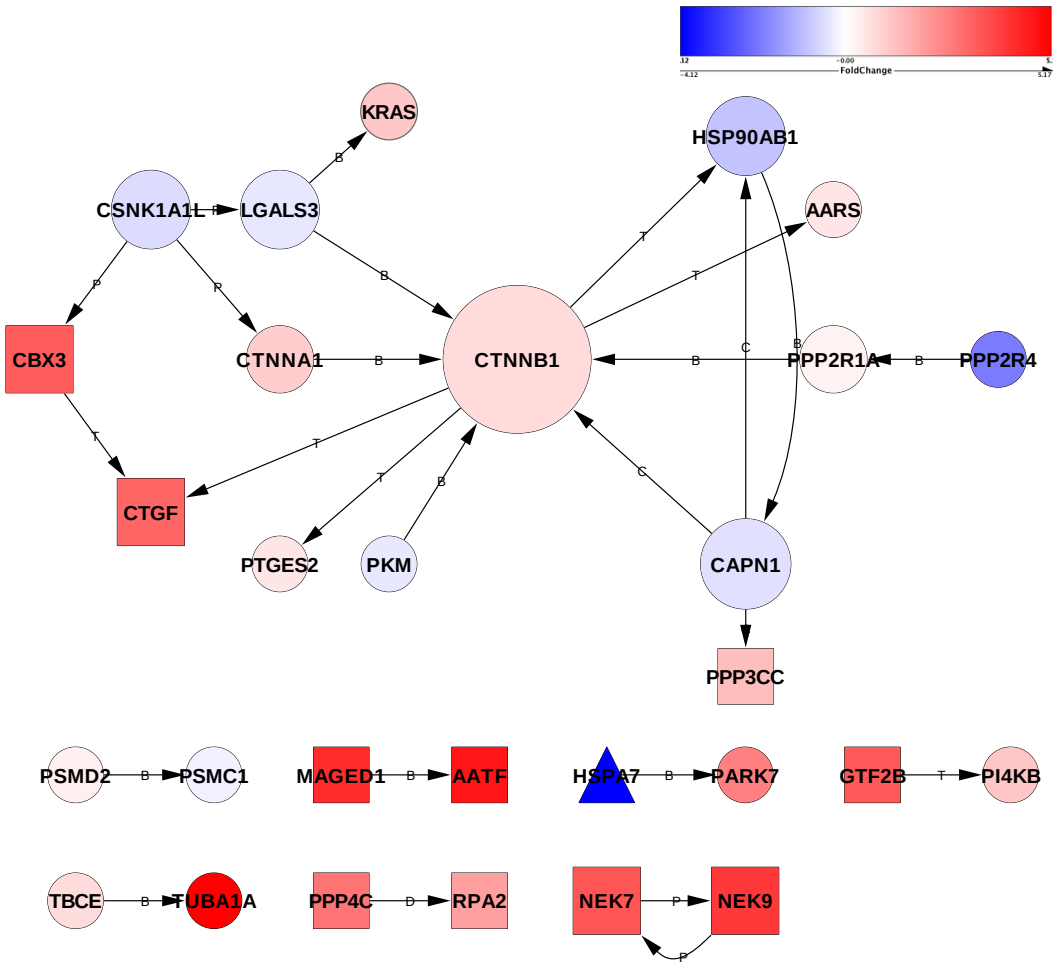


Figure 3.7 Direct protein interaction network of differentially expressed proteins and proteins expressed only in TGF-β1-treated MRC-5 fibroblasts or only in control from the cellular proteome analysis. The size of the node is reflective of the number of interactions and the colour indicates upregulation (red) or downregulation (blue) of proteins in TGF-β_1-treated fibroblasts compared to control. Shape of the node indicates differentially expressed proteins (circle), proteins only in TGF-β_1-treated fibroblasts (square) and only in control (triangle). Edge labels indicate binding (B), cleavage (C), dephosphorylation (D), phosphorylation (P) and transcription regulation (T). Arrow directions indicate which protein is having an effect on (or being affected by) another protein.

3.3.6. Analysis of the MRC-5 fibroblast secretome revealed that many more proteins are secreted after TGF-β_1 stimulation than under normal conditions

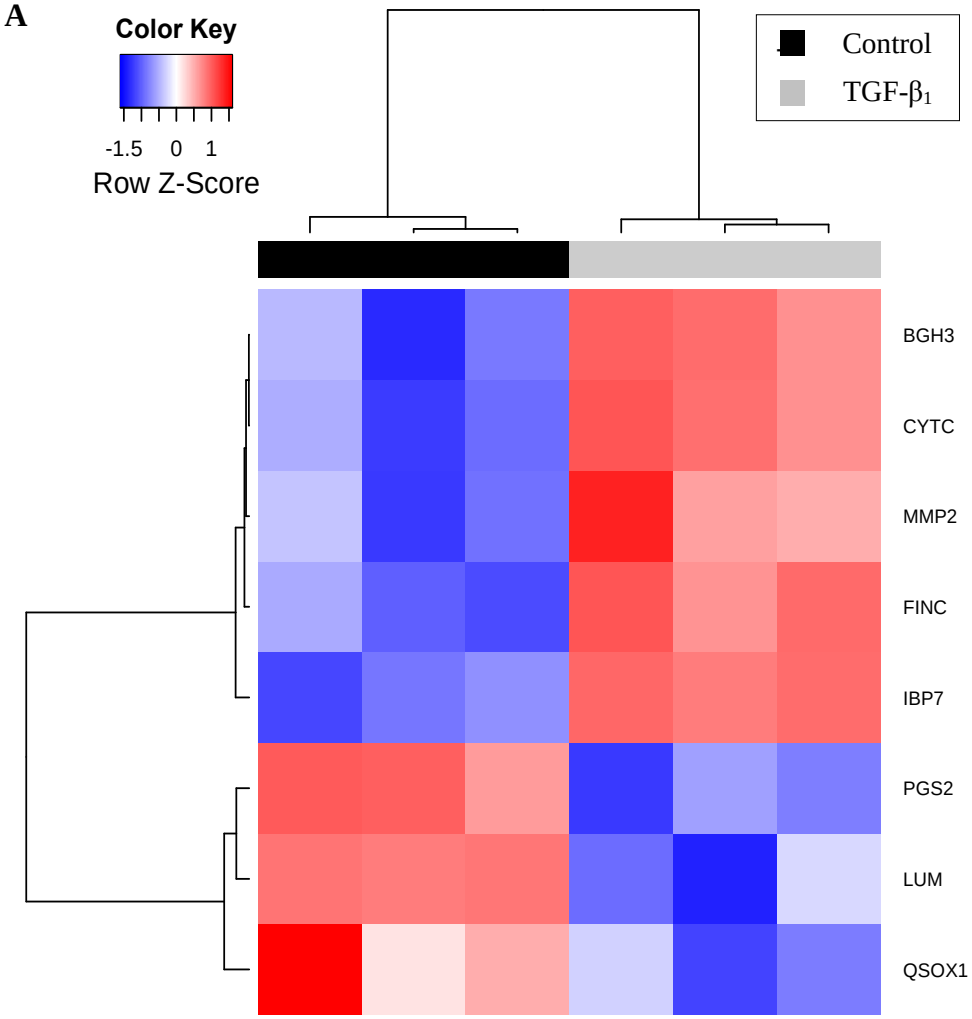
Myofibroblast secretion of extracellular matrix components during wound healing contributes to the creation of a scaffold for tissue repair, and dysregulation of this process is

important in the pathogenesis of fibrotic disease. The fibroblast secretome was analysed in addition to the intracellular proteome in order to generate a more extensive protein expression profile of the fibroblast and myofibroblast, and to identify changes in the secreted proteins between cell states, including ECM proteins and additional proteins important in this process.

In the secretome dataset, eight proteins had a statistically significant difference in expression between control and TGF-β_1-treated fibroblasts by Student's T-Test. The results are reflected in a heat map in **Figure 3.8A,** in which protein expression values are again shown to cluster to condition. **Figure 3.8B** includes the six proteins that were only quantified in TGF-β_1-treated fibroblasts. There were no proteins identified only in control fibroblasts. Heat maps were created using R.

These 14 proteins were taken forward for further analysis. The list of significantly regulated proteins and their corresponding *p*-values can be found in **Appendix B.3-B.4**.

A


Color Key
-1.5 0 1
Row Z-Score
Control
TGF-β1
BGH3
CYTC
MMP2
FINC
IBP7
PGS2
LUM
QSOX1

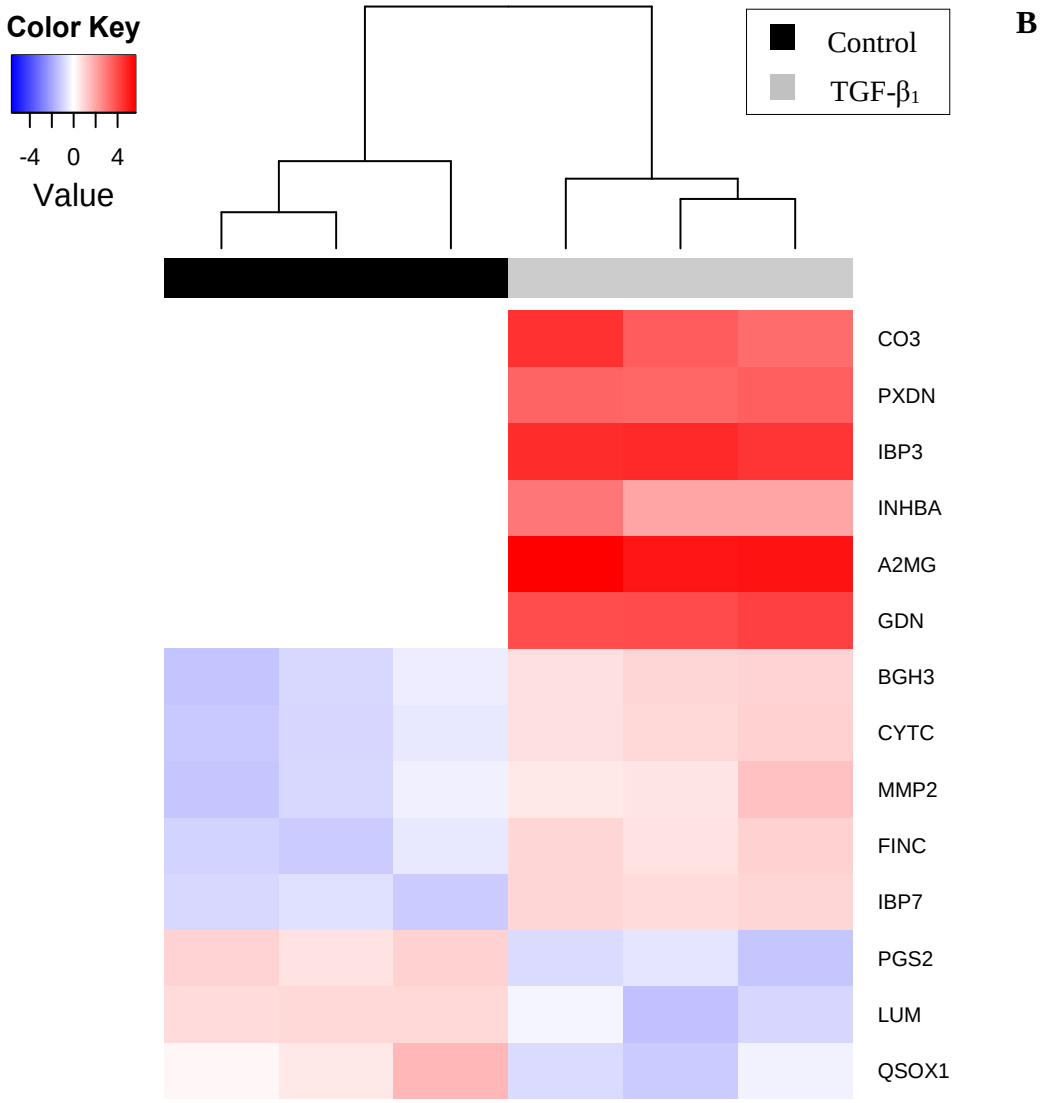


Figure 3.8 Heat maps illustrating the A) differentially expressed proteins and B) differentially expressed proteins as well as proteins expressed only in TGF-β1-treated fibroblasts from the secretome analysis. Hierarchical clustering was performed using Pearson correlation metric to calculate distances and clustered using complete linkage. The expression values of proteins were averaged and scaled across every row to indicate the number of standard deviations above (red) or below the mean (blue), denoted as the row Z-score.

Go analysis was performed on these proteins as previously (see section **3.3.5**). The proteins significantly regulated in the secretome were primarily involved in the regulation of cell growth, differentiation and organisation of the extracellular matrix; these three processes

accounted for 79% of all enriched GO terms for this dataset. Other enriched terms included developmental processes, response to stimulus and peptide cross-linking (**Figure 3.9**). The changes in abundance of these proteins provide a clue as to how the ECM is being organised under TGF-β_1 treatment. Peroxidasin is secreted following treatment; this has previously been shown to increase during TGF-β_1-induced differentiation of fibroblasts and is involved in ECM formation by co-localisation with fibronectin (Peterfi et al., 2009). Fibronectin, which binds several ECM components, is also increased. Matrix metalloproteinase 2 (MMP-2), a collagenase involved in remodelling, along with alpha-2 macroglobulin which can bind and remove MMP-2, is increased with TGF-β_1 stimulation, suggesting an increase in matrix remodelling activity, positively or negatively. Decreases in secretion of collagen-binding proteins lumican and decorin may be because of this remodelling activity. The peptide-crosslinking GO term is associated with decorin and fibronectin.

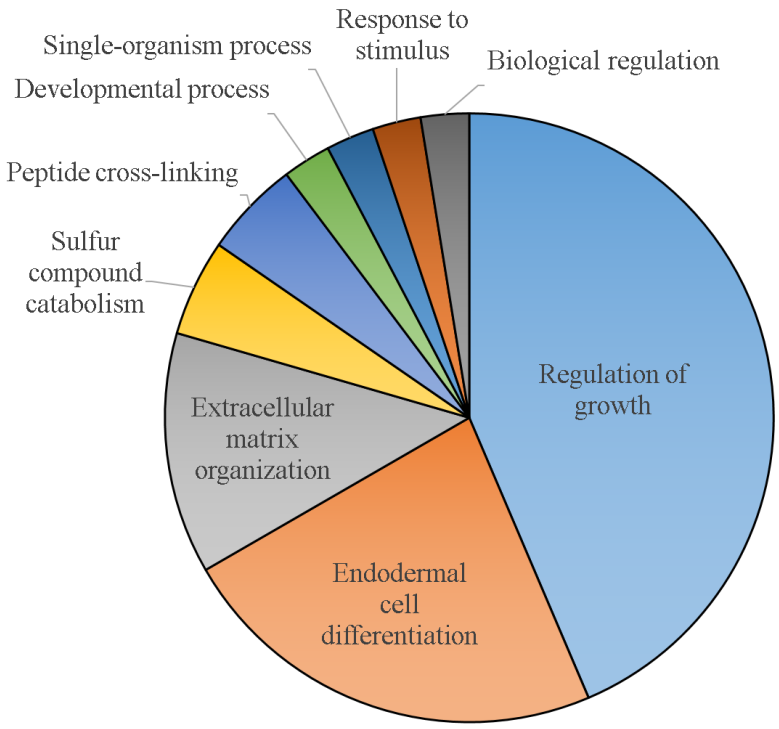


Figure 3.9 Gene ontology enrichment analysis of differentially expressed proteins as well as proteins expressed only in TGF-β1-treated fibroblasts from the secretome analysis. Individual GO terms are combined into related terms using REVIGO that are reflected in the pie chart. Size of the terms is adjusted (log10*p*-value) to reflect the *p*-values.

These proteins were incorporated into the intracellular proteome protein interaction network (see section **3.3.5**) to investigate how the significant fibroblast intracellular and extracellular

proteins interact (**Figure 3.10**). The largest hub generated from this analysis was MMP-2, involved in matrix degradation and remodelling. Metacore analysis showed that β-catenin, the largest cellular protein hub, interacts with MMP-2 through transcriptional regulation. MMP-2 is shown to interact via cleavage with twelve other significantly regulated proteins in the dataset, including the extracellular matrix proteins collagen II-α1, identified with high abundance only in TGF-β_1-treated fibroblasts, peroxidasin, CTGF, and calpain-1.

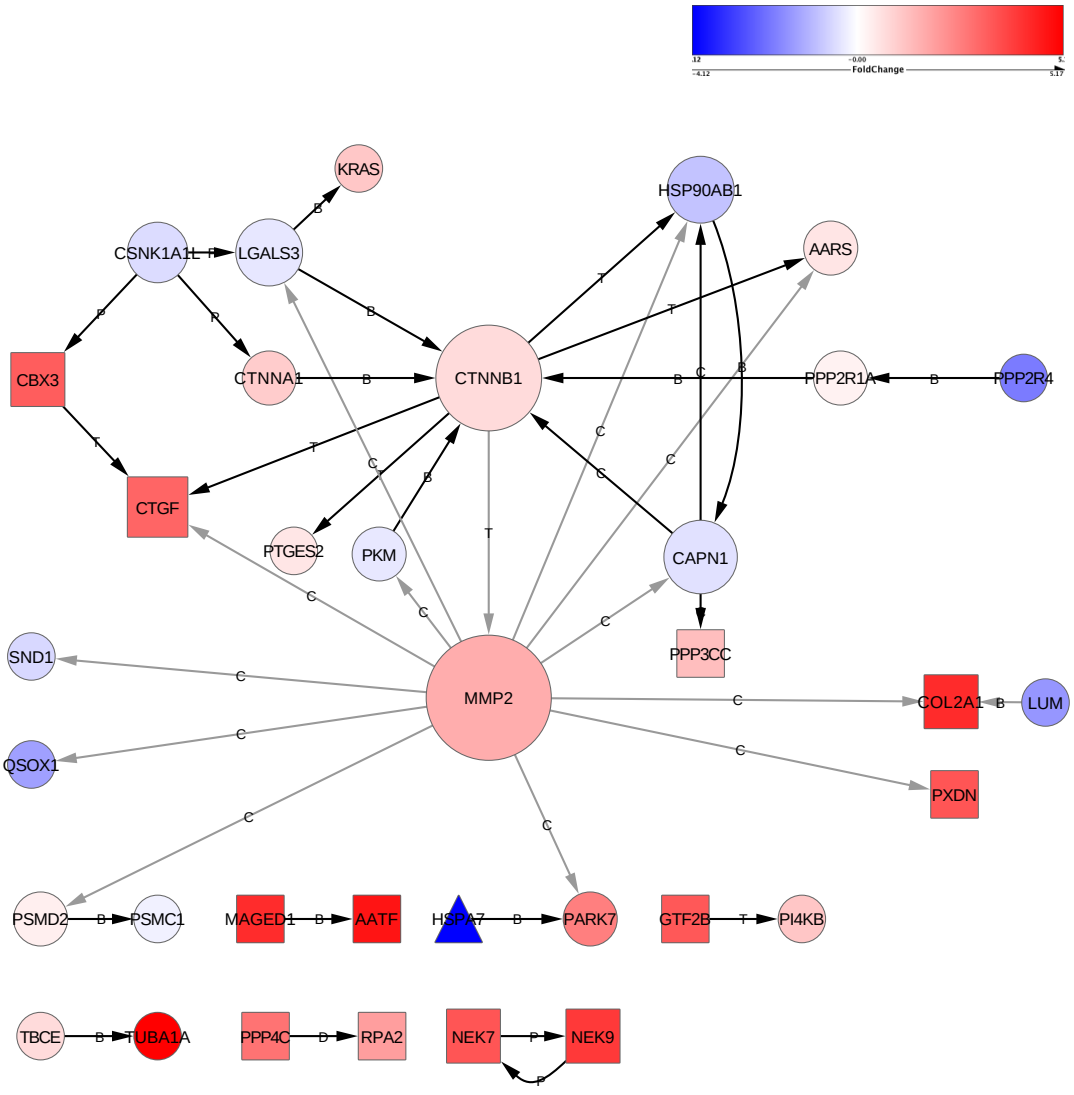


Figure 3.10 Combined intracellular proteome and secretome direct protein interaction network of differentially expressed proteins and proteins expressed only from TGF-β1-treated fibroblasts and those only present from control. The size of the node is reflective of the number of interactions and the colour indicates upregulation (red) or downregulation (blue) in TGF-β_1-treated fibroblasts compared to control. Shape of the node indicates differentially expressed proteins (circle), proteins only in TGF-β_1-treated fibroblasts (square) and only in control (triangle). Edge labels indicate binding (B), cleavage (C), dephosphorylation (D), phosphorylation (P) and transcription regulation (T) and grey edges represent the interactions arising from the addition of proteins from the secretome analysis to the network. Arrow directions indicate which protein is having an effect on (or being affected by) another protein.

β-catenin and MMP-2, the two proteins forming the largest hubs in this network from the intracellular proteome dataset and secretome dataset respectively, were independently analysed by SDS-PAGE and western blotting. Membranes were stripped and re-probed with histone H3 pan as a loading control for β-catenin. Due to the lack of reliable loading controls for secreted proteins, equal volumes of cell-free culture medium taken from equal numbers of cells that were cultured under identical conditions justified equal loading.

Both proteins were found to have increased expression following TGF-β_1 stimulation. A representative image of each result is shown in **Figure 3.11**. This independent verification lends confidence to the validity of these datasets.

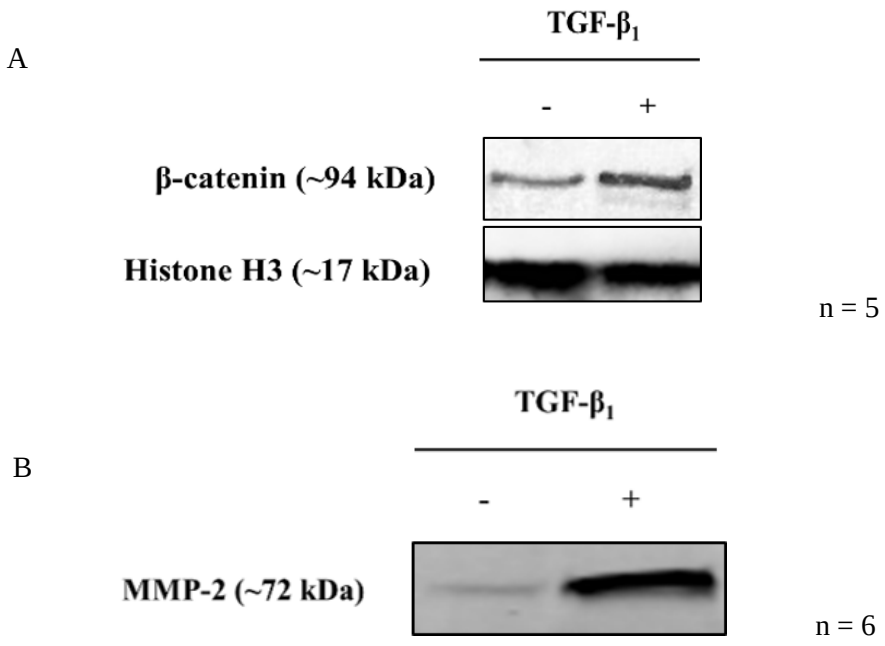


Figure 3.11 Western blot analysis of A) β-catenin and B) MMP-2 from MRC-5 fibroblasts, with and without TGF-β1 treatment. Protein expression was measured after 48 h of fibroblast stimulation with TGF-β_1. Total histone H3 was used as a loading control for β-catenin. Equal cell-free culture medium volumes were used as a loading control for MMP-2.

3.3.7. Fold-change analysis revealed an additional set of differentially regulated proteins

For a more exploratory analysis of the datasets, the criteria were relaxed to include proteins that were quantified in two replicates or more. In this case, 3632 unique proteins quantified

in at least two replicates of untreated and/or TGF-β_1-treated fibroblast lysate samples and 71 proteins quantified in at least two replicates of the secreted protein samples were analysed by fold-change, with a 2-fold cut-off considered as significant.

Using this method of analysis, 500 intracellular proteins had at least a 2-fold change in expression between untreated and TGF-β_1-treated fibroblasts. Nearly 300 proteins were quantified in TGF-β_1-treated fibroblasts only, compared to just 67 only quantified in the control. In the secretome, 12 proteins were differentially expressed by at least 2-fold; 15 were quantified only from TGF-β_1-treated fibroblasts and two were quantified only from control fibroblasts.

Differentially expressed proteins were represented in a protein-protein interaction network using Metacore and Cytoscape software tools by Akul Singhania (**Figure 3.12**). In this network, proteins with differentially regulated expression between fibroblasts and myofibroblasts are represented by a circular node, and proteins identified only in either condition are represented by a square node. The degree of fold-change in expression amongst the proteins is illustrated by the degree of red (upregulated expression) or blue (downregulated expression) colouration of a node. Similarly to the previous networks shown, this network demonstrates interactions between proteins in the dataset generated using a manually-curated database (ThomsonReuters, 2014), and highlights "hubs" of interactions by the increase in size of a node that has many connections.

This protein interaction map allowed the identification of proteins that were part of the same signalling pathway, as the noted interactions between them, due to their roles in the same pathways, caused their representative nodes to cluster together in groups within the network. This analysis revealed that the dataset contained many components of importance to the TGF-β_1 signalling pathways, particularly the non-canonical, non-Smad pathways Ras/MEK/ERK and JNK/p38 MAP kinase signalling pathways, along with an increase in expression of profibrotic growth factors and receptors, such as VEGF-R, FGF-R and CTGF (circled). In the Ras/MEK/ERK pathway for example, N-Ras and K-Ras GTPases had 4.2- and 2.1-fold upregulated expression in TGF-β_1-treated fibroblasts respectively (K-Ras was also significantly regulated according to T-Test (p=0.03)). Raf kinase was detected only in TGF-β_1-treated fibroblasts and MEK1 had upregulated expression of 3.3-fold. These pathways have previously been shown to be involved in the regulation of cell proliferation, differentiation and apoptosis (Roberts and Der, 2007).

Many extracellular matrix proteins also had upregulated expression in TGF-β_1-treated fibroblasts and the secretome. Collagen II-α1 and collagen IV-α1 chains were identified only

in TGF-β_1-treated fibroblasts along with CTGF. Transforming growth factor-β-induced protein has been shown to interact with several extracellular matrix proteins, mainly collagen I, II and IV and fibronectin (Kim et al., 2009), and it had 2.7-fold upregulated expression in TGF-β_1-treated fibroblasts in the exploratory analysis of the cellular proteome, and 4.6-fold upregulated expression in the secretome, in which it was also significantly upregulated according to T-Test (p=0.013). Fibronectin was also upregulated 2-fold in the secretome (p=0.002) along with the matrix metalloproteinase MMP-2, which was upregulated 2.1-fold (p=0.011) following TGF-β_1 treatment.

It is worth noting that pathway analysis tools such as Metacore are mainly curated for cancer, since experimental evidence is used for curation and much work is carried out looking at pathways that are modulated in cancer. This is not optimal for analysing these IPF datasets; however, the aforementioned similarity in pathogenesis between cancer and IPF may mean that it may not be too much of an issue as important pathways are likely to be covered.

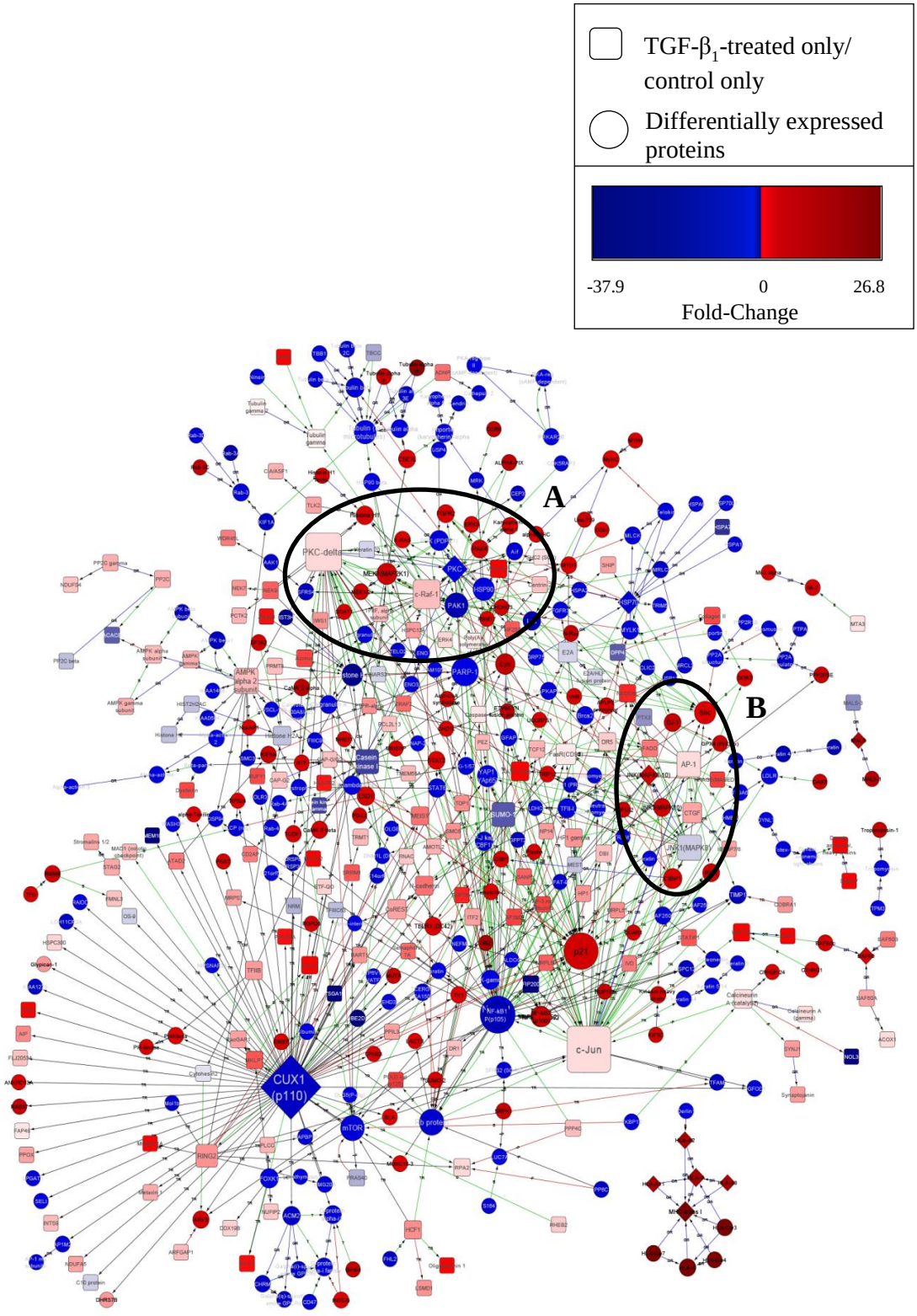


Figure 3.12 Direct interaction network of the proteins with at least a 2-fold difference in expression between untreated and TGF-β1-treated MRC-5 fibroblasts. A and **B** highlight proteins important in the non-canonical TGF-β_1 signalling pathways, plus other growth factors, receptors and extracellular matrix proteins important in fibrosis.

Proteins of particular interest from the fold-change analysis are listed in **Table 3.1** and **Table 3.2.**

Table 3.1 Proteins of interest from the exploratory analysis of the intracellular proteome, with at least a 2-fold increase or decrease in expression.

Differentially expressed proteins between TGF-β_1-treated and control fibroblasts			
Protein accession	**Gene symbol**	**Protein name**	**Log2 fold-change after treatment compared to control**
P10321	HLA-C	HLA class I histocompatibility antigen_ Cw-7 alpha chain	4.81
P30447	HLA-A	HLA class I histocompatibility antigen_ A-23 alpha chain	4.07
P07948	LYN	Tyrosine-protein kinase Lyn	2.73
P53779	MAPK10	Mitogen-activated protein kinase 10	2.67
Q15582	TGFBI	Transforming growth factor-beta-induced protein ig-h3	2.64
Q8IZL9	CDK20	Cyclin-dependent kinase 20	2.62
Q5VW36	FOCAD	Focadhesin	2.60
Q71DI3	HIST2H3A	Histone H3.2	2.36
P22492	HIST1H1T	Histone H1t	1.92
Q02750	MAP2K1	Dual specificity mitogen-activated protein kinase kinase 1	1.87
P38936	CDKN1A	Cyclin-dependent kinase inhibitor 1	1.66
Q9NYF8	BCLAF1	Bcl-2-associated transcription factor 1	1.55
P51452	DUSP3	Dual specificity protein phosphatase 3	1.53
Q8WUP2	FBLIM1	Filamin-binding LIM protein 1	1.51
Q00653	NFKB2	Nuclear factor NF-kappa-B p100 subunit	1.50
Q9Y608	LRRFIP2	Leucine-rich repeat flightless-interacting protein 2	1.48
O96013	PAK4	Serine/threonine-protein kinase PAK 4	1.43
P24821	TNC	Tenascin	1.14
P01116*	KRAS*	GTPase Kras*	1.14
P02452	COL1A1	Collagen alpha-1(I) chain	1.13
Q9GZQ3	COMMD5	COMM domain-containing protein 5	1.10
P29353	SHC1	SHC-transforming protein 1	1.10
Q16832	DDR2	Discoidin domain-containing receptor 2	1.09
O00507	USP9Y	Probable ubiquitin carboxyl-terminal hydrolase FAF-Y	1.08
P35221*	CTNNA1*	Catenin alpha-1*	1.02
P19838	NFKB1	Nuclear factor NF-kappa-B p105 subunit	-4.76
Q13153	PAK1	Serine/threonine-protein kinase PAK 1	-2.77
P06239	LCK	Tyrosine-protein kinase Lck	-2.77

Q9Y2U8	LEMD3	Inner nuclear membrane protein Man1	-1.94
P42226	STAT6	Signal transducer and activator of transcription 6	-1.66
P22607	FGFR3	Fibroblast growth factor receptor 3	-1.59
P09486	SPARC	SPARC	-1.27
O95411	TIAF1	TGFB1-induced anti-apoptotic factor 1	-1.17
P35052	GPC1	Glypican-1	-1.11
Q13671*	RIN1*	Ras and Rab interactor 1*	-1.11
Q68CZ2	TNS3	Tensin-3	-1.02

Proteins in TGF-β_1-treated fibroblasts only

Protein accession	Gene Symbol	Protein name	Log2 Expression
Q6NXT2	H3F3C	Histone H3.3C	5.96
Q02388	COL7A1	Collagen alpha-1(VII) chain	3.96
Q6Q0C0	TRAF7	E3 ubiquitin-protein ligase TRAF7	3.93
O75326	SEMA7A	Semaphorin-7A	3.86
Q13158	FADD	Protein FADD	3.71
Q99570	PIK3R4	Phosphoinositide 3-kinase regulatory subunit 4	3.60
Q8NEM2*	SHCBP1*	SHC SH2 domain-binding protein 1*	3.54
P29279*	CTGF*	Connective tissue growth factor*	3.11
Q9Y2J4	AMOTL2	Angiomotin-like protein 2	3.03
P04049	RAF1	RAF proto-oncogene serine/threonine-protein kinase	2.72
O14763	TNFRSF10B	Tumor necrosis factor receptor superfamily member 10B	2.41
Q9NYR9	NKIRAS2	NF-kappa-B inhibitor-interacting Ras-like protein 2	1.99
P05412	JUN	Transcription factor AP-1	1.90
Q05655	PRKCD	Protein kinase C delta type	1.89
P30453	HLA-A	HLA class I histocompatibility antigen_ A-34 alpha chain	1.69
P25445	FAS	Tumor necrosis factor receptor superfamily member 6	1.24

Protein in control fibroblasts only

Protein accession	Gene Symbol	Protein name	Log2 Expression
Q96BJ3	AIDA	Axin interactor_ dorsalization-associated protein	-3.86
P16188	HLA-A	HLA class I histocompatibility antigen_ A-30 alpha chain	-3.71
Q96B36	AKT1S1	Proline-rich AKT1 substrate 1	-2.82
P45983	MAPK8	Mitogen-activated protein kinase 8	-2.39
P42773	CDKN2C	Cyclin-dependent kinase 4 inhibitor C	-1.98

Table 3.2 Proteins of interest from the exploratory analysis of the secretome, with at least a 2-fold increase or decrease in expression.

Differentially expressed proteins between TGF-β$_1$-treated and control fibroblasts			
Protein accession	**Gene Symbol**	**Protein name**	**Log2 fold-change after treatment compared to control**
Q01995	TAGLN	Transgelin	1.96
P68032	ACTC1	Actin_ alpha cardiac muscle 1	1.87
P21810	BGN	Biglycan	1.85
P23528	CFL1	Cofilin-1	1.57
P12111	COL6A3	Collagen alpha-3(VI) chain	1.45
Q15582*	TGFBI*	Transforming growth factor-beta-induced protein ig-h3*	1.44
P08253*	MMP2*	72 kDa type IV collagenase*	1.10
P08572	COL4A2	Collagen alpha-2(IV) chain	1.02
P02751*	FN1*	Fibronectin*	1.00
Proteins from TGF-β1-treated fibroblasts only			
Protein accession	**Gene Symbol**	**Protein name**	**Log2 Expression**
P01024*	C3*	Complement C3*	3.79
Q92626*	PXDN*	Peroxidasin homolog*	3.43
P17936*	IGFBP3*	Insulin-like growth factor-binding protein 3*	4.62
Q96KK5	HIST1H2AH	Histone H2A type 1-H	0.63
Q9BUD6	SPON2	Spondin-2	1.95
Q02809	PLOD1	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 1	1.48
P14618*	PKM	Pyruvate kinase isozymes M1/M2	2.21
O14556	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	1.24
Q14393	GAS6	Growth arrest-specific protein 6	2.55
P20908	COL5A1	Collagen alpha-1(V) chain	3.07

* = $p<0.05$

In light of the differential regulation of proteins in the MAPK signalling network in fibroblasts responding to TGF-β_1 challenge, an experiment was carried out to find out whether there was an overall pro-proliferative effect following treatment. As previously mentioned, the Raf/Ras/MEK/ERK pathway has roles in regulating cellular proliferation and survival; three of these proteins had significantly upregulated expression with TGF-β_1 treatment and the transcription factor c-Jun was identified only with treatment. The MAPK JNK3 (MAPK10) was upregulated 6-fold, however JNK1 (MAPK8) was only identified in control, thus making it more challenging to speculate on the resulting cellular effects of the

JNK differential regulation. JNK proteins (c-Jun N-terminal kinases) have been reported to be either pro-apoptotic (often in relation to stress) or anti-apoptotic depending on cell type, nature and longevity of the stimulus and the activity of other signalling pathways (Liu and Lin, 2005; Dhanasekaran and Reddy, 2008). In a similar manner, proforms of two proteins of the NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) transcription factor group, p100 and p105, were significantly upregulated and downregulated according to fold-change analysis, respectively. NF-κB has also been shown to have pro-apoptotic or anti-apoptotic effects depending on context and stimuli (Lin et al., 1999; Kaltschmidt et al., 2000). Increases and decreases in the expression of the same groups of proteins makes it more challenging to determine the direction of change of the resulting biological process.

A methylene blue assay was performed to investigate whether proliferation was much higher with TGF-β_1 treatment, or if the treated fibroblasts were proliferating at a similar rate to the control. Fibroblasts were treated in the absence or presence of 10 ng/ml TGF-β_1 for 48 hours. Cells were fixed in formal saline before and after treatment, and stained with methylene blue. Absorbance was measured at 630 nm and differences in absorbance values were compared between 0 hours (T0) and 48 hours (T48) (see section **2.3**).

The results showed that there was a significant difference in growth between 0 and 48 hours, but there was no significant increase or decrease in proliferation with TGF-β_1 stimulation (**Figure 3.13**). This suggests that the MAPK network is possibly antagonised or involved with other signalling pathways that result in different, potentially counteractive effects on proliferation. A larger, clearer effect on proliferation may have been observed had the fibroblasts been exposed to TGF-β_1 for a longer period than 48 hours.

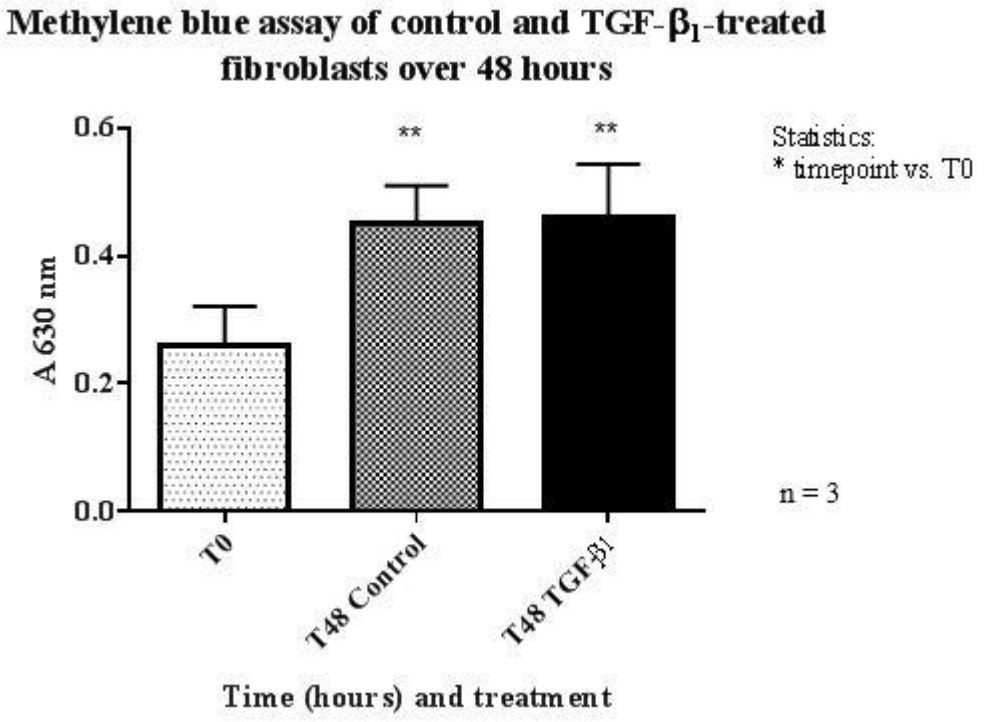


Figure 3.13 Methylene blue assay of MRC-5 fibroblasts. Cells were fixed before treatment at T0 and after 48 hours (T48) with or without treatment with 10 ng/ml TGF-β_1, and stained with methylene blue to determine the rate of cell growth during treatment. Data are expressed as measured absorbance at 630 nm across four replicates. Results were analysed by an unpaired Student's T-Test (T0 vs T48 control p=0.004; T0 vs T48 TGF-β_1 p=0.007).

3.4. Discussion

To evaluate the efficacy of FK228 treatment for IPF patients in a proteomic context, it is important to understand how the proteomes of matrix-secreting differentiated fibroblasts differ from those of undifferentiated fibroblasts. In this study, a method has been developed to globally profile the protein expression of fibroblasts and discover how this differs upon stimulation with the profibrotic growth factor TGF-β_1, in order to study myofibroblast formation at the molecular level. This study utilised the latest proteomics technology to maximise the number of protein identifications from complex whole cell lysates and supernatants, by employing OFFGEL fractionation in combination with nano-UPLC, data-independent acquisition and ion mobility separation. In addition, the Hi3 method of

quantification was used to calculate estimations of absolute amounts of each protein (Silva et al., 2006) in order to compare abundances.

More than 5000 fibroblast proteins were quantified in this study, providing to the author's knowledge the most comprehensive characterisation of TGF-β_1-induced myofibroblast formation to date, having generated both intracellular and extracellular proteomic profiles.

There have been some reports in the literature of studies comparing fibrotic and non-fibrotic lung proteome profiles, specifically and non-specifically to IPF. Malmström *et al.* in 2001 investigated how the proteome changes upon TGF-β_1 treatment in a human fibroblast cell line. Here, proteins were separated by 2D gel electrophoresis and protein spots that differed between the two conditions were in-gel digested and measured using MALDI-TOF-MS. The authors highlighted seven differentially expressed proteins and showed that actin isoforms and tropomyosin were upregulated upon TGF-β_1 treatment (Malmstrom et al., 2001). In a follow up study, the effect of TGF-β_1 on cytoskeletal proteins in fibroblasts was investigated and analysis revealed 18 proteins with significantly increased expression after TGF-β_1 treatment out of a total of 29 proteins identified (Malmstrom et al., 2004b), most of which were associated with the cytoskeleton. These included tropomyosin-α1 chain, caldesmon, calponin-3 and transgelin-2, which also had upregulated expression in this study, although not significantly so. The induction of primarily cytoskeletal proteins following TGF-β_1 treatment correlates well with our findings, as cytoskeleton-based cellular organisation and actin-based protein transport were processes in which many of the proteins differentially expressed in our study were involved.

Korfei *et al.* compared the proteomic profiles of lung tissue from IPF patients and healthy donors to identify proteins potentially involved in overall IPF pathogenesis (Korfei et al., 2011). This group also used 2D gel electrophoresis and MALDI-TOF mass spectrometry, identifying 89 proteins that had differential expression. Fifty-one of those were upregulated in IPF. These were mainly involved in the epithelial cell stress response, particularly endoplasmic reticulum stress and the Unfolded Protein Response (UPR), which is activated upon accumulation of un- or misfolded proteins, resulting in apoptosis of the cell if unsuccessful. Proteins indicative of fibroblast activation and ECM deposition were also identified as upregulated. They identified similar sets of upregulated proteins in IPF when comparing the proteomes of lung tissue from patients with IPF and fibrotic non-specific interstitial pneumonia (fNSIP), another IIP (Korfei et al., 2013). Additionally, when the proteomes of the two IIPs were compared, proteins involved in defence against oxidative stress were upregulated in fNSIP, which may suggest a reason why fNSIP has a better prognosis than IPF.

In this study, gene ontology analysis revealed that the differentially expressed intracellular proteins were mainly involved in cell cycle regulation, protein synthesis and cellular organisation. In context, this suggests that TGF-β_1 causes an increase in the transcription of genes associated with these processes in order to modify the proliferation and differentiation rates of the fibroblast, and the upregulation of organisational cytoskeletal proteins may be linked with the cell's change in morphology towards a differentiated, contractile state. Many differentially expressed secreted proteins had a role in organising or structuring the extracellular matrix, including peptide-crosslinking, which results in a stiffening of the fibrous ECM. Other enriched biological process included regulation of cell growth, response to stimulus and cell differentiation, which suggests involvement of these secreted proteins in the activation of fibroblasts in response to TGF-β_1, and subsequent myofibroblast formation.

β-catenin and MMP-2 were the largest hubs in the interaction network containing all significantly differentially expressed proteins, suggesting a prominent role for these proteins in forming a myofibroblastic phenotype due to their interactions with many proteins indicated to be involved in the process.

There is considered to be complex cross talk between the TGF-β_1 and Wnt/β-catenin signalling pathways (Lam and Gottardi, 2011). It has been demonstrated that in dermal fibroblasts Wnt signalling increases TGF-β_1 expression (Carre et al., 2010) and TGF-β_1 upregulates β-catenin expression at the protein level (Cheon et al., 2004). Small-interfering-RNA (SiRNA) knockdown of β-catenin reverses Wnt-induced increases in TGF-β_1 expression and α-SMA expression in mouse embryonic fibroblasts, suggesting that myofibroblast formation is dependent on Wnt/β-catenin signalling (Carthy et al., 2011). However, forced activation of β-catenin in adult normal lung fibroblasts does not promote expression of α-SMA or of important ECM proteins (Lam et al., 2011). Taken together, this suggests that the cross talk between the two pathways is cell–type and context-dependent (Lam and Gottardi, 2011). This study has shown that β-catenin is significantly upregulated following TGF-β_1 stimulation in a human foetal fibroblastic cell line, and Wnt signalling could be expected to be upregulated during foetal fibroblast growth and development.

A model for β-catenin's role in the pathogenesis of fibrotic disorders suggests that its activation promotes fibroblast proliferation and migration to sites of injury through coregulation of fibrotic gene targets (Lam and Gottardi, 2011), so this protein could be an important biomarker in fibrotic disease. A study by Kim *et al.* (2011) used SiRNA knockdown of β-catenin in a mouse model of idiopathic pulmonary fibrosis. They showed that levels of TGF-β_1 and β-catenin, as well as MMP-2 were increased in tissue homogenate of bleomycin-treated mice, but were all reduced after SiRNA knockdown of β-catenin,

demonstrating that β-catenin influences the expression of MMP-2. In immunohistochemical staining of the IPF lung, MMP-2 has been identified in myofibroblastic foci and surrounding ECM, but not in control lung (Selman et al., 2000). In addition, TGF-β_1 has been shown to increase the expression and activity of MMP-2 in the medium of fibroblasts from fibrous tumours (Lee et al., 2010). The data from this study has shown that this is also the case in foetal pulmonary fibroblasts.

Through the exploratory fold-change analysis, it was found that many different cellular signalling pathways are modulated in response to TGF-β_1; several tyrosine kinase receptors have altered expression levels, such as vascular endothelial growth factor receptors and fibroblast growth factor receptors, with many effector proteins downstream of these receptors showing altered expression. Several of these are involved in more than one signalling pathway. Many differentially regulated proteins were involved in the non-canonical TGF-β_1 signalling pathways Ras/MEK/ERK, JNK and p38 (**Figure 3.14**). These pathways regulate some of the same transcription factors, such as AP-1 transcription factor (Whitmarsh and Davis, 1996). In this dataset, AP-1, specifically the c-Jun protein, was detected only in TGF-β_1-treated fibroblasts. AP-1 is associated with many signalling events, the most prominent being cell proliferation, differentiation and apoptosis (Ameyar et al., 2003).

The fibroblast proliferation rate under challenge by TGF-β_1 was investigated in the light of observed differential abundance changes within groups of proteins, such as JNKs and NF-κB proforms, but no change was observed. As mentioned previously, both of these protein groups have been shown to bring about different effects on growth and apoptosis depending on cellular context. It was shown in lung fibroblasts from patients with scleroderma, another lung disease involving fibrosis, that JNK levels were elevated/constitutive and this contributed to the persistence of a myofibroblastic phenotype (Shi-Wen et al., 2006). NF-κB has also been shown to have cross talk with JNK, for example through anti-apoptotic downregulation of JNK in response to the proinflammatory cytokine tumour necrosis factor-alpha (TNF-α) (Tang et al., 2001; Wullaert et al., 2006). A positive signal from one or more additional signalling pathways other than RAS/MERK/ERK and JNK/p38 may be required before proliferation would increase. Potentially, in this context, more than one protein from these groups needed to have significantly changed in abundance/activity to project the pro-proliferative signal to c-Jun to bring about a change in proliferation, which was not the case. In addition, the signal from these pathways may have been disrupted by other pathways activated or inactivated by TGF-β_1.

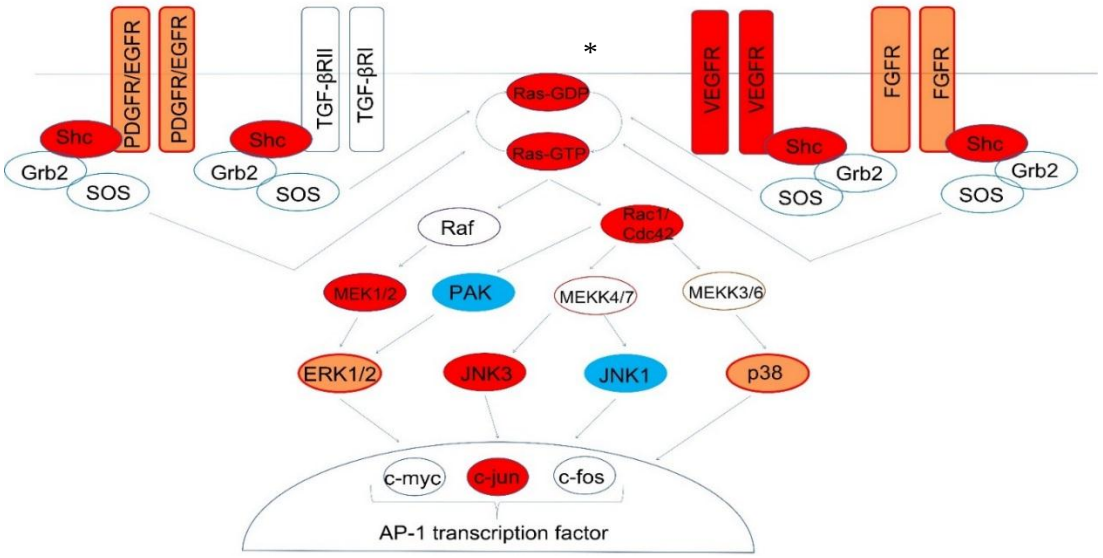


Figure 3.14 A representation of the Ras/MEK/ERK and JNK/p38 signalling pathways. Analysis of the MRC-5 ± TGF-β_1 dataset highlighted pathways involved in TGF-β_1 signal transduction, in particular the non-canonical, non-Smad pathways Ras/MEK/ERK and JNK/p38 that have previously been shown to be involved in the regulation of cell proliferation, differentiation and apoptosis. Red = significantly upregulated following TGF-β_1 treatment; blue = significantly downregulated following TGF-β_1 treatment; orange = identified but not a DEP; white = not identified. * = $p < 0.05$ compared to control.

The Jun/Fos heterodimer of AP-1 has been shown to act together with the Smad3/Smad4 heterodimer in the canonical TGF-β_1 signalling pathway for transcriptional activation, indicating a convergence of Smad-dependent and independent signalling (Zhang et al., 1998). Interestingly, no Smad proteins were detected in this study, with focus instead on the non-canonical pathways. The protein expression of α-SMA was also not significantly increased; preliminary experiments using RT-qPCR and western blotting showed that mRNA expression had increased more than 4-fold after 48 hours of TGF-β_1 stimulation and that protein expression had also increased. It was detected in the cell lysates by mass spectrometry; however, it had no significant change in expression. In the cell, α-smooth muscle actin is involved in fibril formation for the contraction of the myofibroblast, and it is the fibril formation detected in the intact cell that is more commonly used as a biomarker for myofibroblast formation rather than an increase in abundance of the protein itself (Hinz et al., 2007). Therefore, it is more an increase in protein activity (increased cross-linking, cell contraction) than an increase in expression that one may expect to see following TGF-β_1-treatment.

Differential regulation of cell cycle proteins including cyclin-dependent kinases and their inhibitors suggests disruption in cell cycle checkpoints, which would also affect cellular proliferation. The presence of 10% FBS in the culture medium during treatment may also have encouraged proliferation of both untreated and treated fibroblasts to a point at which any pro-proliferative effects of TGF-β_1 may not be obvious (treatment conditions of the fibroblasts for the methylene blue assay were identical to those for the proteomic experiment to reflect the outcome of treatment).

FBS was used in the proteomic experiments to promote cell survival and to prevent cellular stress caused by the experimental setup, however, when analysing the medium by mass spectrometry in preliminary experiments this large amount of serum masked the lower abundant proteins also in the medium, resulting in non-detection. Therefore, in the analysis of the secretome, the fibroblasts were cultured in serum-free media during treatment and harvesting of the secreted proteins, to prevent this from occurring. Initial experiments suggested that the cells were not overly stressed when cultured in these conditions, with no significant cell death or change in morphology (data not shown); however, they proliferated much less and were therefore cultured to a slightly higher confluence before treatment was added in comparison with the cell lysate experiments (~70% confluence as opposed to 50-60%, respectively). The secretome is also a less complex sample type, thus in principal there would be less chance of peptide ion masking by a high concentration of other ions in the sample, resulting in more protein identifications.

The myofibroblastic model used in this study has some limitations due to its experimental setup, in terms of using it to study fibrosis. The protocol was designed for the subsequent studies of primary fibroblasts and their molecular response to a histone deacetylase inhibitor (see **Chapter 4**), therefore profibrotic conditions were not optimal here. The model could have been improved by using low serum in the medium so that the fibroblasts would be less proliferative and more secretory. Treating the cells with TGF-β_1 after a longer culture period, when the cells would be more confluent in the culture plate, would also hinder their proliferation. The addition of L-ascorbic acid, an essential cofactor for fibrillar collagen production (Chojkier et al., 1989) to the culture medium would have promoted extracellular matrix secretion and deposition, and thus would more closely mimic a fibrotic environment and be a better model for a fibrotic lung than the one used in this study, which modelled the more basic fibroblast differentiation process.

Another factor to consider is the impact of the cell line used. The MRC-5 fibroblast cell line is embryonic, and therefore not an ideal cell line for modelling fibrotic disease such as IPF, which has a late onset. One may expect to see differences in the expression of

developmental proteins (such as β-catenin) between these cells and cells from IPF patients. However, it is still a suitable cell line for studying simple fibroblast differentiation, which was the primary aim of this study.

In summary, the work in this chapter presents an in-depth characterisation of myofibroblast formation at the proteomic level by mass spectrometry. The analysis of both the fibroblast and the TGF-β_1-stimulated "myofibroblast" protein expression profiles, both within the cell lysate and within the culture medium, has allowed for extensive intra- and extracellular characterisation of the fibroblast during differentiation. It has enabled comparison between the two cell states to assess the differences in proteins present and the abundance of those proteins, and has highlighted biological processes and signalling pathways that are modulated after cellular differentiation. This method for profiling cells and supernatants can be used for studying diseases such as idiopathic pulmonary fibrosis, and the findings from this study may help identify important mechanisms in fibrotic disease progression, bearing in mind that TGF-β_1 is just one of many fibroblast-influencing signalling molecules involved in this process. For example, in this study, β-catenin and MMP-2 were both highlighted as proteins with important roles during fibroblast differentiation, and both are considered key players in fibrosis.

Based upon the work presented in this chapter, these methods have been used to analyse the cellular proteomes of primary IPF fibroblasts and non-IPF fibroblasts, for the identification of candidate protein biomarkers of IPF pathogenesis, as discussed in the following chapter.

4. Results - Characterisation of the idiopathic pulmonary fibrosis fibroblast proteome with or without treatment with the histone deacetylase inhibitor FK228

4.1. Abstract

The aetiology of idiopathic pulmonary fibrosis, a chronic, fibrosing type of interstitial lung disease, is unknown and treatment options are limited, however, its pathogenesis shares some similarities with cancer. Based on these similarities, possible therapeutics may include anti-cancer drugs. Romidepsin, or FK228, is a histone deacetylase inhibitor (HDACI) approved for use as an anti-cancer drug that has been shown to have anti-proliferative effects on IPF fibroblasts, and therefore may be a candidate for IPF therapy.

The aim of this investigation was to characterise the proteome of IPF fibroblasts in order to identify candidate disease biomarkers, and to identify potential indirect drug targets of FK228 within these cells and monitor their expression. After preliminary experiments to determine an appropriate dose, IPF and non-IPF fibroblasts were challenged with FK228 for 48 hours. Proteins were isolated from whole cell lysates, digested and fractionated before analysis by UPLC-HDMS^{E} with ion mobility enabled.

Over 6500 proteins were identified in total from fibroblasts by mass spectrometry analysis. One hundred and twenty two proteins had significant differences in abundance between IPF and non-IPF primary fibroblasts and herein will be referred to as “potential candidate” IPF protein biomarkers – these proteins cannot be considered biomarkers without extensive independent validation. The majority of them were involved in gene transcription. Over 200 proteins from IPF fibroblasts had a considerable change in abundance when these cells were treated with FK228, indicating that this drug induces a large effect on gene transcription and protein synthesis; some of these proteins were also similarly affected by FK228 in non-IPF fibroblasts.

Proteins identified as potential IPF biomarker candidates were further analysed for differences in expression following FK228 treatment. Future work is to validate a selected panel of these differentially expressed proteins, which could be used in future clinical trials of FK228 treatment in IPF.

4.2. Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease whose pathogenic mechanism is hypothesized to involve an aberrant wound healing response, including the dysregulated proliferation and differentiation of pulmonary fibroblasts following alveolar epithelial injury, and subsequent excessive deposition of extracellular matrix in the interstitium, leading to respiratory failure. The cause is unknown and treatment options are limited.

The pathogenesis of IPF has been noted to share some similarities with cancer pathogenesis, such as uncontrolled cell proliferation, gene expression changes, delayed apoptosis and dysregulated signalling within certain transduction pathways (Vancheri et al., 2010). In light of these observations it is hypothesized that therapeutic strategies designed to treat cancer may have a degree of effectiveness in IPF (Vancheri, 2012). This is already being put into clinical practice with Nintedanib, an approved anti-cancer triple angiokinase inhibitor that has also recently been approved by the FDA as a treatment for IPF, along with another anti-fibrotic/anti-inflammatory compound, Pirfenidone (Karimi-Shah and Chowdhury, 2015). Nintedanib has been shown to reduce angiogenesis in cancer by blocking vascular endothelial growth factor receptor (VEGF-R) (Woodcock *et al.*, 2013) as well as platelet-derived growth factor receptor (PDGF-R) and fibroblast growth factor receptor (FGF-R), inhibiting the binding of their corresponding profibrotic ligands (Rolfo et al., 2013). In IPF it has been shown that by blocking these receptors, Nintedanib interferes with fibrosis *in vitro* by inhibiting fibroblast proliferation, TGF-β-induced differentiation and ECM secretion, in addition to having anti-fibrotic activity in animal models of fibrosis (Wollin et al., 2015). In clinical trials it was also shown to reduce the decline in lung function of IPF patients (Richeldi *et al*., 2014b).

Histone deacetylases (HDACs) are also considered to be potential drug targets for cancer (Choudhary et al., 2009). Acetylation is a protein post translational modification whereby an acetyl group is added to a residue of a protein; HDACs remove acetyl groups from residues such as histone tails, which causes suppression of transcription due to the resulting inaccessibility of DNA to transcription factors (VanderMolen et al., 2011). Genes controlling the cell cycle and apoptosis, important dysregulated processes in both cancer and IPF, are some of the 2-5% of genes estimated to be regulated by histone acetyltransferases (HATs) and HDACs (Emanuele et al., 2008). The HDAC mechanism of action is blocked by HDAC inhibitors, which have been shown to have anticancer activity (Balakin et al., 2007). FK228, also known as Romidepsin and trade name Istodax® was discovered in the early

1990s (Ueda et al., 1994; VanderMolen et al., 2011; Nakajima et al., 1998) and was identified as a Class I HDAC inhibitor. It was approved for treatment of cutaneous T-cell lymphoma (CTCL) in 2009 by the U.S. FDA, and more recently for treatment of peripheral T-cell lymphoma (Glass and Viale, 2013). FK228 acts as a pro-drug; its disulphide bond is reduced as it enters the cell, enabling it to interact with the active sites of HDACs (VanderMolen et al., 2011). It can cause cell cycle arrest, reduce cell proliferation and induce cell death (Piekarz and Bates, 2004).

As mentioned previously, a study by Davies *et al*. (2011) demonstrated that FK228 inhibits proliferation and differentiation of pulmonary fibroblasts from IPF patients in a time and concentration-dependent manner, further supporting the hypothesis that anti-cancer therapeutics can be used for the treatment of IPF. They also showed that it suppresses the expression of alpha-smooth muscle actin, collagen III and HDAC4, which is required for myofibroblast formation (Davies *et al*., unpublished). Further work demonstrated that there was little effect of the HDAC inhibitor on the proliferation of alveolar type 2 epithelial cells isolated from IPF lung biopsy. This indicated little toxicity and that epithelial cell growth in the lung would not be inhibited by this drug if administered to a patient. In addition, they demonstrated that FK228 suppresses bleomycin-induced pulmonary fibrosis in mice, with decreases in the expression of collagen III and collagen I following treatment.

This study follows on from this work by aiming to identify candidate protein biomarkers of IPF in fibroblasts. In clinical trials, data from biomarkers that represent a disease signature are often used to monitor disease and verify treatment efficacy, and can be used as a decision-making tool. There are currently no validated protein biomarkers specifically pertaining to IPF (Vij and Noth, 2012). In this study, the cellular proteomes of IPF fibroblasts and non-IPF fibroblasts were analysed, and differences in protein expression levels were examined between these two datasets in order to identify candidate biomarkers of IPF. These datasets were also compared to data from IPF and non-IPF fibroblasts treated with FK228 to investigate the effects of this HDAC inhibitor on protein abundance and identify additional biological processes that are affected.

For treatment of the primary fibroblasts for proteomic analysis, a dose of FK228 had to be determined that would produce a large enough effect to be able to clearly establish which proteins had had their expression levels modified, but would not be cytotoxic. A series of preliminary experiments were carried out to assess the levels of cell viability and proliferation at a range of concentrations of FK228 to decide an optimum dose, and to determine how this HDAC inhibitor brings about these changes. IPF and non-IPF primary pulmonary fibroblasts were then treated with the optimal concentration of FK228 for 48

hours, and the protein expression of these cells was compared with those of corresponding untreated fibroblasts by proteomic profiling using UPLC-HDMS^{E}.

More than 6500 proteins were identified in this study and over 6300 were quantified, from which a potential candidate IPF biomarker panel could be constructed, and analysed for changes in their expression after treatment with FK228.

4.3. Results

Surgical lung biopsies taken with ethical approval from five patients with a confirmed diagnosis of IPF or from three donors without IPF were used to establish fibroblast cultures. Lung biopsy samples were cut into approximately 2x2 mm sections; pieces of tissue were added to wells of a 6-well culture plate containing Dulbecco's modified Eagle medium (10% foetal bovine serum (FBS)) and scratched into the bottom of the wells. Medium was changed after 7 days of incubation and then replaced every 2-3 days for 2 weeks. Outgrown fibroblasts were dissociated from the wells and cultured in a 75 cm^2 tissue culture flask. At 70-80% confluence cells were collected, counted and split into several 75 cm^2 culture flasks to increase cell number (see section **2.2**).

4.3.1. The metabolic activity of IPF fibroblasts after FK228 treatment decreases with dose above doses of 1 nM concentration

An MTS assay was performed to determine the metabolic activity of IPF fibroblasts after FK228 treatment at a range of concentrations, in order to estimate viability and to determine an optimum dose at which to treat the fibroblasts for subsequent proteomic experiments. The MTS assay is a colorimetric method for measuring metabolic activity, which is determined by the extent of MTS conversion to a formazan product by dehydrogenase enzymes of metabolically active cells in the presence of phenazine methosulfate (PMS). The amount of formazan measured by absorbance at 492 nm is proportional to the number of viable, metabolically active cells. Two time points were chosen to look at the effect of FK228: 48 hours, the proposed time-point for later experiments (approximately one population doubling) and 144 hours, to show its long-term effect.

Fibroblast cultures at ~60% confluence were treated with a range of concentrations of FK228: 0 nM, 0.1 nM, 1 nM, 5 nM, 10 nM and 50 nM. At 48 hours or 144 hours post treatment 20 µl MTS/PMS solution was added to the fibroblasts and cells were incubated for

3 hours at 37ºC, 5% CO_2. Absorbance was read at 490 nm and absorbance values were compared over the FK228 concentration range (see section **2.4**).

As shown in **Figure 4.1**, the metabolically activity of IPF fibroblasts decreased above concentrations of 1 nM FK228 at both time-points. The proportion of metabolically active cells compared to control was similar for the two time-points at concentrations of FK228 up to 5 nM, suggesting that they could tolerate low concentrations for a sustained period.

The percentage of metabolically active fibroblasts at the highest dose of 50 nM FK228 compared to control was approximately 60% and 20% at 48 hours and 144 hours respectively.

From the results of the MTS assay, 2 nM was chosen as the dose of FK228 at which to continue the preliminary experiments, as the IPF fibroblasts were approximately 90% as metabolically active and thus estimated to be 90% as viable/proliferative at this concentration after 48 hours of FK228 treatment as the control. It was hypothesized that at this chosen concentration the fibroblasts would show a large change in protein expression over 48 hours but without a large increase in cellular stress or toxicity.

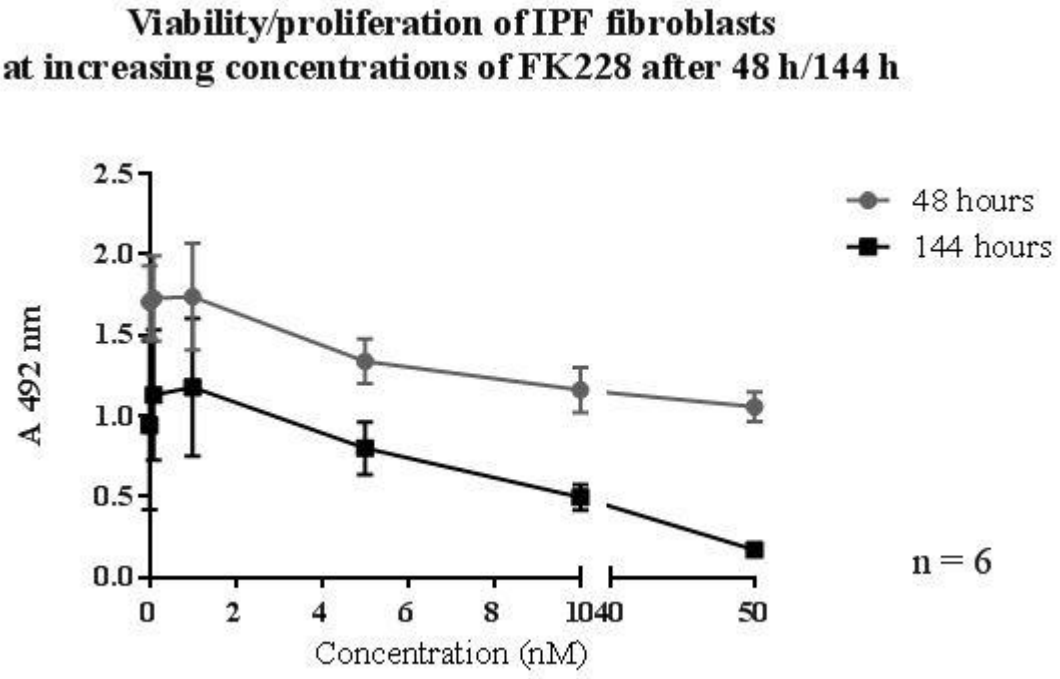


Figure 4.1 IPF fibroblast viability/proliferation after 48 hours/144 hours of treatment with a range of concentrations of FK228. IPF fibroblasts were treated with 0, 0.1, 1, 5, 10 and 50 nM FK228 and their metabolic activity was measured using an MTS assay at two time-points.

4.3.2. The IPF fibroblast cell cycle begins to arrest after 48 hours of FK228 treatment

A decrease in metabolic activity could indicate cell cycle arrest, as growth-arrested cells are less metabolically active. Analysis of DNA content was performed on primary IPF and non-IPF fibroblasts, either untreated or treated with 2 nM FK228, to assess whether cell growth was arrested upon treatment or whether cell death was occurring, in order to explain why the number of metabolically active cells decreased with FK228 stimulation. The percentage of cells in G1 phase and in S phase was calculated by measuring DNA content by flow cytometry.

Fibroblasts at ~60% confluence were treated with 2 nM FK228 or left untreated for 48 hours. Cells were harvested by trypsin-EDTA and the cell pellets were resuspended in 1 ml phosphate buffered saline (PBS). A master mix of PBS, 5 mg/ml PI and 4 mg/ml RNase was added to each sample in a FACS tube, and samples were incubated at 37°C for 30 minutes. Cells were analysed for their DNA content using the FACS Calibur flow cytometer and CellQuest™ software (see section **2.5**).

The percentage of cells in S phase decreased with FK228 treatment (**Figure 4.2**), resulting in decreased proliferation, and a slight accumulation of cells in G1 phase was observed, overall indicating growth arrest in FK228-treated cells. A larger effect was observed in non-IPF fibroblasts than IPF fibroblasts; this may be due to non-IPF cells being more proliferative regardless of treatment. There is conflicting evidence as to whether IPF fibroblasts are less or more proliferative than normal lung fibroblasts (Ramos et al., 2001; Moodley et al., 2003). If they were more proliferative it would mean that more cells would reach G1 phase of the cell cycle and arrest growth during the 48 hours of FK228 stimulation. These results are also demonstrated in flow cytometry histograms in **Figure 4.3**. Cell cycle analysis of FK228-treated IPF fibroblasts compared to control by flow cytometry has also been carried out with similar results by Dr Franco Conforti (Brooke Laboratory, Faculty of Medicine, Southampton University Hospital, University of Southampton) (Davies *et al*., unpublished).

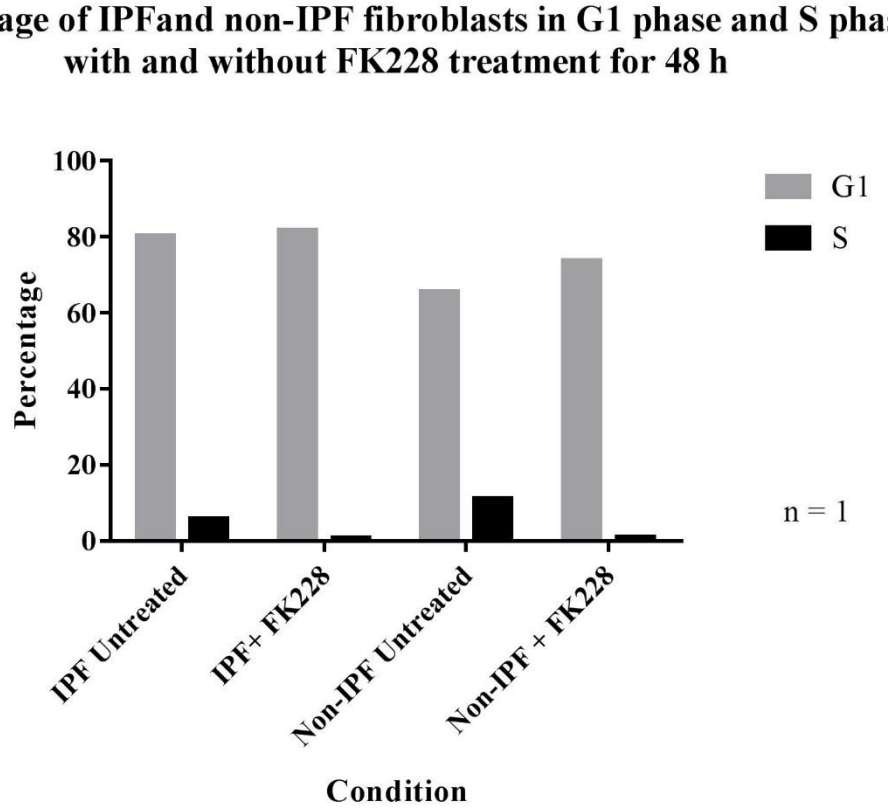


Figure 4.2 The percentage of cells in G1/S phase with or without FK228 treatment.
The proportion of cells in S phase decreased with FK228 treatment, and a higher proportion of cells were in G1 phase after 48 hours of FK228 treatment compared to control.

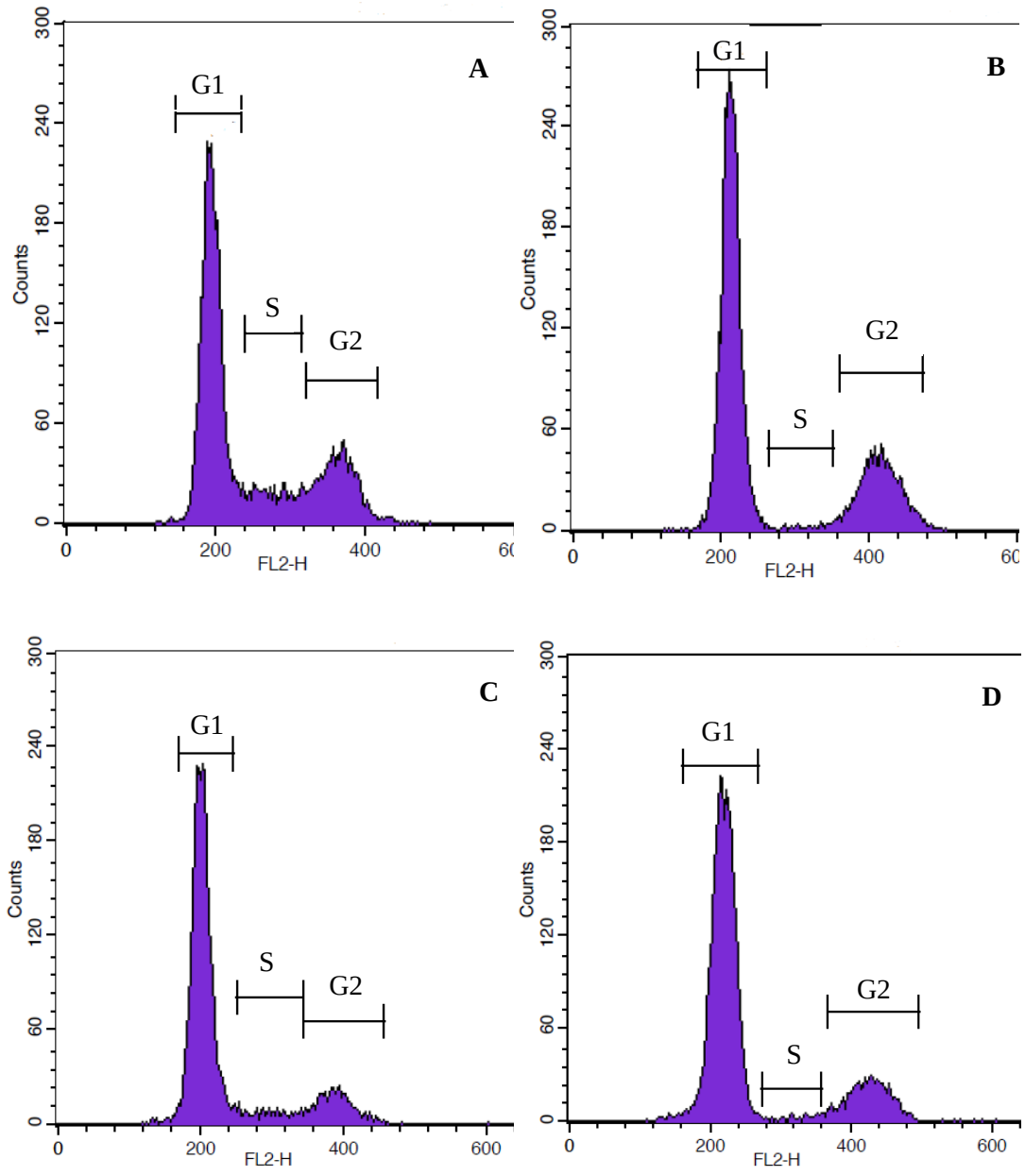


Figure 4.3 Flow cytometry histograms of cells in G1 phase, S phase and G2 phase of the cell cycle, untreated or after treatment with FK228. A clear reduction of the number of cells in S phase can be observed between untreated non-IPF fibroblasts (A) and FK228-treated non-IPF fibroblasts (B). This reduction can also be observed between untreated IPF cells (C) and FK228-treated IPF cells (D).

4.3.3. The non-IPF fibroblasts were more proliferative than the IPF fibroblasts, but proliferation decreased upon treatment with FK228 in both conditions

The growth rates of the primary IPF and non-IPF lung fibroblast cultures used in this study were determined prior to proteomic analysis. Fibroblasts were further assessed for changes in proliferation, both between disease states and in the presence or absence of a 48-hour challenge with 2 nM FK228. Fibroblasts were cultured per condition for 96 hours total; $4.5x10^5$ cells were seeded per condition and treatment was carried out 48 hours before cell counting in trypan blue stain.

Over the course of the culture period, the untreated IPF primary cells almost doubled twice, indicating a doubling time of approximately 48 hours (**Figure 4.4**). However, the number of 2 nM FK228-treated IPF cells had only doubled once, suggesting that the incorporation of FK228 into the cells at 48 hours had induced growth arrest. The number of untreated non-IPF cells over the 96-hour culture time had multiplied almost three times since seeding; their rate of growth and proliferation was higher than that of the untreated IPF fibroblasts, thus their doubling time was shorter. However, the number of FK228-treated cells counted after 48 hours of treatment was less than half compared to untreated non-IPF fibroblasts. It appears that some cells may have doubled twice before treatment due to the shorter doubling time, but proliferation was slowed/arrested shortly after. The difference in cell number following FK228 treatment for both IPF and non-IPF fibroblasts was found to be statistically significant by Student's T Test, as was the difference in number of untreated IPF and non-IPF fibroblasts after the culture period.

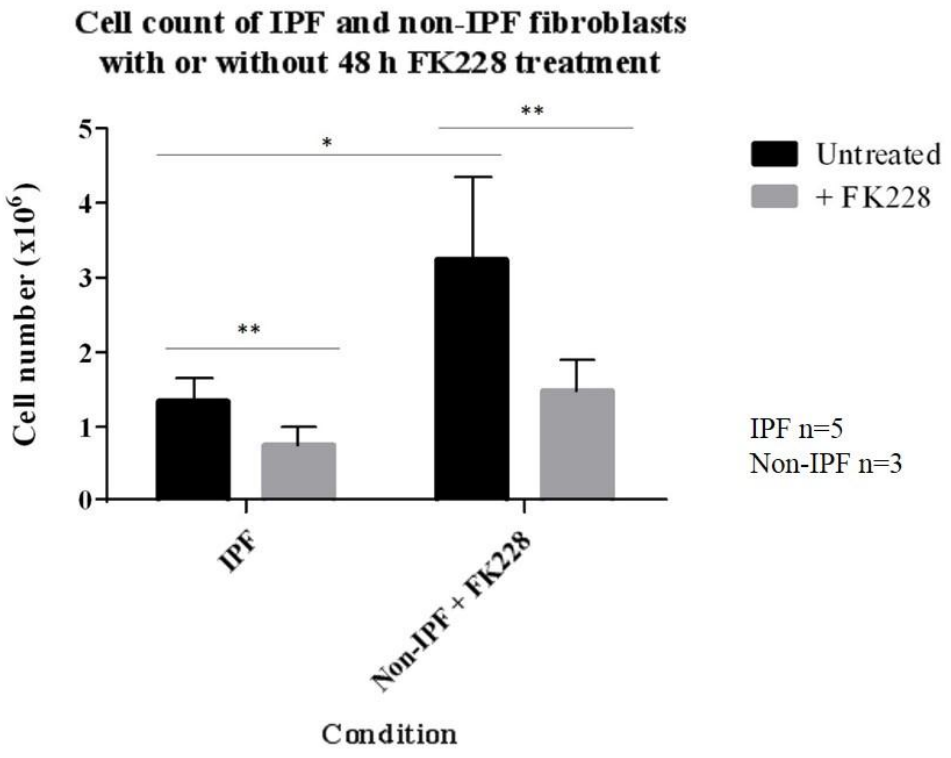


Figure 4.4 Number of primary IPF and non-IPF fibroblasts after 48 hours of stimulation with 2 nM FK228 compared to control. $4.5x10^5$ cells were seeded per condition. Upon harvesting cells were counted using trypan blue stain as outlined in section **2.2.2** to assess the rate of proliferation. Results were analysed by Student's T-Test (IPF untreated vs. IPF + FK228 p=0.001, IPF untreated vs. non-IPF untreated p=0.009, non-IPF untreated vs. non-IPF + FK228 p=0.049).

4.3.4. The acetylation of lysine residues increases following treatment with FK228

Acetylation of proteins within the IPF/non-IPF ± FK228 fibroblast samples was analysed by western blotting prior to mass spectrometry analysis, to validate that the treatment was having the expected effect in increasing the level of acetylation by blocking the action of HDACs. Cell lysates were subjected to SDS polyacrylamide gel electrophoresis and western blotting for the detection of acetylated lysine. Membranes were stripped and re-probed for β-actin as a loading control (see section **2.14**).

Acetylation of histone lysine residues increased in both IPF and non-IPF fibroblasts after treatment in three biological replicates, indicating that the fibroblasts had responded to treatment in the expected manner. An interesting finding is that the basal level of histone

acetylation was lower in IPF fibroblasts than non-IPF fibroblasts. **Figure 4.5** shows a representative image of the increased lysine acetylation of histones upon FK228 treatment. These results validate the histone deacetylase inhibition mechanism of the drug.

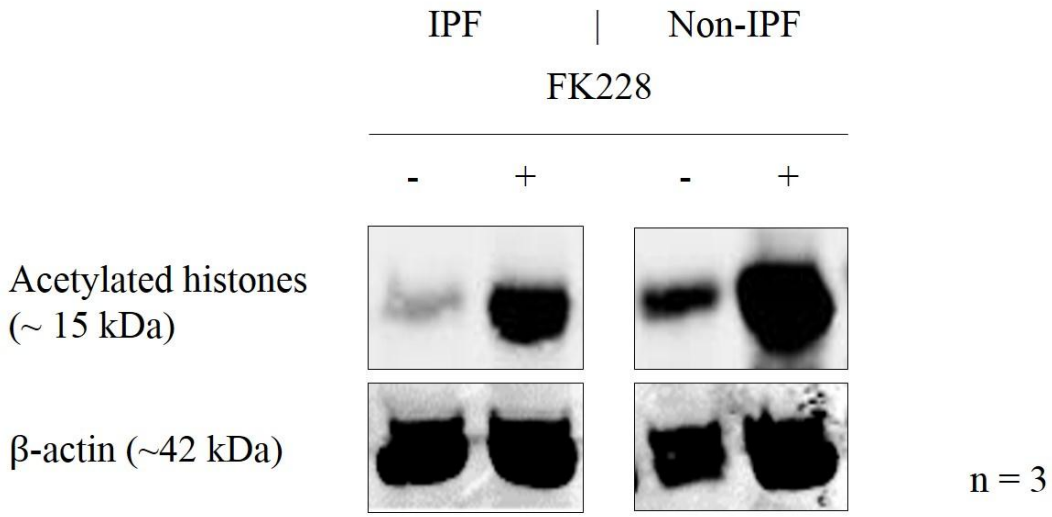


Figure 4.5 Western blot of histone lysine acetylation expression in both IPF and non-IPF lung fibroblasts, with and without FK228 treatment. The expression of acetylated lysines was analysed by western blotting at 48 hours after FK228 treatment, with β-actin as a loading control.

4.3.5. Mass spectrometry analysis of IPF and non-IPF fibroblasts revealed significant differences in their cellular proteomes, which were further modified after FK228 treatment

The proteomic profiles of primary IPF and non-IPF lung fibroblasts were generated using a quantitative, mass spectrometry-based proteomic approach, for the identification of potential candidate protein biomarkers that following rigorous independent validation, could be used in the diagnosis and monitoring of IPF disease progression. Both IPF and non-IPF fibroblasts were treated with the histone deacetylase inhibitor FK228 to determine changes in global protein expression and identify biological processes and signalling pathways that are perturbed with treatment. A previously established method for fibroblast lysate sample preparation for mass spectrometry analysis (see **Chapter 3**) was used to identify and quantify a high number of proteins in order to compare expression levels across all samples.

As illustrated in **Figure 4.6**, at 60% confluence IPF pulmonary fibroblast cultures from 5 patient biopsies and fibroblast cultures from 3 non-IPF donor biopsies were treated in the absence or presence of 2 nM FK228 for 48 hours. Media were collected and cells were harvested using trypsin-EDTA, counted in trypan blue stain as described in section **2.2.2**, washed 4 times in PBS and lysed in 0.1% sodium dodecyl sulfate using a sonication probe (see section **2.6**).

100 µg protein was precipitated by methanol/chloroform extraction prior to reduction for 1 hour, 60 ºC with 1 mM DTT, alkylation for 45 minutes, RT, in the dark with 5.5 mM iodoacetamide, followed by digestion. Proteins were digested firstly with 2 µg endoproteinase Lys-C for 4 hours at 37ºC then 2 µg trypsin overnight at 37ºC. Enolase and ClpB internal protein digest standards were spiked into each sample at 300 fmol concentration before peptides were separated according to isoelectric point by OFFGEL fractionation into 12 peptide fractions, according to manufacturer's instructions. Each fraction was purified using a C18 Empore 96-well solid phase extraction plate to remove residual salts, buffers and contaminants before lyophilisation and resuspension in loading buffer A (3% acetonitrile + 0.1% formic acid) for mass spectrometry analysis (see section **2.9**).

Fractions were analysed by UPLC-HDMS^E. Half of each fraction was injected and peptides were separated by liquid chromatography using a NanoACQUITY UPLC system with a C18 reverse-phase column, at a flow rate of 300 nL/minute over a 3-50% gradient of buffer B (80% acetonitrile/dH_2O + 0.1% formic acid) over 90 minutes. Peptide ions were sprayed into a Waters Synapt G2-S system operating in positive ion mode, with ion mobility enabled before fragmentation. Data was collected in MS^E mode of acquisition, alternating between low energy (5V) and high energy (15V - 45V ramp) scans. Glu-fibrinopeptide (m/z = 785.8426, 100 fmol/µl) was used as LockMass and was sampled every 60 seconds for calibration (see section **2.10**).

Raw data files were processed using Protein Lynx Global Server (PLGS) version 3.0. Data were searched against the human UniProt database using an Ion Accounting algorithm in PLGS 3.0.2 (see section **2.11**). Estimates of absolute quantification were calculated using the Hi3 method of quantification (Silva et al., 2006).

In total, 6544 unique proteins were identified from five different IPF fibroblast biopsy cultures and three different non-IPF biopsy cultures. 6374 proteins were associated with 3 or more unique tryptic peptides and could be quantified.

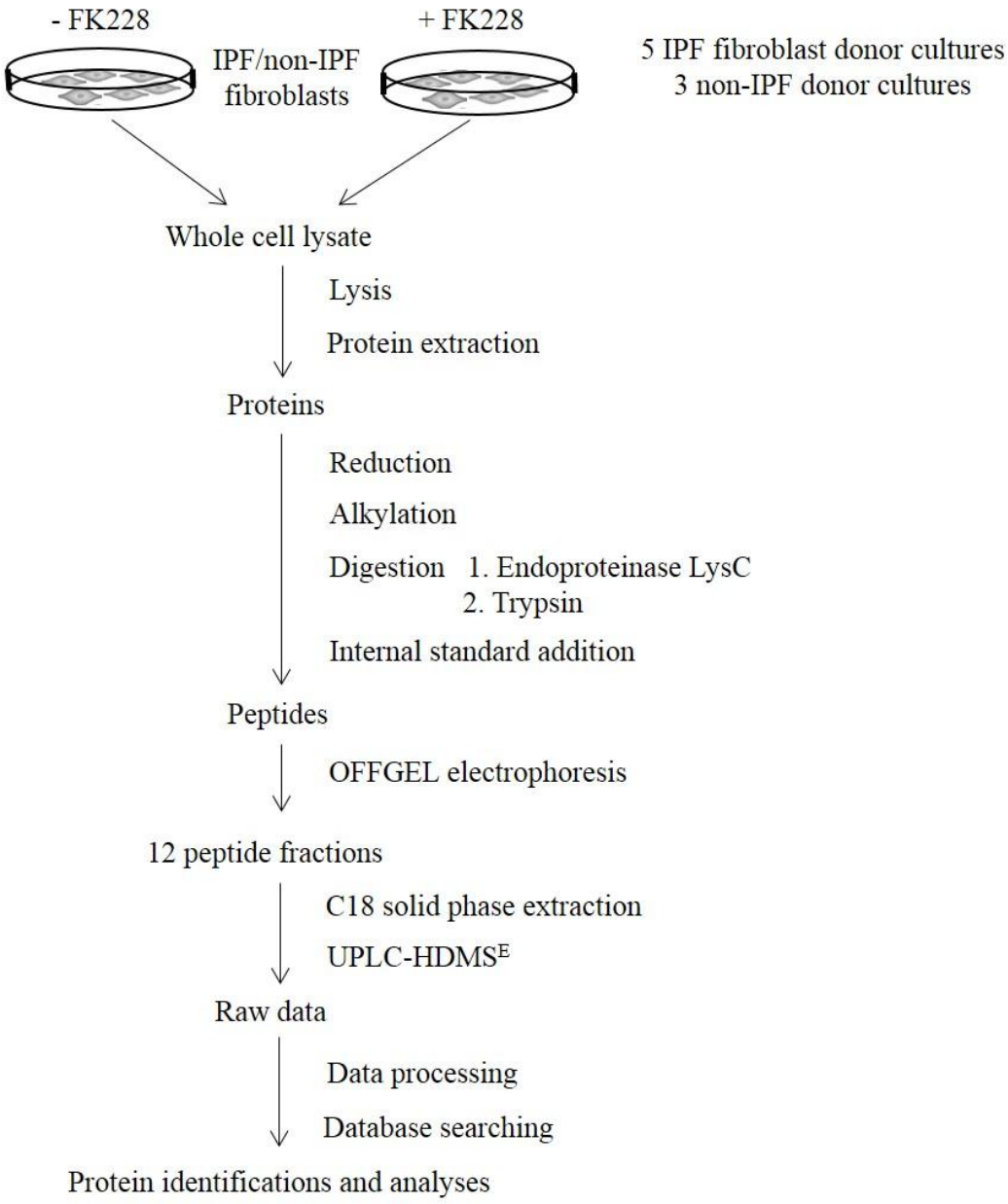


Figure 4.6 A representation of the method workflow used to identify and quantify the proteome of non-IPF fibroblasts and IPF fibroblasts, untreated or treated with 2 nM FK228. Primary IPF fibroblasts were treated with 2 nM FK228 for 48 h and compared against untreated IPF fibroblasts and non-IPF fibroblasts as controls. Protein lysates were reduced, alkylated and digested into peptides using endoproteinase Lys-C and trypsin. Enolase internal reference standard (Waters) was spiked into the digests for Hi3 quantification. Peptides were fractionated by OFFGEL electrophoresis (Agilent). Peptide fractions were separated by UPLC and eluted peptides were sprayed directly into a Synapt G2-S mass spectrometer (Waters) operating in MS^E mode with ion mobility enabled. Data were processed and searched using PLGS 3.0/3.0.2 (Waters) prior to statistical analyses and pathway analyses. Each experiment was performed in triplicate.

A series of analyses were carried out on this dataset, which following normalisation (see section **2.12** and **Appendix A.1**) was split into three sets of comparisons, to be discussed in the following order:

- Untreated IPF fibroblasts vs. untreated non-IPF fibroblasts
  - in order to identify candidate protein biomarkers of IPF
- Untreated IPF fibroblasts vs. IPF fibroblasts treated with FK228
  -to examine the proteomic effects of the drug on IPF fibroblasts
- Untreated non-IPF fibroblasts vs. non-IPF fibroblasts treated with FK228
  -to understand whether the proteomic effects that occur in IPF fibroblasts with FK228 treatment also occur in non-IPF fibroblasts

Quantification using the Hi3 method allowed the direct comparison of protein expression levels across independent donors. Proteins identified across 3 or more replicates of a condition were analysed for differential expression using either a two-tailed unpaired or paired Student's T-Test depending on the dataset. *p*-values < 0.05 were considered significant. Proteins identified in at least three replicates of one condition and in none of the other condition were also considered significant. Proteins were also analysed for differential expression according to fold-change, to reveal the direction of change (up- or down-regulation).

4.3.5.1. Proteomic comparison of IPF and non-IPF fibroblasts

Global protein expression from the cell lysates of untreated IPF and non-IPF fibroblasts were compared for the identification of potential protein biomarkers of this disease, which are currently lacking in clinical practice. The roles of the differentially expressed proteins in the cell between diseased and non-IPF states could also highlight processes that are dysregulated in IPF fibroblasts, which could be important targets for future therapies.

In this comparison of IPF and non-IPF fibroblasts, 1333 proteins were eligible for analysis by Student's T-Test. Of these, 57 were found to be significantly different in abundance ($p<0.05$). Forty-seven of those had overall decreased expression in IPF compared to non-IPF control and 10 showed increased expression overall. Eleven proteins were identified in all three primary non-IPF donor cultures, but in none of the primary IPF donor cultures; conversely, 54 proteins were identified in at least three IPF donor cultures but none of the non-IPF donor cultures (122 significant proteins total). The full list of proteins with significant differential expression can be found in **Table 4.1**.

The list of proteins found to be significant by T-Test are reflected in a heat map in **Figure 4.7**, which shows the direction of change in abundance for each protein (red = increased, blue = decreased).

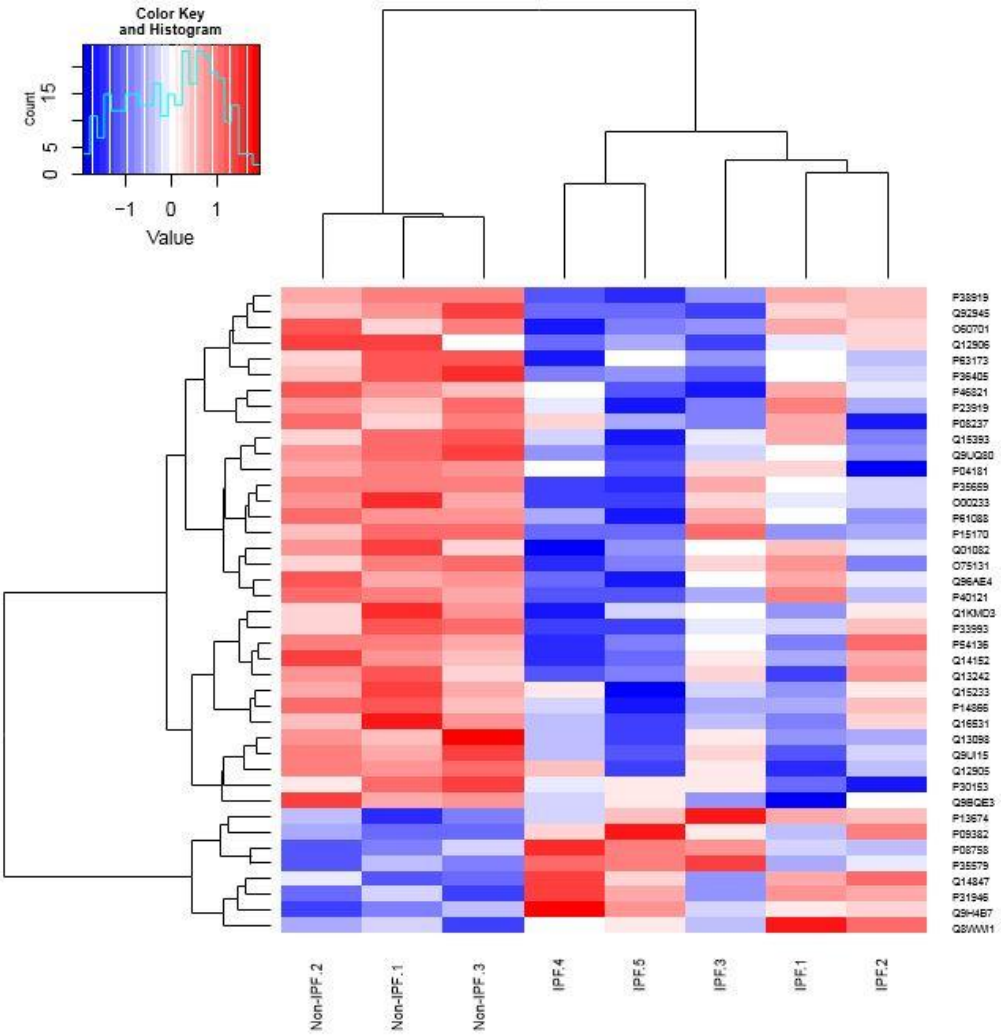


Figure 4.7 Heat map illustrating the differentially expressed proteins between untreated IPF and non-IPF fibroblasts. Heat maps were generated in R using ggplots and d3 heatmap packages. The expression values of proteins were averaged and scaled across every row to indicate the number of standard deviations above (red) or below the mean (blue). Distance between experiments and proteins was calculated and hierarchical clustering was performed using ward linkage.

Protein expression values clustered to condition in this heat map; the three non-IPF primary donor cultures are distinguishable from the IPF fibroblast donor cultures as the expression

values of the proteins between the non-IPF donor cultures are very similar in the same direction of abundance, which is in stark contrast to the IPF donor cultures in the majority of instances. All IPF primary donor cultures cluster together, however, the expression values for each protein shows significant variation between donors. In some cases, a protein that shows decreased expression in IPF in the majority of donor cultures may have increased expression in one or two donor cultures. These results suggest a great deal of variation in gene expression/protein expression of certain proteins, not only between diseased and non-IPF lung fibroblasts, but also between fibroblasts from different IPF patients, potentially due to dysregulation of transcription/translation and thus gene expression.

The proteins with significant differences in abundance plus the proteins only identified in either non-IPF fibroblasts or IPF fibroblasts were analysed for their role within the cell by gene ontology (GO) analysis. This was performed using GOrilla (Eden et al., 2009) to identify GO terms significantly overrepresented for the differentially expressed proteins as well as proteins identified only in one condition (Chen et al., 2009). GO terms with a *p*-value of <0.05 were considered significant. The resulting GO terms were collapsed into categories of related terms using REVIGO (Supek et al., 2011). Enriched gene ontology terms were summarized into those shown in **Figure 4.8**.

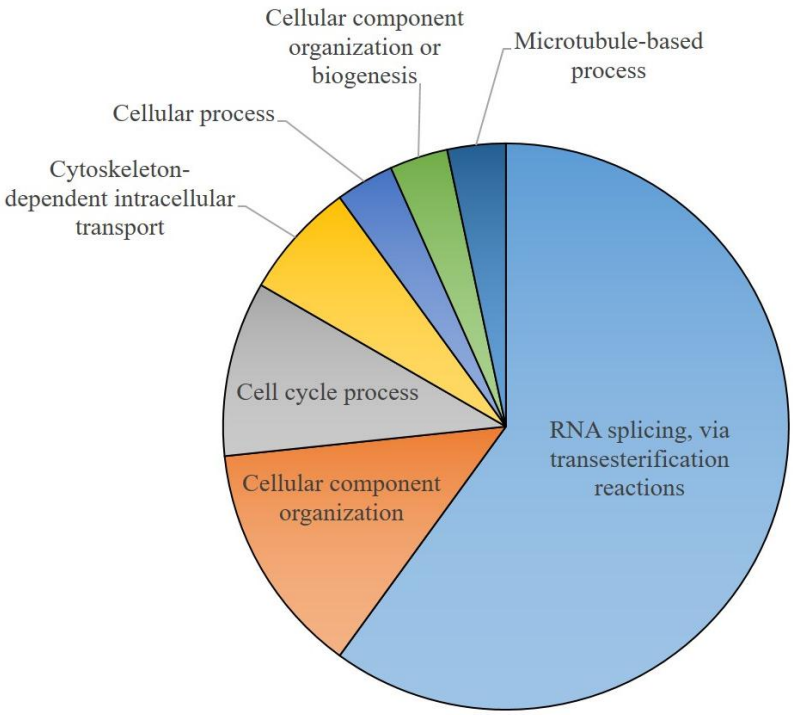


Figure 4.8 Gene ontology enrichment analysis of differentially expressed proteins as well as proteins expressed only in IPF fibroblasts and only in non-IPF fibroblasts. Individual GO terms were combined into related terms using REVIGO that are reflected in the pie chart. Size of the terms is adjusted (log10*p*-value) to reflect the *p*-values.

The most enriched GO term for the statistically significant protein dataset was "RNA splicing, via transesterification reactions". Each term represented is a summarized GO term from several GO terms in order to form a succinct list. This specific umbrella term covers seventeen other enriched terms such as nucleoside triphosphate metabolic process, mRNA metabolic process, RNA processing, RNA splicing and post-transcriptional regulation of gene expression; all relate to transcription and transcriptional regulation. These data indicate that many of the proteins that have a large difference in expression in IPF compared to non-IPF fibroblasts have roles in gene expression and protein synthesis and the regulation of these processes. Additional enriched GO terms include cellular component organisation, cell cycle processes and cytoskeleton-based processes. Changes in these processes in IPF fibroblasts may be being driven by profibrotic cytokines.

4.3.5.2. Candidate protein biomarkers of IPF

The 122 proteins found to be significantly differentially expressed between IPF and non-IPF fibroblasts, either with a *p*-value of <0.05 or only identified in one condition, are listed

below in **Table 4.1** as potential candidate biomarkers of IPF. These proteins have not yet undergone extensive validation and therefore at present this data just provides an indication of proteins that should be considered for further analysis of biomarker suitability.

Fold-change values (positive values indicate an increase in abundance in IPF and vice versa) and *p*-values are listed.

Table 4.1 List of identified potential IPF biomarker candidates from proteomic analyses.

Proteins with a significant difference in expression between IPF and non-IPF fibroblasts			
Protein Name	**Protein Description**	**Fold Change in IPF compared to non-IPF fibroblasts**	**T-Test *p*-value**
Q96DU9	Polyadenylate-binding protein 5	36.170	0.048
Q9H4B7	Tubulin beta-1 chain	9.732	0.018
P05026	Sodium/potassium-transporting ATPase subunit beta-1	2.374	0.030
P31946	14-3-3 protein beta/alpha	2.188	0.028
P13674	Prolyl 4-hydroxylase subunit alpha-1	2.161	0.016
Q8WWI1	LIM domain only protein 7	1.967	0.049
Q14847	LIM and SH3 domain protein 1	1.873	0.036
P08758	Annexin A5	1.573	0.029
P09382	Galectin-1	1.494	0.011
P35579	Myosin-9	1.284	0.018
P14866	Heterogeneous nuclear ribonucleoprotein L	-1.257	0.015
P30153	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	-1.350	0.041
Q01082	Spectrin beta chain_ non-erythrocytic 1	-1.373	0.047
Q92945	Far upstream element-binding protein 2	-1.395	0.031
Q14152	Eukaryotic translation initiation factor 3 subunit A	-1.419	0.033
P08237	6-phosphofructokinase_ muscle type	-1.423	0.044
P04181	Ornithine aminotransferase_ mitochondrial	-1.437	0.049
Q12906	Interleukin enhancer-binding factor 3	-1.486	0.047
P46821	Microtubule-associated protein 1B	-1.513	0.040
Q9BQE3	Tubulin alpha-1C chain	-1.531	0.016

Q15233	Non-POU domain-containing octamer-binding protein	-1.531	0.024
Q6IBS0	Twinfilin-2	-1.556	0.004
Q13098	COP9 signalosome complex subunit 1	-1.574	0.036
Q15393	Splicing factor 3B subunit 3	-1.592	0.027
Q12874	Splicing factor 3A subunit 3	-1.593	0.004
Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2	-1.598	0.029
P54136	Arginine--tRNA ligase_ cytoplasmic	-1.617	0.047
O60701	UDP-glucose 6-dehydrogenase	-1.644	0.043
P63173	60S ribosomal protein L38	-1.661	0.021
Q14980	Nuclear mitotic apparatus protein 1	-1.705	0.009
Q12905	Interleukin enhancer-binding factor 2	-1.732	0.024
P38919	Eukaryotic initiation factor 4A-III	-1.747	0.041
Q9Y2X3	Nucleolar protein 58	-1.749	0.044
P61088	Ubiquitin-conjugating enzyme E2 N	-1.772	0.019
P23919	Thymidylate kinase	-1.781	0.036
Q9Y2S7	Polymerase delta-interacting protein 2	-1.797	0.001
Q9UQ80	Proliferation-associated protein 2G4	-1.883	0.001
Q13242	Serine/arginine-rich splicing factor 9	-1.927	0.047
P36405	ADP-ribosylation factor-like protein 3	-1.939	0.007
O75131	Copine-3	-1.947	0.036
Q16531	DNA damage-binding protein 1	-1.963	0.030
P15170	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A	-2.004	0.024
O15355	Protein phosphatase 1G	-2.038	0.024
P40121	Macrophage-capping protein	-2.069	0.024
Q96AE4	Far upstream element-binding protein 1	-2.135	0.032
O00233	26S proteasome non-ATPase regulatory subunit 9	-2.170	0.015
P25205	DNA replication licensing factor MCM3	-2.199	0.024
P33993	DNA replication licensing factor MCM7	-2.207	0.021
Q92974	Rho guanine nucleotide exchange factor 2	-2.227	0.030
Q14738	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform	-2.542	0.031
Q13630	GDP-L-fucose synthase	-2.556	0.038
O43765	Small glutamine-rich tetratricopeptide repeat-	-2.571	0.018

	containing protein alpha		
P35659	Protein DEK	-2.676	0.024
P48509	CD151 antigen	-2.751	0.030
Q15645	Pachytene checkpoint protein 2 homolog	-2.877	0.046
P08648	Integrin alpha-5	-3.341	0.010
Q9UI15	Transgelin-3	-4.863	0.006

Proteins identified in IPF fibroblasts only		**Log2 Expression**
Q9NZU5	LIM and cysteine-rich domains protein 1	4.51
O75347	Tubulin-specific chaperone A	5.78
O76014	Keratin_ type I cuticular Ha7	2.90
O60361	Putative nucleoside diphosphate kinase	3.93
P48751	Anion exchange protein 3	3.67
Q9H3K6	BolA-like protein 2	2.89
Q13233	Mitogen-activated protein kinase kinase kinase 1	3.54
Q96EL3	39S ribosomal protein L53_ mitochondrial	5.06
Q96C36	Pyrroline-5-carboxylate reductase 2	2.67
P31513	Dimethylaniline monooxygenase [N-oxide-forming] 3	4.66
Q9HAU4	E3 ubiquitin-protein ligase SMURF2	4.33
P10412	Histone H1.4	4.64
O43617	Trafficking protein particle complex subunit 3	3.94
O15078	Centrosomal protein of 290 kDa	6.08
Q96DA2	Ras-related protein Rab-39B	3.04
Q9BQS8	FYVE and coiled-coil domain-containing protein 1	5.00
Q92572	AP-3 complex subunit sigma-1	3.80
Q9Y3C6	Peptidyl-prolyl cis-trans isomerase-like 1	3.68
Q15772	Striated muscle preferentially expressed protein kinase	6.18
P07942	Laminin subunit beta-1	4.76
P05534	HLA class I histocompatibility antigen_ A-24 alpha chain	1.14
P09496	Clathrin light chain A	5.11
Q8NB90	Spermatogenesis-associated protein 5	4.13
O60313	Dynamin-like 120 kDa protein_ mitochondrial	3.76
Q8NB37	Parkinson disease 7 domain-containing protein 1	3.14
O15484	Calpain-5	3.55
Q9BZV1	UBX domain-containing protein 6	3.27
Q8IUE6	Histone H2A type 2-B	5.43
Q14789	Golgin subfamily B member 1	6.27
O60573	Eukaryotic translation initiation factor 4E type 2	3.08
Q6YP21	Kynurenine--oxoglutarate transaminase 3	3.84
Q8TAD7	Overexpressed in colon carcinoma 1 protein	5.46
Q8WZA9	Immunity-related GTPase family Q protein	3.49
P0C7P4	Putative cytochrome b-c1 complex subunit Rieske-like protein 1	4.47
Q9BRP8	Partner of Y14 and mago	3.39
Q8N2U0	UPF0451 protein C17orf61	3.36
Q96T58	Msx2-interacting protein	4.61

Q9UBT6	DNA polymerase kappa	3.04
P48668	Keratin_ type II cytoskeletal 6C	5.22
Q8NHP1	Aflatoxin B1 aldehyde reductase member 4	2.39
Q8N9V7	Testis- and ovary-specific PAZ domain-containing protein 1	5.29
Q5TBB1	Ribonuclease H2 subunit B	6.56
Q6DN03	Putative histone H2B type 2-C	4.94
O75326	Semaphorin-7A	3.94
Q9NP61	ADP-ribosylation factor GTPase-activating protein 3	4.86
Q6P587	Acylpyruvase FAHD1_ mitochondrial	3.25
Q9NRF8	CTP synthase 2	3.76
Q7Z739	YTH domain family protein 3	2.87
Q502W7	Coiled-coil domain-containing protein 38	5.11
A9Z1Z3	Fer-1-like protein 4	2.02
P02461	Collagen alpha-1(III) chain	4.56
Q8WUP2	Filamin-binding LIM protein 1	3.61
Q5T7N2	LINE-1 type transposase domain-containing protein 1	2.58
Q9NSK0	Kinesin light chain 4	2.75

Proteins identified in non-IPF fibroblasts only		**Log2 Expression**
Q8IXM3	39S ribosomal protein L41_ mitochondrial	2.53
P32189	Glycerol kinase	3.35
Q99805	Transmembrane 9 superfamily member 2	5.75
P48454	Serine/threonine-protein phosphatase 2B catalytic subunit gamma isoform	3.02
Q9GZP4	PITH domain-containing protein 1	2.69
O14530	Thioredoxin domain-containing protein 9	3.72
P57740	Nuclear pore complex protein Nup107	4.81
Q96EL2	28S ribosomal protein S24_ mitochondrial	4.23
P00374	Dihydrofolate reductase	2.10
P49247	Ribose-5-phosphate isomerase	3.17
Q99811	Paired mesoderm homeobox protein 2	4.39

As mentioned previously, a protein can only be identified in a sample within a dataset if it was present in amounts above the limit of detection of the mass spectrometer. Therefore, a protein not being identified in a sample could be due to its presence in very low amounts in the sample, or could be because it is not soluble or amenable to measurement by mass spectrometry, rather than it being absent from a condition. This could explain how proteins such as collagen III, known to be present in normal fibroblasts, has been noted as only present in IPF fibroblasts in this experiment.

Following these analyses, IPF fibroblasts were then examined for differences in protein expression with and without treatment with the HDAC inhibitor FK228, to study its molecular effects. The potential candidate protein biomarkers of IPF identified in this study

were also investigated in FK228-treated IPF fibroblasts, to find out what effect the drug has on these particular proteins with aberrant expression in this disease.

4.3.5.3. Proteomic comparison of untreated and FK228-treated IPF fibroblasts

IPF fibroblasts from five different patients were profiled at the protein level, untreated and with 48 hours of FK228 treatment. By doing so the effects of FK228 on protein abundance could be examined, which would reveal the extent of the effects on gene regulation of this HDAC inhibitor. The roles of the differentially expressed proteins were studied to identify any currently unknown processes in the cell that are affected by this drug in addition to changes in transcriptional regulation, growth, proliferation and differentiation.

Proteins identified with estimates of absolute quantification in three or more primary donor cultures of a condition were analysed by Student's T-test or fold-change analysis. In this comparison, 1302 proteins were eligible for analysis by paired Student's T-Test. Of these, 105 were found to have significantly different expression ($p<0.05$). The number of proteins with increased or decreased expression in this dataset was approximately equal. 104 proteins were identified in at least 3 untreated primary donor cultures and not in any after FK228 treatment, and just 7 proteins were only detected after FK228 in three or more donor cultures. The full lists of proteins with significant differential expression can be found in **Appendix C.1-C.3**.

The results are reflected in a heat map in **Figure 4.9**, which shows the direction of change in abundance for each protein (red = increased, blue = decreased). Only proteins from this list that were identified in all samples are shown here (55 proteins).

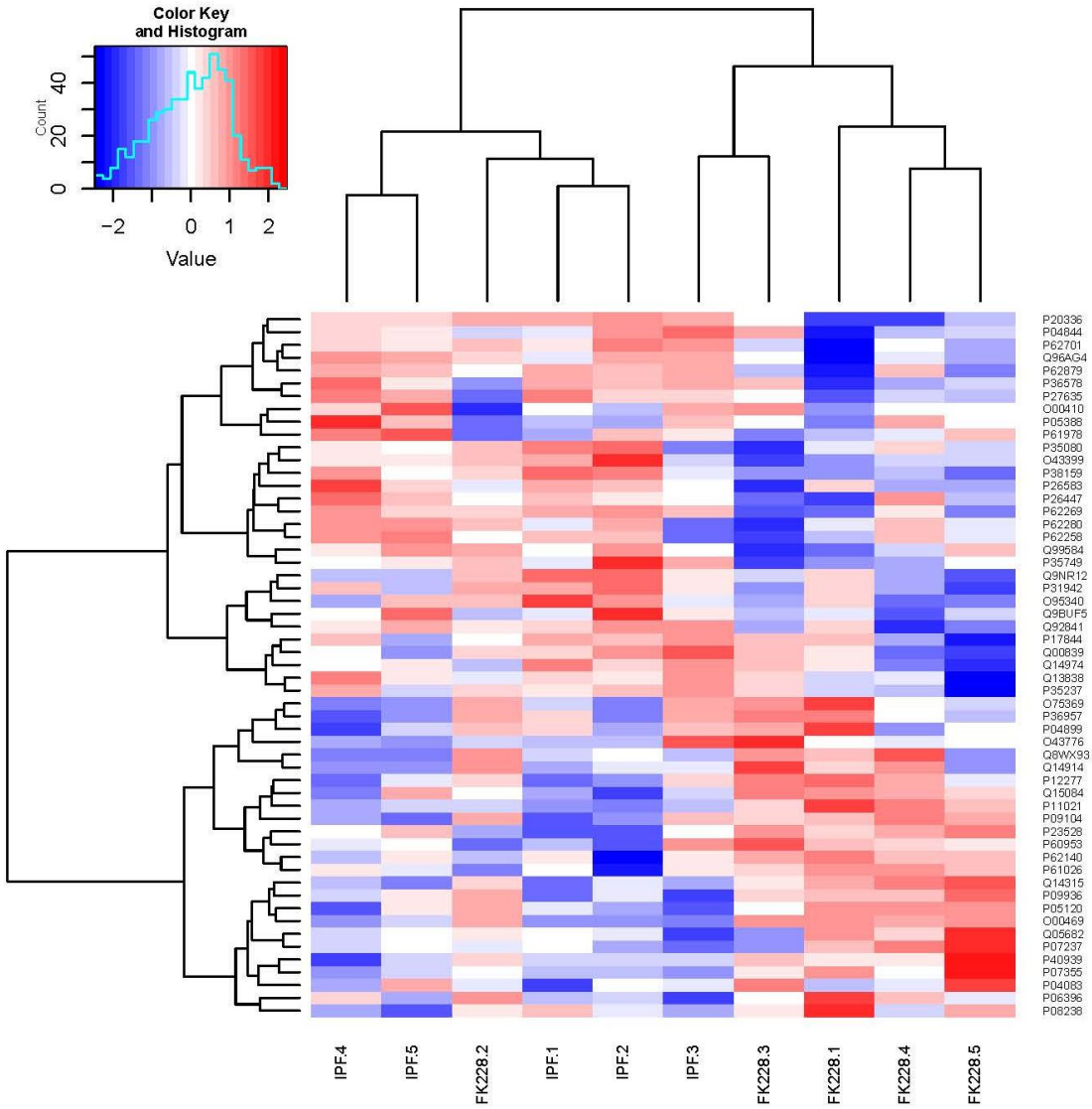


Figure 4.9 Heat map depicting the differentially expressed proteins between untreated and FK228-treated IPF fibroblasts. Heat maps were generated in R using ggplots and d3 heatmap packages. The expression values of proteins were averaged and scaled across every row to indicate the number of standard deviations above (red) or below the mean (blue). Distance between experiments and proteins was calculated and hierarchical clustering was performed using ward linkage.

The heat map reveals that protein expression values do not cluster to condition overall; 4 out of 5 of each condition cluster together but one sample of each condition has clustered to the opposite condition. FK228-treated fibroblasts from IPF patient 2 (FK228.2) have clustered with the untreated IPF samples; this may potentially be because FK228 did not have as much of an effect in fibroblasts from this patient as the others, and therefore the fibroblast proteome resembles more an untreated fibroblast than a treated fibroblast. The fact that it did

not cluster the closest to its untreated counterpart (IPF.2) within the group suggests than there was at least a slight effect on the expression of these proteins with treatment.

Both untreated and FK228-treated fibroblast samples from IPF patient 3 (IPF.3 and FK228.3) cluster to themselves within the FK228-treated group. This suggests that many of these proteins from untreated fibroblasts from this patient do not have similar expression as those of the other untreated IPF fibroblasts. This could perhaps correlate with a difference in disease progression or another clinical variable. The fibroblasts from this patient biopsy could be less fibrotic than fibroblasts from the other IPF donor biopsies, as this sample clusters more closely with other FK228-treated fibroblasts than untreated fibroblasts, assuming that the fibroblasts treated with FK228 have become less fibrotic.

Between the samples that have clustered to condition, there is still large amount of variability in the abundance of the proteins in this heat map between primary donor cultures. This again suggests differences in transcriptional regulation, gene expression and protein synthesis among IPF patients, resulting in a specific outcome on the expression of various proteins for each patient after treatment with FK228.

The proteins with significant differences in abundance plus the proteins only identified in three or more primary donor cultures of one condition were analysed for their role within the cell by gene ontology analysis, to identify biological processes in the cell that may be affected by FK228 treatment. Enriched gene ontology terms were summarized into those shown in **Figure 4.10**.

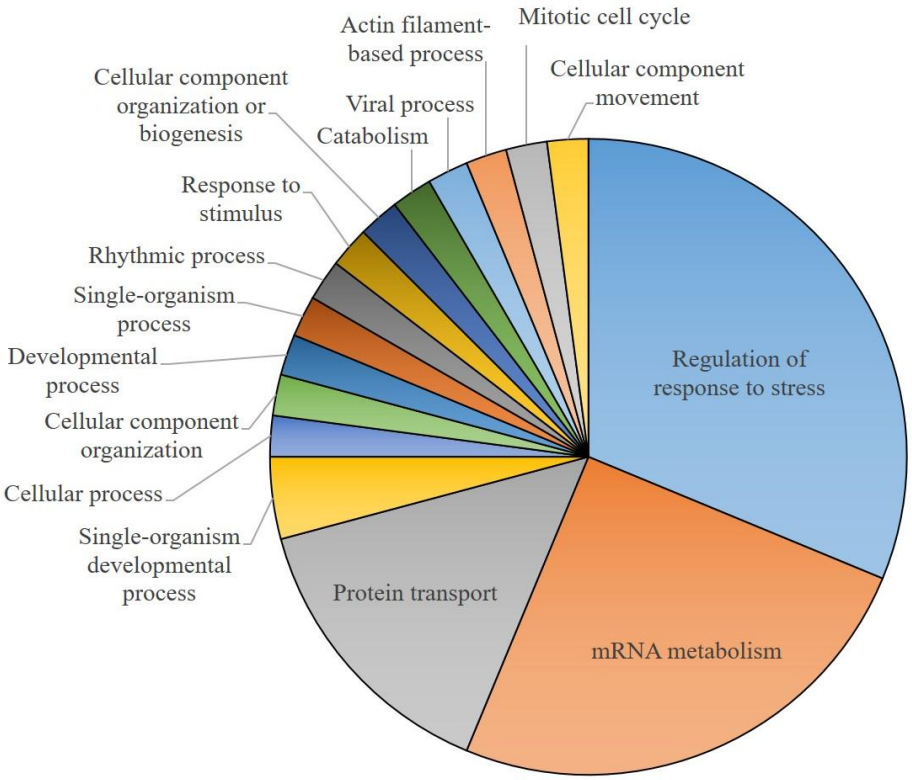


Figure 4.10 Gene ontology enrichment analysis of differentially expressed proteins as well as proteins expressed only in untreated IPF fibroblasts and only in FK228-treated IPF fibroblasts. Individual GO terms were combined into related terms using REVIGO that are reflected in the pie chart. Size of the terms is adjusted (log10*p*-value) to reflect the *p*-values.

This analysis revealed that many processes could be affected by changes in the expression of certain proteins in response to FK228 treatment. The most enriched GO terms for these proteins were regulation of response to stress and mRNA metabolism. It would be expected that these terms should be enriched in this dataset as FK228 may impact the expression of a potentially large number of genes. It has been previously mentioned that regulation of acetylation state has been reported to modify the expression of up to approximately 5% of genes, a proportion of which could be modified by FK228. Additionally, this drug may also cause changes in the acetylation state of non-histone proteins which could bring about other, off-target effects, such as changes in DNA-binding efficacy of transcription factors (Pang and Zhuang, 2010). The protein transport process would also be affected by changes in gene transcription and protein synthesis.

Additional enriched GO terms include cellular component organisation/movement, actin-based process and mitotic cell cycle changes, which could be affected in response to the anti-proliferative, matrix production-inhibiting effects of FK228. This would result in a less-myofibroblastic phenotype, including a reduction in expression of α-SMA and collagen III, which has been demonstrated to occur *in vitro* in IPF fibroblasts treated with FK228 by Davies *et al.* (unpublished).

4.3.5.4. Proteomic comparison of untreated and FK228-treated non-IPF fibroblasts

Non-IPF fibroblast proteins from three different primary donor cultures were analysed by mass spectrometry, untreated and with 48 hours of FK228 treatment, to find out whether the proteomic changes that occur in IPF fibroblasts with FK228 stimulation are also observed in fibroblasts from non-IPF patients.

Proteins identified with estimates of absolute quantification in all six samples were analysed by Student's T-test and fold-change analysis. One thousand, two hundred and thirty proteins were analysed by paired Student's T-Test. Of these, 84 were found to have significantly different expression ($p<0.05$). Fourteen proteins were identified in all three untreated primary donor cultures and not in any after FK228 treatment, and 17 proteins were only detected after FK228 in all three cultures. The full list of proteins with significant differential expression can be found in **Appendix C.4-C.6**.

From a similar number of proteins analysed to the number that could be compared between untreated and FK228-treated IPF fibroblasts (1302 in comparison to 1230) over one hundred less proteins changed in abundance significantly (either by T-Test or present only in one condition) in non-IPF fibroblasts than IPF fibroblasts (115 in non-IPF fibroblasts, 216 in IPF fibroblasts). This suggests that the effect on non-IPF fibroblasts may be less severe, possibly due to their transcriptional regulation being under tighter control to start with.

To test whether FK228 causes similar protein expression changes in both IPF and non-IPF fibroblasts, the significantly differentially expressed proteins from the two datasets were compared. This comparison was fairly limited as only proteins quantified in three or more primary donor cultures of non-IPF/IPF ± FK228 fibroblasts were included in the aforementioned T-Test analyses from which the protein lists were obtained (105 proteins $p<0.05$ in **4.3.5.3**, 84 proteins $p<0.05$ in **4.3.5.4**).

There were 12 proteins that had a *p*-value of less than 0.05 in both non-IPF and IPF ± FK228 datasets, and interestingly all of them had the same direction of change in abundance after FK228 treatment, suggesting (from a small number of comparisons) that FK228 exerts

certain effects on protein expression in fibroblasts regardless of disease state. The results of this analysis are shown below in **Table 4.2**.

Table 4.2 List of proteins significant according to Student's T-Test in both IPF ± FK228 and non-IPF ± FK228 datasets, and their fold change in abundance.

Protein Accession	Protein name	IPF ± FK228 fold change (+FK228: untreated)	Non-IPF ± FK228 fold change (+FK228: untreated)
O75369	Filamin-B	1.77	1.76
P04818	Thymidylate synthase	only present in untreated fibroblasts	only present in untreated fibroblasts
P43243	Matrin-3	-1.92	-1.94
P55145	Mesencephalic astrocyte-derived neurotrophic factor	only present in untreated fibroblasts	-3.00
Q00839	Heterogeneous nuclear ribonucleoprotein U	-1.42	-1.23
Q14914	Prostaglandin reductase 1	1.73	1.64
Q8NFJ5	Retinoic acid-induced protein 3	only present in FK228-treated fibroblasts	only present in FK228-treated fibroblasts
Q9BUF5	Tubulin beta-6 chain	-1.50	-1.45
Q9NPQ8	Synembryn-A	-1.48	-1.61
Q9UHD9	Ubiquilin-2	only present in untreated fibroblasts	-2.48
Q9UI42	Carboxypeptidase A4	only present in FK228-treated fibroblasts	only present in FK228-treated fibroblasts
Q9Y2D5	A-kinase anchor protein 2	2.28	1.9

4.3.5.5. Potential IPF biomarker candidates with significantly different expression following FK228 treatment

The differentially expressed proteins between IPF and non-IPF fibroblasts, listed in **Table 4.1**, were analysed for further changes in expression when IPF fibroblasts were treated with FK228. Proteins identified as candidate biomarkers of IPF fibroblasts and proteins with at least a 2-fold change in expression between IPF and non-IPF fibroblasts, that also had a significant change in expression with and without treatment (either p-value <0.05, fold-change of 2-fold or more, or only identified in 3+ primary cultures of either untreated or FK228-treated IPF fibroblasts) are listed in **Table 4.3**. These proteins are particularly interesting as not only could they potentially be used to diagnose and monitor the disease, they could also be used in clinical trials for FK228 in IPF patients to help inform on the efficacy of the drug and help clinicians to be able to recognise if the drug is having the

expected effect. This would be the ideal end goal of a biomarker study, however, as previously discussed, extensive independent validation would be needed to truly class any of these proteins as biomarker candidates.

Table 4.3 List of identified candidate protein biomarkers of IPF that are significantly modulated in expression with FK228 treatment.

Potential candidate IPF biomarkers modulated by FK228 treatment	
Proteins identified in IPF fibroblasts that were not identified in FK228-treated IPF fibroblasts or non-IPF fibroblasts	
Protein Accession	**Protein name**
Q96C36	Pyrroline-5-carboxylate reductase 2
Q9HAU4	E3 ubiquitin-protein ligase SMURF2
O43617	Trafficking protein particle complex subunit 3
O15078	Centrosomal protein of 290 kDa
O60573	Eukaryotic translation initiation factor 4E type 2
Q8WZA9	Immunity-related GTPase family Q protein
Q9UBT6	DNA polymerase kappa
P48668	Keratin_ type II cytoskeletal 6C
Q8N9V7	Testis- and ovary-specific PAZ domain-containing protein 1
Q5TBB1	Ribonuclease H2 subunit B
O75326	Semaphorin-7A
Q9NP61	ADP-ribosylation factor GTPase-activating protein 3
Q7Z739	YTH domain family protein 3
Q502W7	Coiled-coil domain-containing protein 38
A9Z1Z3	Fer-1-like protein 4
P02461	Collagen alpha-1(III) chain
Q8WUP2	Filamin-binding LIM protein 1
Q5T7N2	LINE-1 type transposase domain-containing protein 1
Q9NSK0	Kinesin light chain 4

Proteins with increased expression in IPF, not detected following FK228 treatment					
Protein Accession	**Protein Name**	**IPF/Non-IPF Fold Change**	***p*-value if applicable**	**IPF-/+FK228 Fold Change**	***p*-value if applicable**
Q9Y2T7	Y-box-binding protein 2	2.113		N/A	
Proteins with increased expression in IPF and decreased expression following FK228 treatment					
Protein Accession	**Protein Name**	**IPF/Non-IPF Fold Change**	***p*-value if applicable**	**IPF-/+FK228 Fold Change**	***p*-value if applicable**
Q01546	Keratin_ type II cytoskeletal 2 oral	8.066		-2.394	

P63267	Actin_ gamma-enteric smooth muscle	6.466		-3.312	
P63241	Eukaryotic translation initiation factor 5A-1	6.410		-2.512	
P08962	CD63 antigen	3.967		-2.503	
P02538	Keratin_ type II cytoskeletal 6A	3.376		-2.447	$3.3x10^{-4}$
P08779	Keratin_ type I cytoskeletal 16	2.908		-2.369	
Q01130	Serine/arginine-rich splicing factor 2	2.654		-2.202	
P46779	60S ribosomal protein L28	2.279		-1.736	
Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4	2.149		-2.174	0.043
P62879	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	2.001		-2.225	0.006

Proteins with decreased expression in IPF and increased expression following FK228 treatment

Protein Accession	**Protein Name**	**IPF/Non-IPF Fold Change**	***p*-value if applicable**	**IPF-/+FK228 Fold Change**	***p*-value if applicable**
Q6IBS0	Twinfilin-2	-1.556	0.004	1.242	
P61088	Ubiquitin-conjugating enzyme E2 N	-1.772	0.019	1.764	0.04
P43490	Nicotinamide phosphoribosyltransferase	-2.096		2.358	
P05787	Keratin_ type II cytoskeletal 8	-2.237		2.229	
P62820	Ras-related protein Rab-1A	-2.461		4.323	
P09104	Gamma-enolase	-3.238		10.094	0.016
Q9NRR5	Ubiquilin-4	-3.430		2.623	0.006
Q9UI15	Transgelin-3	-4.863		3.031	0.029

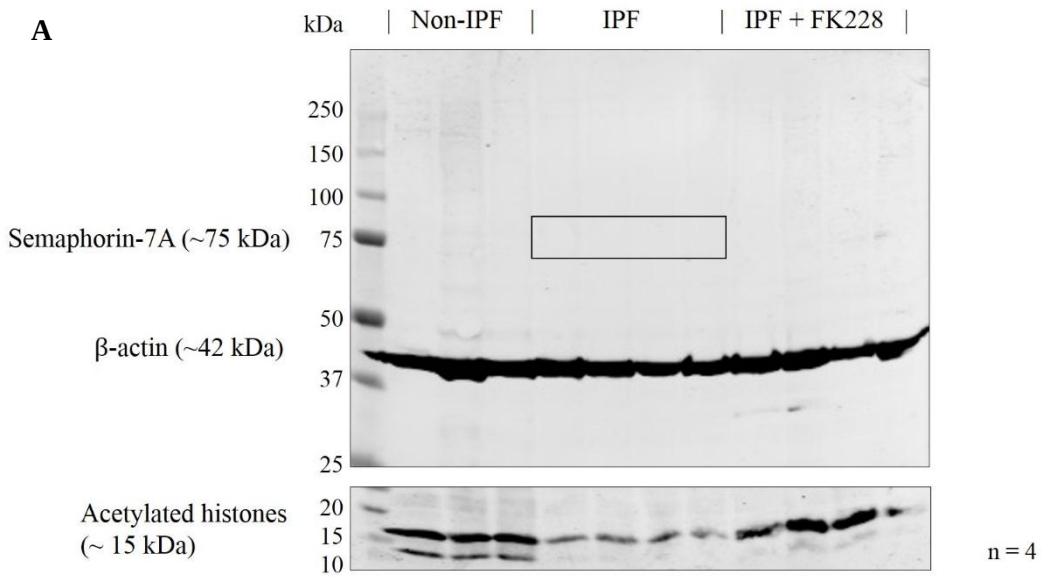
4.3.5.6. Further work is required for independent validation of potential candidate biomarkers

Two proteins that were identified in untreated IPF fibroblasts but not in non-IPF fibroblasts or IPF fibroblasts treated with FK228, Semaphorin-7a and Filamin-binding LIM protein 1

(FBLIM-1), were analysed by SDS-PAGE and western blotting for independent validation as potential IPF biomarkers. However, results were inconclusive.

No signal was seen for Semaphorin-7a, even when the anti-Semaphorin-7a antibody was used at 10x the recommended concentration (a representative image is shown in **Figure 4.11A**, with the box in the centre demonstrating the area in which bands were expected to be observed). Probing the blot for β-actin demonstrated the presence of protein and equal loading between lanes, and probing for acetylated lysine showed that untreated IPF fibroblast proteins had the least amount of acetylation, and the expected increase in protein acetylation in FK228-treated samples was also observed. This indicates that the experiment was performed successfully, and the non-detection of Semaphorin-7a is likely to either be due to an antibody issue or instability of the protein. This could be corrected by generating fresh cell lysates or using a different anti-Semaphorin-7a antibody.

Western blotting for the validation of FBLIM-1 as a potential biomarker was also inconclusive due to non-specific binding of the anti-FBLIM-1 antibody or degradation of the protein. Instead of the expected result of one band per lane at ~50 kDa, bands were also present at ~37 kDa and 15 kDa (**Figure 4.11B**). In addition, the protein was only expected to be detected in untreated IPF fibroblast samples, but bands were observed across the entire blot. Therefore, further work is needed in order to successfully validate FBLIM-1 as a candidate biomarker.



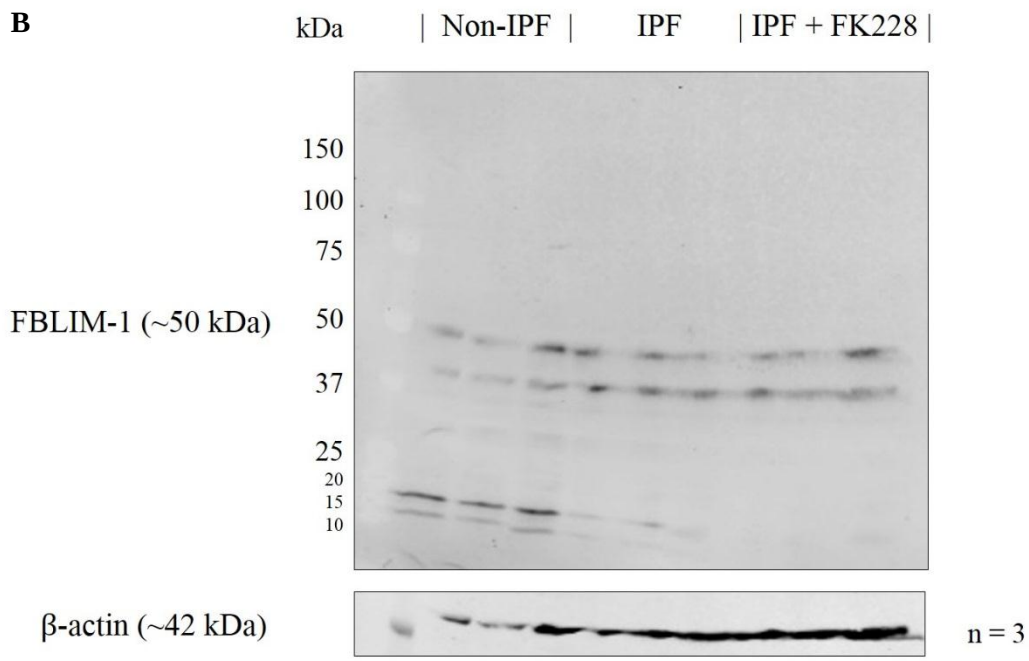


Figure 4.11 Western blot of A) Semaphorin-7a and acetylated lysine expression and B) FBLIM-1 expression in untreated non-IPF fibroblasts, and in IPF fibroblasts with and without FK228 treatment. The expression of Semaphorin-7a, acetylated lysine and FBLIM-1 was analysed by western blotting, with β-actin as a loading control.

4.4. Discussion

There is a great demand to find a treatment for idiopathic pulmonary fibrosis that will improve the quality of life of patients, slow or inhibit the progression of the disease at any stage and prolong survival. Neither Pirfenidone nor Nintedanib, the two pharmacologic agents so far approved for treatment by the FDA and the European Medicines Agency, are a cure for IPF. These treatments have been shown to improve the decline in lung function (forced vital capacity) of patients over 1 year and reduce mortality rate (Wani et al., 2015), but neither of them stop disease progression, and both have several adverse effects. In addition, these drugs are very expensive, expected to cost approximately $100,000 per patient per year in the USA (Tzouvelekis et al., 2015) and it is not yet known whether they will be cost-effective. However, due to unmet need in this disease at present it is expected that Pirfenidone and Nintedanib will be widely used regardless of their disadvantages (Tzouvelekis et al., 2015). It is imperative that alternate treatment options continue to be investigated for the possibility of halting the progression of, or even curing, IPF.

Due to the aforementioned similarities to the pathogenesis of cancer in IPF, attention is turning towards anti-cancer therapeutics in the rigorous search for treatment, including histone deacetylase inhibitors. In addition to the work discussed here, studies employing HDACIs in fibrosis include:
- investigations into the effects of trichostatin A (TSA) on fibrotic renal fibroblasts, in which TSA suppressed cell proliferation and differentiation (Pang et al., 2009)
- studies using TSA or valproic acid in polycystic kidney disease, in which serum-induced ciliary disassembly was blocked, and kidney cyst formation and decline in murine lung function was suppressed, respectively (Pugacheva et al., 2007; Cao et al., 2009)
- a study of HDAC inhibitors in cardiac hypertrophy, which was significantly reduced by TSA and valproic acid administration (Kee et al., 2006)

There have been several studies examining the effects of HDAC inhibitors other than FK228 in IPF, to study the link between fibroblast differentiation and HDAC activity (Pang and Zhuang, 2010), although to the author's knowledge there are no published proteomic studies. In one report, TSA inhibited TGF-β_1-mediated expression of α-SMA and collagen I in normal lung fibroblasts (Guo et al., 2009). In another, the HDAC inhibitor suberoylanilide hydroxamic acid (SAHA) inhibited TGF-β_1-induced myofibroblast formation and serum-induced fibroblast proliferation (Wang et al., 2009). Zhang *et al.* also used SAHA to show that the increased expression of collagen III-α1 in IPF fibroblasts is reduced upon treatment with a HDACI, in both human cells and in a murine model (Zhang et al., 2013b). In 2015, Korfei *et al*. demonstrated a strong elevation of almost all HDAC enzymes in myofibroblasts from IPF lungs compared to control, and treatment of IPF fibroblasts with the pan-HDACI panobinostat/LBH589 significantly down-regulated the expression of genes involved in ECM production, cell proliferation and survival (Korfei et al., 2015). Panobinostat also recently gained FDA approval as a cancer treatment, for multiple myeloma (Laubach et al., 2015).

In 2012, Davies *et al*. showed that the HDAC inhibitor spiruchostatin A inhibited the proliferation of both primary control and IPF fibroblasts, in a concentration- and time-dependent manner with little cytotoxicity, and suppressed TGF-β_1-stimulated myofibroblast differentiation and collagen gene expression (Davies et al., 2012). As mentioned previously, subsequent work studying the effects of FK228, which is structurally similar to spiruchostatin A, found that this HDACI had similar effects *in vitro* and was even more potent at lower concentrations (Davies, 2011), and was also effective in suppressing fibrosis *in vivo* (Davies *et al*., unpublished).

FK228 is an ideal potential candidate for IPF therapy, not only due to its demonstrated anti-proliferative effects, but also because it has already been approved for human use in cancer therapy (VanderMolen et al., 2011), making the approval process much quicker if it is shown to be suitable as an IPF treatment. FK228 was isolated in 1994 (Ueda et al., 1994) and preliminary investigations into the compound revealed potent cytotoxicity against several types of cancer cell lines. It was shown to cause cell cycle arrest as well as having histone deacetylase inhibitory activity (Nakajima et al., 1998). FK228 contains a disulphide bond which is reduced by glutathione in cells. The free thiol group of the reduced compound interacts with zinc ions in the active site of histone deacetylases, blocking substrate access (VanderMolen et al., 2011; Furumai et al., 2002). It was shown to inhibit class I HDACs more strongly than Class II HDACs, and in the same study it was also demonstrated to be a natural prodrug, stable in medium or serum in its inactive form but becoming active after uptake into cells due to intracellular reducing activity. This suggested that circulation of FK228 after *in vivo* administration would be stable and that the drug would become reactive after uptake into cells, for example tumour cells (Furumai et al., 2002).

Results from a phase I trial of FK228 in 2001 suggested that this prodrug could be an effective treatment for T-cell lymphoma after all patients with cutaneous T-cell lymphoma (CTCL) had a partial response, and a patient with peripheral T-cell lymphoma had a complete response. Results also showed increased histone acetylation in cells obtained from patients after treatment (Piekarz et al., 2001). A subsequent Phase II trial (Piekarz et al., 2009) included 71 patients with CTCL. As in the Phase I trial, FK228 was administered at a dose of 14 mg/m^2 as a 4-hour intravenous infusion, on days 1, 8, and 15 of a 28-day cycle, for an average of 4 cycles. 34% of patients responded, with 4 complete responses and 20 partial responses. The median time to a response was 2 months, lasting for a median of 13.7 months. Some patients responded for over 3 years, even after discontinuation of treatment. The drug was also well tolerated. These results combined with similar results from a parallel Phase II study of 96 patients with CTCL (Whittaker et al., 2010) led to FDA approval in 2009 of FK228 for treatment of patients with advanced CTCL (Jain and Zain, 2011). In 2011 it also gained FDA approval for the treatment of peripheral T-cell lymphoma (Barbarotta and Hurley, 2015).

To the author's knowledge, there are just two published proteomic studies of FK228 in cancer. One study investigated FK228-induced cytotoxicity-associated proteins in lung cancer cells using 2D gel electrophoresis and MS/MS (Chen et al., 2008). From 45 protein spots with a significant difference in intensity, 27 proteins were identified by mass spectrometry, which were involved in multiple biological processes including transcriptional regulation, metabolism and cytoskeletal organisation, as was also found in this study. The

second study analysed global protein expression changes in a bladder cancer cell line in response to FK228 using LC-MS/MS (Li et al., 2016). Protein identification was based on at least two unique peptides and proteins were relatively quantified. 3471 proteins were identified as being differentially expressed by at least 2-fold following FK228 treatment in this study, and they were associated with 13 biological processes, the most enriched being metabolism. The results of these two studies and the study described in this chapter indicate that FK228 has similar proteomic effects on tumour cell and IPF fibroblast biological processes, which is likely to be due to similar disease pathogenic characteristics.

This project follows on from the work carried out by Davies *et al*. described above, studying the effects of FK228 in IPF fibroblasts. This study aimed to identify candidate protein biomarkers of idiopathic pulmonary fibrosis, which could also be monitored during FK228 treatment at clinical trial. Biomarkers are becoming useful not only for disease diagnosis, but also for monitoring progression, guiding therapeutic selection and predicting/profiling a response to treatment in clinical trial, giving insight into expected effects (de Lemos and McGuire, 2011).

In order to identify these biomarker candidates, IPF fibroblasts were characterised at the proteomic level, with and without FK228 treatment. Fibroblasts play a major role in the wound healing response, which is hypothesized to be dysregulated in IPF, and were therefore chosen as the initial source of potential protein biomarkers; in the lung they migrate to sites of injury and differentiate into contractile myofibroblasts that deposit extracellular matrix (ECM) components into the interstitium, to form a scaffold for tissue repair and wound closure (Hinz, 2007). Fibroblasts are therefore an appropriate source for linking biomarkers to IPF pathogenesis (Kaarteenaho and Lappi-Blanco, 2015). A report estimated in 2013 that approximately a third of patients require biopsy to achieve a diagnosis (Kaarteenaho, 2013), and therefore cells from lung tissue of a relatively high number of patients could be obtained for the study of a more physiologically relevant sample type than other, less invasively-obtained sample types such as sputum, BAL fluid or serum.

Initial experiments to assess the viability and proliferation rate of IPF fibroblasts after FK228 treatment at a range of doses, and to assess the cytotoxicity and assign an optimum dose for following experiments, showed that cellular metabolic activity decreased in a dose-dependent manner at doses higher than 1 nM, however, they remained >60% metabolically active after 48 hours of treatment even at high concentrations of the HDAC inhibitor. 2 nM FK228 was determined to be the optimum dose as an effect on metabolism/proliferation and potentially viability could be observed, although not large enough for potential cell death to affect experiments. Previous reports have calculated the IC_{50} of FK228 in IPF fibroblasts to

be 2 nM (Davies, 2011). There were significantly fewer cells counted after 48 hours of 2 nM FK228 treatment for both IPF and non-IPF fibroblasts, indicating either an anti-proliferative effect or cell death. FACS cell cycle analysis showed that fibroblasts were arresting in G1 phase of the cell cycle, indicating that the lower non-IPF and IPF fibroblast cell numbers after treatment was due to growth arrest following incorporation of the drug rather than excessive cell death. These results suggest that at this dose FK228 causes a sizeable change in fibroblast gene and protein expression to arrest proliferation without being excessively cytotoxic. These findings correlate with previous reports in which FK228 was shown to promote cell cycle arrest (Nakajima et al., 1998) (Davies *et al.*, unpublished).

The proteomes of fibroblasts from 5 primary IPF donor cultures and 3 non-IPF primary donor cultures were then analysed. Fibroblasts were treated with 2 nM FK228 for 48 hours and the cell lysates were analysed by western blotting and mass spectrometry. Western blotting results showed that lysine acetylation increased with FK228 treatment, suggesting that the drug had successfully inhibited histone deacetylase activity. Overall, 6374 unique proteins were identified by UPLC-HDMS^{E} in the cell lysates with estimates of absolute quantification, from which a potential candidate biomarker panel could be constructed.

There are relatively few reports in the literature pertaining to proteomic analyses in IPF. Most studies focus on bronchoalveolar lavage (BAL) fluid samples from patients, using 2D electrophoresis (Lenz et al., 1993; Wattiez et al., 2000; Magi et al., 2002; Rottoli et al., 2005) and more recently a gel-free quantitative approach (Foster et al., 2015) which quantified over 1000 immuno-depleted proteins by HDMS^{E}. Korfei *et al*. performed mass spectrometry on lung tissue samples from IPF patients, as well as NSIP patients and non-IIP controls, as outlined in **Chapter 3**, identifying 89 proteins with differential expression between IPF and healthy lungs (Korfei et al., 2011; Korfei et al., 2013). Most recently, Carleo *et al*. (2016) compared protein expression from the BAL fluid of patients with familial and sporadic IPF using 2D electrophoresis and mass spectrometry. They identified 88 protein spots differently expressed between the two conditions.

In this study, several methods of separation were employed in order to maximise proteome coverage from whole cell lysates, without the use of gels. These included OFFGEL fractionation, nano-UPLC and ion mobility separation. Using a data-independent mode of acquisition (MS^{E}) reduced selection bias and thus allowed the study of low abundant proteins, and a label-free approach for quantification allowed comparison of protein expression levels across independent samples.

One hundred and twenty-two proteins were identified as potential candidate biomarkers of IPF from fibroblasts, using significant differences in expression by Student's T-Test and

presence or absence from a condition as identifiers. As discussed in **Chapter 3**, there are no standard statistical tests that are always applied to proteomic data, so each case must be judged individually. Fold-change was also calculated in each case to reveal the direction of change (an increase or decrease in expression).

Protein expression appeared relatively uniform between the three non-IPF donor samples analysed, however, there was a great deal of variation in individual protein expression between all IPF donor samples, suggesting dysregulation of gene expression related to these proteins, not only in IPF compared to non-IPF control conditions, but also between patients. As discussed previously (**Chapter 1**), the progression and presentation of IPF in a patient can vary from person to person, and dysregulation in transcription and translation could be linked to this. Differences in factors such as patient age, sex and smoking status may also contribute to the variation in expression of certain proteins and progression of the disease, however this information was not available and so it was not possible to use these co-variables in the analyses.

The hypothesis that dysregulation of transcription is a major factor in IPF is supported by gene ontology analysis of the candidate biomarker dataset, as many enriched terms were associated with transcriptional regulation. According to this analysis, many of the proteins with significant expression differences in IPF and non-IPF fibroblasts play a role in gene expression and its regulation within the cell.

Additional enriched GO terms include cellular component organisation, cell cycle processes and cytoskeleton-based processes. Differences in expression of proteins involved in these processes in IPF may be driven by profibrotic cytokines as the cell takes on a myofibroblastic phenotype. These GO terms were also enriched in the dataset of proteins with statistically different expression in MRC-5 fibroblasts treated with the profibrotic cytokine transforming growth factor-beta compared to control (see **Chapter 3**). This indicates that pathogenic changes within the primary cells are similar to the changes in the regulation of biological processes that occur during TGF-β_1-induced myofibroblast formation, which is interesting as this process occurs aberrantly in IPF. This information provided by the primary fibroblasts increases the credibility of the myofibroblastic model.

There was considerable variability in individual protein expression between IPF fibroblasts with and without FK228, with the clustering analysis of differentially expressed proteins unable to distinguish between treated and untreated samples in every case. One possible explanation for this is that the histone deacetylase inhibitor causes changes in transcriptional regulation with varying outcomes on protein expression, due to these fibroblasts already

showing differences in transcriptional regulation between patients to begin with, as discussed earlier.

Overall, FK228 appeared to have a large effect on the IPF fibroblasts at 2 nM concentration, with over 200 proteins having a significant difference in abundance with and without treatment, and a wide range of biological processes potentially being affected by changes in expression of these proteins. The drug had similar effects in non-IPF fibroblasts. FK228 treatment significantly affected the expression of some of the potential candidate biomarkers of IPF; in some cases in which a protein had increased expression in IPF fibroblasts or was only detected in IPF fibroblasts compared to non-IPF fibroblasts, FK228 treatment caused their expression to decrease towards non-IPF fibroblast levels, demonstrating the potential for some of these proteins as biomarkers of response to therapy. Examples of this include collagen III, Filamin-binding LIM protein 1 (FBLIM-1, shown to promote stress fibre formation (Takafuta et al., 2003)) and γ-smooth muscle actin. The reduction in expression of these proteins following treatment is additional evidence of inhibition of myofibroblast formation and matrix deposition caused by this drug.

If the acetylation states of candidate biomarkers were also to be examined in response to FK228, it may show whether this HDACI also alters the functions and/or interactions of these proteins by increasing acetylation, as well as having an effect on their abundance. The development of a method to assess the off-target acetylation effects of FK228 is discussed in **Chapter 6**.

Successful validation of the potential IPF biomarkers identified in this study would add them to a list of candidate biomarkers already in the literature. As previously stated, there are currently no validated molecular biomarkers specifically for IPF. A comprehensive review of current candidate IPF biomarkers was put together by Flynn *et al.* in 2015. They noted that the area of discovery of potential biomarkers has shifted over time, according to changes in hypotheses concerning IPF pathogenesis, discussed in **Chapter 1**. Many candidate molecular biomarkers are epithelial cell-derived, along with several that are involved in matrix remodelling. The suggestion that inflammation may play a part in IPF pathogenesis rather than being the cause has prompted the candidature of several inflammatory markers also. The full list is shown in **Figure 4.12**.

Interestingly, none of the proteins listed in this review associated with matrix remodelling were identified in this study, with the exception of periostin which was identified in one untreated IPF primary donor culture. These proteins are more likely to be found in the fibroblast secretome rather than the intracellular proteome due to their roles in remodelling. Semaphorin-7a, listed under the T-cell mechanistic category in this list, is a membrane

protein involved in integrin-mediated signalling and in regulating cell migration and immune responses. It also promotes activation of the protein kinase FAK1 and subsequent phosphorylation of MAPK1 and MAPK3 (van Rijn et al., 2016), important molecules in the Smad-independent TGF-β_1 signalling pathway. It has been shown to have increased expression in tumour cells compared to normal cells, and to promote tumour growth (Garcia-Areas et al., 2014). This protein was also identified in this study as a potential candidate biomarker of IPF, not being identified in any non-IPF primary donor cultures but identified in at least three primary IPF cultures, and no longer detected in any of the IPF primary donor cultures after FK228 treatment (**Table 4.3**), making it a potential candidate biomarker for use in clinical trials. Attempts to validate these results by western blotting were unfortunately unsuccessful due to issues with the primary antibody. If this result could be validated in IPF fibroblasts there would be great potential for this protein as a candidate biomarker. A study investigating the effects of Semaphorin-7a in a cancer cell line demonstrated cell cycle arrest in G1 phase and reduced secretion of matrix proteins including MMPs in Semaphorin-7a knockdown cells (Saito et al., 2015), therefore causing a reduction in the expression of this protein may be one way that FK228 causes growth arrest and decreased matrix deposition in IPF fibroblasts.

Mechanistic category	Candidate biomarker
Alveolar epithelial cell damage/dysfunction	KL-6
	SP-A + SP-D
Extracellular signals	OPN
	VEGF
Fibrogenesis/matrix remodeling	MMP7
	LOXL2
	Periostin
	Fibrocytes
Immune dysregulation/inflammation	CCL18
	YKL-40
	IL-8
	ICAM-1
T-cells	Sema7a
	CD28
B-cells and autoimmunity	anti-HSP70
	BLyS
	CXCL13
Genetic mutations	MUC5B
	Telomere shortening
	TOLLIP

Figure 4.12 List of current candidate biomarkers in the literature for idiopathic pulmonary fibrosis. Adapted from Flynn et al. (2015) with permission from Dove Medical Press Ltd. Abbreviations: KL-6, Krebs von den Lungen 6; SP, surfactant protein; OPN, osteopontin; VEGF, vascular endothelial growth factor; MMP7, matrix metalloproteinase-7; LOXL2, lysyl oxidase-like 2; CCL18 chemokine (C-C motif) ligand 18; IL-8, interleukin-8; ICAM-1, intercellular adhesion molecule-1; Sema7a, semaphorin 7a; HSP70, heat shock protein 70; CXCL13, chemokine (C-X-C motif) ligand 13; MUC5B, Mucin 5B; TOLLIP, toll-interacting protein.

In summary, the proteomes of primary IPF and non-IPF fibroblasts have been profiled and investigated for changes in response to the histone deacetylase inhibitor FK228, resulting in the identification of over 6500 proteins and 122 proteins with differential expression between IPF and non-IPF fibroblasts. A thorough search of the literature has revealed that this is the first time that pulmonary fibroblast lysates from IPF patient biopsies have been profiled by mass spectrometry. This study also used a data-independent mode of acquisition to calculate estimates of absolute quantification for nearly all of the proteins detected, in order to provide an in-depth proteomic characterisation of the IPF fibroblast and to compare this profile with that of non-IPF fibroblasts and identify proteins with differential expression. In addition, the aforementioned experiments performed by Davies *et al.*, and this follow-on proteomics study are the first reports of investigating FK228 as a potential therapy for IPF.

The results from these studies hold much promise for getting this drug to clinical trial for IPF.

Any potential biomarkers identified in this study will need successful independent validation before they can be considered for clinical use, as well as rigorous testing for sensitivity, specificity and reproducibility. The implementation will also need to be cost-effective (Flynn et al., 2015). Thus, there is still a long way to go in terms of translating any identified potential biomarkers to clinical practice, as demonstrated by the published list of candidate biomarkers for IPF above, none of which have been implemented in the clinic. KL-6 was approved by the Japanese Health Insurance Program in 1999 as a diagnostic biomarker for interstitial lung diseases (ILDs), but not specifically for IPF. It is also not typically used as a biomarker outside of the Japanese population (Horimasu et al., 2012). Methods for validation of some of the potential biomarkers in this study in the first instance could include western blotting, immunohistochemistry and MRM assays. Efforts have been made in this study to validate FBLIM-1 in addition to semaphorin-7a by western blotting, but at present the results have proved inconclusive. Validation of these proteins and additional potential biomarkers will be carried out as a follow up to this study.

The method developed in **Chapter 3** and employed in this study for sample preparation for mass spectrometry has potential for use in investigating many other diseases, including other types of ILD. This would be important for demonstrating whether any identified potential candidate biomarkers from this study are specific to IPF and not any other IIPs such as NSIP, which can be difficult to distinguish from IPF. It could also be applied for profiling other cell types that may be interesting to study in IPF. Furthermore, it could also be adapted for studying protein expression in other culture systems. Following this study, the proteomes of IPF fibroblasts grown in an *in vitro* 3D culture system, thought to be more physiological than fibroblasts grown on cell culture plates, were analysed by UPLC-HDMS^{E}, first of all to find out whether it is possible to analyse this kind of sample efficiently by mass spectrometry, and to discover how protein expression within IPF fibroblasts differs between the two culture systems.

5. Results - Proteomic characterisation of a 3D *in vitro* IPF fibroblast culture model

5.1. Abstract

The pathogenesis of idiopathic pulmonary fibrosis is thought to involve dysregulated interactions between alveolar epithelial cells and fibroblasts during the wound healing response, leading to excessive production and deposition of extracellular matrix (ECM) components in the interstitium. Studies in the literature have described the employment of a 3D *in vitro* culture system as a novel method of studying ECM production by fibroblasts. This type of culture system may be useful as a model for fibroblastic foci and matrix in IPF. It has been demonstrated that fibroblasts taken from biopsies from IPF patients can be cultured using this system and that they deposit a collagen-rich matrix.

This study aimed to investigate whether this 3D tissue-like model could be analysed effectively using a proteomic approach, given its complex nature; this is the first time that this model has been analysed using proteomics. It also aimed to characterise the proteome of the IPF fibroblasts, and to an extent the ECM, grown in this type of culture system. Work is currently underway to evaluate the histone deacetylase inhibitor FK228 as a treatment option for IPF, therefore these fibroblasts were treated with this drug to examine its effect on the proteome and compare them to fibroblasts grown previously in cell culture monolayers. Proteins were extracted from fibroblasts grown in the 3D culture model, and protein expression was compared between two different time-points (48 hours and 3 weeks), with and without FK228 treatment, and to that of IPF fibroblasts grown on typical culture plastic.

Over 4900 unique proteins were isolated and identified from samples generated from fibroblasts grown in the 3D culture system, indicating that the presence of deposited matrix does not greatly hinder the efficiency of lysis and digestion for analysis by mass spectrometry. After 6 weeks of culture and 3 weeks of chronic FK228 treatment, 62 proteins had had their expression significantly modulated by FK228, most playing a role in metabolism. Comparison of protein expression between fibroblasts from this model and those grown in 2D culture showed that many more ECM proteins could be analysed from the 3D model. These results show that this type of sample can be analysed effectively by mass spectrometry, and they add to previous indications that IPF fibroblasts grown in this manner could be a suitable model for studying the extracellular matrix in the pathogenesis of IPF.

5.2. Introduction

Idiopathic pulmonary fibrosis is a progressive interstitial lung disease of unknown cause, which results in excessive deposition of extracellular matrix components. This causes stiffening of the pulmonary tissue and impairs gas exchange. The ECM is composed of fibrous proteins, proteoglycans and glycosaminoglycans, and is important in maintaining homeostasis under normal conditions, as it forms a scaffold so that inflammatory and fibrogenic mediators can rapidly respond to injury and re-epithelialisation can occur. Components of the matrix such as collagen, fibronectin and laminin are produced by myofibroblasts, which also produce enzymes capable of degrading these components. The production and secretion of remodelling proteins such as proteolytic matrix metalloproteinases (MMPs) and their inhibitors ensures a close regulation of matrix composition and turnover, for efficient resolution of wound healing (Clarke et al., 2013). MMPs are also secreted by epithelial cells and alveolar macrophages, particularly when responding to injury, and MMPs have been shown to be upregulated in the IPF lung (Pardo and Selman, 2012). Collagen cross-linking enzymes such as lysyl oxidases, secreted by myofibroblasts, are also important in modifying and regulating the ECM. Lysyl oxidase (LOX) has recently been identified as a potential biomarker of IPF in the bronchoalveolar lavage fluid of IPF patients (Davies et al, unpublished), and agents that block LOXL2, another member of the LOX family, have entered clinical trials as a potential IPF treatment (Clarke et al., 2013).

As well as cells regulating the ECM in the lung, it has been shown that the ECM can regulate cellular behaviour. For example, it has been shown that culturing both IPF and normal lung fibroblasts on matrices of varying stiffness changes their rate of proliferation and contraction (Marinkovic et al., 2013). It has also been demonstrated that lung matrix stiffness increases six-fold after bleomycin injury in mice compared to normal, and with it fibroblasts move from a quiescent state to an increasingly proliferative and secretory phenotype (Liu et al., 2010). This has important implications for studies of fibrosis involving fibroblasts cultured on plastic or glass, as carried out in **Chapter 4**, as some observed cellular responses may be influenced by these substrates, resulting in altered ECM and cell behaviour (Clarke et al., 2013). Normal lung tissue stiffness has a Young's modulus of approximately 1.96 kPa, and IPF lung has been measured at approximately 16.5 kPa, which is significantly higher than normal; however, this is much lower than the Young's modulus of tissue culture plastic, measuring at 2-4 GPa (Booth et al., 2012). This may impact significantly on cell behaviour. However, fibroblast cell lines or explants from biopsy grown

as monolayers on cell culture plastic is a very common method of studying fibrosis because of limitations with available animal models. There is currently no *in vivo* model that can fully replicate the features of IPF. Most commonly used is the bleomycin-induced fibrosis model in murine lungs, which has aided understanding of pathogenic mechanisms, but does not share the progressive nature or irreversibility of IPF. In addition, many agents that show antifibrotic effects in this murine model do not do so in humans (Moeller et al., 2008), therefore new methods of modelling fibrosis are greatly warranted.

Recently, novel methods of studying the extracellular matrix of fibrotic cells have been developed. Dr Mark Jones from the Brooke laboratory in the Faculty of Medicine, University of Southampton NHS foundation trust has recently developed a 3D *in vitro* fibroblast model of ECM production using primary IPF fibroblasts. Dr Jones has demonstrated that when cultured in this 3D system, IPF fibroblasts produce a collagen-rich matrix after 5 weeks (**Figure 5.1)** (Jones, pers comm).

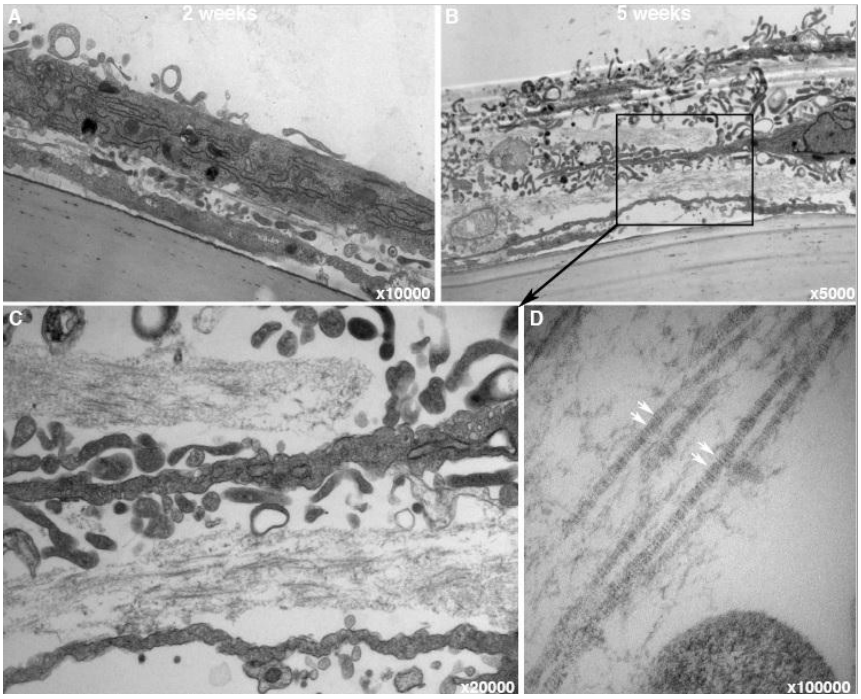


Figure 5.1 Formation of a multi-layer cell structure following long-term culture in the 3D in vitro system. Parenchymal lung fibroblasts were cultured for up to 2 or 5 weeks then processed for transmission electron microscopy. A.) At 2 weeks, 1-2 layers of cells can be observed B.) After 5 weeks, layers of cells are visible, with incorporation of ECM in the extracellular space (C) including fibrillar collagen (D, white arrows). Taken with permission from Dr Mark Jones, Brooke Laboratory, University of Southampton NHS foundation trust (pers comm).

This model was developed using the principles of a culture method described by Shamis et al. (2012), who investigated differences in ECM produced by induced pluripotent stem cells and parental dermal fibroblasts. This study cultured fibroblasts on porous hanging cell culture inserts rather than directly on tissue culture plastic, for a period of 5 weeks, with L-ascorbic acid-2-phosphate added to the culture medium. This compound is an essential cofactor for fibrillar collagen production, stimulating collagen gene expression in fibroblasts (Chojkier et al., 1989) and promoting their reorganisation into a "3D tissue-like substance" (Hata and Senoo, 1989).

An aim of this study was to find out whether the fibroblasts and matrix cultured in this 3D *in vitro* system could be analysed effectively using a proteomic approach. Previously, proteomic profiles of IPF fibroblasts grown as monolayers on plastic have been generated, identifying on average almost 2360 proteins per primary donor culture (**Chapter 4**). Potential difficulties with attempting to analyse the 3D model using proteomics include possible inefficiencies in the lysis and digestion of this kind of sample, due to its tissue-like structure and the large amount of collagen, which can be difficult to solubilise and digest effectively. If good proteome coverage could be achieved, which was the case, the data could then be used to characterise the IPF fibroblast proteome in this culture model, which could be compared with that of IPF fibroblasts grown as monolayers.

IPF fibroblast monolayers have also been analysed for protein expression changes in response to FK228, a histone deacetylase inhibitor currently being evaluated as a potential IPF treatment (**Chapter 4**). Therefore, the IPF fibroblasts grown in the novel 3D culture system were also subjected to FK228 treatment to see if the effects of this drug were comparable between culture methods. Fibroblasts were treated with FK228 either for 48 hours as in the previous study (**Chapter 4**) or for 3 weeks to evaluate the effects of chronic treatment in a more matrix-rich culture. Samples were prepared for proteomic profiling using an adapted version of previously developed methods (**Chapters 3 and 4**).

Over 4700 unique proteins were identified in the 3D model samples with estimates of absolute quantification over both time-points and conditions. An average of 2170 proteins were quantified from untreated fibroblasts cultured for 3 weeks in this system, and almost 1500 proteins were quantified from untreated fibroblasts cultured for 6 weeks. Many proteins identified from these samples were extracellular matrix proteins. These results were compared with those from fibroblasts grown in monolayer culture to assess similarities and differences in matrix expression.

5.3. Results

A surgical lung biopsy taken with ethical approval from a patient with a confirmed diagnosis of IPF (IPF patient 1) was used to establish fibroblast cultures. The sample was cut into approximately 2x2 mm sections and pieces of tissue were added to wells of a 6-well culture plate containing Dulbecco's modified Eagle medium (10% foetal bovine serum (FBS), phenol red) and scratched into the bottom of the wells. Medium was changed after 7 days of incubation and replaced every 2-3 days for 2 weeks. Any outgrown fibroblasts were dissociated from the wells and cultured in a 75 cm^2 tissue culture flask. At 70-80% confluence cells were collected, counted and split into several 75 cm^2 culture flasks to increase cell number (see section **2.2**).

Hanging cell culture inserts were added to 24-well plates. 900 µl complete DMEM (10% FBS) was added to each well, and 200 µl was added to each insert. Plates were incubated for 3 hours to equilibrate the membranes. All medium was removed and IPF fibroblasts were seeded into the inserts at $1.5x10^4$ cells/insert in 200 µl medium, with 900 µl medium added to the wells. Medium was removed after 24 hours and replaced with equivalent volumes of DMEM (10% FBS) supplemented with adenine (0.18 mM), HEPES (8 mM), insulin (5 µg/ml), hydrocortisone (0.5 µg/ml), EGF (10 ng/ml), F-12 (1 µl/ml, mixed at a 3:1 ratio) and 10 µg/ml L-ascorbic acid 2-phosphate; this was used throughout the rest of the culture period. Medium was replaced every 3 days of culture.

5.3.1. Fibroblasts grown in the 3D *in vitro* culture system form a tissue-like substance

In contrast to the typical behaviour of IPF fibroblasts that grow as monolayers on tissue culture plastic (**Figure 5.2A**), after 3 weeks of culture, the IPF fibroblasts in the 3D culture system appeared to be growing on top of each other and forming attachments to each other in such a manner that individual cells were indistinguishable (**Figure 5.2B**). This behaviour was observed to continue through to 6 weeks of culture (**Figure 5.2C**). This "tissue-like" substance formed by the IPF fibroblasts was clearly observed following incubation with trypsin-EDTA during cell harvesting (**Figure 5.2D**). The fibroblasts slowly detached from the hanging cell culture insert without detaching from each other, the layers of cells having formed one "sheet" of cells that peeled from the edge of the plate, and was aspirated as a single clump of cells.

After 3 weeks of culture, fibroblast medium was removed and replaced with fresh medium with 2 nM FK228 in 50% of inserts, and fresh medium containing the equivalent volume of

DMSO in the remaining inserts. After 48 hours, half of the untreated cells and FK228-treated cells were harvested using trypsin-EDTA. Cell pellets were washed 4 times in phosphate buffered saline before storage at -80°C until further use. The remaining fibroblasts were cultured for a further 3 weeks, with 50% of fibroblasts subjected to chronic treatment every 3 days with fresh FK228-containing medium. All cells were then harvested in the same manner as previously, and were stored at -80°C until further use.

After 3 weeks of culture and 48 hours of treatment with FK228, the IPF fibroblasts in the 3D culture system appeared slightly perturbed in comparison with the uniform appearance of the control (**Figures 5.2E and 5.2B**, respectively) with areas of denser patches of cells. After 6 weeks of culture and of chronic FK228 treatment every three days in the second half of the culture period, most fibroblasts had detached from the hanging cell culture insert in the majority of cases but remained in contact with each other, similar to following trypsin incubation (**Figure 5.2F**). However, instead of forming a "sheet-like, tissue-like" substance they formed a very dense ball of cells. This ball of cells remained attached to the insert. The frequent addition of fresh FK228 to the culture medium over time appeared to disrupt adhesion and matrix proteins in contact with the insert, and cause changes in fibroblasts interactions and matrix organisation.

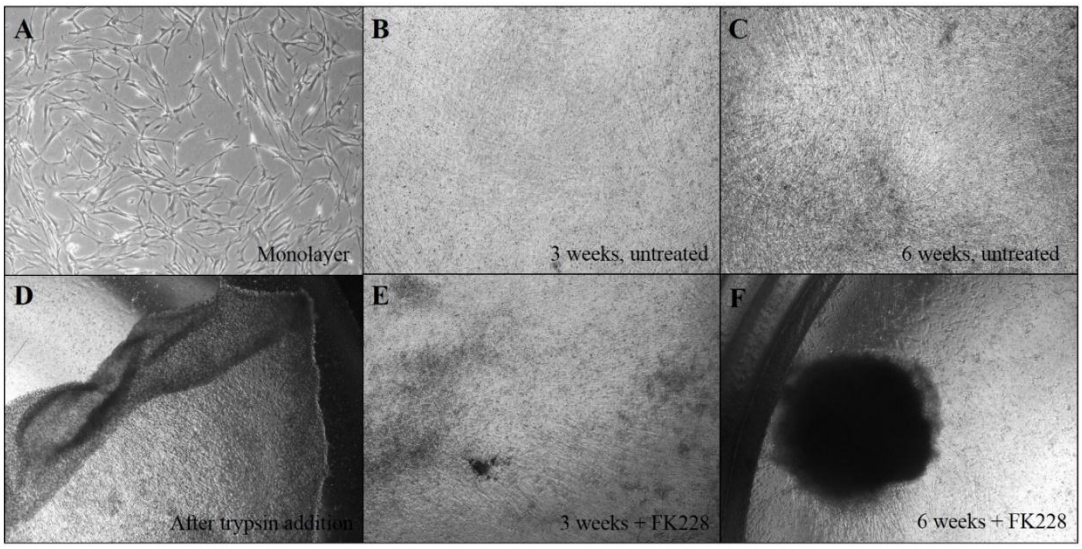


Figure 5.2 The appearance of fibroblasts grown in the 3D model culture system for 3 or 6 weeks, with and without FK228, in comparison with fibroblasts grown as monolayers. Representative microscopic images of IPF fibroblasts **A)** cultured on plastic for the proteomic experiment performed in **Chapter 4 B)** grown for 3 weeks in the 3D culture system, without treatment **C)** grown for 6 weeks in the 3D culture system, without treatment **D)** after 10 minutes of incubation with trypsin-EDTA for sample harvesting **E)** grown for 3 weeks in the 3D culture system, with 48 hours of FK228 treatment and **F)** grown for 6 weeks in the 3D culture system, with 3 weeks of FK228 treatment, administered every 3 days.

5.3.2. Extensive proteome coverage was achieved of IPF fibroblasts grown in the 3D *in vitro* culture system

As reflected in **Figure 5.3**, cells were lysed in 0.1% sodium dodecyl sulfate in an ultrasonic bath followed by a sonication probe. Fibroblasts from individual hanging cell culture inserts were pooled in pairs to yield approximately 100 µg protein per sample. The protein was precipitated by methanol/chloroform extraction prior to reduction for 1 hour, RT with 1 mM DTT, alkylation for 45 minutes, RT, in the dark with 5.5 mM iodoacetamide, followed by digestion. Proteins were digested firstly with 2 µg endoproteinase Lys-C for 4 hours at RT then 2 µg trypsin overnight at RT. Enolase and ClpB internal standards were spiked into each sample at 300 fmol and peptides were separated by OFFGEL fractionation into 12 peptide fractions, according to manufacturer's instructions. Each fraction was purified using a C18 Empore 96-well solid phase extraction plate before evaporation to completion and resuspension in loading buffer A (3% acetonitrile + 0.1% formic acid) for mass spectrometry analysis. The experiment was performed in triplicate.

Fractions were analysed by UPLC-HDMS^E. Half of each fraction was injected and peptides were separated by liquid chromatography using a NanoACQUITY UPLC system with a C18 reverse-phase column at a flow rate of 300 nL/minute over a gradient of 90 minutes of 3-50% buffer B (80% acetonitrile/dH_2O + 0.1% formic acid). Peptide ions were sprayed into a Waters Synapt G2-S system operating in positive ion mode, with ion mobility enabled prior to fragmentation. Data was collected in MS^E mode of acquisition, alternating between low energy (5V) and high energy (15V - 45V ramp) scans. Glu-fibrinopeptide (m/z = 785.8426, 100 fmol/µl) was used as LockMass and was sampled every 60 seconds for calibration.

Raw data files were processed using Protein Lynx Global Server (PLGS) version 3.0. Data were searched against the human UniProt database (downloaded 29/11/2013) using an Ion Accounting algorithm in PLGS 3.0.2. Estimates of absolute quantification were calculated using the Hi3 method of quantification (Silva et al., 2006) and values were normalised as described in **Appendix A.1**.

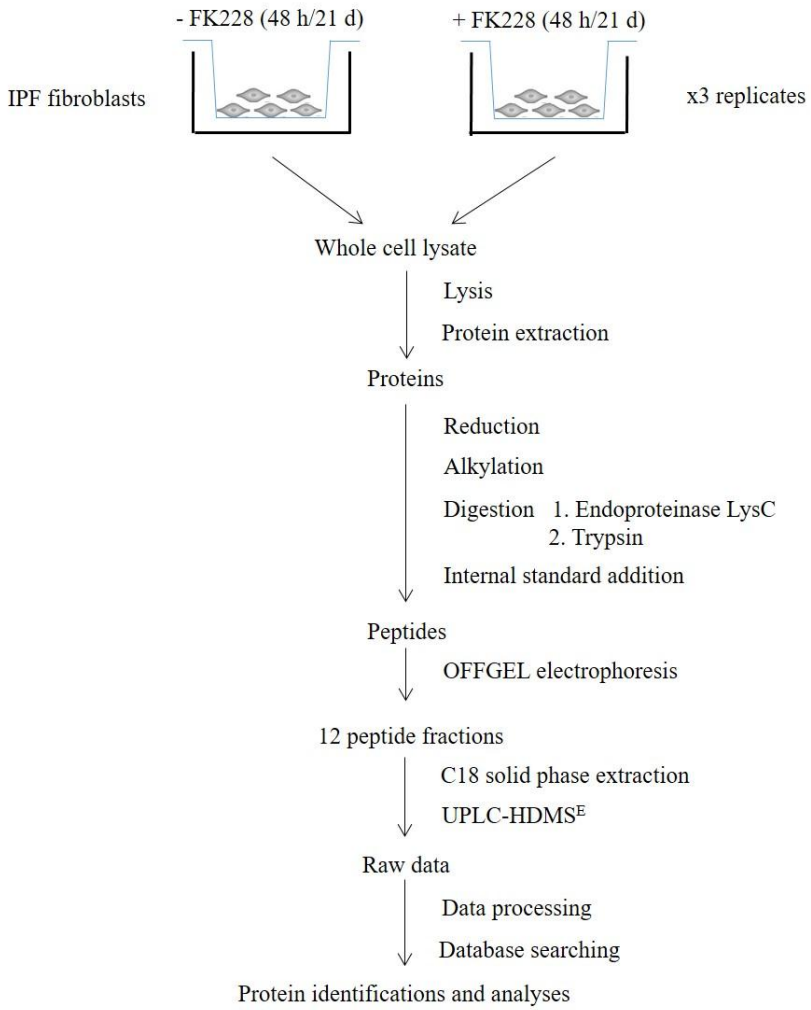


Figure 5.3 A representation of the method workflow used to identify and quantify the proteome of IPF fibroblasts grown in a 3D in vitro culture model, left untreated or treated with 2 nM FK228. Primary IPF fibroblasts were treated with 2 nM FK228 for 48 h or 21 days and compared against untreated IPF fibroblasts at each time point. Protein lysates were reduced, alkylated and digested into peptides using endoproteinase Lys-C and trypsin. Enolase and ClpB internal reference standards (Waters) were spiked into the digests for Hi3 quantification. Peptides were fractionated by OFFGEL electrophoresis (Agilent) and separated by UPLC, and eluted peptides were sprayed directly into a Synapt G2-S mass spectrometer (Waters) operating in MS^E mode with ion mobility enabled. Data were processed and searched using PLGS 3.0/3.0.2 prior to statistical analyses and pathway analyses. Each experiment was performed in triplicate.

In total, 4913 unique proteins were identified from all IPF fibroblast samples grown in the 3D *in vitro* culture system by mass spectrometry analysis; 4794 of these were quantified with estimates of absolute amounts using the Hi3 method of quantification. On average,

without FK228 treatment, 2170 proteins were quantified from IPF fibroblasts grown for 3 weeks in the 3D culture system, and 1447 proteins were quantified from IPF fibroblasts grown for 6 weeks. Slightly fewer identifications from fibroblasts grown for 6 weeks in the 3D system may be due to the enhanced "tissue-like" structure previously described formed after a longer culture period and increased production and deposition of a collagen-rich matrix, which could lead to difficulties in cell lysis and protein digestion during mass spectrometry sample preparation.

In comparison, the number of quantified proteins from IPF fibroblasts explanted from the same patient biopsy (IPF patient 1) but cultured as a monolayer on tissue culture plastic totalled 2718 (**Chapter 4**). However, only a single biological replicate was analysed for this particular monolayer culture. Therefore, a wider look was taken at expression data from IPF monolayers cultured from different patient biopsy explants (the IPF fibroblast proteomic data in **Chapter 4**, n=5), to determine the average number of proteins identified from IPF fibroblasts grown using this type of culture method. This analysis revealed that the average number of proteins quantified from IPF fibroblast monolayer culture is 2358, with a range of 1500 proteins.

The number of quantified proteins for all three culture conditions studied here are shown in **Figure 5.4**. There is large overlap in the number of protein identifications between monolayers and fibroblasts grown for 3 weeks in the 3D culture system. The mean number of proteins identified from fibroblasts grown for 6 weeks in the 3D culture system is lower than the fewest number identified in the monolayer culture (1447 and 1686 proteins, respectively). At present, in the Brooke laboratory IPF fibroblasts are typically cultured for the full 6 weeks in the 3D *in vitro* system in order for the fibroblasts to deposit a substantial ECM for analysis, therefore these samples are the most important to study. These data suggest that the use of a more physiologically relevant model than monolayer culture for global protein profiling of IPF fibroblasts can be performed at the expense of a much less in-depth proteomic characterisation of the cells.

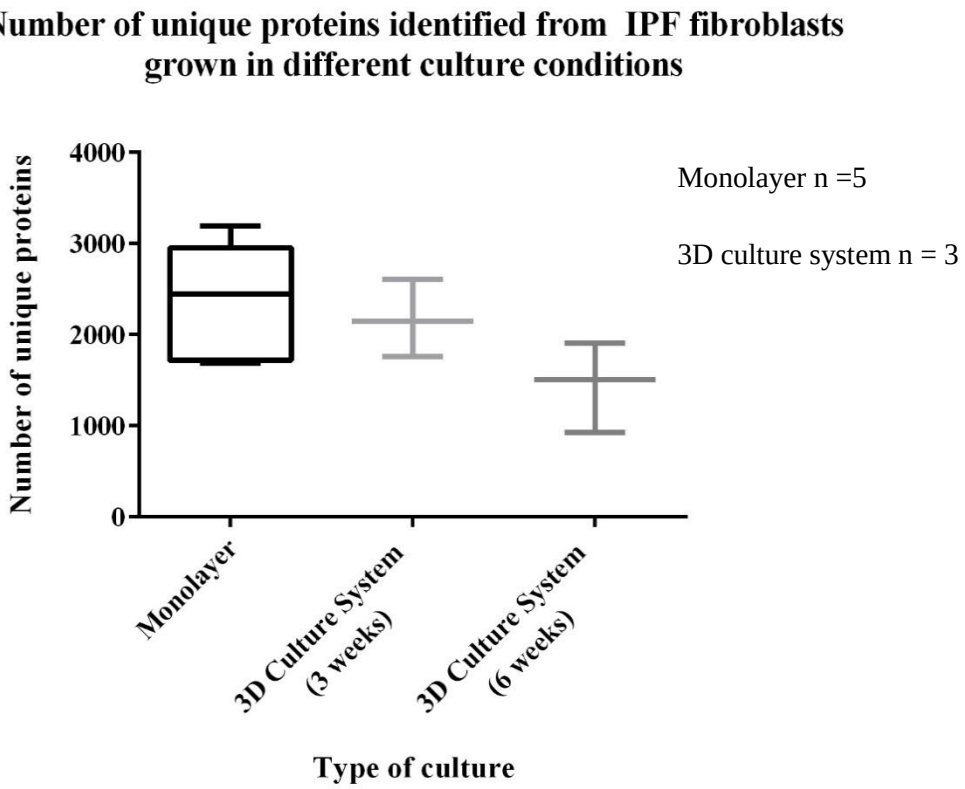


Figure 5.4 The number of unique proteins identified with estimates of absolute quantification in biological replicates of IPF fibroblasts grown either on plastic as monolayers, or in the 3D culture system for either 3 weeks or 6 weeks, all untreated. Fibroblasts explanted from biopsies from five IPF patients were grown as monolayer cultures, and three replicate fibroblast explant cultures from one patient biopsy were grown for either 3 weeks or 6 weeks in a 3D *in vitro* culture system, before protein extraction, digestion and analysis by mass spectrometry.

5.3.2.1. Characterisation of IPF fibroblasts cultured in the 3D model system revealed that there are large proteomic differences in comparison with IPF fibroblasts cultured in a monolayer system

To assess whether the 3D *in vitro* culture system would be a more worthwhile model for studying IPF fibroblasts at the molecular level than the typical monolayer method of culture, global protein expression levels of fibroblasts in the two culture systems were compared between models.

Six weeks of culture in the 3D system is the time-point developed in the Brooke laboratory by Dr Jones as the proposed most representative model, with characteristics such as matrix

formation in the extracellular space. Therefore, in the first instance, the expression levels of proteins extracted from the primary fibroblast donor culture "IPF patient 1" grown in monolayer culture and analysed by mass spectrometry (n=1, **Chapter 4**) were compared to those of proteins isolated from fibroblasts from the same donor, but cultured for 6 weeks in the 3D culture model (n=3). This was performed using fold-change analysis, with a 2-fold cut-off for significance. Only proteins quantified in at least two replicates of both 3D culture model conditions were included. As only one replicate of the IPF patient 1 monolayer culture was analysed, only proteins also identified in at least one other IPF monolayer culture were used in the analysis (2471 out of 2718 proteins).

This analysis revealed that out of 1003 proteins compared for differences in expression, 189 proteins had upregulated expression in the 3D model culture and 255 proteins had downregulated expression, by at least 2-fold. This equates to 44% of proteins analysed having a significant difference in abundance between the two culture systems, a marked difference in the cellular proteome. Since these proteins were extracted from fibroblasts from the same donor, this shows that the manner in which fibroblasts are cultured has a large effect on their intracellular protein expression, which is reflected in the differences in behaviour. This emphasises the importance of developing a model that accurately reflects the pathology of IPF, in order to study it effectively *in vitro*.

Gene ontology (GO) analysis was performed using GOrilla to identify GO terms significantly overrepresented for the differentially expressed proteins as well as proteins only in one condition (Chen et al., 2009). GO terms with a *p*-value of <0.05 were considered significant. The resulting GO terms were collapsed into categories of related terms using REVIGO.

GO analysis revealed that the proteins with upregulated expression in the 3D model culture were associated with 220 enriched GO terms, and the proteins with downregulated expression were associated with 225 enriched GO terms, which have been summarized into distinct terms in **Figure 5.5**. The majority of the proteins with upregulated expression (**A**) or downregulated expression (**B**) have roles in various areas of cellular metabolism; many proteins involved in translation/protein synthesis had lower expression in the 3D model.

The differences in metabolic profile between the fibroblasts in the two culture systems is likely due to differing cellular needs and functions between cells growing as a tissue and cells growing in a monolayer. Almost all of the proteins found to be downregulated in the 3D model culture and involved in translation were ribosomal proteins and eukaryotic translational initiation factors. One might expect that protein synthesis would be increased in this culture system due to the synthesis of matrix, but these results suggest that there are

more translation and protein synthesis processes occurring in fibroblasts cultured as a monolayer.

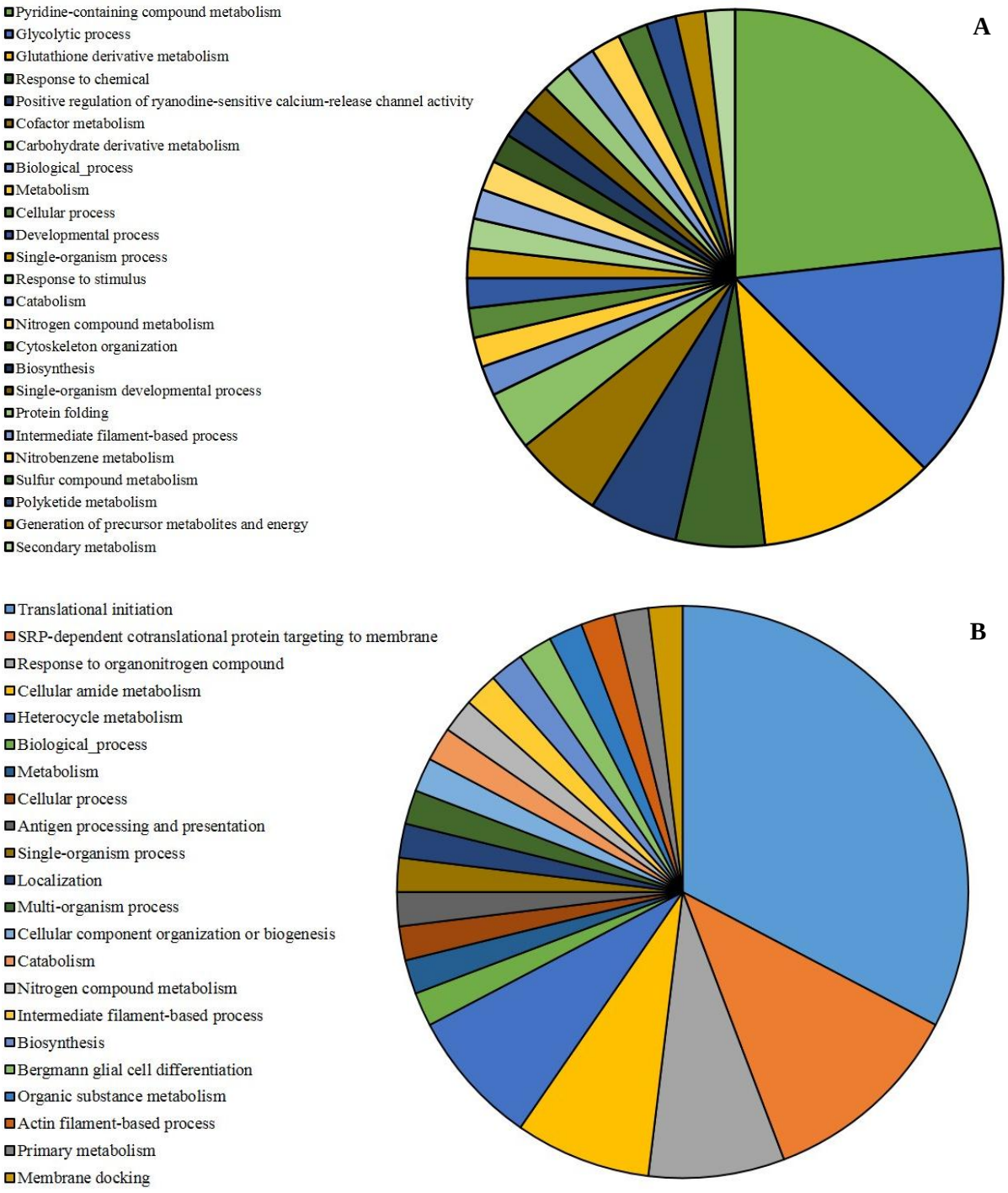


Figure 5.5 Gene ontology enrichment analysis of proteins with A.) upregulated expression in 3D culture or B.) downregulated expression in 3D culture when comparing protein abundance to IPF fibroblasts grown in monolayer culture. Individual GO terms were combined into related terms using REVIGO that are reflected in the pie chart. Size of the terms is adjusted (log10*p*-value) to reflect the *p*-values.

Four hundred and sixty-nine proteins were quantified from the IPF monolayer that were not identified in any replicates of the 3D model-cultured fibroblasts grown for either 3 or 6 weeks. Gene ontology analysis of these proteins highlighted 140 enriched GO terms within this set of proteins, represented in a pie chart in **Figure 5.6**. The most enriched GO term was RNA processing, a term encompassing related processes including transcription and mRNA metabolism. Many of these proteins were also involved in metabolism. These results correlate with the previous analysis, in that there are significant differences in cellular metabolism and gene expression/protein synthesis between the fibroblasts grown in monolayers and in the 3D model.

This analysis also highlights the organisation of molecules/components within the cell and cell cycle processes as biological processes associated with the proteins only identified in the monolayers. Differences in cellular component organisation between fibroblasts in the two systems could be linked to differences in the movement of important molecules around and out of the cell, such as the rate of secretion of molecules, including matrix proteins. It could also include internal rearrangement for cell division/proliferation, which is also likely to be differentially regulated between the two conditions due to the differences in substrate stiffness and thus extracellular cues.

Some of the differences in protein expression and cell behaviour could also be due to differences in the medium supplements between the two culture systems. For example, the addition of insulin and EGF to the 3D culture medium would promote cell proliferation in this model, and the monolayer culture medium contains double the amount of FBS, which contains growth-promoting factors. The viability of the cells within the layers in the 3D model is also unknown as this was not assessed; it is possible that some of the cells were dying, which would have an effect on overall protein expression of the samples.

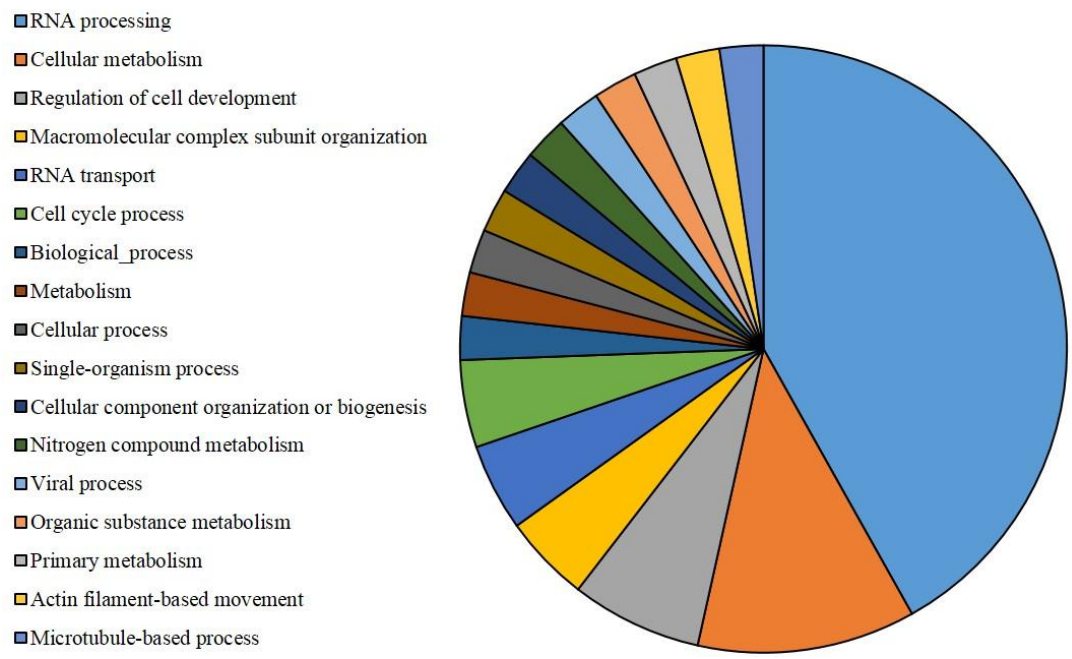


Figure 5.6 Gene ontology enrichment analysis of proteins only quantified in IPF fibroblasts grown in vitro as monolayers and not detected in fibroblasts cultured in the 3D model system. Individual GO terms were combined into related terms using REVIGO that are reflected in the pie chart. Size of the terms is adjusted (log10*p*-value) to reflect the *p*-values.

5.3.2.2. The proteins with the largest differences in expression between 3 weeks and 6 weeks of IPF fibroblast culture in the 3D culture system were involved in protein synthesis, transport and cell-cell adhesion

Differences in protein expression were compared between IPF fibroblasts cultured for 6 weeks in the 3D model system and those cultured for 3 weeks, to find out how the proteome changes as the cells begin to form layers with matrix deposited in the extracellular space. Since three replicates of the same primary donor culture across both of these conditions were analysed by mass spectrometry, statistical analysis could be performed on proteins quantified in all samples using Student's T-Test. *p*-values < 0.05 were considered significant. Proteins identified in three replicates of one condition and in none of the other condition were also considered significant.

Out of 584 proteins quantified in all 3D model samples and analysed by T-Test, 26 had a significant difference in expression ($p<0.05$). This result is reflected in a heat map in **Figure**

5.7, which shows clustering of the data to condition, and within each condition, the expression values are either all increased or decreased in comparison to the other condition (red = increased, blue = decreased). Since each replicate was generated from the same donor culture, one would expect to see more uniformity between biological replicates than was observed between monolayers in **Chapter 4**. Most of these proteins had decreased expression after an extra 3 weeks of culture. In addition to these significantly differently expressed proteins, 72 proteins were detected only in fibroblasts grown for 3 weeks, and just three were detected only after 6 weeks of culture. This could be because fewer proteins were identified after 6 weeks of fibroblast culture than after 3 weeks, rather than a reflection of fewer proteins being present in these samples. The full lists of proteins with significant differential expression can be found in **Appendix D.1-D.3**.

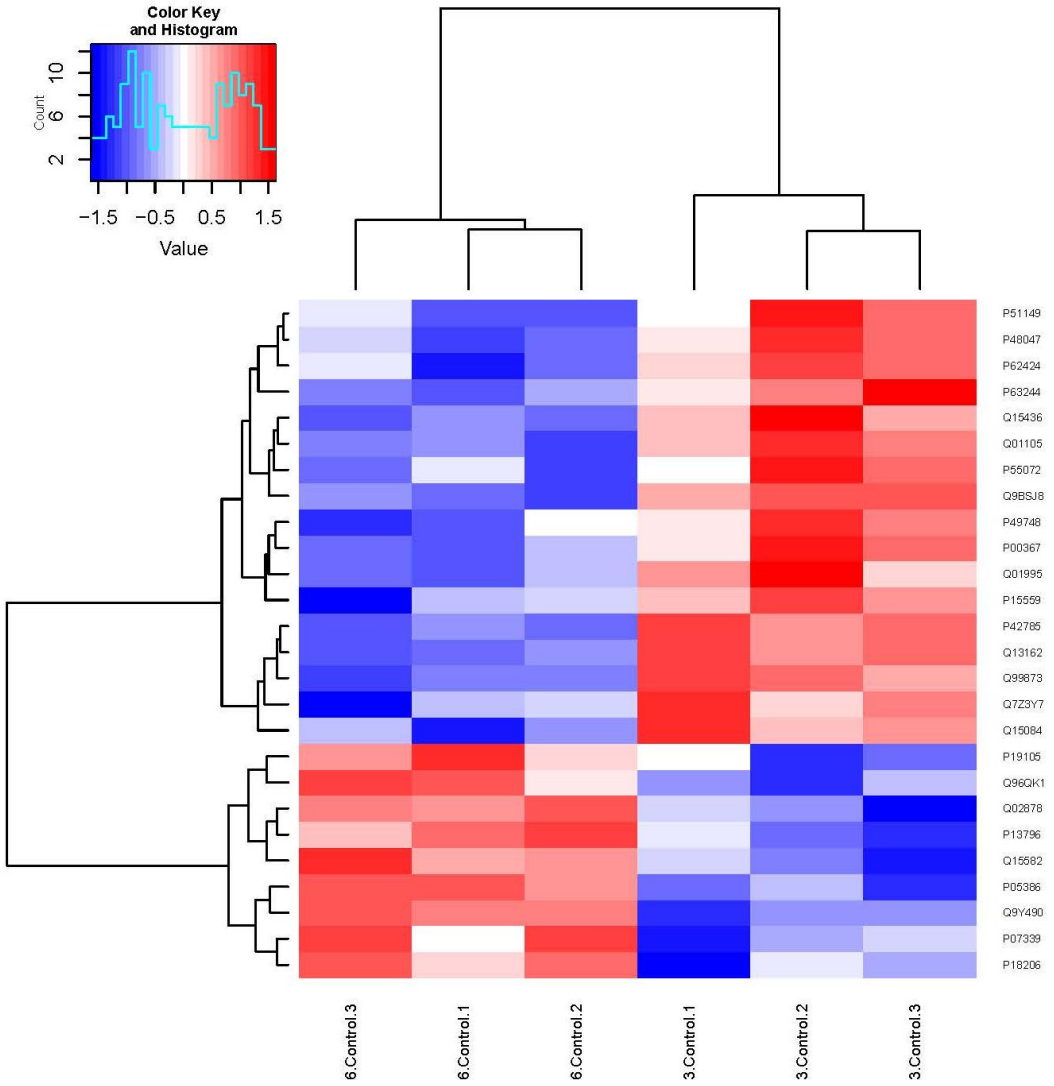


Figure 5.7 Heat map illustrating the expression levels of the proteins differentially expressed between IPF fibroblasts grown for 3 weeks and 6 weeks in the 3D in vitro culture system. Heat maps were generated in R using ggplots and d3 heatmap packages. The expression values of proteins were averaged and scaled across every row to indicate the number of standard deviations above (red) or below the mean (blue). Distance between experiments and proteins was calculated and hierarchical clustering was performed using ward linkage.

The 26 proteins with significantly different expression were analysed to define functional roles in the cell by gene ontology analysis as performed in section **5.3.2.1**. Over half of the enriched GO terms for this dataset were related to protein synthesis and transport (**Figure 5.8**). This may reflect changes in the rate of production and secretion of matrix proteins

between 3 weeks and 6 weeks. The majority of the proteins related to these GO terms had downregulated expression after 6 weeks of culture. This could indicate matrix components having already been deposited at this time-point, and the present ECM being maintained rather than new components being synthesized to help establish the matrix, as would be occurring at 3 weeks.

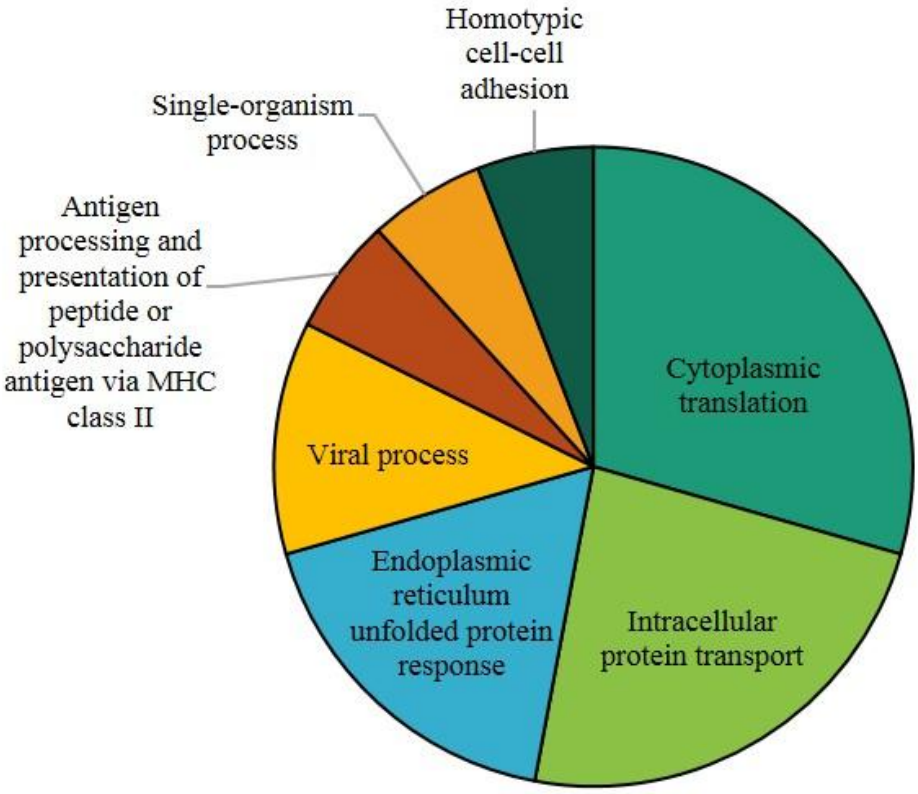


Figure 5.8 Gene ontology enrichment analysis of proteins with significantly different expression levels (p<0.05) between 3 and 6 weeks of culture in the 3D model system. Individual GO terms were combined into related terms using REVIGO that are reflected in the pie chart. Size of the terms is adjusted (log10p-value) to reflect the p-values.

An interesting result is the enrichment of the GO term "homotypic cell-cell adhesion" within this dataset. This term was associated with three adhesion-related proteins: myosin light chain 12a, talin-1 and vinculin. All three proteins showed increased expression in fibroblasts grown for 6 weeks, at which point they had formed a tissue-like structure, adhering strongly to each other in layers.

Also observed was an increase in transforming growth factor beta-induced protein after 6 weeks of culture. As discussed in **Chapter 3**, TGF-β_1 is an important profibrotic molecule whose effects include the upregulation of collagen.

5.3.2.3. A higher number of extracellular matrix proteins were identified from IPF fibroblasts in the 3D model system than from monolayer culture

As previously noted, the advantage of using the 3D culture system as a model for IPF pathogenesis is that the fibroblasts appear more effective at producing and depositing collagenous matrix within the extracellular spaces. Therefore, the 3D model and monolayer-combined datasets were analysed for the presence or absence of extracellular matrix proteins as reported in the literature (Hynes and Naba, 2012). Overall, more ECM proteins were identified in the 3D model than the monolayers. Out of 47 matrix proteins identified in at least one sample across the dataset, 41 were identified after 3 weeks of culture in the 3D model system, 34 were identified after 6 weeks, and 27 ECM proteins were identified from the monolayer culture. The results are shown in **Table 5.1**. A list of expression values for these ECM proteins can be found in **Appendix D.4**.

Table 5.1 List of extracellular matrix-related proteins identified from IPF fibroblasts grown in 3D and monolayer cultures.

Protein accession	Protein name	Identified in fibroblasts from culture system (✓/×)		
		3D system (3 weeks)	3D system (6 weeks)	Monolayer
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	✓	✓	×
P21810	Biglycan	✓	✓	×
Q05682	Caldesmon	✓	✓	✓
P02452	Collagen alpha-1(I) chain	✓	✓	✓
P02461	Collagen alpha-1(III) chain	✓	✓	✓
P12109	Collagen alpha-1(VI) chain	✓	✓	✓
Q02388	Collagen alpha-1(VII) chain	✓	✓	×
P27658	Collagen alpha-1(VIII) chain	✓	✓	×
Q99715	Collagen alpha-1(XII) chain	✓	✓	✓
P39059	Collagen alpha-1(XV) chain	✓	✓	×
P39060	Collagen alpha-1(XVIII) chain	×	✓	×
Q86Y22	Collagen alpha-1(XXIII) chain	✓	×	×
P08123	Collagen alpha-2(I) chain	✓	✓	✓
P08572	Collagen alpha-2(IV) chain	✓	✓	×
P05997	Collagen alpha-2(V) chain	✓	✓	×

P12110	Collagen alpha-2(VI) chain	✓	✓	✓
P12111	Collagen alpha-3(VI) chain	✓	✓	✓
A8TX70	Collagen alpha-5(VI) chain	×	✓	×
Q9Y5P4	Collagen type IV alpha-3-binding protein	✓	×	×
P07585	Decorin	✓	✓	✓
Q16610	Extracellular matrix protein 1	×	×	✓
P35555	Fibrillin-1	✓	✓	×
P02751	Fibronectin	✓	✓	✓
P07942	Laminin subunit beta-1	✓	×	✓
P55268	Laminin subunit beta-2	×	×	✓
P51884	Lumican	✓	×	×
P50281	Matrix metalloproteinase-14	✓	✓	✓
Q9ULZ9	Matrix metalloproteinase-17	×	✓	×
O60882	Matrix metalloproteinase-20	✓	×	×
P01033	Metalloproteinase inhibitor 1	✓	✓	✓
P35625	Metalloproteinase inhibitor 3	×	×	✓
Q14112	Nidogen-2	✓	✓	×
Q15063	Periostin	✓	×	×
Q8NBJ5	Procollagen galactosyltransferase 1	✓	✓	✓
Q02809	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 1	✓	✓	✓
O00469	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 2	✓	✓	✓
O60568	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 3	✓	✓	✓
Q32P28	Prolyl 3-hydroxylase 1	✓	✓	✓
Q8IVL6	Prolyl 3-hydroxylase 3	✓	×	✓
P13674	Prolyl 4-hydroxylase subunit alpha-1	✓	✓	✓
O15460	Prolyl 4-hydroxylase subunit alpha-2	✓	✓	✓
Q7Z4N8	Prolyl 4-hydroxylase subunit alpha-3	✓	×	✓
P24821	Tenascin	✓	✓	×
P07996	Thrombospondin-1	✓	✓	✓
P35442	Thrombospondin-2	✓	×	×
Q15582	Transforming growth factor-beta-induced protein ig-h3	✓	✓	✓
P04004	Vitronectin	✓	×	×
	Total proteins identified out of 47 =	41	34	27

An important difference between the matrix proteins identified between the two methods of culture was the number of collagen proteins identified. Since collagen is the main component of matrix, it provides an indication of the extent of ECM deposition and the efficacy of ECM analysis. Fifteen collagen proteins were identified overall, and all but two were present in the 3-week 3D culture, all but one was present in the 6-week 3D culture, but only seven were identified in the monolayer culture. An interesting observation is that elastin, another major ECM fibre, was not detected in either type of culture method. Neither was it identified in any other IPF fibroblast monolayers previously studied. This could be because it is difficult to digest, and indicates that the method for sample preparation may require further optimisation. The glycoproteins tenascin, nidogen and biglycan were present in the 3D system at both time-points but were not identified in the monolayer culture. Tenascin has previously been shown to be present in fibroblastic foci of IPF patients (Jones, pers comm). Other ECM proteins not detected in the monolayer include periostin and fibrillin-1. Only two proteins, laminin subunit beta-2 and metalloproteinase inhibitor 3, were identified only in the IPF monolayer sample.

Figure 5.9 shows a heat map of the presence/absence and differences in expression of the extracellular proteins identified in at least two of the samples involved in the analysis which shows the direction of change in abundance for each protein (red = increased, blue = decreased, black = not detected). Eleven proteins were identified across the whole dataset; one replicate of the 6-week 3D model fibroblast condition contained a high proportion of missing data for ECM proteins. Several proteins in the monolayer sample showed much lower expression compared to the rest of the dataset; these were predominately collagen VI proteins (P12109-11), plus procollagen-lysine 2-oxoglutarate 5-dioxygenase 1and 2 (PLOD1 and PLOD2), involved in collagen cross-linking. In contrast, collagen I expression was generally higher in fibroblasts in the monolayer compared to the 3D model. It has previously been reported that fibroblasts grown within a collagen matrix (as is formed in this culture model) have increased collagen VI gene expression compared to monolayer cultures, and collagen I gene expression is decreased (Hatamochi et al., 1989) and these results support this finding. All three collagen VI proteins also had increased expression after 6 weeks in the 3D culture when compared to 3 weeks. The most highly expressed ECM protein in fibroblasts from monolayer culture compared to the 3D model was collagen XII, which associates with collagen I (Ricard-Blum, 2011). Transforming growth factor beta-induced protein, which increased between the two time-points in the 3D model, was expressed at a much lower level in the monolayer culture, indicating that the ECM may influence the expression of this protein.

The lack of detection for some of these proteins may be an artefact of the culture system. One factor that must be considered in this kind of analysis is that in both experiments only the fibroblasts were analysed by mass spectrometry, and not the secretome, which is more likely to contain extracellular matrix proteins, especially in terms of the monolayer culture. Secretome analysis would greatly enhance the information given by this study. The ECM proteins identified here are most likely intracellular, synthesized but not yet secreted, due to using trypsin for cell harvesting. The nature of the 3D model and the manner in which it was harvested, by detaching from the substrate as one sheet, likely means that some of the deposited ECM between cell layers was also collected and analysed, resulting in more ECM protein identifications. For example, lumican was present in a sample taken from the 3D model, however a previous study (**Chapter 3**) has shown that in fibroblast monolayer cultures this protein can be found in the secretome rather than the cellular proteome.

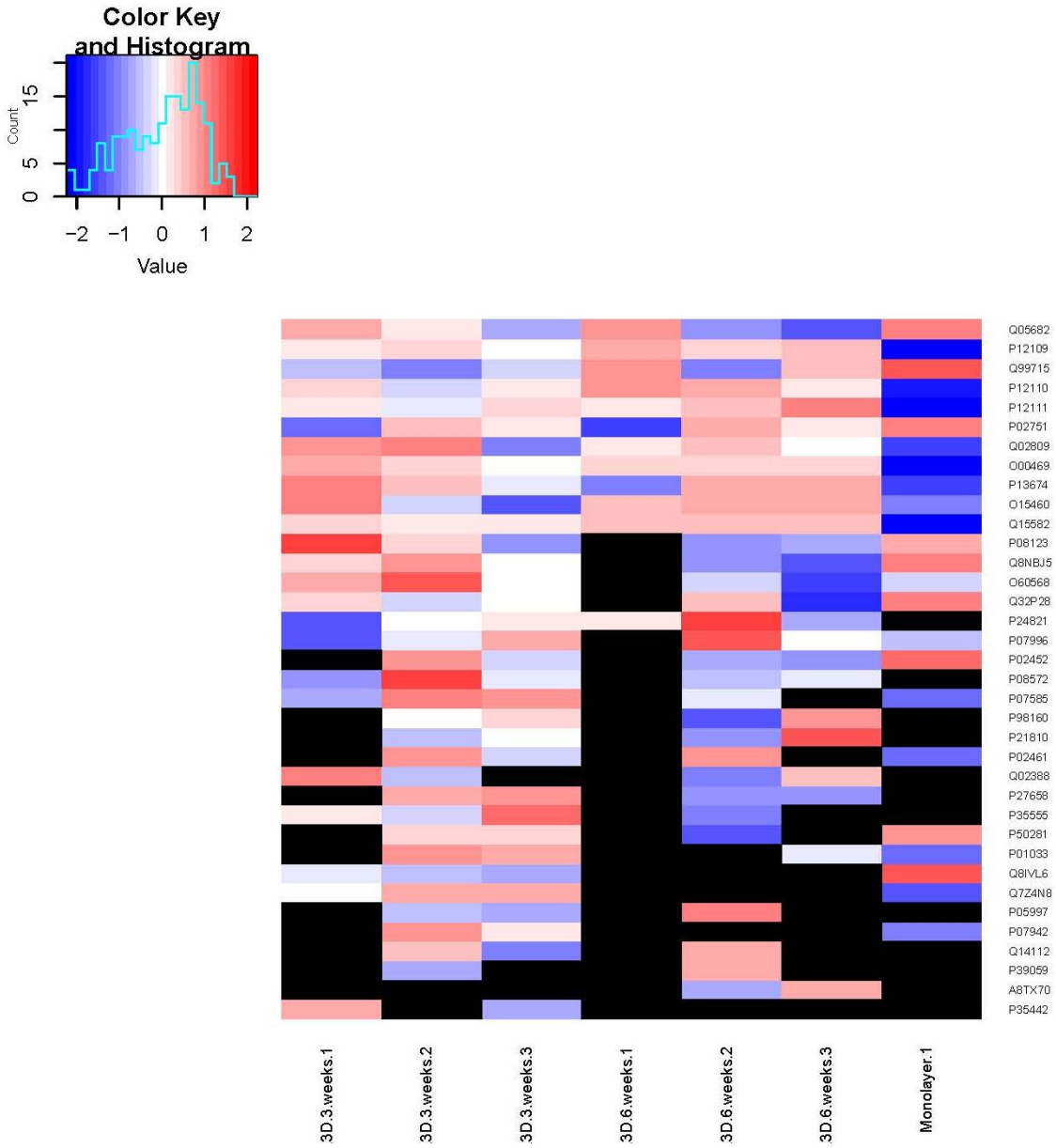


Figure 5.9 Heat map illustrating the expression levels of the extracellular matrix proteins expressed in IPF fibroblasts grown for 3 weeks and 6 weeks in the 3D in vitro culture system and in the monolayer culture system. Heat maps were generated in R using ggplots and d3 heatmap packages. The expression values of proteins were averaged and scaled across every row to indicate the number of standard deviations above (red) or below the mean (blue). Distance between experiments and proteins was calculated and hierarchical clustering was performed using ward linkage. Black indicates non-detection of a protein within a sample.

The results from the comparisons of general protein expression and ECM protein expression between 3D and monolayer culture suggest that the 3D system is more informative about the extracellular matrix as more proteins can be measured, but for a full characterisation of this, additional secretome analysis would be more informative. If the main interest is the molecular profile of the fibroblasts themselves, for example for studying a basic molecular response to treatment, more protein identifications are gained from monolayer culture, but their expression may be less reflective of pathogenesis than the 3D tissue-like culture; the data from the latter system may be of more value from this perspective despite fewer numbers of proteins being identified.

5.3.3. Acute and chronic FK228 treatment cause different proteomic effects in IPF fibroblasts grown in the 3D culture system

IPF fibroblasts grown for both 3 weeks and 6 weeks in the 3D culture system were treated with the histone deacetylase inhibitor FK228, to elucidate its effects on protein expression in a different culture model to the monolayer culture studied in **Chapter 4**. Comparison of protein profiles between fibroblast cultures in the two systems were used to find out whether the different cellular properties influenced by the two culture conditions would result in different molecular responses to FK228. It was thought that after 6 weeks total of culture in the 3D system, at which time fibroblasts had incorporated themselves into a tissue-like structure, FK228 may not be taken up into the cells as easily and exert an effect that could easily be observed with only 48 hours of treatment. Therefore, in order to be able to compare protein expression in this type of culture and the monolayer culture, cells cultured for 3 weeks in the 3D system at the beginning of tissue formation were treated with FK228 for 48 hours, and the more physiological 6-week culture was treated for a longer time period. The latter is also more realistic in terms of patient treatment since patients with T-cell lymphoma are currently treated chronically, once a week for four weeks at a time.

5.3.3.1. The acetylation of lysine residues increases following treatment with FK228 at both time-points

Prior to proteomic analysis, all cell lysates were analysed by western blotting for histone acetylation to check the efficacy of FK228, as FK228 is known to cause an increase in the acetylation of histone H3 by blocking the action of histone deacetylases. Cell lysates were subjected to SDS polyacrylamide gel electrophoresis and western blotting for the detection of acetylated lysine with β-actin as a loading control.

Figure 5.10 shows increases in histone lysine acetylation after both 48 hours and 3 weeks of FK228 treatment, at 3 weeks and 6 weeks of fibroblast culture in the 3D system, respectively, validating the HDAC inhibitory mechanism of FK228 in these fibroblast cultures.

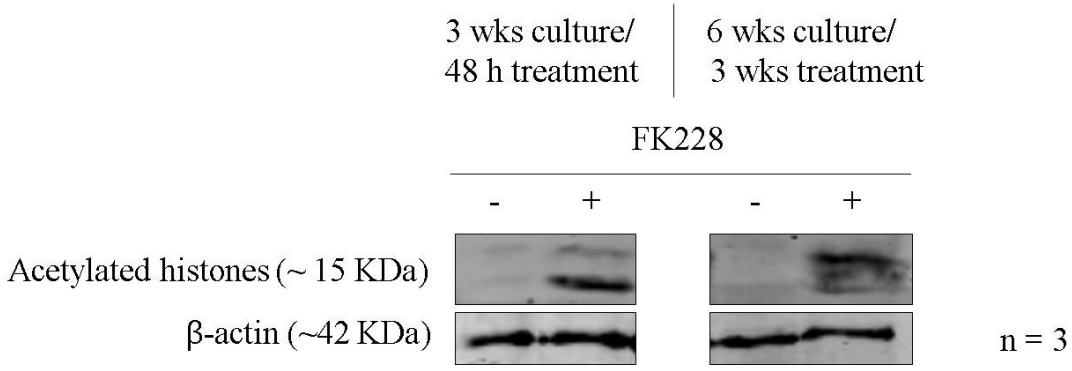


Figure 5.10 Western blot of histone lysine acetylation in IPF fibroblasts after both 48 hours and 3 weeks of FK228 treatment. The expression of acetylated lysine residues was analysed by western blotting after 3 weeks of culture and 48 hours of FK228 treatment, and after 6 weeks of culture and 3 weeks of FK228 treatment, with β-actin as a loading control.

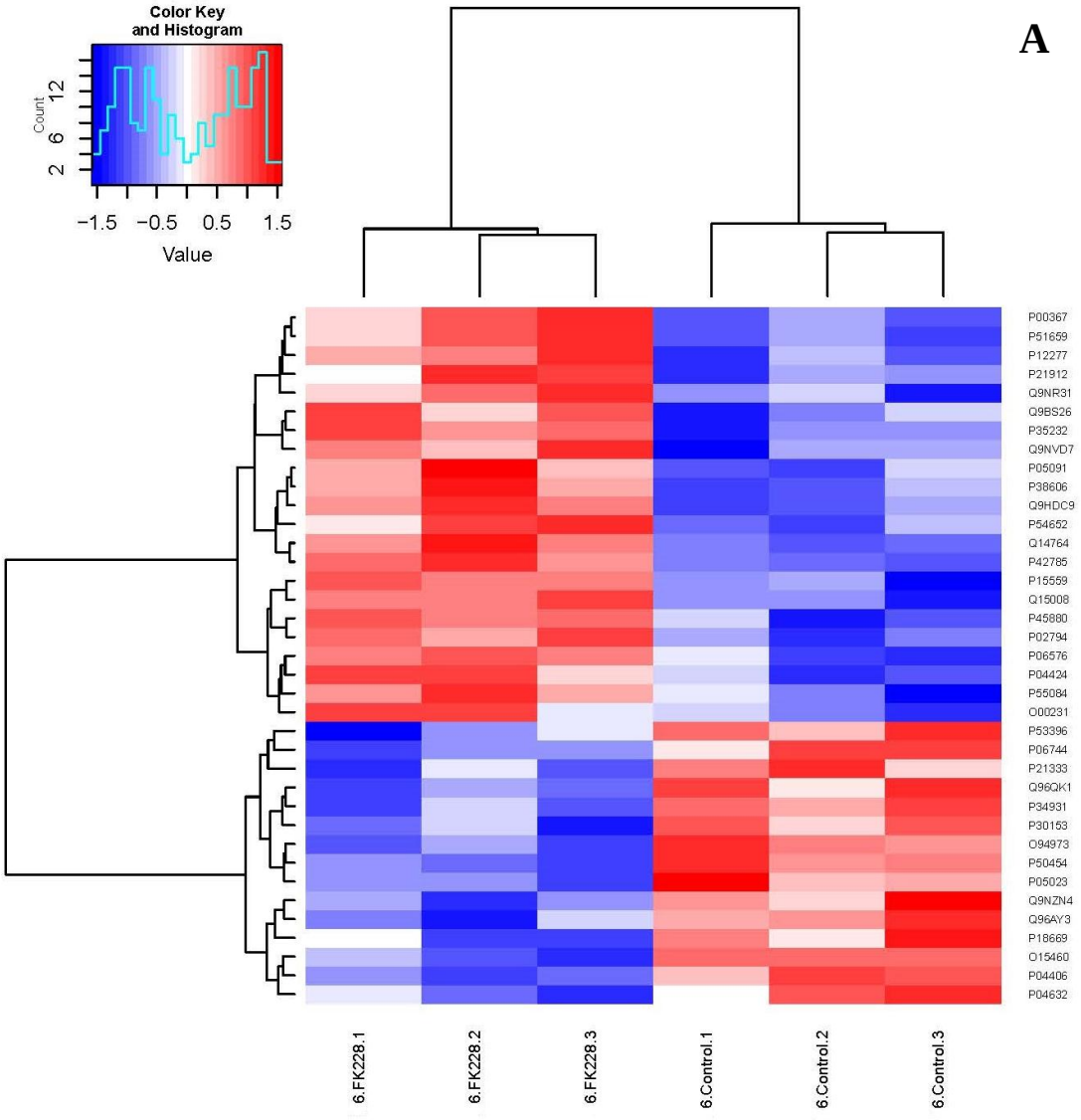
5.3.3.2. Acute and chronic FK228 treatment affected the expression of different IPF fibroblast proteins in the 3D model

The protein expression of IPF fibroblasts cultured for 6 weeks in the 3D system, both untreated and treated for 3 weeks with FK228, was analysed by mass spectrometry, with three biological replicates of each condition. Seven proteins were identified in all three replicates of the untreated samples without detection in any of the FK228-treated samples, and 18 proteins were identified in all three replicates of the FK228-treated samples without detection in any untreated samples.

From 477 proteins identified with quantification across all six samples and analysed by unpaired Student's T-Test, 37 had a significant difference in expression ($p<0.05$). The heat map in **Figure 5.11A** shows that the replicates cluster to condition, and that within this list of proteins, more proteins had increased expression after treatment than decreased expression (red = increased, blue = decreased).

The same analysis performed on data from three biological replicates of IPF fibroblasts cultured for 3 weeks in the 3D system, with or without FK228 treatment, found 18 proteins that had a significant difference in expression, from 715 analysed. It might be expected that

fewer proteins would change significantly in expression after just 48 hours of treatment compared to three weeks. However, 47 proteins that were present in all untreated fibroblasts were not detected in any FK228-treated samples, indicating there was a greater response to acute treatment than the T-Test results would suggest. There were no proteins present only in all treated samples and not in the untreated samples. Also in contrast to the protein expression changes after chronic treatment, the majority of proteins with a significant difference in expression after acute treatment decreased in expression (**Figure 5.11B**). The full lists of proteins with significant differential expression after FK228 treatment at both time-points can be found in **Appendix D.5-D.9**.



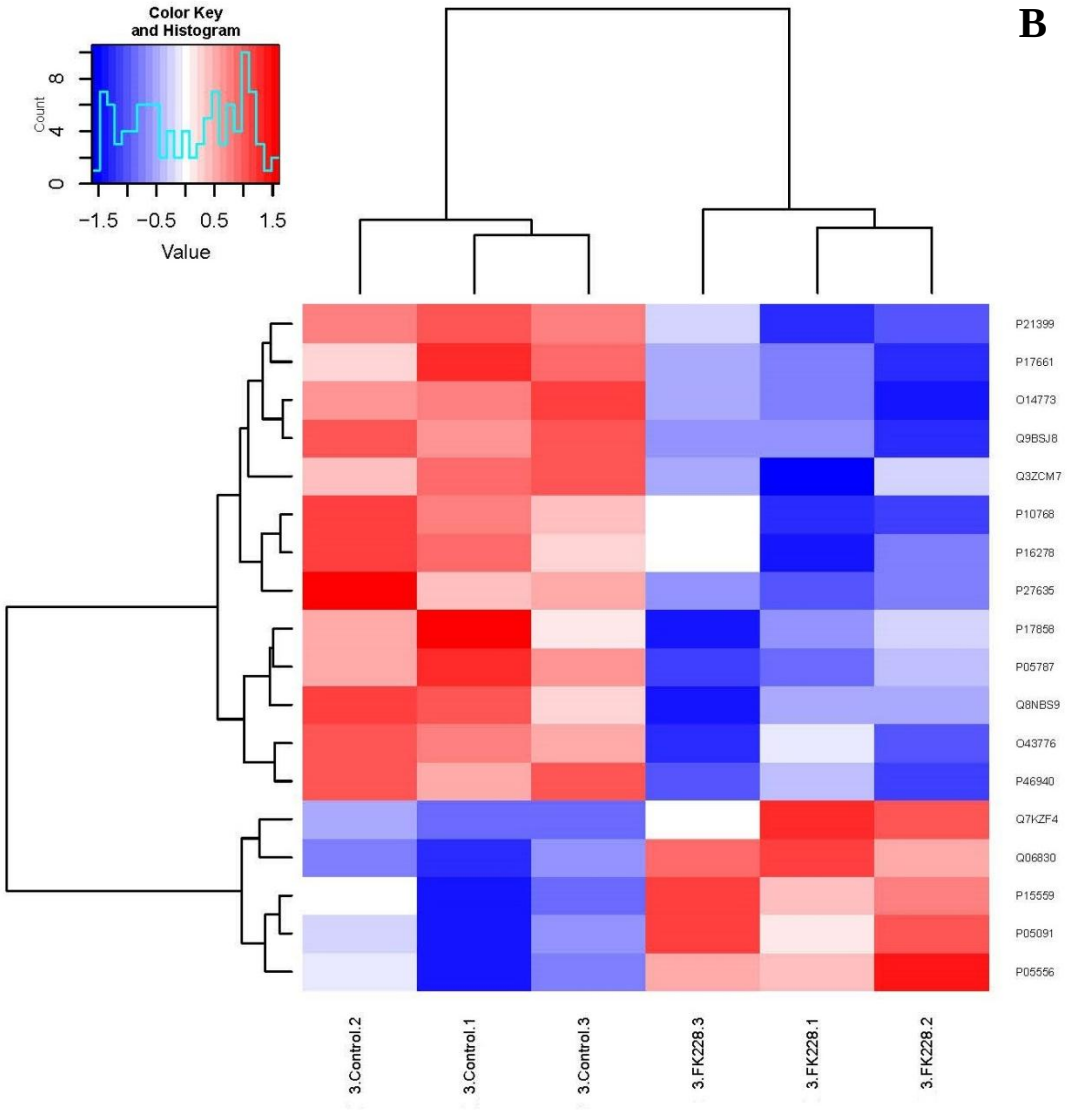


Figure 5.11 Heat maps illustrating the expression levels of the proteins differentially expressed between IPF fibroblasts grown in the 3D in vitro culture system for A) 6 weeks, with or without 3 weeks of FK228 treatment and B) 3 weeks, with or without 48 hours of FK228 treatment. Heat maps were generated in R using ggplots and d3 heatmap packages. The expression values of proteins were averaged and scaled across every row to indicate the number of standard deviations above (red) or below the mean (blue). Distance between experiments and proteins was calculated and hierarchical clustering was performed using ward linkage.

Fifty-three unique proteins had a significant difference in expression across the two datasets (p<0.05). Only 2 proteins were common to both: NAD(P)H dehydrogenase [quinone] 1 (NQO1) and aldehyde dehydrogenase, mitochondrial (ALDH2). This was despite over 400

proteins being included in both T-Test analyses (only proteins identified in all three replicates of both untreated and FK228-treated conditions were included in T-Test analysis in both cases – 765 additional proteins were analysed in one dataset but not the other due to missing data/non-detection and were therefore not included in this comparison). After both acute and chronic treatment, their expression significantly increased. Five proteins out of the fifty-three also had significantly different expression between three and six weeks (section **5.3.2.2**). This implies that more than 90% had stable expression between these time-points, and the differences in expression were due to FK228 stimulation rather than differences in culture period.

Each of the fifty-three proteins that had significantly different expression after FK228 treatment in either dataset was examined for its expression in the opposite dataset, to find out whether a different dosage period had caused a similar change in expression (albeit not a significant change, except for NQO1 and ALDH2, as already discussed). Seventy percent of these proteins were either upregulated or downregulated in expression at both time-points, leaving 30% whose expression was modified in the opposite manner. This suggests that FK228 causes a similar alteration in expression of the majority of these proteins during both acute and chronic treatment; however, the expression of some proteins is affected differently depending on the duration of treatment.

5.3.3.3. Different proteins were significantly affected by 48 hours of FK228 treatment in the 3D culture system compared to the monolayer culture

To elucidate whether the same proteins were affected by FK228, the proteins found to have significantly different expression after 48 hours of treatment in the 3D culture system (including those only detected in 3 replicates of one condition) were compared with the proteins found to have significantly different expression after 48 hours of treatment in the monolayer culture (see section **4.3.5.3**).

Out of 654 proteins that were analysed by T-Test or only identified in one condition in both datasets, only four were identified in both sets of significant proteins: 60S ribosomal proteins L10, heterogeneous nuclear ribonucleoprotein A0, asparagine-tRNA ligase, cytoplasmic and NQO1. The former two proteins were significantly downregulated in both the monolayer cultures (including IPF patient 1) and the 3D model cultures after FK228 treatment. As previously discussed, NQO1 was significantly upregulated after treatment in both culture systems. Asparagine-tRNA ligase, cytoplasmic significantly increased after treatment in the monolayer cultures, including IPF patient 1, but significantly decreased in the 3D model cultures. The proteins other than NQO1 have roles in post-transcriptional

regulation and translation, and they may be altered in expression due to changes in transcription caused by the histone deacetylation-inhibition activity of FK228.

These results indicate that the type of culture system used to culture cells can impact protein expression, even changing the outcomes of cell treatment as a consequence of differences in cellular properties, in this case monolayer formation and tissue formation.

5.4. Discussion

There is an unmet need for a treatment for idiopathic pulmonary fibrosis that could slow or inhibit the progression of disease and improve patient survival. FK228 is a candidate therapy due to its anti-proliferative effects on matrix-secreting fibroblasts. However, in order for IPF pathogenesis and drug treatment to be studied effectively, an appropriate model must be chosen to mimic the disease as closely as possible, in order to understand the biology and predict the effects of treatment in patients. Standard practices of culturing primary cells on tissue culture plastic in monolayers and using animal models for studying IPF are lacking in their ability to reproduce certain features of the disease and to predict clinical responses to therapies.

The bleomycin-induced fibrosis *in vivo* model is the most frequently used animal model of fibrosis, and has been for over four decades (Moore et al., 2013). It has been applied to various animals, with mice being the most commonly used, and can be delivered in multiple ways. Lung-specific delivery causes alveolar epithelial cell injury by DNA stand breaks, resulting in induction of oxidative stress, inflammation, and then fibrosis (Moore and Hogaboam, 2008). However, the histopathology of bleomycin-induced fibrosis is not consistent with the hallmark usual interstitial pneumonia (UIP) pattern of IPF, and the fibrosis can resolve in some cases (Degryse and Lawson, 2011). Other animal models can involve the overexpression of cytokines such as TGF-β_1, TNF-α and IL-1β or mutation/deficiency of genes associated with familial interstitial pneumonia, such as *SFTPC*, *SFTPA2*, *TERT* and *TERC* (Moore *et al.*, 2013). These kinds of models are useful for studying particular signalling pathways and cellular interactions associated with IPF, however unfortunately none of them fully mimic its progressive, irreversible, fibrotic nature. Animal models of fibrosis are also inaccurate in reflecting human responses to treatment, with many drugs inhibiting fibrosis in rodents but failing to do so in clinical trials (Moeller et al., 2008).

Monolayer culture of cells taken from 3D tissues can lead to unnatural cell behaviour, which stems from them experiencing a different microenvironment. In monolayer culture, only a part of a cell's membrane interacts with either neighbouring cells or the ECM, with the remainder exposed to the culture medium; these conditions are not representative of their native tissue (Tibbitt and Anseth, 2009). Atypical cell behaviour presents an additional challenge for modelling a complex, diseased tissue such as the IPF lung. 3D scaffolds are therefore being investigated in order to better model the cellular microenvironment.

3D *in vitro* models such as hydrogel scaffolds have been developed to study fibroblast responses to extracellular matrix during wound healing. Hydrogels are cross-linked polymer networks with a high water content, to which cells are added and cultured. They can be composed of naturally derived matrix components such as collagen, or synthetic materials such as polyethylene glycol (PEG) (Smithmyer et al., 2014). Hydrogels containing native ECM components are better mimics for the biophysical and biochemical properties of a tissue, however the resulting complexity can make it challenging to define which cellular functions are being promoted by which cues (Cushing and Anseth, 2007). Inherent lack of reproducibility between batches of this kind of hydrogel can mean that cells behave differently between cultures due to differing signals to and from their environment (Tibbitt and Anseth, 2009). Hydrogels made from synthetic materials are highly reproducible and mechanical factors such as stiffness can be easily monitored, however these lack endogenous factors that play a large role in cellular function in tissue. These kinds of models allow the study of the fibroblast response to their microenvironment, but analysing the matrix produced endogenously by the fibroblasts in culture would be incredibly difficult.

Due to the unmet need to be able to study both the extracellular matrix and its effect on fibroblasts in a physiologic and disease-specific way, Booth *et al.* (2012) reported the development of acellular lung matrices for use as a model culture system. Decellularisation of normal and IPF lungs using detergents, salts and DNase resulted in a substrate that retains the architecture of the lung and the stiffness and composition of the matrix, and can be analysed by proteomics. The matrices can also be reseeded with fibroblasts to analyse their response to the ECM. They also showed that fibroblasts reseeded onto IPF-derived matrices behave differently to fibroblasts reseeded onto normal lung-derived matrices, including having increased α-SMA expression. An advantage of this culture system is that it is fully humanized, with the matrix being completely human lung-derived.

The 3D *in vitro* model devised by Dr Mark Jones in the Brooke laboratory, University of Southampton NHS Foundation Trust, and used in this proteomic study has the advantage of the ECM being fully humanized. Most importantly, the ECM is produced by the fibroblasts

cultured in the 3D system; therefore, this model allows the study of a) the matrix itself b) fibroblasts cultured in a more representative microenvironment than found in 2D culture c) the fibroblasts' response to their own synthesized matrix, and d) the production and secretion of matrix components by the fibroblasts over time.

This 3D *in vitro* culture system was originally adapted from a culture method previously used in a study by Shamis *et al.* (2012), in which the culture system was used to compare the characteristics of matrix produced by induced pluripotent stem cell-derived fibroblasts to that of the matrix produced by parental dermal fibroblasts, due to their potential for use in regenerative medicine. Culturing the fibroblasts long-term (5 weeks) on a porous polycarbonate membrane with L-ascorbic acid-2-phosphate added to the culture medium enabled the organisation of a 3D matrix and formation of tissue, whilst maintaining a stable fibroblast morphology. L-ascorbic acid-2-phophate has previously been shown to cause a significant increase in collagen synthesis without causing any change in abundance of non-collagen proteins (Murad et al., 1981). The accumulation of collagen promotes reorganisation of fibroblasts into a multilayer of cells, forming a tissue-like substance (Hata and Senoo, 1989) and the synthesis of a thick, organised 3D matrix; these structures are described as "self-assembled" (Pouyani et al., 2009). In the study by Shamis *et al.,* L-ascorbic acid-2-phophate elevated the expression of collagen in monolayer culture, and increased the amount of matrix produced by the fibroblasts in the 3D model.

This model has the potential for contributing to mechanistic studies of disease pathogenesis, such as IPF. Due to an interest in the role played by the ECM in IPF pathogenesis, this culture method was adapted by Dr Jones for studying extracellular matrix produced by IPF fibroblasts. Work performed by Dr Jones *et al.* has shown that there are differences in cell proliferation, fibroblastic foci formation, gene expression, MMP production and collagen deposition between IPF and non-IPF fibroblasts cultured in this 3D system. It has also been shown that this model can also be decellularised, and non-IPF fibroblasts behave differently when cultured on this IPF fibroblast-derived matrix in comparison with their own synthesized matrix (Jones et al., 2014).

The work presented in this chapter has expanded on the concept of using this 3D *in vitro* culture system as a model of matrix deposition by fibroblasts in IPF pathogenesis. This investigation aimed primarily to find out whether this novel type of culture system for studying IPF fibroblasts could be analysed effectively by proteomics, for characterisation of the IPF fibroblast proteome after culture in this model, and potentially the matrix also, due to the manner in which the sample was harvested (see section **5.3.1**). It also aimed to find out how the fibroblast proteome changes from the beginning of matrix deposition to the

formation of a multi-layered structure. Having already studied the proteome of IPF fibroblasts grown as monolayers (see **Chapter 4**), an additional aim was to elucidate the main differences in proteome between the two culture systems, in order to help evaluate the 3D system as a worthwhile model. In the 3D culture system, the fibroblasts are incubated in more complex media than standard 2D culture media, due to the addition of several supplements, such as hydrocortisone, EGF and insulin. This is to ensure that the cells continue to grow and proliferate in long-term culture. As previously discussed, the optimum culture period is 6 weeks, which is more time-consuming and costly in terms of consumables than 2D culture. Whilst fibroblasts in the monolayer culture may exhibit certain behaviour different from their native tissue due to their 2D microenvironment, potentially rendering results from these studies less representative, this type of culture system is relatively inexpensive, and the cells can be cultured for just 5 days post-seeding into culture plates, making experiments more easily reproducible.

The number of proteins quantified from fibroblasts cultured in the 3D system for 3 weeks was comparable to that from fibroblasts grown as monolayers. In addition, there was no significant difference in the number of proteins quantified from fibroblasts grown for 6 weeks in the 3D system and from fibroblasts grown in monolayer culture. This suggests that the proteome of fibroblasts cultured in the 3D system for the optimum time-period can be analysed almost equally as effectively by mass spectrometry as the proteome from fibroblasts cultured using standard monolayer culture. During sample preparation there was some difficulty in breaking apart the multi-layered structure in order to extract and digest the proteins, but the increased difficulty in lysis did not appear to hinder the final results. This indicates that the increased complexity of the 3D model does not cause a disadvantage in terms of proteomic study, which has not been performed before on this type of sample.

The 6-week culture period, despite being used in previous studies, may have not been ideal for this study as the cells within the layers may have been dying, or hypoxic by this time – the images in **Figure 5.2** show cells that appear perturbed by week 6, even in the control. As previously stated, viability assays were not performed on these cultures; these should be carried out to check whether cell death is occurring. Perhaps a 5-week, or even a 4-week culture period would have been more suitable, to avoid toxicity induced solely by culture conditions.

Almost half of the proteins that could be compared from IPF fibroblasts grown in the 3D system and the monolayer culture had at least a 2-fold difference in expression. The fibroblasts cultured in each system were from the same donor biopsy; taken together this suggests that the microenvironment and extracellular matrix has a big impact on protein

expression and thus cellular function, as observed through differences in phenotype. The most enriched gene ontology terms associated with these proteins were related to metabolism, likely due to the different energy needs between the two types of cultures, as the fibroblasts respond to different environments. For example, cells cultured on a stiff culture substrate would have different rates of proliferation to cells that are proliferating but also organising themselves into a layered tissue structure, as well as synthesizing and remodelling extracellular matrix. The energy needed for these functions are provided by various levels of molecular degradation and protein turnover, processes in which many of the proteins found to have a difference in expression between the two systems are involved in. Differences in protein turnover could be due to differences in cross-linking of the matrix/matrix remodelling. It is worth considering that although this proteomic analysis provides information on the differences in protein abundance between the two culture methods, it does not give an indication of potential differences in relation to protein cross-linking, which could be as important as differences in the expression of the proteins themselves.

It is difficult to draw conclusions from this comparison at present, as only one replicate of the monolayer fibroblast culture was analysed and included. Further work would be to generate at least two more samples of this primary donor culture in monolayers and analyse them by mass spectrometry, to achieve a more reliable comparison between the two culture systems.

A higher number of extracellular matrix proteins were identified in the 3D model data than in the monolayer culture data, indicating that this is likely to be a more physiologically relevant model for ECM analysis. This is most likely due to more matrix being synthesized in response to the microenvironment and to the presence of L-ascorbic acid 2-phosphate in the culture medium. As previously mentioned, these proteins are likely to be intracellular; therefore, this information provides a snapshot of what proteins are being produced before they are secreted. Due to the tissue-like nature of the 3D model, proteins deposited in the extracellular spaces between layers are also likely to have been harvested and analysed, increasing the amount identified by mass spectrometry in this system. To fully characterise the matrix, the secretome of both culture systems could also be profiled by analysing the culture medium by mass spectrometry. However, the 3D culture model would require optimisation for this as currently FBS is added to the culture medium, and previous secretomic experiments with FBS present in culture media have resulted in a very low number of secreted protein identifications by mass spectrometry. Proteomic analysis of the decellularised 3D model would provide additional ECM information to the data presented in this chapter in terms of exclusively identifying the different proteins produced by IPF

fibroblasts and their levels of expression. In addition, it would also be fairer to compare protein expression from fibroblasts in this 3D culture system to that of fibroblasts grown in monolayers with the addition of L-ascorbic acid 2-phosphate and the other 3D culture medium supplements to the culture medium. This could increase the amount of extracellular matrix proteins produced by these fibroblasts.

The proteomic analysis carried out so far suggests that the 3D model is more informative of the ECM produced by IPF fibroblasts, and the monolayer culture is more informative of the fibroblast proteome due to slightly more proteins being identified overall (although this information could be less representative of IPF tissue). Therefore, this indicates that the type of culture method used for an experiment depends on the experiment's main goal.

The same principal also applies to which 3D culture system should be used for an experiment, if 3D culture is the method of choice. As described above, the 3D system used in this study is a, productive method to analyse the extracellular matrix produced by resident fibroblasts, and these fibroblasts' behavioural response to their own deposited matrix. However, the acellular lung matrix culture system developed by Booth *et al.* previously described is an effective way of analysing the fibroblast response to a substrate which retains the properties of matrices that have been synthesised and deposited in a fibrotic lung. These matrices were produced in response to *in vivo* environmental cues and thus are more physiologically similar to those found in an IPF lung than the matrices produced in culture in the model used for this study. The latter model contains fibrotic fibroblasts responding to a very different environment – cues designated by the culture medium, incubation and the plastic ware.

A limitation of the Booth model however, is that fibroblasts seeded onto the decellularised lung matrices are very unlikely to be from the same patient that provided the matrices, and thus may respond differently to how the patient's fibroblasts would have *in vivo* or *in vitro*. Therefore, whilst it is ideal for studying responses to lung matrices synthesized in vivo, it cannot be used to study the fibroblast response to their own synthesized matrix, which is a benefit of the model used in this study.

This 3D *in vitro* model can also aid in studies of potential therapies, as the work presented here has demonstrated. An additional aim of this investigation was to examine the proteomic effects of the histone deacetylase inhibitor FK228 on IPF fibroblasts cultured in this system, to examine whether the effects are similar to what has previously been observed in monolayer culture.

FK228 affected the expression of different proteins after acute and chronic treatment. Less than 10% of those greatly affected had significantly different expression in untreated fibroblasts at both time-points, indicating that these differences between acute and chronic treatment were largely due to the drug and were not impacted by the different culture periods. Similar numbers of proteins were significantly affected at the two time-points; 62 at 6 weeks (37 $p<0.05$, 25 only detected in one condition) and 65 at 3 weeks (18 $p<0.05$, 47 only detected in one condition), but only two proteins had significantly different expression in both datasets.

However, in both datasets, many of these significantly differentially regulated proteins were involved in metabolism (see **Appendix D.10**). This is similar to that observed in monolayer cultures after acute treatment, although this analysis included donor cultures additional to IPF patient 1 used in the 3D model. The range of biological processes that were affected by FK228 treatment was much greater in monolayer cultured-IPF fibroblasts (see **Chapter 4**). In addition, different proteins were affected by FK228 in monolayer culture and in 3D culture. FK228 appeared to have a much larger effect on protein expression in the monolayer culture, with over 100 proteins having significantly different expression with treatment ($p<0.05$) and an additional 104 proteins only being detected in one condition. This may be due to more proteins being identified in at least three samples of the monolayers and therefore being eligible for T-Test. However, it does raise questions about the extent of the molecular effect of FK228 in different models of IPF. If the 3D culture model is more physiologically relevant than monolayer culture in terms of modelling IPF tissue and fewer proteins were significantly affected by treatment, FK228 may have less of an effect than previously thought when applied to tissue. The multi-layered structure formed in the 3D culture may have created difficulties in drug penetration into the cells. The different microenvironments created by the two culture systems may alter the susceptibility to FK228. This information may help in deciding a dose for clinical trials, and narrowing down the potential candidate biomarkers that would be useful in monitoring treatment efficacy. For example, of over 650 proteins that were analysed by T-Test or only identified in one condition in both the 3D model and monolayer datasets, only four changed significantly in abundance in both, demonstrating that not only do cells behave differently in 2D and 3D culture, but they also respond differently to external stimuli such as drug treatment. This highlights the importance of developing a culture system or model for use in experiments that is representative of the disease and can efficiently predict responses to drug treatment prior to clinical trial, in order to find a treatment that can be successfully taken through to approval.

Both culture systems highlighted NQO1 as a protein whose expression is significantly increased following FK228 treatment. In the 3D culture system, the expression of NQO1 decreased between three and six weeks as the 3D tissue formed, but increased with FK228 after both 48 hours and 3 weeks of treatment. NQO1 is an enzyme responsible for scavenging free radicals to protect cells against oxidative stress, which can occur due to ECM accumulation, cell death and increased production of ROS during fibrosis (Usunier et al., 2014), and is thought to be a risk factor for cellular injury in IPF. A study by Artaud-Macari *et al*. in 2013 demonstrated that NQO1 mRNA expression is significantly downregulated in IPF fibroblasts compared to healthy controls (Artaud-Macari et al., 2013). In addition, an activator of NQO1 has been shown to attenuate bleomycin-induced fibrosis in mice (Oh et al., 2013). Similarly, the study in **Chapter 4** found than NQO1 protein expression was downregulated 1.6-fold in IPF fibroblasts compared to non-IPF fibroblasts. After FK228 treatment, expression increased 3-fold in the IPF patient 1 monolayer, and 1.5-fold across all monolayer cultures analysed, reaching significance in T-Test analysis (p<0.038). Increasing the expression of NQO1 in IPF fibroblasts using FK228 could reduce fibrosis and increase protection against ROS already produced during the established fibrosis, relieving oxidative stress. NQO1 is therefore an ideal potential candidate biomarker for response to FK228 in clinical trials.

Differences in proteomic results concerning FK228 in the two culture systems suggests that one culture system alone may not be sufficient in predicting biomarkers of IPF and of the fibroblastic response to FK228. Comparing IPF and non-IPF lung fibroblasts cultured in this 3D *in vitro* system using a proteomic approach would be an interesting avenue of investigation for IPF biomarker discovery, together with previous work presented in this thesis using differences in protein expression between IPF and non-IPF fibroblast monolayer cultures to identify potential biomarkers.

6. Results - Characterisation of the fibroblast acetylome – method development

6.1. Abstract

Acetylation, the process by which an acetyl group is added to a residue of a protein, is one of the most common protein post-translational modifications. The "acetylome" encompasses all the acetylated proteins in an organism, at a specific point in time, under certain conditions. Enzymatic regulation of acetylation occurs through the antagonistic actions of histone acetyltransferases and histone deacetylases; by regulating acetylation, these enzymes regulate the expression of particular genes. This occurs through changing the interactions between (non)acetylated histones and local DNA within chromatin to alter the accessibility of the DNA to transcriptional machinery. Acetylation also occurs to non-histone proteins involved in important biological processes.

FK228 is a potent inhibitor of histone deacetylases. By modification of gene expression, it can cause inhibition of cell growth and proliferation. FK288 is a potential candidate therapy for idiopathic pulmonary fibrosis (IPF), a devastating progressive respiratory disease. In IPF, an unknown stimulus causes repetitive injury to lung epithelial cells and differentiated fibroblasts aberrantly accumulate in response. Their secretion of extracellular matrix components into the interstitium leads to impaired gas exchange. FK228 has been shown to inhibit the aberrant proliferation and differentiation of IPF patient lung fibroblasts *in vitro*.

The aim of this study is to investigate changes in acetylation caused by FK228, including off-target sites. The first step to achieving this was the development and optimisation of a method that allows acetylated proteins or peptides to be identified from fibroblast lysates. Fibroblasts were chosen for investigation due to their major role in IPF pathogenesis. In preliminary experiments, just six acetylated peptides were identified from MRC-5 fibroblast cell lysates by immunoprecipitation of acetylated peptides and mass spectrometry analysis; following method development, 48 acetylated peptides have been identified to date.

The end goal is to be able to characterise the primary IPF fibroblast acetylome, its similarities and differences to that of non-IPF fibroblasts and its modulation after treatment with FK228. This may improve our mechanistic understanding of IPF and/or lead to the discovery of new drug targets. The approach may also be extended to studying other forms of post-translational modification, such as carbonylation or ubiquitination.

6.2. Introduction

Acetylation is the reversible addition of an acetyl group to a compound. It is a highly common type of protein post-translational modification (PTM). The addition of acetyl groups is performed by histone acetyltransferases (HATs), enzymes that function antagonistically to histone deacetylases (HDACs), which remove acetyl groups. Histone acetylation is the best characterised type of acetylation. The acetylation of lysine residues of core histones within nucleosomes has been associated with gene regulation through transcriptional activation, as acetylation causes a decrease in affinity of histone N-terminal domains for DNA and a subsequent change in chromatin structure, making local chromatin more accessible to transcriptional machinery (Kuo and Allis, 1998).

Protein acetylation was first noted in the 1960's (Allfrey et al., 1964), but it was only in recent years that the analysis of global acetylation of proteins within a cell has been achieved using various approaches (Smith and Workman, 2009). Much work has been carried out to identify substrates of Sirtuin3, a mitochondrial deacetylase, by analysing acetylation changes in its absence (Sol et al., 2012; Rardin et al., 2013). In a study by Zhang *et al.* the acetylome of *Escherichia coli* was profiled, identifying 8 times the number of acetylation sites than previously reported (1070 sites) (Zhang et al., 2013a). In 2009, Choudhary *et al.* successfully profiled the acetylome of three human cell lines and identified over 3600 acetylation sites, including sites that changed in acetylation state following treatment with the histone deacetylase inhibitor suberoylanilide hydroxamic acid (SAHA) (Choudhary et al., 2009). This allowed monitoring of the acetylation-modulating effects of a drug in a global manner. In 2014 Choudhary *et al.* published an article that included a list of large-scale proteomic studies of acetylation that had been carried out up to that point in various cell types and species. They listed 22 studies, with the number of acetylation sites identified in each, which reached into the thousands in three quarters of the studies (Choudhary et al., 2014). They deemed recent breakthroughs as the "age of acetylomics".

FK228, or Romidepsin, is a potent inhibitor of histone deacetylases, which along with SAHA has been approved as an anti-cancer treatment. FK228 is currently under evaluation as part of this thesis as a candidate for treatment of the progressive interstitial lung disease idiopathic pulmonary fibrosis, whose pathogenesis shares characteristics with cancer (Vancheri et al., 2010) (see **Chapter 4**). As shown in **Chapters 3 and 4,** studies are underway to globally profile fibroblast differentiation at the protein level, which occurs aberrantly in IPF, in order to investigate changes in expression and signalling pathways upon myofibroblast formation. This has been extended to characterisation of the IPF fibroblast

itself and the identification of potential candidate biomarkers of IPF through comparison with fibroblasts from non-IPF donor biopsies. Analysis of protein expression from cells treated with FK228 has allowed for the identification of changes in response to the drug, the monitoring of potential IPF biomarkers, and the identification of biological processes and signalling pathways that alter after FK228 treatment, which are all important observations for characterising the cellular response.

The aim of this study is to focus specifically on the acetylome of the IPF fibroblast, and observe how it alters in response to treatment with the histone deacetylase inhibitor FK228. This work would contribute to the knowledge of the effects of FK228 on general protein abundance. It has previously been shown that FK228 causes a dose-dependent increase in the acetylation of histone H3 in primary IPF fibroblasts (Davies, 2011) and this investigation aims to identify other off-target effects by generating a global profile of IPF fibroblast acetylation sites and pinpointing those that are modulated by FK228. In order to study this, the first step was to devise and develop a method to detect acetylation within fibroblast lysates. Acetylated proteins or peptides were enriched from cell lysates by immunoprecipitation (IP) before analysis by mass spectrometry. Enrichment of acetylated peptides from cell lysates is necessary to reduce sample complexity, to prevent non-detection due to a background of non-acetylated peptides. Studying the acetylome may lead to the identification of novel FK228 drug targets and may form a foundation for methods to analyse other forms of PTM, for example carbonylation or ubiquitination, for an even further in depth analysis of FK228 effects in IPF.

6.3. Results

6.3.1. Western blotting of cell lysates for acetylated lysine residues before and after immunoprecipitation showed an enrichment of acetylated proteins after IP

Acetylated proteins from untreated MRC-5 fibroblasts were principally subjected to immunoprecipitation using anti-acetyl-lysine agarose followed by western blotting for acetylation, to investigate the efficacy of acetylated protein-binding to the beads. Anti-acetyl-lysine agarose was washed 3x with PBS-T, 2x with 0.1 M NaH_2PO_4 + 1M NaCl and 1x with PBS-T, with centrifugation at 1036 xg for 5 minutes between washes. Cell lysate was added to the beads and was incubated overnight at 4ºC on a rotator. Proteins were eluted from the beads in Laemmli buffer. 15 µg protein aliquots taken before protein

immunoprecipitation, after incubation and after elution were subjected to SDS-PAGE, followed by western blotting for acetyl lysine, with β-actin as a loading control.

Figure 6.1 shows that the basal level of lysine acetylation in the cell lysate (**lane A**) is relatively low. This basal level decreases in the supernatant during IP (**lane B**), indicating depletion of acetylated proteins from the supernatant and instead binding to the anti-acetyl-lysine agarose. The eluted sample in **lane C** showed increased signal intensity for histone lysine acetylation over the total cell lysate in **A**, and several other acetylated proteins were present with molecular weights higher than histone molecular weight after IP. The β-actin loading control indicates that the majority of the sample contained non-acetylated proteins that did not bind and thus were left in the supernatant during IP (lane **B**, in which the β-actin band is of a similar intensity to that in the total cell lysate in **A** but is reduced in **C**). The results indicate that the eluate was enriched for acetylation, due to less total protein being loaded but more acetylation detected.

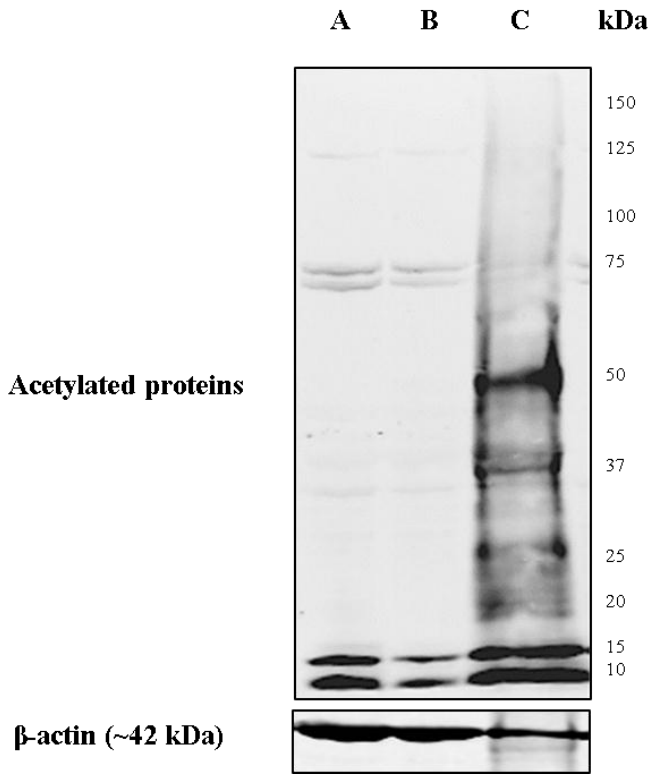


Figure 6.1 Western blot of acetylated lysine residues of proteins from MRC-5 fibroblast lysates before, during and after IP. A = cell lysate, B = supernatant during IP, C = eluted acetylated proteins from anti-acetyl lysine agarose. β-actin was used as a loading control.

6.3.2. Data from MRC-5 fibroblast acetylated peptide enrichment experiments showed that there was a positive trend in the number of unique acetylation identifications throughout method development

Two different methods of sample preparation involving immunoprecipitation were developed in order to study the acetylome of MRC-5 fibroblasts. The first, which is from here on referred to as the Choudhary method, was based on the protocol from the study by Choudhary et al. (2009) and underwent three rounds of development, with the help of Dr Caterina Folisi. The second, which is from here on referred to as the Li method, was based on the protocol published by Li et al. (2013) and was tested once. The differences in the two methods and the modifications to the Choudhary method are as follows:

*Choudhary method (see section **2.9.3** for more information)*
MRC-5 fibroblasts were cultured to produce approximately 2 mg protein. Cells were harvested using trypsin-EDTA, counted using trypan blue stain, washed 4 times in phosphate buffered saline and lysed in 0.1% sodium dodecyl sulfate using a sonication probe. Following cell lysis, proteins were reduced with DTT (1/25 w/w) for 1 h at 60°C and alkylated with iodoacetamide (1/5 w/w) for 45 minutes in the dark at RT. Proteins were digested overnight at 37°C with sequencing grade modified trypsin (1/50 w/w) at 37°C.

Anti-acetyl-lysine agarose was used for IP at a ratio of 2 mg protein: 50 µl agarose. Agarose was washed 3x with PBS-T, 2x with 0.1 M NaH_2PO_4 + 1M NaCl and 1x with PBS-T, and centrifuged at 1036 xg for 5 minutes between washes. Cell lysate was added to the beads and the mixture was incubated overnight at 4ºC on a rotator. After incubation beads were washed 4x in PBS-T and eluted in 1% TFA. Purification of peptides was performed using C18 SpinTips. Eluates were concentrated to completion and peptides were resuspended in buffer A (98% dH_2O/acetonitrile + 0.1% formic acid) for mass spectrometry analysis.

Peptides were separated by liquid chromatography using a NanoACQUITY UPLC system with a C18 reverse-phase column over a gradient of 90 minutes at 3-50% buffer B (80% acetonitrile/dH_2O + 0.1% formic acid), and sprayed directly into a Waters Synapt G2-S system, with ion mobility enabled. Data was processed and searched against the human UniProt database using Waters PLGS 3.0/3.0.2 software before analyses. This method workflow is illustrated in **Figure 6.2**.

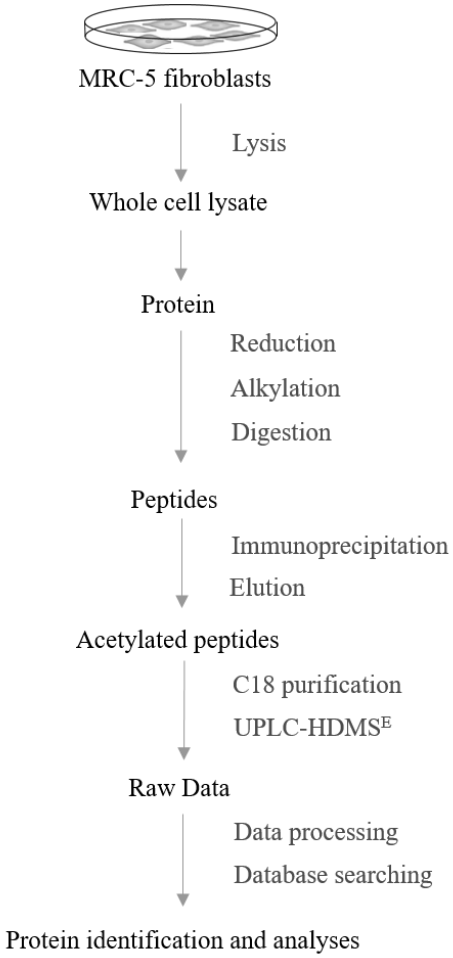


Figure 6.2 A representation of the original Choudhary method workflow used for the identification of acetylated peptides from untreated MRC-5 fibroblast cell lysates. Protein lysates from MRC-5 fibroblasts were reduced, alkylated and trypsin-digested into peptides. Peptides were enriched for acetylation by immunoprecipitation using anti-acetyl-lysine agarose beads (Immunechem) and eluted in 1% TFA. Peptides were separated by UPLC and sprayed directly into a Synapt G2-S mass spectrometer (Waters) operating in MS^E mode with ion mobility. Data were processed and searched using PLGS 3.0/3.0.2 (Waters) before analysis.

Choudhary method modification 1

The original Choudhary method was adapted in the IP stage, to use a Protein A immunoprecipitation kit (Roche) to enrich for acetylated proteins. Digested proteins were pre-cleared with Protein A-agarose for 3 h, 4°C with mixing, to remove non-specific proteins. Beads were collected and anti-acetyl-lysine antibody was added to the supernatant and incubated overnight at 4°C with mixing. Protein A-agarose was then added to each sample and incubated overnight at 4°C with mixing. Agarose beads were collected and

washed twice with wash buffer 1, twice with wash buffer 2 and once with 1 ml wash buffer 3, according to manufacturer's instructions (see section **2.1.2** for components of wash buffers 1, 2 and 3). Peptides were eluted from the beads with 1% TFA.

Additional method modifications were made for the removal of the contaminate polyethylene glycol (PEG), which was identified in the mass spectra of the sample analysed using the original method. Peptides were separated into 12 fractions according to their isoelectric point using an Agilent 3100 OFFGEL fractionator and IPG Strips, 13 cm, pH 3–10 according to manufacturer's instructions. Peptides were focused for 20 kVh. Peptides were purified using an Empore C18 96-well solid phase extraction plate. Fractions were resuspended in 98% dH2O/acetonitrile + 0.1% formic acid (buffer A) for mass spectrometry analysis. Due to the increased number of injections necessary following fractionation, peptides were separated by liquid chromatography over a shorter, 60-minute gradient of 3-50% buffer B (80% acetonitrile/dH_2O + 0.1% formic acid).

Choudhary method modification 2

Following a larger number of acetylated peptides having been identified using the original method of enrichment, anti-acetyl-lysine agarose beads were used for the remainder of method development. Method modification 2 included two additional wash steps of the agarose with PBS-T and an extra elution with 1% TFA. This method also included separating the peptides according to isoelectric point by OFFGEL fractionation and purifying using an Empore C18 96-well solid phase extraction plate. Peptides were also separated by liquid chromatography over the shorter 60-minute gradient of 3-50% buffer B (80% acetonitrile/dH_2O + 0.1% formic acid).

Choudhary method modification 3

After also identifying PEG in the mass spectra following these modifications, the Choudhary method plus modification 2 was further adapted to include methanol/chloroform precipitation of proteins to concentrate and purify them from the lysate, in order to remove more contaminants than previously, and resuspension in 6M urea to re-solubilise the proteins before digestion. Peptides were resuspended in "immune-affinity precipitation buffer" (IAP, see section **2.1.2**) before IP and eluted from the beads in 0.1% TFA rather than 1% TFA.

Papers published by Sol *et al.* (2012) and Choudhary *et al.* (2009) demonstrated the use of an IAP buffer for immunoprecipitation using the same anti-acetyl lysine agarose beads, and elution of peptides at a lower acidic concentration. The IAP buffer may be more compatible with the beads and a less acidic elution buffer could lower the possibility of PEG shedding from the beads, if this was the cause.

Treatment of fibroblasts with 2 nM FK228 was also carried out to be able to compare the number of acetylated peptides following treatment. At 60% confluence during culture, cells were treated in the absence or presence of 2 nM FK228 for 48 hours before harvesting using trypsin-EDTA. The final Choudhary method is illustrated in **Figure 6.3**.

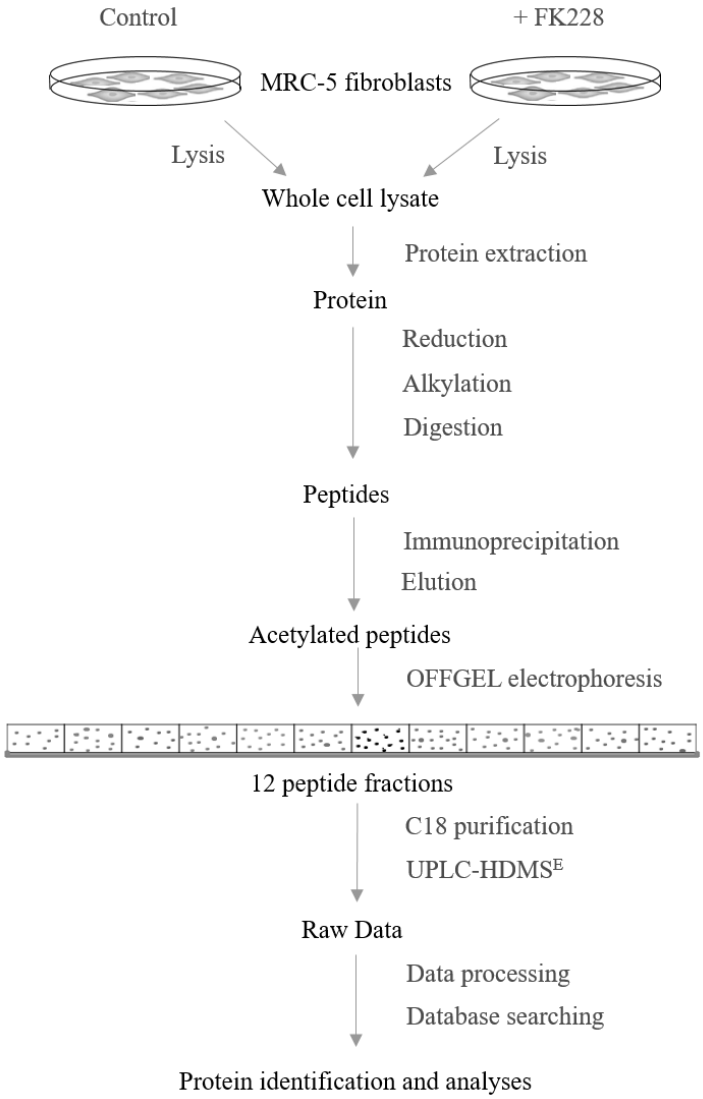


Figure 6.3 A representation of the final Choudhary method workflow used for the identification of acetylated peptides from FK228-treated and untreated MRC-5 fibroblast cell lysates after three rounds of optimisation. MRC-5 fibroblasts were treated with 2 nM FK228 for 48 h and compared against untreated fibroblasts as controls. Protein lysates were reduced, alkylated and trypsin-digested into peptides. Peptides were enriched for acetylation by immunoprecipitation using anti-acetyl lysine agarose beads, washed with IAP buffer and eluted from the beads with 0.1% TFA, before being subjected to isoelectric focusing (IEF) by OFFGEL electrophoresis; each fraction was separated further by UPLC before HDMS^E analysis.

*Li method (see section **2.9.4** for more information)*

Cell culture, harvesting and sample preparation was performed according to Li *et al.* (2013). MRC-5 fibroblasts were cultured to produce approximately 10-20 mg protein. Cells were lysed by scraping into urea lysis buffer (see section **2.1.2/2.7**). Cell lysate was sonicated and cleared by centrifugation. Proteins were reduced with 1/278 volume of 1.25 M DTT for 30 minutes, RT, followed by alkylation with 1/10 volume of iodoacetamide for 15 minutes, RT in the dark. The lysate was diluted 3-fold with 20 mM HEPES, pH 8.0 and digested with 1/100 volume of 1 mg/ml trypsin-TPCK in 1 mM HCl overnight at RT. Proteins were purified using a Sep-Pak® C18 Plus Long Cartridge and eluted in 40% acetonitrile + 0.1% TFA. The eluate was stored at -80°C overnight then lyophilised for 48 hours.

Peptides were resuspended in IAP buffer and IP was carried out as described above. Peptides were eluted in 0.15% TFA, and the eluate was purified using C18 SpinTips as described in the original method. Peptides were resuspended in 98% dH2O/acetonitrile + 0.1% formic acid (buffer A) for mass spectrometry analysis. Peptides were separated by liquid chromatography using a NanoACQUITY UPLC system (Waters) with a C18 reverse-phase column over a 90-minute gradient of 3-50% buffer B (80% acetonitrile/dH_2O + 0.1% formic acid), and sprayed directly into a Waters Synapt G2-S system with ion mobility enabled. Data was processed and searched against the human UniProt database using Waters PLGS 3.0/3.0.2 software before analyses. This method is illustrated in **Figure 6.4**.

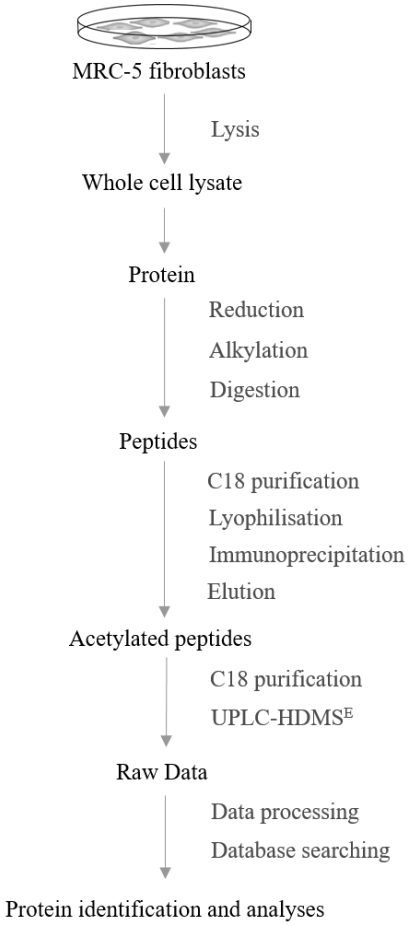


Figure 6.4 A representation of the Li method workflow used for enrichment of acetylated peptides from untreated MRC-5 fibroblast cell lysates. MRC-5 fibroblasts were lysed in urea lysis buffer, and proteins were reduced, alkylated and trypsin-digested into peptides. Peptides were purified then enriched for acetylation using anti-acetyl lysine agarose beads (Immunechem), washed with IAP buffer and eluted with 0.15% TFA; peptides were analysed by UPLC- HDMS^E and data was processed and searched in PLGS 3.0/3.0.2.

The data from each round of method development was analysed for a.) the number of proteins identified b.) the percentage of proteins that contained at least one identified acetylated lysine residue (i.e. how enriched the sample was for acetylated peptides) and c.) the total number of unique acetylated peptides for each experiment. This information was used to track the progress towards increased coverage of acetylated peptides throughout the Choudhary method optimisation.

In some cases, acetylated peptides were assigned to more than one protein by the processing software due to sequence homology of that peptide, with no additional peptides of that

protein identified to distinguish between the proteins. Due to the risk of false positives and inaccurate conclusions when studying post-translational modifications (Kim et al., 2016), it is important to be aware that not all identifications of acetylation may be true, on top of the FDR percentage set for database searching.

The assigned b and y peptide fragment ions of each peptide can be used to manually assess whether or not each acetylated peptide is a confident assignment. Fragment ions with a charge on the N-terminus or C-terminus are indicated with a "b" and "y" respectively, and the number associated with each ion represents the number of amino acids in the fragment. For example, a peptide with fragment ions reported as “b8b9” indicates that two fragment ions were identified, one eight amino acids in length and one nine amino acids in length, and both are fragments ions at the N-terminal end of the peptide. The difference in mass between consecutive fragment ions indicates an amino acid.

The more b and y ions identified, the more likely that the identification is true as more amino acids of a peptide sequence can be accurately determined, although b and y ions with potential modifications (indicated by ° and *) would lower confidence due to in most cases not having also identified the unmodified ion, which could indicate a false positive result.

The position of the identified b and y ions along the peptide sequence also can indicate whether or not the identification is likely to be true; many consecutive fragment ions determining one part of the amino acid sequence of a peptide and no other parts would be less likely to be a true identification as many amino acids in the peptide sequence have not been accurately determined. For example, if an identified acetylated peptide is 28 amino acids in length and the associated fragment ions are “b2b3b4b7”, individual amino acids have been not accurately determined for three-quarters of the peptide sequence, which could include the acetylated lysine residue, and therefore this would not be a confident identification.

The lists of proteins and acetylated peptides identified in each experiment can be found in **Appendix E**, along with the amino acid sequence of the peptide (the acetylated lysine residue identified is highlighted in red) and the b/y fragment ions for that sequence.

The results for all proteomic experiments for this study are summarised in **Table 6.1**. The degree of enrichment and number of acetylated peptides for each experiment are compared in **Figure 6.5**.

Table 6.1 Results of the analysis of data from each experiment during method development.

Experiment	No. proteins identified	No. proteins with acetylated peptides	% proteins with acetylated peptides	No. unique acetylated peptides	No. unique acetylated sites
Choudhary 1.1	39	11	28.21	6	6
Choudhary 1.2	4	3	75.00	3	4
Choudhary 1.3	10	4	40.00	15	15
Choudhary 1.4 Control	16	8	50.00	13	13
Choudhary 1.4 + FK228	16	11	68.75	32	31
Li 2.1	151	39	25.83	48	54

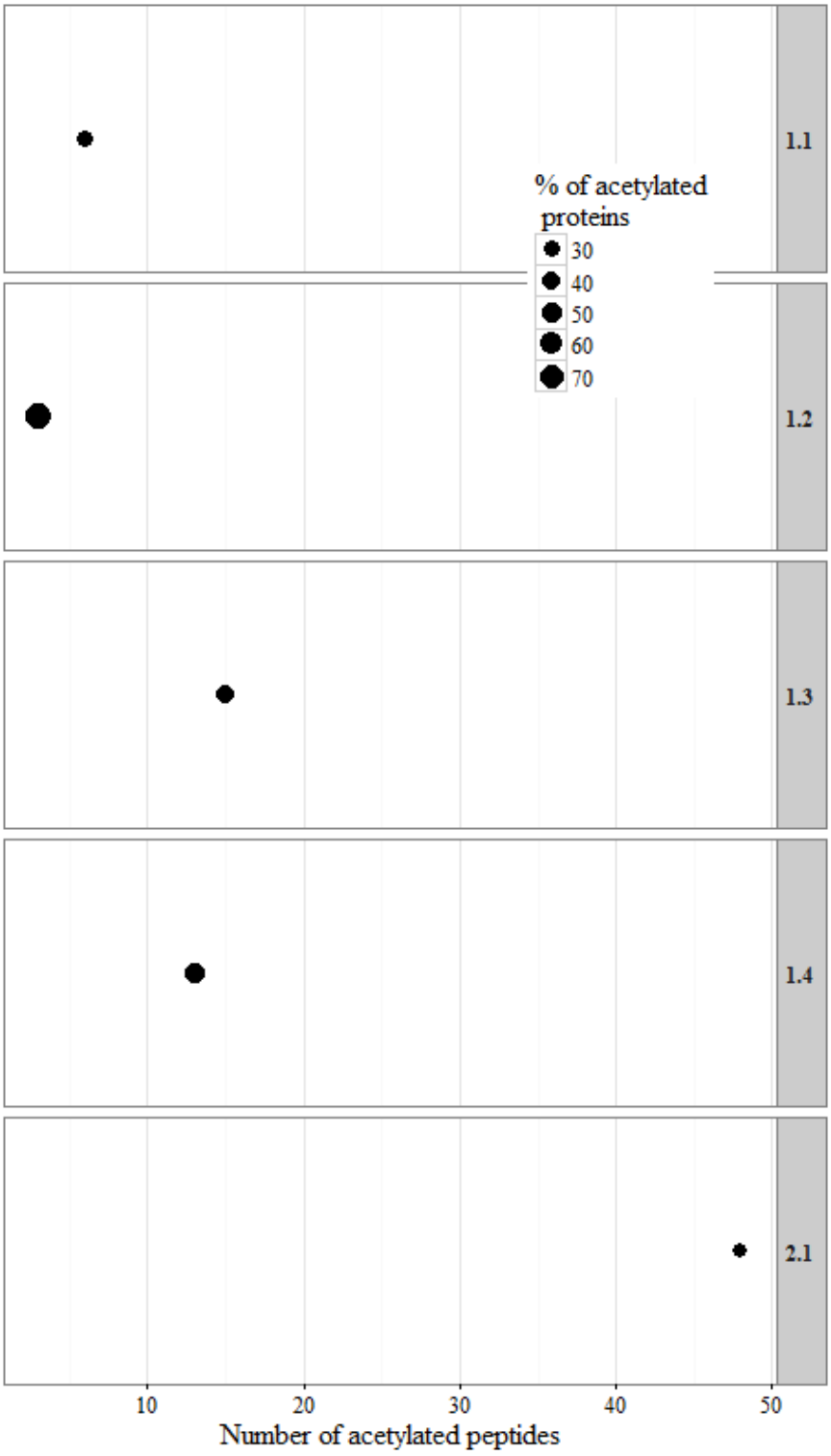


Figure 6.5 The number of unique acetylated peptides identified in untreated fibroblasts throughout method development, in relation to the degree of enrichment for acetylation within each sample. Data points are sized according to the percentage of proteins identified that were associated with an acetylated peptide in each sample. 1.1-1.4 = Choudhary method; 2.1 = Li method.

Using the original Choudhary method for sample preparation and immunoprecipitation of acetylated peptides, 39 proteins were identified, but only six peptides in this sample were acetylated. The subsequent use of Protein A agarose and an anti-acetyl-lysine antibody for IP decreased the yield of both total proteins and acetylated peptides compared to the original method in which anti-acetyl-lysine agarose conjugate was used, therefore subsequent methods used the latter option for IP. There was a dramatic increase in the degree of enrichment using Choudhary method modification 1; however, this was due to three out of only 4 proteins identified containing an acetylated lysine residue.

The third experiment (Choudhary method modification 2) was an improvement on both the first and second in terms of the number of acetylated peptides being identified (15 total unique), although fewer proteins were identified by this method than by the original.

The results from the last optimisation of the Choudhary method showed a higher degree of specificity for acetylation than the previous experiment. Fifty percent of identified proteins contained at least one acetylated lysine residue on a peptide compared to 40%. A similar number of unique acetylated peptides were identified (13 compared to 15). However, modifications to the method did not increase the number of acetylated peptides towards a similar number that were identified from cell lysates in the literature.

Using the Li method for the enrichment of acetylated peptides, there was a dramatic increase in the number of unique acetylated peptides identified compared to the Choudhary method. In this experiment there were 48 unique acetylated peptides identified, an increase of 320%. However, the specificity for acetylation fell to 25.8%, the lowest of all experiments; 151 proteins were identified and 39 of these contained at least one acetylated lysine residue. Overall, the Li method yielded the highest number of unique acetylated peptides, but at the expense of IP specificity for acetylation.

The last round of optimisation of the Choudhary method included treatment of the MRC-5 fibroblasts with 2 nM FK228 for comparison with an untreated control, to find differences in the extent of acetylation. Prior to immunoprecipitation of peptides, protein from each trypsin-EDTA-harvested cell lysate was also subjected to western blotting for acetyl-lysine and acetyl histone H3, with β-actin as a loading control. Acetylation of histone H3 increased with FK228 treatment as previously reported by Davies *et al*. (unpublished), suggesting that FK228 had successfully inhibited histone deacetylation (**Figure 6.6**).

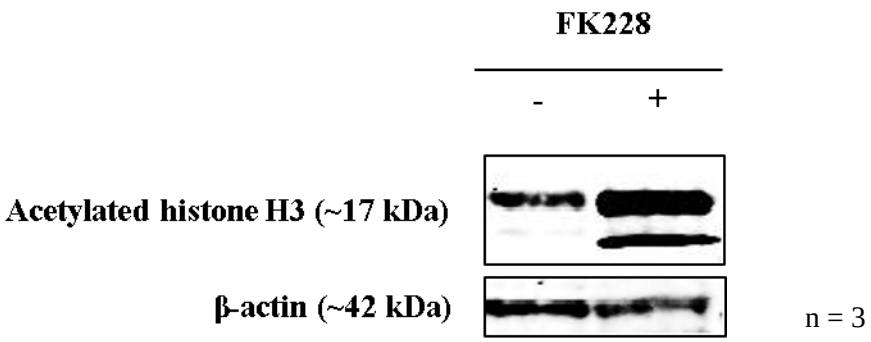


Figure 6.6 Western blot of acetylated histone H3 from MRC-5 fibroblasts with and without FK228 treatment. MRC-5 fibroblasts were left untreated or were treated with 2 nM FK228 for 48 h, and proteins were analysed by western blotting for acetylated histone H3, with β-actin as a loading control.

Global lysine acetylation was also analysed in response to treatment. Lysine acetylation increased in several other non-histone proteins following FK228 treatment (bands at ~130 kDa, 62 kDa, 50 kDa, 42 kDa, 37 kDa and 25 kDa) (**Figure 6.7**).

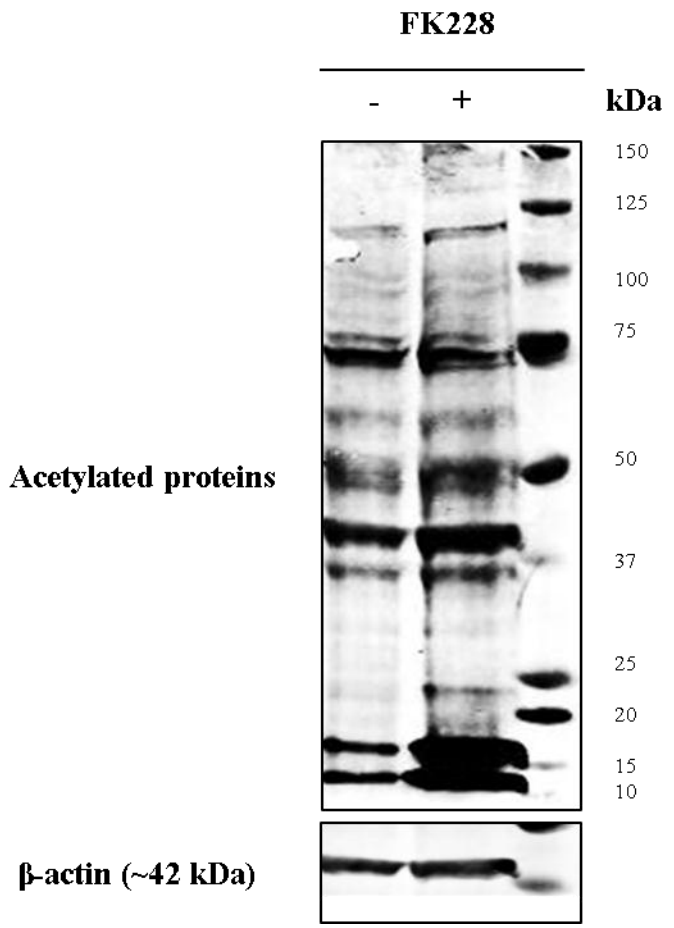


Figure 6.7 Western blot of acetylated lysine residues of proteins from MRC-5 fibroblasts with and without FK228 treatment. MRC-5 fibroblasts were treated with 2 nM FK228 for 48 h and proteins were analysed by western blotting for acetylated lysine, with β-actin as a loading control.

Following the western blot results, cell lysates were subjected to immunoprecipitation and analysis by UPLC-HDMS^{E}. Sixteen proteins were identified in both conditions; eight proteins had at least one acetylated peptide in the control, and eleven in the FK228-treated sample, indicating that this sample was more enriched for acetylation. Thirteen unique acetylated peptides were identified in the control, and this more than doubled in the FK228-treated sample, with 32 unique acetylated peptides identified. This is illustrated in **Figure 6.8**. The large increase in the number of acetylated peptides compared to protein number was in part due to 2 proteins, WD repeat-containing protein 43 and HIV Tat-specific factor 1, having 7 and 14 unique acetylated peptides identified respectively. HIV Tat-specific factor 1 contained 21 acetylated lysine residues on 14 acetylated peptides. This protein is a general transcription factor that plays a role in transcriptional elongation (Li and Green, 1998).

These results suggest that a higher proportion of cellular proteins may be acetylated after FK228 treatment, and some may potentially have more than one acetylation site.

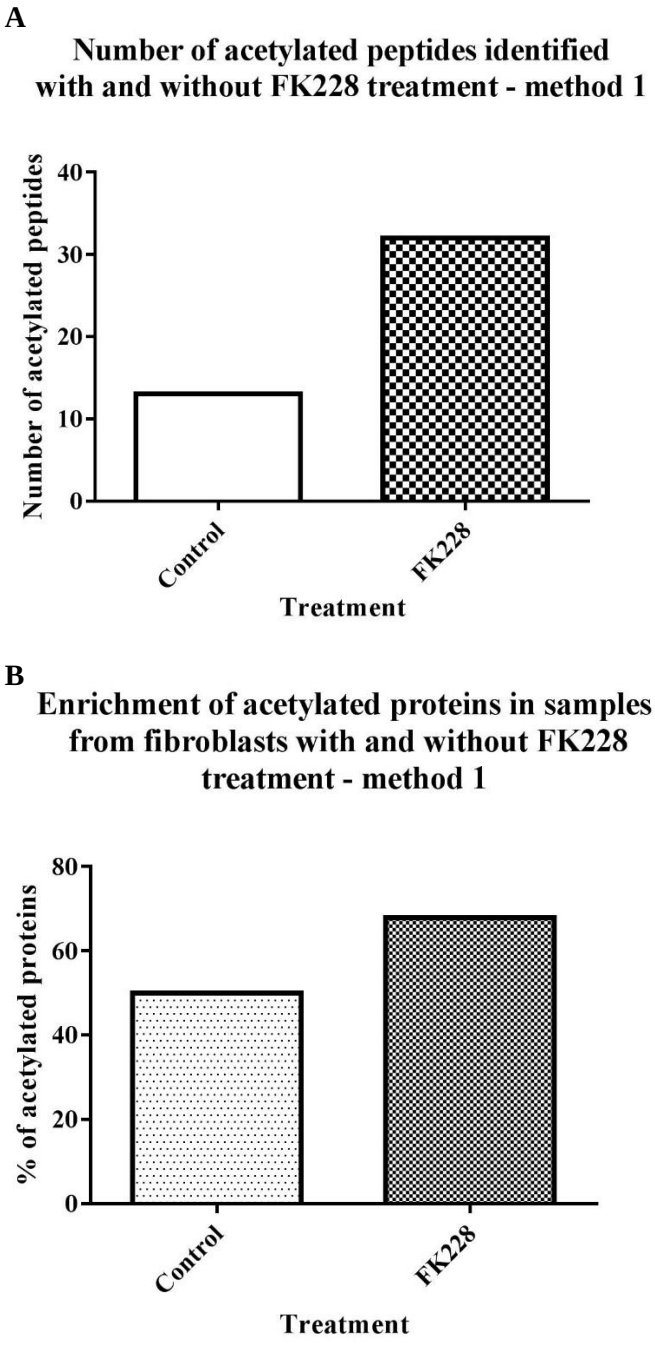


Figure 6.8 A.) The number of unique acetylated peptides identified from fibroblasts with and without FK228 treatment and B.) The degree of enrichment of acetylated peptides in samples from control and FK228-treated fibroblasts, prepared using the Choudhary method + modification 3. MRC-5 fibroblasts were treated with 2 nM FK228 for 48 h or left untreated before IP and mass spectrometry analysis.

6.4. Discussion

FK228 is a histone deacetylase inhibitor with potential as a treatment for idiopathic pulmonary fibrosis. This is attributed to its inhibitory effects on the excessive proliferation and differentiation of fibroblasts (Davies *et al*., unpublished), which contributes to the pathogenesis of this disease through subsequent deposition of extracellular matrix in the lung interstitium (Fernandez and Eickelberg, 2012). These inhibitory effects are due to transcriptional regulation of genes involved in important biological processes including the cell cycle (Nakajima et al., 1998), through changes in the acetylation of histone and non-histone proteins. Recent technological advancements in proteomics has allowed for the "acetylome" or acetylation profiles of cells to be characterised. The end goal of this study is to apply these types of methods to IPF fibroblasts and to analyse their acetylome, to reveal site-specific targets of this drug.

Prior to this study, a second workflow was used to search the data generated in **Chapter 4** concerning the global protein expression of IPF fibroblasts with and without FK228 treatment, which included acetylated lysine as a variable modification in the search criteria in order to look at the acetylation states of the proteins identified (data not shown). However, the 4% false discovery rate applied to the search, which was used for general global expression profiling to ensure that fewer true identifications are missed, meant that many proteins identified as acetylated were very likely to be false positives. In addition, peptide modifications are typically low abundant and may not have been identified due to the complex background of non-modified proteins in these samples. Thirdly, this manner of acetylation analysis is not quantitative, as quantification was performed at the protein level for these samples using an internal digest standard, rather than at the peptide level. Therefore, the only information that could be of use using this method of acetylation analysis was qualitative, which is less informative.

Differences in acetylation of proteins between untreated and FK228-treated fibroblasts could be quantitatively measured by employing the stable-isotope labelling with amino acids in cell culture (SILAC) technique. Using this method, cells from one condition would be cultured with natural amino acids, and the other with heavy isotope-labelled amino acids. Both samples would be mixed 1:1 and analysed in one MS run. The ratio of intensities of the individual light and heavy-labelled peptides can be used for peptide level quantification (Ong et al., 2002). This method would allow the detection and quantification of global changes in acetylation in response to FK228.

As SILAC is an expensive technique in terms of both cost and time (Chaube, 2014), initial experiments focused on the qualitative analysis of acetylated peptides, initially in a fibroblast cell line. The aforementioned study by Choudhary et al. (2009) showed that the number of acetylation sites identified without affinity enrichment was 60-fold lower than when samples were enriched. Acetylated peptides were therefore enriched from untreated and later FK228-treated, MRC-5 fibroblast lysates to maximise the opportunity for their detection against a complex background of high abundant, non-acetylated peptides. Enrichment was performed using an immunoprecipitation technique, for the most part based upon agarose-conjugated anti-acetyl-lysine antibodies, and the resulting peptides were analysed by UPLC-HDMS^{E}.

Initial testing of the anti-acetyl-lysine agarose by western blotting of immunoprecipitated proteins showed more lysine acetylation in the IP eluate as opposed to the cell lysate, suggesting an enrichment of acetylated proteins in the eluate. Western blotting of fibroblast cell lysates following FK228 treatment showed that FK228 had had the expected effect of increasing lysine acetylation.

However, mass spectrometry analysis of peptide IP eluates showed low yields of acetylated peptides throughout method development, with 48 being the maximum number of unique acetylated peptide identifications in one sample. These results are very different from those of Choudhary *et al.*, who reported identifying 3600 acetylation sites on 1750 proteins in a human acute myeloid leukaemia cell line, with and without SAHA treatment, using this method of enrichment. This group used the same reagents for the IP and elution, and an almost identical method for sample preparation to the Choudhary method + modification 3 described here, including protein extraction and OFFGEL fractionation steps. It is therefore difficult to isolate the component of the method associated with such major differences in the number of acetylated proteins identified between the two studies. One key point is that untreated MRC-5 fibroblasts may show limited acetylation, leading to few acetylated peptide identifications, but given that this number did not then increase dramatically following treatment with a potent histone deacetylase inhibitor using the same method for sample preparation, this seems unlikely to be the case.

Another point to note is that the report by Choudhary *et al*. does not state the amount of starting material used, which could be substantially greater that the 2 mg protein used for the Choudhary method for this study. Further investigation of the literature following all experiments using the Choudhary method highlighted that a larger quantity of starting protein may increase the yield of acetylated peptides; Li *et al*. (2013) described a method that uses 10-20 mg of protein for enrichment, noting that they have been unsuccessful at

immune-purifying acetylated peptides with less than 1-2 mg of starting protein. Using this method (denoted the Li method), there was a large increase in the number of unique acetylated peptides identified. However, the number did not exceed 50 peptides, questioning whether the higher yield of acetylated peptides justifies the large amount of starting material. It is likely that many more acetylated peptides would have been identified using the Choudhary method with this amount of starting material, which is approximately a 10x larger amount.

Many of the modifications to the Choudhary method were primarily to eliminate polyethylene glycol (PEG), which was present in the majority of experiments. One of the main uses of PEG is for the stabilisation of antibodies, but it is one of the most common and problematic contaminants in mass spectrometry and it can be introduced into samples through use of organic solvent, sample buffers and detergents, all commonly used in the laboratory (Hodge et al., 2013). PEG co-eluting from the LC column with peptide ions of interest can cause ion suppression, thus substantially lowering the number of peptide identifications (Ramanathan, 2008). In these experiments, it is currently thought that it was introduced through the agarose bead solution, present to stabilise the antibody, or due to "shedding" from the beads if incubation or elution was performed in too harsh conditions such as the elution buffer being too acidic. The method used for this analysis consisted of steps previously used in sample preparation for mass spectrometry analyses with the exception of the immunoprecipitation step, with no detectable PEG contamination; through the process of elimination the agarose beads or antibody are assumed to be the cause of this.

From these method optimisation studies, it can be concluded that a method that does not require the use of agarose beads is likely to cause much fewer difficulties, both with respect to acetylated peptide enrichment and mass spectrometry analysis. It is essential that the enrichment be sufficiently optimised before any quantitative difference in acetylation profile can be measured.

If these problems can be resolved, this method could be used for the enrichment of acetylated peptides from IPF and non-IPF primary fibroblasts, to analyse acetylome differences, and the SILAC technique could be applied for quantification. These fibroblasts could also be treated with FK228 to analyse the change in levels of acetylation across the acetylome in response to this HDAC inhibitor. It is hypothesized that there would be an increase in the number of acetylated peptides or more acetylation per peptide in response to FK228 due to inhibition of the removal of acetyl groups by histone deacetylases. Choudhary *et al*. reported that the HDACIs they studied were very specific, only upregulating approximately 10% of the acetylation sites identified by more than 2-fold, so it would be

really interesting to find out if FK228 also possesses this quality in terms of affecting global acetylation and gene expression. They also showed that in three different cell lines, many of their identified acetylation sites were found to be present on proteins with roles in important biological processes including the cell cycle and transcription, both of which are affected by FK228 treatment in IPF fibroblasts (see **Chapter 4**). Therefore, future studies may be able to show that this is also the case in IPF fibroblasts, and that some of these acetylation sites are affected by FK228 treatment, which would highlight important targets of FK228 and tie these studies together effectively.

7. Final Discussion and Future Work

The work presented in this thesis has made significant steps in identifying novel candidate biomarkers of idiopathic pulmonary fibrosis, and characterising the fibroblast response to FK228 treatment.

Firstly, myofibroblast formation has been characterised at the proteomic level by mass spectrometry (**Chapter 3**), a process that occurs aberrantly in IPF. The analysis of protein expression in both the fibroblast and the myofibroblast model, both intracellularly and extracellularly, has allowed for fully comprehensive characterisation of both inside and outside of the cell during differentiation. It has also enabled quantitative comparison between the two cell states, for assessment of the differences in proteins present, the abundance of those proteins, and the biological processes and signalling pathways that are modulated after cellular differentiation.

This study characterised lung fibroblasts with extensive coverage, and observed a greater number of differentially expressed proteins in response to the profibrotic cytokine transforming growth factor beta-1 than any other study reported in the literature to date, by a substantial margin. In addition, the large number of differentially expressed proteins reported here illustrates the extent to which the proteome changes as a fibroblast differentiates, and highlights β-catenin and MMP-2 as key players of the myofibroblast.

A notable result from this study was that the proteins with significantly different expression following TGF-β_1 treatment were enriched for the non-canonical, Smad-independent TGF-β_1 signalling pathways, RAS/MEK/ERK and JNK/p38, with little representation of the Smad pathway. Previous studies have focused on Smads in TGF-β_1 signal transduction (Kottmann et al., 2012; Lepparanta et al., 2012; Thannickal et al., 2003) and their role as effectors is not disputed by their non-detection in these proteomic experiments, as the abundance of these proteins could have been below the limit of detection. Although it may be assumed that cells would increase the synthesis of a particular protein in order to increase its activity, the activity of a protein does not necessarily correlate to abundance. Rather, these results indicate that the influence of the Smad-independent pathways on TGF-β_1 signalling is important. The induction of α-SMA and collagen expression by TGF-β_1 has been demonstrated to be dependent on p38 MAP kinase and on ERK (Sakai and Tager, 2013; Hu et al., 2006) and as previously mentioned, it has been shown that Smad-dependent and –independent signalling pathways converge (Zhang et al., 1998), and influence similar biological processes through transcriptional activation. TGF-β_1 signalling in terms of

myofibroblast formation should not be viewed as signal transduction through one main pathway, but instead, as described in this thesis, as a signalling network induced by TGF-β_1 during fibroblast differentiation. The increase in abundance of several proteins involved in the MAP kinase cascades in this study suggests that an upregulation of these pathways, which regulate the transcription of genes involved in cell proliferation, differentiation and apoptosis, resulting in a myofibroblastic phenotype, may serve as an additional indicator for myofibroblast identification. This would be along with the current indicators - an increase in phosphorylation of Smad proteins, an increase in collagen III expression and the presence of α-smooth muscle actin in fibrils.

IPF fibroblasts were analysed using a proteomic approach to identify potential biomarkers of the disease. Fibroblasts were used as the potential source of biomarkers since they play a major role in IPF pathogenesis, so that those identified were more likely to be pathogenesis-relevant. Other IPF biomarker studies have used sources such as serum, sputum or BAL fluid, so that if translated to the clinic, the samples could be obtained repeatedly using a less invasive method than biopsy (Kaarteenaho and Lappi-Blanco, 2015). Obtaining BAL fluid can however impose a risk to the patient as it can trigger an acute exacerbation (Sakamoto et al., 2012), and blood samples are obtained relatively distant from the lung, so it may be difficult to link potential biomarkers to the disease. Furthermore, studies that have used BAL fluid, serum and sputum to identify biomarkers have identified several candidates, but none have been validated and translated into clinical use (Flynn et al., 2015). Analysing the expression of proteins from BALF and serum from the same patients whose fibroblasts have been analysed in this study would reveal whether any of these potential candidate biomarkers in the fibroblasts are also elevated/decreased in the BALF and/or serum compared to non-IPF controls. This would give more credibility to these proteins as a candidate IPF biomarker, and would make them more accessible than fibroblasts isolated from biopsy.

Investigations of the primary IPF fibroblast proteome, both in 2D and 3D culture (**Chapters 4 and 5**) allowed for the identification of over 100 differentially expressed proteins between IPF and non-IPF fibroblasts and thus potential candidate protein biomarkers of IPF, the primary aim outlined in this thesis. Over 300 proteins had a considerable change in expression following treatment with the potential IPF therapeutic FK228, across the two culture systems. Further work involves validating some of these differentially expressed proteins by other proteomic methods, such as western blotting, or a targeted mode of mass spectrometry such as MRM, or by immunohistochemistry.

The study discussed in **Chapter 5** reports the first time that profiles of IPF fibroblasts grown in a 3D culture system such as this have been studied. 3D model systems enable the study of disease and testing of therapeutics in a more physiologically relevant manner. A large difference in protein expression was observed between these IPF fibroblasts and fibroblasts of the same origin but grown in 2D culture as monolayers. Candidate biomarker identification should therefore also be performed in the 3D system by growing non-IPF fibroblasts in 3D culture and identifying protein expression differences between these cells and IPF fibroblasts. Proteins independently listed as potential candidate biomarkers of IPF and/or the IPF fibroblast response to FK228 in both culture systems are more likely to be reliable and thus have a higher chance of being validated for use in the clinic. An example of this is NQO1, a protein with significantly affected expression following FK228 treatment in both 2D and 3D culture. NQO1 is likely to be more reliable as a candidate biomarker of the FK228 cellular response than any identified in only one culture system, since neither system is truly representative of *in vivo* pathogenesis and the change in expression of this protein occurs independently of the culture system chosen.

An additional source of yet unexplored biomarkers in these studies is the IPF fibroblast secretome, which should be analysed due to the importance of extracellular matrix component secretion in IPF pathogenesis. MRC-5 fibroblasts treated with the profibrotic growth factor TGF-β_1 secreted many more proteins than the control, including ECM proteins. Fibroblast differentiation and matrix deposition are increased in IPF, so it is likely that many more secreted proteins would be detected from IPF fibroblasts than from non-IPF controls. Additionally, the large differences in protein expression observed intracellularly between fibroblasts from the same donor but grown in different culture systems suggests that their secretomes would also be very different. Secretome analysis would also contribute to a more comprehensive proteomic profile of the ECM in both systems, which is important because of its influence on cell behaviour.

More dramatic expression differences could have been observed by pre-treating IPF/non-IPF fibroblasts with TGF-β_1 prior to analysis, in both culture systems, as this would enhance their fibrotic properties. Fibroblast responses to FK228 treatment might also be amplified. This was not carried out in the first instance, as the main aim of these studies was to characterise a typical IPF fibroblast proteome, which would in theory already more fibrotic than a non-IPF fibroblast, and to find out how it responds to FK228 treatment.

Both studies would benefit from increasing the number of primary donor cultures (n=20 or more). One donor culture was analysed in the 3D system, and there was a considerable amount of variability in protein expression between the five 2D donor cultures analysed.

Due to IPF being a heterogeneous disease, as shown in this study (**Chapter 4**) and others, with progression varying between patients (Raghu et al., 2011), it is likely that increasing the number of different donor cultures for analysis would introduce even more variation, but it may make it easier to highlight overall molecular patterns common to IPF patient fibroblasts.

IPF fibroblast treatment with FK228 was shown to inhibit proliferation rather than cause cell death (**Chapter 4**), and reduce collagen production (Davies *et al*., unpublished). In order to be successful as a treatment for IPF patients, this drug should also not cause damage to other cell types, such as the alveolar epithelial cells already damaged by the disease. Dr Franco Conforti has recently shown that AT2 cell number is not affected by FK228 treatment and that the lamellar bodies of AT2 cells treated with FK228 are not visibly different from untreated cells (Davies *et al*., unpublished). His preliminary work even suggests that AT1 epithelial cells may transdifferentiate into AT2 type epithelial cells following treatment, which would replenish AT2 tissue in the alveolus and contribute to tissue repair (Conforti, pers comm), increasing the likelihood of this drug being effective as a therapy.

The evidence presented in this thesis suggests that fibroblast cell cycle progression is inhibited by FK228; therefore, at an equivalent concentration, it is likely that the lungs of an FK228-treated IPF patient would still show the hallmarks of IPF, including fibroblastic foci, but there would not be a further increase in the accumulation of fibroblasts. It is difficult to speculate on a possible decrease in scar tissue in IPF lungs following drug treatment given the type of samples analysed. The monolayer study (**Chapter 4**) did not show a significant decrease in any ECM proteins produced after FK228 treatment; however, the secretomes of these IPF donor cultures were not analysed, which would be more likely to answer this specific question since the intracellular proteome can only give an idea of the proteins produced, rather than secreted and deposited. A significant decrease in a few ECM proteins could be demonstrated after FK228 treatment in **Chapter 5** from IPF fibroblasts grown in 3D culture, but whether these proteins were deposited in the ECM layers of the model, or were intracellular could not be determined. Secretomic analysis of TGF-β_1-treated MRC-5 fibroblasts showed that the protein expression of the matrix metalloproteinase MMP2 was significantly increased compared to control; secretomic analysis of IPF fibroblasts with and without FK228 treatment may show a decrease in fibrous ECM proteins and an increase in ECM remodelling and degradation proteins following treatment. This would indicate whether the stiff matrix already formed in the lungs of an IPF patient at the time of diagnosis could be degraded by FK228, which would contribute to softening of lung tissue and even lung repair, in addition to AT1 cells transdifferentiating into AT2 cells. At present, evidence suggests that FK228 treatment would halt progression rather than cure, but there are exciting

avenues to investigate that could show that this drug could be even more effective than currently thought.

If FK228 was effective in halting disease progression, a large reduction in lung function decline may be observed over time in IPF patients, similarly to the effects of Pirfenidone and Nintedanib. Depending on the extent of disease progression at the time of diagnosis, if a patient's lung architecture was largely intact and lung fibroblast cell cycle/ECM production was arrested upon treatment, FK228 could be effective in prolonging survival. It is difficult to say at such an early stage of drug evaluation whether FK228 would be more successful than current therapeutic options, but the results from all the studies of FK228 as a possible IPF treatment show great promise.

There are several limitations to bottom-up proteomic studies of cell lysates such as those discussed in this thesis; the primary limitations are potential difficulties in the first steps of sample preparation: cell lysis, protein extraction, and protein digestion. These can be challenging as samples need to be treated in such as manner so as not to cause problems in downstream mass spectrometry analysis, whilst trying to maximise protein extraction and digestion efficiency in order to identify and quantify as many proteins as possible for proteome profiling.

Cells are often lysed using detergents for many kinds of experiments, however detergents are not compatible with LC-MS. In the protocols used in these studies, this limitation has been circumvented by adding just 0.1% SDS to each sample for lysis, which is then removed during the methanol/chloroform protein extraction step. To compensate for the use of such little detergent, the cells are also subjected to sonication, performed on ice, which shears the cells (including DNA) using ultrasonic waves (Brown and Audet, 2008). This commonly-used method of cell lysis is very effective, and avoids the addition of other potential contaminants.

The hydrophobic nature of many proteins can hinder their extraction as they are water-insoluble, therefore the addition of SDS is still necessary as it efficiently solubilizes proteins. This is especially important in the studies discussed here as fibroblast extracellular matrix proteins such as fibrillar collagen are particularly difficult to solubilise (Pflieger et al., 2006). The low concentration of SDS necessary in the protocols within this thesis may unfortunately not fully solubilise all proteins present, which would prevent their downstream analysis.

The subsequent methanol/chloroform precipitation step included in the protocol in these studies to remove SDS and other contaminants presents a similar challenge, as all proteins in

the sample, including the hydrophobic proteins, need to be re-solubilised before digestion. Because of this, proteins are resuspended in a buffer with a high urea content. Urea breaks protein disulphide bonds and increases protein solubility (Hummon et al., 2007). However, trypsin is inactive at high concentrations of urea (Harris, 1956) and urea is incompatible with mass spectrometry, therefore the buffer is diluted 4 times before trypsin digestion and samples are purified prior to MS analysis.

Trypsin is an excellent cleaver of proteins at arginine residues, but can often miss cleavages at lysine residues. Reports have shown the ratio of missed lysine sites to missed arginine sites to be 5:1-6:1 regardless of the sample analysed (Saveliev et al., 2013). This is not so problematic in the analysis of simple mixtures, as there are fewer opportunities for missed cleavages, but in complex samples such as cell lysates, inefficient digestion could result in false or fewer protein identifications. To compensate for this, and for the inactivity of trypsin in a high urea environment, endoproteinase Lys-C was also used for protein digestion in the studies discussed in this thesis. This enzyme retains activity in high urea concentrations, and efficiently cleaves lysine residues. An initial digestion period with endoproteinase Lys-C results in a higher amount of simpler peptide chains which can be more efficiently digested using trypsin.

Other limitations of proteomic studies are the sensitivity and resolution of the mass spectrometer during detection of peptide ions, especially in complex biological samples such as cell lysates from which thousands of proteins could potentially be identified. Efficient cell lysis, protein isolation and digestion are all important for introducing as many tryptic peptides into the mass spectrometer for analysis as possible, but the sensitivity of the instrument can limit the detection of low abundant peptides and therefore identification of proteins, and the resolution can hinder the ability to distinguish between peptide ions of similar *m/z* and thus make confident identifications. To reduce the impact these limitations could have in these studies, several separation techniques were employed such as isoelectric focusing and liquid chromatography, plus ion mobility and TOF within the mass spectrometer itself. This ensures that fewer peptide ions arrive at the detector at once, so that as many distinct peptide ions can be detected as possible, without missing low abundant ions. In addition, fragmenting all peptide ions without selection using the MS^E method of analysis aids the detection of low abundant proteins. However, this method is limited by the efficiency of peptide ion fragment matching back to their parent precursor ions.

In summary, there are several limitations to proteomic studies of complex biological samples, but in the studies within this thesis, steps have been taken to ensure that the impact of these limitations is kept as low as possible.

The final study in this thesis aimed to develop a method that would allow for quantitative analysis of changes in the fibroblast acetylome in response to FK228 treatment (**Chapter 6**). This is important because modulating the acetylation state of histone proteins regulates the expression of many genes, and acetylation of non-histone proteins can dramatically change their characteristics, interactions and function (Spange et al., 2009). Identifying these modified proteins and their function could provide much clearer insight into how the fibroblasts are modulated by FK228 treatment. The aim of this study was to identify all proteins whose acetylation state is modified by FK228 treatment, the number and identity of which is currently unknown. However, few peptides were identified in these experiments, and fewer were acetylated. This indicates that the immunoprecipitation technique used in this study is not efficient for enriching acetylated peptides.

A novel approach for the quantitative analysis of acetylation is presented below, which avoids performing immunoprecipitation; this would circumvent issues such as PEG contamination and masking of peptides, the use of antibodies/agarose, and such a large amount of starting material. This novel approach would require isobaric labelling for quantification using mass spectrometry, such as iTRAQ labelling (see section **1.7.6**). As iTRAQ labels react with primary amines on the N-termini of peptides and lysine residues, they will label most peptides (Trinh et al., 2013) and are ideal for studying acetylation, as described below.

Figure 7.1 illustrates how this method is proposed to work. Tryptic peptides containing both acetylated and non-acetylated residues are labelled with an iTRAQ reagent (115 in this example). The label reacts with the primary amine on the N-termini and the non-acetylated lysine residues. A HDAC enzyme is then added to the sample, which deacetylates all acetylated lysine residues, setting them up to be labelled using an ITRAQ reagent of different reporter ion mass, such as 113.

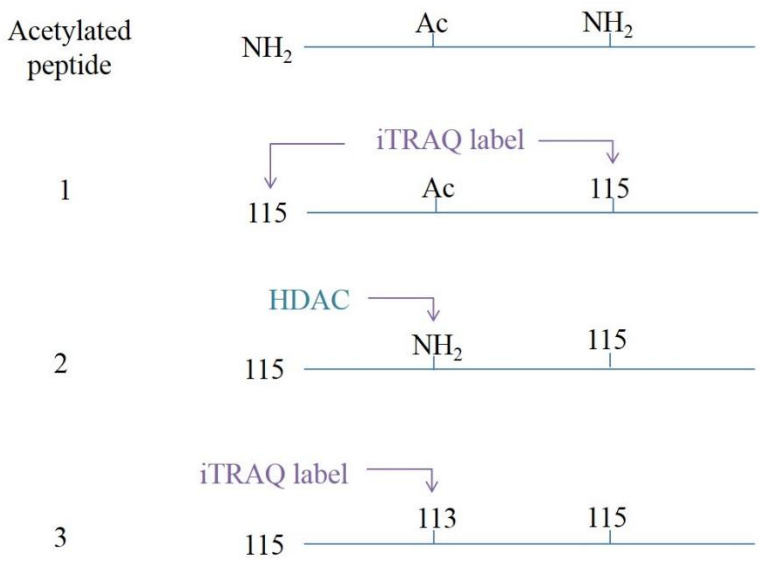


Figure 7.1 A representation of a novel method to analyse acetylated peptides. Primary amines on the N-terminus and non-acetylated lysine residues of a peptide are labelled with an iTRAQ reagent, with a reporter ion mass of 115 (**1**). A HDAC enzyme removes the acetyl group from any acetylated lysine residues (**2**) which are then labelled with a second iTRAQ reagent, with a reporter mass of 113 (**3**). Ac = acetylated lysine residue.

When analysed by mass spectrometry there would be no difference in mass between the iTRAQ reagents in the survey scan, but upon fragmentation the labels would report different masses; therefore, the fragmentation spectrum of an acetylated peptide would show the relative intensities of the reporter ions at the chosen reporter ion mass, according to the extent of acetylation (**Figure 7.2A**). This method can be adapted in order to study more than one acetylation state, for example, before and after fibroblast treatment with FK228. Cell lysate digests could be prepared separately and peptides labelled in the above manner, using iTRAQ reagents with reporter ions of different masses (for example 114 and 116) for the FK228-treated sample. Following iTRAQ labelling the samples would be mixed 1:1 and analysed on the mass spectrometer in a single run, distinguished by their reporter ions of different mass upon fragmentation. The same peptide from both samples could be studied for its degree of acetylation with and without treatment, as the number of acetylated sites can be relatively quantified by the intensity of the corresponding reporter ion (**Figure 7.2B**).

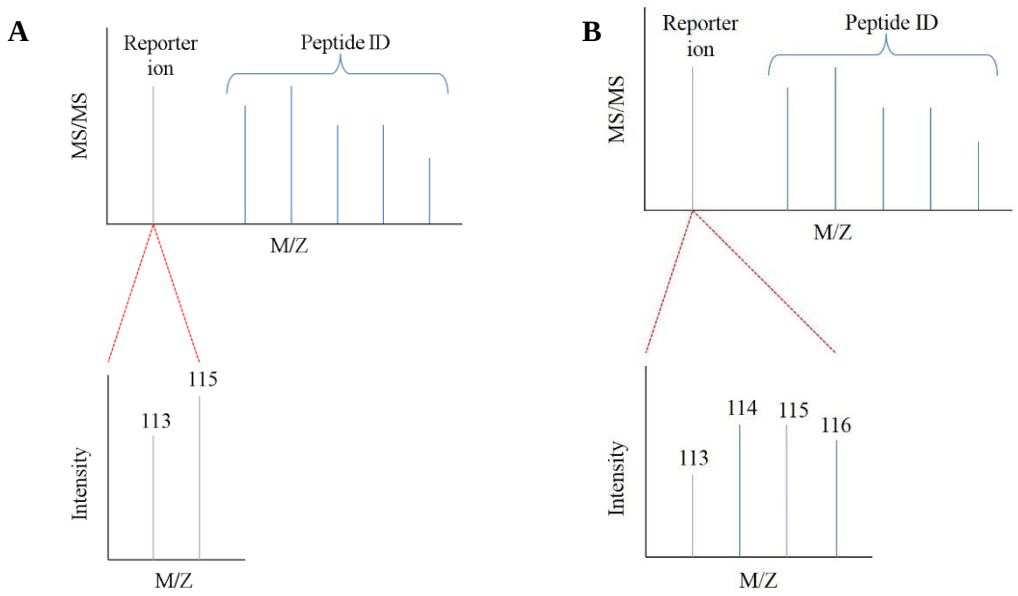


Figure 7.2 A representation of the MS/MS spectrum of a hypothetical iTRAQ-labelled acetylated peptide. The peptide can be identified by the peptide ions detected, and A.) the acetylation of the peptide can be quantified by the intensities of the reporter ions that fragment from the acetylated/non-acetylated lysine labels; B.) the difference in acetylation on the same peptide between two samples can be relatively quantified by the relative intensities of the two reporter ions of different mass. Adapted from Gingras et al., 2007.

An advantage of this method is that the whole sample is analysed by mass spectrometry, meaning that the number of peptides analysed is not dependent on an earlier step in sample preparation that may not have isolated all possible acetylated peptides (such as immunoprecipitation). Another advantage from a clinical perspective is that only a small amount of starting material is needed in order to achieve full coverage of labelling. A risk is that full coverage is not achieved, leaving acetylated and non-acetylated lysine residues unlabelled, so the amount of starting material would require optimisation; however if the ratio of label to substrate is suitable the chemical labelling itself is very efficient (Nikolov, 2012). Other disadvantages are the cost of the iTRAQ reagents that can be several thousand pounds, which potentially limits the number of experiments that can be performed, and the dependence on high fragmentation efficiency (Nikolov, 2012). Another option could be to use a Tandem Mass Tag (TMT) system (ThermoFisherScientific), which is another form of isobaric labelling. Anti-TMT antibodies are available for this system, which makes the option of enriching samples for TMT-tagged acetylated peptides available, without it being a requirement. However, TMT reagents are also expensive.

This proposed approach for studying the acetylome has not been tested before; therefore, future work would be to find out if it could be used for this purpose. Initial experiments would test the efficacy of an in-house-synthesized HDAC enzyme by adding it to a sample of synthesised acetylated peptides and checking the extent of acetylation with and without the HDAC. The synthesised acetylated peptides could then be spiked into a cell lysate digest to re-check the efficacy of the HDAC in a more complex sample containing both acetylated and non-acetylated peptides. If these experiments were successful, this method could be applied to the cell lysates of IPF and non-IPF fibroblasts to determine changes in the acetylation profiles of fibroblasts from IPF patients, which could also be analysed following treatment with FK228 for changes in acetylation.

7.1 Final conclusions

The work presented within this thesis has identified over one hundred and twenty differentially expressed proteins between IPF and non-IPF lung fibroblasts, and with further validation, some of these could be taken forward as candidate biomarkers of IPF for potential use in the clinic as diagnostic tools. It has shown that the main molecular differences are found at the point of transcriptional control, with IPF fibroblast protein expression varying greatly between patients. This may help explain why patients progress differently, and why the search for biomarkers for this disease has so far been so challenging.

This work has also contributed to current knowledge of the biological processes and signalling pathways that are differentially regulated during fibroblast differentiation, which occurs aberrantly in IPF pathogenesis; thus, it has highlighted processes and pathways that are likely to be modulated in fibrotic disease. It has shown that the expression of both β-catenin and MMP-2 is significantly increased in myofibroblasts, within the cell and the secretome respectively, and that these proteins are key interactors with other proteins that are differentially regulated during myofibroblast formation.

In addition, the fibroblast response to the histone deacetylase inhibitor FK228 has been characterised at the molecular level, and this work has shown that this response can vary between time-points and fibroblast culture methods. Following extensive validation, some of the potential candidate biomarkers of the FK228 response revealed here could be taken forward for use in clinical trials to evaluate this drug as a therapeutic for IPF. It is hoped that FK228 would be successful in slowing/halting disease progression in patients, and will be approved for use as an effective IPF treatment, increasing patient survival and quality of life.

8. References

Abe, R., Donnelly, S. C., Peng, T., Bucala, R. & Metz, C. N. 2001. Peripheral blood fibrocytes: Differentiation pathway and migration to wound sites. *Journal of Immunology,* 166**,** 7556-7562.

Aebersold, R. & Mann, M. 2003. Mass spectrometry-based proteomics. *Nature,* 422**,** 198-207.

Alder, J. K., Chen, J. J. L., Lancaster, L., Danoff, S., Su, S. C., Cogan, J. D., Vulto, I., Xie, M. Y., Qi, X. D., Tuder, R. M., Phillips, J. A., Lansdorp, P. M., Loyd, J. E. & Armanios, M. Y. 2008. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 13051-13056.

Allfrey, V. G., Faulkner, R. & Mirsky, A. E. 1964. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A,* 51**,** 786-94.

Amara, N., Goven, D., Prost, F., Muloway, R., Crestani, B. & Boczkowski, J. 2010. NOX4/NADPH oxidase expression is increased in pulmonary fibroblasts from patients with idiopathic pulmonary fibrosis and mediates TGFbeta1-induced fibroblast differentiation into myofibroblasts. *Thorax,* 65**,** 733-8.

Ameyar, M., Wisniewska, M. & Weitzman, J. B. 2003. A role for AP-1 in apoptosis: the case for and against. *Biochimie,* 85**,** 747-52.

Antoniou, K. M. & Cottin, V. 2012. The challenge of acute exacerbation of pulmonary fibrosis. *Respiration,* 83**,** 13-6.

Antoniou, K. M. & Wells, A. U. 2013. Acute exacerbations of idiopathic pulmonary fibrosis. *Respiration,* 86**,** 265-74.

Armanios, M. Y., Chen, J. J., Cogan, J. D., Alder, J. K., Ingersoll, R. G., Markin, C., Lawson, W. E., Xie, M., Vulto, I., Phillips, J. A., 3rd, Lansdorp, P. M., Greider, C. W. & Loyd, J. E. 2007. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med,* 356**,** 1317-26.

Armeanu, S., Pathil, A., Venturelli, S., Mascagni, P., Weiss, T. S., Gottlicher, M., Gregor, M., Lauer, U. M. & Bitzer, M. 2005. Apoptosis on hepatoma cells but not on primary hepatocytes by histone deacetylase inhibitors valproate and ITF2357. *J Hepatol,* 42**,** 210-7.

Artaud-Macari, E., Goven, D., Brayer, S., Hamimi, A., Besnard, V., Marchal-Somme, J., Ali, Z. E., Crestani, B., Kerdine-Romer, S., Boutten, A. & Bonay, M. 2013. Nuclear factor erythroid 2-related factor 2 nuclear translocation induces myofibroblastic dedifferentiation in idiopathic pulmonary fibrosis. *Antioxid Redox Signal,* 18**,** 66-79.

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., Zeger, S. L. & Grp, B. D. W. 2001. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics,* 69**,** 89-95.

Balakin, K. V., Ivanenkov, Y. A., Kiselyov, A. S. & Tkachenko, S. E. 2007. Histone deacetylase inhibitors in cancer therapy: latest developments, trends and medicinal chemistry perspective. *Anticancer Agents Med Chem,* 7**,** 576-92.

Banerjee, S. & Mazumdar, S. 2012. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem,* 2012**,** 282574.

Barbarotta, L. & Hurley, K. 2015. Romidepsin for the Treatment of Peripheral T-Cell Lymphoma. *J Adv Pract Oncol,* 6**,** 22-36.

Bataineh, M., Scott, A. C., Fedorak, P. M. & Martin, J. W. 2006. Capillary HPLC/QTOF-MS for characterizing complex naphthenic acid mixtures and their microbial transformation. *Anal Chem,* 78**,** 8354-61.

Baumgartner, K. B., Samet, J. M., Coultas, D. B., Stidley, C. A., Hunt, W. C., Colby, T. V. & Waldron, J. A. 2000. Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. *Am J Epidemiol,* 152**,** 307-15.

Baumgartner, K. B., Samet, J. M., Stidley, C. A., Colby, T. V. & Waldron, J. A. 1997. Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 155**,** 242-8.

Blackburn, K., Mbeunkui, F., Mitra, S. K., Mentzel, T. & Goshe, M. B. 2010. Improving Protein and Proteome Coverage through Data-Independent Multiplexed Peptide Fragmentation. *Journal of Proteome Research,* 9**,** 3621-3637.

Bond, N. J., Shliaha, P. V., Lilley, K. S. & Gatto, L. 2013. Improving qualitative and quantitative performance for MS(E)-based label-free proteomics. *J Proteome Res,* 12**,** 2340-53.

Bonella, F., Stowasser, S. & Wollin, L. 2015. Idiopathic pulmonary fibrosis: current treatment options and critical appraisal of nintedanib. *Drug Des Devel Ther,* 9**,** 6407-19.

Booth, A. J., Hadley, R., Cornett, A. M., Dreffs, A. A., Matthes, S. A., Tsui, J. L., Weiss, K., Horowitz, J. C., Fiore, V. F., Barker, T. H., Moore, B. B., Martinez, F. J., Niklason, L. E. & White, E. S. 2012. Acellular normal and fibrotic human lung matrices as a culture system for in vitro investigation. *Am J Respir Crit Care Med,* 186**,** 866-76.

Brown, A. W., Kaya, H. & Nathan, S. D. 2015. Lung transplantation in IIP: A review. *Respirology*.

Brown, R. B. & Audet, J. 2008. Current techniques for single-cell lysis. *J R Soc Interface,* 5 Suppl 2**,** S131-8.

Bucala, R., Spiegel, L. A., Chesney, J., Hogan, M. & Cerami, A. 1994. Circulating fibrocytes define a new leukocyte subpopulation that mediates tissue repair. *Mol Med,* 1**,** 71-81.

Calandra, L. 2014. *Boehringer Ingelheim's investigational therapy nintedanib* receives the first FDA breakthrough therapy designation in IPF* [Online]. 55216 Ingelheim am Rhein

GERMANY: Boehringer Ingelheim. Available: http://www.boehringer-ingelheim.com/news/news_releases/press_releases/2014/16_july_2014_ipf.html [Accessed 23/07/14 2014].

Cao, Y., Semanchik, N., Lee, S. H., Somlo, S., Barbano, P. E., Coifman, R. & Sun, Z. 2009. Chemical modifier screen identifies HDAC inhibitors as suppressors of PKD models. *Proc Natl Acad Sci U S A,* 106**,** 21819-24.

Carleo, A., Bargagli, E., Landi, C., Bennett, D., Bianchi, L., Gagliardi, A., Carnemolla, C., Perari, M. G., Cillis, G., Armini, A., Bini, L. & Rottoli, P. 2016. Comparative proteomic analysis of bronchoalveolar lavage of familial and sporadic cases of idiopathic pulmonary fibrosis. *J Breath Res,* 10**,** 026007.

Carre, A. L., James, A. W., MacLeod, L., Kong, W., Kawai, K., Longaker, M. T. & Lorenz, H. P. 2010. Interaction of wingless protein (Wnt), transforming

growth factor-beta1, and hyaluronan production in fetal and postnatal fibroblasts. *Plast Reconstr Surg,* 125**,** 74-88.

Carthy, J. M., Garmaroudi, F. S., Luo, Z. & McManus, B. M. 2011. Wnt3a induces myofibroblast differentiation by upregulating TGF-beta signaling through SMAD2 in a beta-catenin-dependent manner. *PLoS One,* 6**,** e19809.

Chaube, R. 2014. Absolute quantitation of post-translational modifications. *Front Chem,* 2**,** 58.

Chen, G., Li, A., Zhao, M., Gao, Y., Zhou, T., Xu, Y., Du, Z., Zhang, X. & Yu, X. 2008. Proteomic analysis identifies protein targets responsible for depsipeptide sensitivity in tumor cells. *J Proteome Res,* 7**,** 2733-42.

Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research,* 37**,** W305-W311.

Cheon, S. S., Nadesan, P., Poon, R. & Alman, B. A. 2004. Growth factors regulate beta-catenin-mediated TCF-dependent transcriptional activation in fibroblasts during the proliferative phase of wound healing. *Exp Cell Res,* 293**,** 267-74.

Chojkier, M., Houglum, K., Solis-Herruzo, J. & Brenner, D. A. 1989. Stimulation of collagen gene expression by ascorbic acid in cultured human fibroblasts. A role for lipid peroxidation? *J Biol Chem,* 264**,** 16957-62.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V. & Mann, M. 2009. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science,* 325**,** 834-840.

Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. 2014. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol,* 15**,** 536-50.

Clarke, D. L., Carruthers, A. M., Mustelin, T. & Murray, L. A. 2013. Matrix regulation of idiopathic pulmonary fibrosis: the role of enzymes. *Fibrogenesis Tissue Repair,* 6**,** 20.

Coker, R. K., Laurent, G. J., Shahzeidi, S., Lympany, P. A., du Bois, R. M., Jeffery, P. K. & McAnulty, R. J. 1997. Transforming growth factors-beta 1, -beta 2, and -beta 3 stimulate fibroblast procollagen production in vitro but are

differentially expressed during bleomycin-induced lung fibrosis. *Am J Pathol,* 150**,** 981-91.

Conte, E., Gili, E., Fagone, E., Fruciano, M., Iemmolo, M. & Vancheri, C. 2014. Effect of pirfenidone on proliferation, TGF-beta-induced myofibroblast differentiation and fibrogenic activity of primary human lung fibroblasts. *Eur J Pharm Sci,* 58**,** 13-9.

Cottin, V., Brillet, P. Y., Nunes, H. & Cordier, J. F. 2007. Combined pulmonary fibrosis and emphysema. *Presse Medicale,* 36**,** 936-944.

Cox, J., Michalski, A. & Mann, M. 2011. Software lock mass by two-dimensional minimization of peptide mass errors. *J Am Soc Mass Spectrom,* 22**,** 1373-80.

Cushing, M. C. & Anseth, K. S. 2007. Materials science. Hydrogel cell cultures. *Science,* 316**,** 1133-4.

Davies, E. R., Ganesan, A., Packham, G., Thatcher, T., Sime, P. J., O'Reilly, K. M. A. & Davies, D. E. 2011. The Depsipeptide HDAC Inhibitor FK228 (Romidepsin) Has Anti-Fibrotic Properties In Fibrotic Primary Pulmonary Fibroblasts. *American Journal of Respiratory and Critical Care Medicine,* 183.

Davies, E. R., Haitchi, H. M., Thatcher, T. H., Sime, P. J., Kottmann, R. M., Ganesan, A., Packham, G., O'Reilly, K. M. A. & Davies, D. E. 2012. Spiruchostatin A Inhibits Proliferation and Differentiation of Fibroblasts from Patients with Pulmonary Fibrosis. *American Journal of Respiratory Cell and Molecular Biology,* 46**,** 687-694.

de Godoy, L. M. F., Olsen, J. V., de Souza, G. A., Li, G. Q., Mortensen, P. & Mann, M. 2006. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biology,* 7.

de Lemos, J. A. & McGuire, D. K. 2011. Biomarkers in clinical trials: can we move from fortune telling to disease profiling? *Circulation,* 124**,** 663-5.

Degryse, A. L. & Lawson, W. E. 2011. Progress toward improving animal models for idiopathic pulmonary fibrosis. *Am J Med Sci,* 341**,** 444-9.

Derynck, R. & Zhang, Y. E. 2003. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature,* 425**,** 577-84.

Desmouliere, A., Geinoz, A., Gabbiani, F. & Gabbiani, G. 1993. 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,* 122**,** 103-11.

Dhanasekaran, D. N. & Reddy, E. P. 2008. JNK signaling in apoptosis. *Oncogene,* 27**,** 6245-51.

Doneanu, C. E., Xenopoulos, A., Fadgen, K., Murphy, J., Skilton, S. J., Prentice, H., Stapels, M. & Chen, W. 2012. Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *MAbs,* 4**,** 24-44.

Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics,* 10**,** 48.

Emanuele, S., Lauricella, M. & Tesoriere, G. 2008. Histone deacetylase inhibitors: apoptotic effects and clinical implications (Review). *Int J Oncol,* 33**,** 637-46.

Enomoto, T., Usuki, J., Azuma, A., Nakagawa, T. & Kudoh, S. 2003. Diabetes mellitus may increase risk for idiopathic pulmonary fibrosis. *Chest,* 123**,** 2007-11.

Evans, C., Noirel, J., Ow, S. Y., Salim, M., Pereira-Medrano, A. G., Couto, N., Pandhal, J., Smith, D., Pham, T. K., Karunakaran, E., Zou, X., Biggs, C. A. & Wright, P. C. 2012. An insight into iTRAQ: where do we stand now? *Anal Bioanal Chem,* 404**,** 1011-27.

Fahim, A., Crooks, M. & Hart, S. P. 2011a. Gastroesophageal reflux and idiopathic pulmonary fibrosis: a review. *Pulm Med,* 2011**,** 634613.

Fahim, A., Dettmar, P. W., Morice, A. H. & Hart, S. P. 2011b. Gastroesophageal Reflux and Idiopathic Pulmonary Fibrosis: A Prospective Study. *Medicina-Lithuania,* 47**,** 200-205.

Fala, L. 2015. Ofev (Nintedanib): First Tyrosine Kinase Inhibitor Approved for the Treatment of Patients with Idiopathic Pulmonary Fibrosis. *Am Health Drug Benefits,* 8**,** 101-4.

Fernandez, I. E. & Eickelberg, O. 2012. New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis. *Lancet,* 380**,** 680-688.

Flynn, M., Baker, E. & Kass, D. J. 2015. Idiopathic pulmonary fibrosis biomarkers: clinical utility and a way of understanding disease pathogenesis. *Current Biomark Find,* 5**,** 21-33.

Foss, F., Horwitz, S., Pro, B., Prince, H. M., Sokol, L., Balser, B., Wolfson, J. & Coiffier, B. 2016. Romidepsin for the treatment of relapsed/refractory peripheral T cell lymphoma: prolonged stable disease provides clinical benefits for patients in the pivotal trial. *J Hematol Oncol,* 9**,** 22.

Foster, M. W., Morrison, L. D., Todd, J. L., Snyder, L. D., Thompson, J. W., Soderblom, E. J., Plonk, K., Weinhold, K. J., Townsend, R., Minnich, A. & Moseley, M. A. 2015. Quantitative proteomics of bronchoalveolar lavage fluid in idiopathic pulmonary fibrosis. *J Proteome Res,* 14**,** 1238-49.

Furumai, R., Matsuyama, A., Kobashi, N., Lee, K. H., Nishiyama, M., Nakajima, H., Tanaka, A., Komatsu, Y., Nishino, N., Yoshida, M. & Horinouchi, S. 2002. FK228 (depsipeptide) as a natural prodrug that inhibits class I histone deacetylases. *Cancer Res,* 62**,** 4916-21.

Garcia-Areas, R., Libreros, S., Amat, S., Keating, P., Carrio, R., Robinson, P., Blieden, C. & Iragavarapu-Charyulu, V. 2014. Semaphorin7A promotes tumor growth and exerts a pro-angiogenic effect in macrophages of mammary tumor-bearing mice. *Front Physiol,* 5**,** 17.

Garcia-Sancho, C., Buendia-Roldan, I., Fernandez-Plata, M. R., Navarro, C., Perez-Padilla, R., Vargas, M. H., Loyd, J. E. & Selman, M. 2011. Familial pulmonary fibrosis is the strongest risk factor for idiopathic pulmonary fibrosis. *Respiratory Medicine,* 105**,** 1902-1907.

Gardet, A., Zheng, T. S. & Viney, J. L. 2013. Genetic architecture of human fibrotic diseases: disease risk and disease progression. *Front Pharmacol,* 4**,** 159.

Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G. Z., Gorenstein, M. V., Bateman, R. H. & Langridge, J. I. 2009. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics,* 9**,** 1683-95.

Glass, E. & Viale, P. H. 2013. Histone deacetylase inhibitors: novel agents in cancer treatment. *Clin J Oncol Nurs,* 17**,** 34-40.

Graham, D. R. M., Elliott, S. T. & Van Eyk, J. E. 2005. Broad-based proteomic strategies: a practical guide to proteomics and functional screening. *Journal of Physiology-London,* 563**,** 1-9.

Gribbin, J., Hubbard, R. & Smith, C. 2009. Role of diabetes mellitus and gastro-oesophageal reflux in the aetiology of idiopathic pulmonary fibrosis. *Respiratory Medicine,* 103**,** 927-931.

Gribbin, J., Hubbard, R. B., Le Jeune, I., Smith, C. J. P., West, J. & Tata, L. J. 2006. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Thorax,* 61**,** 980-985.

Gui, C. Y., Ngo, L., Xu, W. S., Richon, V. M. & Marks, P. A. 2004. Histone deacetylase (HDAC) inhibitor activation of p21WAF1 involves changes in promoter-associated proteins, including HDAC1. *Proc Natl Acad Sci U S A,* 101**,** 1241-6.

Guo, W., Shan, B., Klingsberg, R. C., Qin, X. & Lasky, J. A. 2009. Abrogation of TGF-beta1-induced fibroblast-myofibroblast differentiation by histone deacetylase inhibition. *Am J Physiol Lung Cell Mol Physiol,* 297**,** L864-70.

Hanke, S., Besir, H., Oesterhelt, D. & Mann, M. 2008. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *Journal of Proteome Research,* 7**,** 1118-1130.

Harris, J. I. 1956. Effect of urea on trypsin and alpha-chymotrypsin. *Nature,* 177**,** 471-3.

Hata, R. & Senoo, H. 1989. L-ascorbic acid 2-phosphate stimulates collagen accumulation, cell proliferation, and formation of a three-dimensional tissuelike substance by skin fibroblasts. *J Cell Physiol,* 138**,** 8-16.

Hatamochi, A., Aumailley, M., Mauch, C., Chu, M. L., Timpl, R. & Krieg, T. 1989. Regulation of collagen VI expression in fibroblasts. Effects of cell density, cell-matrix interactions, and chemical transformation. *J Biol Chem,* 264**,** 3494-9.

Heldin, C. H., Miyazono, K. & ten Dijke, P. 1997. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature,* 390**,** 465-71.

Hinz, B. 2007. Formation and function of the myofibroblast during tissue repair. *Journal of Investigative Dermatology,* 127**,** 526-537.

Hinz, B., Phan, S. H., Thannickal, V. J., Galli, A., Bochaton-Piallat, M. L. & Gabbiani, G. 2007. The myofibroblast - One function, multiple origins. *American Journal of Pathology,* 170**,** 1807-1816.

Hisatomi, K., Mukae, H., Sakamoto, N., Ishimatsu, Y., Kakugawa, T., Hara, S., Fujita, H., Nakamichi, S., Oku, H., Urata, Y., Kubota, H., Nagata, K. & Kohno, S. 2012. Pirfenidone inhibits TGF-beta1-induced over-expression of collagen type I and heat shock protein 47 in A549 cells. *BMC Pulm Med,* 12**,** 24.

Hodge, K., Have, S. T., Hutton, L. & Lamond, A. I. 2013. Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J Proteomics,* 88**,** 92-103.

Honda, E., Yoshida, K. & Munakata, H. 2010. Transforming Growth Factor-beta Upregulates the Expression of Integrin and Related Proteins in MRC-5 Human Myofibroblasts. *Tohoku Journal of Experimental Medicine,* 220**,** 319-327.

Horimasu, Y., Hattori, N., Ishikawa, N., Kawase, S., Tanaka, S., Yoshioka, K., Yokoyama, A., Kohno, N., Bonella, F., Guzman, J., Ohshimo, S. & Costabel, U. 2012. Different MUC1 gene polymorphisms in German and Japanese ethnicities affect serum KL-6 levels. *Respir Med,* 106**,** 1756-64.

Hu, Y., Peng, J., Feng, D., Chu, L., Li, X., Jin, Z., Lin, Z. & Zeng, Q. 2006. Role of extracellular signal-regulated kinase, p38 kinase, and activator protein-1 in transforming growth factor-beta1-induced alpha smooth muscle actin expression in human fetal lung fibroblasts in vitro. *Lung,* 184**,** 33-42.

Hubbard, R., Lewis, S., Richards, K., Johnston, I. & Britton, J. 1996. Occupational exposure to metal or wood dust and aetiology of cryptogenic fibrosing alveolitis. *Lancet,* 347**,** 284-289.

Hubner, N. C., Ren, S. & Mann, M. 2008. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics,* 8**,** 4862-72.

Hummon, A. B., Lim, S. R., Difilippantonio, M. J. & Ried, T. 2007. Isolation and solubilization of proteins after TRIzol extraction of RNA and DNA from patient material following prolonged storage. *Biotechniques,* 42**,** 467-70, 472.

Hutchinson, J., Fogarty, A., Hubbard, R. & McKeever, T. 2015. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J,* 46**,** 795-806.

Hynes, R. O. & Naba, A. 2012. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol,* 4**,** a004903.

Idiopathic Pulmonary Fibrosis Clinical Research, N., Raghu, G., Anstrom, K. J., King, T. E., Jr., Lasky, J. A. & Martinez, F. J. 2012. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med,* 366**,** 1968-77.

Jain, S. & Zain, J. 2011. Romidepsin in the treatment of cutaneous T-cell lymphoma. *J Blood Med,* 2**,** 37-47.

Janossy, J., Ubezio, P., Apati, A., Magocsi, M., Tompa, P. & Friedrich, P. 2004. Calpain as a multi-site regulator of cell cycle. *Biochem Pharmacol,* 67**,** 1513-21.

Jones, M. G., Conforti, F., Montfort, T., Smart, D., Andriotis, O., Maher, T. M., Thurner, P., Mahajan, S., Richeldi, L., O'Reilly, K. M. A. & Davies, D. E. 2014. Characterisation of a novel in vitro model of fibroblastic foci. *International Colloquium on Lung and Airway Fibrosis.*

Kaarteenaho, R. 2013. The current position of surgical lung biopsy in the diagnosis of idiopathic pulmonary fibrosis. *Respir Res,* 14**,** 43.

Kaarteenaho, R. & Lappi-Blanco, E. 2015. Tissue is an issue in the search for biomarkers in idiopathic pulmonary fibrosis. *Fibrogenesis Tissue Repair,* 8**,** 3.

Kaltschmidt, B., Kaltschmidt, C., Hofmann, T. G., Hehner, S. P., Droge, W. & Schmitz, M. L. 2000. The pro- or anti-apoptotic function of NF-kappaB is determined by the nature of the apoptotic stimulus. *Eur J Biochem,* 267**,** 3828-35.

Karimi-Shah, B. A. & Chowdhury, B. A. 2015. Forced vital capacity in idiopathic pulmonary fibrosis--FDA review of pirfenidone and nintedanib. *N Engl J Med,* 372**,** 1189-91.

Kee, H. J., Sohn, I. S., Nam, K. I., Park, J. E., Qian, Y. R., Yin, Z., Ahn, Y., Jeong, M. H., Bang, Y. J., Kim, N., Kim, J. K., Kim, K. K., Epstein, J. A. & Kook, H. 2006. Inhibition of histone deacetylation blocks cardiac hypertrophy induced by angiotensin II infusion and aortic banding. *Circulation,* 113**,** 51-9.

Khalil, N., Parekh, T. V., O'Connor, R., Antman, N., Kepron, W., Yehaulaeshet, T., Xu, Y. D. & Gold, L. I. 2001. Regulation of the effects of TGF-beta 1 by activation of latent TGF-beta 1 and differential expression of TGF-beta

receptors (T beta R-I and T beta R-II) in idiopathic pulmonary fibrosis. *Thorax,* 56**,** 907-15.

Kim, B. Y., Olzmann, J. A., Choi, S. I., Ahn, S. Y., Kim, T. I., Cho, H. S., Suh, H. & Kim, E. K. 2009. Corneal Dystrophy-associated R124H Mutation Disrupts TGFBI Interaction with Periostin and Causes Mislocalization to the Lysosome. *Journal of Biological Chemistry,* 284**,** 19580-19591.

Kim, D. S., Collard, H. R. & King, T. E., Jr. 2006. Classification and natural history of the idiopathic interstitial pneumonias. *Proc Am Thorac Soc,* 3**,** 285-92.

Kim, M. S., Zhong, J. & Pandey, A. 2016. Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics,* 16**,** 700-14.

Kim, T. H., Kim, S. H., Seo, J. Y., Chung, H., Kwak, H. J., Lee, S. K., Yoon, H. J., Shin, D. H., Park, S. S. & Sohn, J. W. 2011. Blockade of the Wnt/beta-catenin pathway attenuates bleomycin-induced pulmonary fibrosis. *Tohoku J Exp Med,* 223**,** 45-54.

King, T. E., Jr., Pardo, A. & Selman, M. 2011. Idiopathic pulmonary fibrosis. *Lancet,* 378**,** 1949-1961.

Kliment, C. R. & Oury, T. D. 2010. Oxidative stress, extracellular matrix targets, and idiopathic pulmonary fibrosis. *Free Radical Biology and Medicine,* 49**,** 707-717.

Knight, D. A. & Holgate, S. T. 2003. The airway epithelium: structural and functional properties in health and disease. *Respirology,* 8**,** 432-46.

Konstantinopoulos, P. A., Vandoros, G. P. & Papavassiliou, A. G. 2006. FK228 (depsipeptide): a HDAC inhibitor with pleiotropic antitumor activities. *Cancer Chemother Pharmacol,* 58**,** 711-5.

Korfei, M., Schmitt, S., Ruppert, C., Henneke, I., Markart, P., Loeh, B., Mahavadi, P., Wygrecka, M., Klepetko, W., Fink, L., Bonniaud, P., Preissner, K. T., Lochnit, G., Schaefer, L., Seeger, W. & Guenther, A. 2011. Comparative proteomic analysis of lung tissue from patients with idiopathic pulmonary fibrosis (IPF) and lung transplant donor lungs. *J Proteome Res,* 10**,** 2185-205.

Korfei, M., Skwarna, S., Henneke, I., MacKenzie, B., Klymenko, O., Saito, S., Ruppert, C., von der Beck, D., Mahavadi, P., Klepetko, W., Bellusci, S., Crestani, B., Pullamsetti, S. S., Fink, L., Seeger, W., Kramer, O. H. &

Guenther, A. 2015. Aberrant expression and activity of histone deacetylases in sporadic idiopathic pulmonary fibrosis. *Thorax,* 70**,** 1022-32.

Korfei, M., von der Beck, D., Henneke, I., Markart, P., Ruppert, C., Mahavadi, P., Ghanim, B., Klepetko, W., Fink, L., Meiners, S., Kramer, O. H., Seeger, W., Vancheri, C. & Guenther, A. 2013. Comparative proteome analysis of lung tissue from patients with idiopathic pulmonary fibrosis (IPF), non-specific interstitial pneumonia (NSIP) and organ donors. *J Proteomics,* 85**,** 109-28.

Kottmann, R. M., Kulkarni, A. A., Smolnycki, K. A., Lyda, E., Dahanayake, T., Salibi, R., Honnons, S., Jones, C., Isern, N. G., Hu, J. Z., Nathan, S. D., Grant, G., Phipps, R. P. & Sime, P. J. 2012. Lactic Acid Is Elevated in Idiopathic Pulmonary Fibrosis and Induces Myofibroblast Differentiation via pH-Dependent Activation of Transforming Growth Factor-beta. *American Journal of Respiratory and Critical Care Medicine,* 186**,** 740-751.

Kramer, G., Moerland, P. D., Jeeninga, R. E., Vlietstra, W. J., Ringrose, J. H., Byrman, C., Berkhout, B. & Speijer, D. 2012. Proteomic analysis of HIV-T cell interaction: an update. *Front Microbiol,* 3**,** 240.

Kuo, M. H. & Allis, C. D. 1998. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays,* 20**,** 615-626.

Kuzyk, M. A., Parker, C. E., Domanski, D. & Borchers, C. H. 2013. Development of MRM-based assays for the absolute quantitation of plasma proteins. *Methods Mol Biol,* 1023**,** 53-82.

Lam, A. P., Flozak, A. S., Russell, S., Wei, J., Jain, M., Mutlu, G. M., Budinger, G. R., Feghali-Bostwick, C. A., Varga, J. & Gottardi, C. J. 2011. Nuclear beta-catenin is increased in systemic sclerosis pulmonary fibrosis and promotes lung fibroblast migration and proliferation. *Am J Respir Cell Mol Biol,* 45**,** 915-22.

Lam, A. P. & Gottardi, C. J. 2011. beta-catenin signaling: a novel mediator of fibrosis and potential therapeutic target. *Curr Opin Rheumatol,* 23**,** 562-7.

Lama, V. N. & Phan, S. H. 2006. The extrapulmonary origin of fibroblasts: stem/progenitor cells and beyond. *Proc Am Thorac Soc,* 3**,** 373-6.

Lane, C. S. 2005. Mass spectrometry-based proteomics in the life sciences. *Cellular and Molecular Life Sciences,* 62**,** 848-869.

Lange, V., Picotti, P., Domon, B. & Aebersold, R. 2008. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology,* 4.

Latsi, P. I., du Bois, R. M., Nicholson, A. G., Colby, T. V., Bisirtzoglou, D., Nikolakopoulou, A., Veeraraghavan, S., Hansell, D. M. & Wells, A. U. 2003. Fibrotic idiopathic interstitial pneumonia - The prognostic value of longitudinal functional trends. *American Journal of Respiratory and Critical Care Medicine,* 168**,** 531-537.

Laubach, J. P., Moreau, P., San-Miguel, J. F. & Richardson, P. G. 2015. Panobinostat for the Treatment of Multiple Myeloma. *Clin Cancer Res,* 21**,** 4767-73.

Lawson, W. E., Loyd, J. E. & Degryse, A. L. 2011. Genetics in pulmonary fibrosis--familial cases provide clues to the pathogenesis of idiopathic pulmonary fibrosis. *Am J Med Sci,* 341**,** 439-43.

Leask, A. & Abraham, D. J. 2004. TGF-beta signaling and the fibrotic response. *FASEB J,* 18**,** 816-27.

Lee, C. H., Hong, C. H., Yu, H. S., Chen, G. S. & Yang, K. C. 2010. Transforming growth factor-beta enhances matrix metalloproteinase-2 expression and activity through AKT in fibroblasts derived from angiofibromas in patients with tuberous sclerosis complex. *Br J Dermatol,* 163**,** 1238-44.

Lenz, A. G., Meyer, B., Costabel, U. & Maier, K. 1993. Bronchoalveolar lavage fluid proteins in human lung disease: analysis by two-dimensional electrophoresis. *Electrophoresis,* 14**,** 242-4.

Lepparanta, O., Sens, C., Salmenkivi, K., Kinnula, V. L., Keski-Oja, J., Myllarniemi, M. & Koli, K. 2012. Regulation of TGF-beta storage and activation in the human idiopathic pulmonary fibrosis lung. *Cell and Tissue Research,* 348**,** 491-503.

Lewis JK WJ, S. G. 2000. Matrix-assisted Laser Desorption/Ionization Mass Spectrometry in Peptide and Protein Analysis. *Encyclopedia of Analytical Chemistry,* R.A.Meyers edition**,** 5880-94.

Ley, B., Brown, K. K. & Collard, H. R. 2014. Molecular biomarkers in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol,* 307**,** L681-91.

Li, G. Z., Vissers, J. P., Silva, J. C., Golick, D., Gorenstein, M. V. & Geromanos, S. J. 2009. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics,* 9**,** 1696-719.

Li, Q. Q., Hao, J. J., Zhang, Z., Hsu, I., Liu, Y., Tao, Z., Lewi, K., Metwalli, A. R. & Agarwal, P. K. 2016. Histone deacetylase inhibitor-induced cell death in bladder cancer is associated with chromatin modification and modifying protein expression: A proteomic approach. *Int J Oncol,* 48**,** 2591-607.

Li, X. Y. & Green, M. R. 1998. The HIV-1 Tat cellular coactivator Tat-SF1 is a general transcription elongation factor. *Genes Dev,* 12**,** 2992-6.

Li, Y., Silva, J. C., Skinner, M. E. & Lombard, D. B. 2013. Mass Spectrometry-Based Detection of Protein Acetylation. *Sirtuins: Methods and Protocols,* 1077**,** 81-104.

Lin, B., Williams-Skipp, C., Tao, Y., Schleicher, M. S., Cano, L. L., Duke, R. C. & Scheinman, R. I. 1999. NF-kappaB functions as both a proapoptotic and antiapoptotic regulatory factor within a single cell type. *Cell Death Differ,* 6**,** 570-82.

Liu, F., Mih, J. D., Shea, B. S., Kho, A. T., Sharif, A. S., Tager, A. M. & Tschumperlin, D. J. 2010. Feedback amplification of fibrosis through matrix stiffening and COX-2 suppression. *J Cell Biol,* 190**,** 693-706.

Liu, J. & Lin, A. 2005. Role of JNK activation in apoptosis: a double-edged sword. *Cell Res,* 15**,** 36-42.

Lomas, N. J., Watts, K. L., Akram, K. M., Forsyth, N. R. & Spiteri, M. A. 2012. Idiopathic pulmonary fibrosis: immunohistochemical analysis provides fresh insights into lung tissue remodelling with implications for novel prognostic markers. *International Journal of Clinical and Experimental Pathology,* 5**,** 58-71.

Magi, B., Bini, L., Perari, M. G., Fossi, A., Sanchez, J. C., Hochstrasser, D., Paesano, S., Raggiaschi, R., Santucci, A., Pallini, V. & Rottoli, P. 2002. Bronchoalveolar lavage fluid protein composition in patients with sarcoidosis and idiopathic pulmonary fibrosis: a two-dimensional electrophoretic study. *Electrophoresis,* 23**,** 3434-44.

Maher, T. M. 2013. PROFILEing idiopathic pulmonary fibrosis: rethinking biomarker discovery. *Eur Respir Rev,* 22**,** 148-52.

Maher, T. M., Evans, I. C., Bottoms, S. E., Mercer, P. F., Thorley, A. J., Nicholson, A. G., Laurent, G. J., Tetley, T. D., Chambers, R. C. & McAnulty, R. J. 2010. Diminished prostaglandin E2 contributes to the apoptosis paradox in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med,* 182**,** 73-82.

Maher, T. M., Wells, A. U. & Laurent, G. J. 2007. Idiopathic pulmonary fibrosis: multiple causes and multiple mechanisms? *European Respiratory Journal,* 30**,** 835-839.

Malmstrom, J., Larsen, K., Malmstrom, L., Tufvesson, E., Parker, K., Marchese, J., Williamson, B., Hattan, S., Patterson, D., Martin, S., Graber, A., Juhasz, P., Westergren-Thorsson, G. & Marko-Varga, G. 2004a. Proteome annotations and identifications of the human pulmonary fibroblast. *Journal of Proteome Research,* 3**,** 525-537.

Malmstrom, J., Lindberg, H., Lindberg, C., Bratt, C., Wieslander, E., Delander, E. L., Sarnstrand, B., Burns, J. S., Mose-Larsen, P., Fey, S. & Marko-Varga, G. 2004b. Transforming growth factor-beta 1 specifically induce proteins involved in the myofibroblast contractile apparatus. *Mol Cell Proteomics,* 3**,** 466-77.

Malmstrom, J., Westergren-Thorsson, G. & Marko-Varga, G. 2001. A proteomic approach to mimic fibrosis disease evolvement by an in vitro cell line. *Electrophoresis,* 22**,** 1776-1784.

Marinkovic, A., Liu, F. & Tschumperlin, D. J. 2013. Matrices of physiologic stiffness potently inactivate idiopathic pulmonary fibrosis fibroblasts. *Am J Respir Cell Mol Biol,* 48**,** 422-30.

Mason, R. J. 2006. Biology of alveolar type II cells. *Respirology,* 11 Suppl**,** S12-5.

Mason, R. J. & Williams, M. C. 1977. Type II alveolar cell. Defender of the alveolus. *Am Rev Respir Dis,* 115**,** 81-91.

Mejia, M., Carrillo, G., Rojas-Serrano, J., Estrada, A., Suarez, T., Alonso, D., Barrientos, E., Gaxiola, M., Navarro, C. & Selman, M. 2009. Idiopathic pulmonary fibrosis and emphysema: decreased survival associated with severe pulmonary arterial hypertension. *Chest,* 136**,** 10-5.

Mi, H., Muruganujan, A. & Thomas, P. D. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res,* 41**,** D377-86.

Miyake, Y., Sasaki, S., Yokoyama, T., Chida, K., Azuma, A., Suda, T., Kudoh, S., Sakamoto, N., Okamoto, K., Kobashi, G., Washio, M., Inaba, Y., Tanaka, H. & Japan Idiopathic Pulmonary Fibrosis Study, G. 2005. Case-control study of medical history and idiopathic pulmonary fibrosis in Japan. *Respirology,* 10**,** 504-9.

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

Molyneaux, P. L. & Maher, T. M. 2013. The role of infection in the pathogenesis of idiopathic pulmonary fibrosis. *Eur Respir Rev,* 22**,** 376-81.

Moodley, Y. P., Scaffidi, A. K., Misso, N. L., Keerthisingam, C., McAnulty, R. J., Laurent, G. J., Mutsaers, S. E., Thompson, P. J. & Knight, D. A. 2003. Fibroblasts isolated from normal lungs and those with idiopathic pulmonary fibrosis differ in interleukin-6/gp130-mediated cell signaling and proliferation. *Am J Pathol,* 163**,** 345-54.

Moore, B., Lawson, W. E., Oury, T. D., Sisson, T. H., Raghavendran, K. & Hogaboam, C. M. 2013. Animal models of fibrotic lung disease. *Am J Respir Cell Mol Biol,* 49**,** 167-79.

Moore, B. B. & Hogaboam, C. M. 2008. Murine models of pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol,* 294**,** L152-60.

Murad, S., Grove, D., Lindberg, K. A., Reynolds, G., Sivarajah, A. & Pinnell, S. R. 1981. Regulation of collagen synthesis by ascorbic acid. *Proc Natl Acad Sci U S A,* 78**,** 2879-82.

Nakajima, H., Kim, Y. B., Terano, H., Yoshida, M. & Horinouchi, S. 1998. FR901228, a potent antitumor antibiotic, is a novel histone deacetylase inhibitor. *Exp Cell Res,* 241**,** 126-33.

Nakayama, S., Mukae, H., Sakamoto, N., Kakugawa, T., Yoshioka, S., Soda, H., Oku, H., Urata, Y., Kondo, T., Kubota, H., Nagata, K. & Kohno, S. 2008. Pirfenidone inhibits the expression of HSP47 in TGF-beta1-stimulated human lung fibroblasts. *Life Sci,* 82**,** 210-7.

Nathan, S. D. & Meyer, K. C. 2014. IPF clinical trial design and endpoints. *Curr Opin Pulm Med,* 20**,** 463-71.

Navaratnam, V., Fleming, K. M., West, J., Smith, C. J., Jenkins, R. G., Fogarty, A. & Hubbard, R. B. 2011. The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax,* 66**,** 462-7.

Nikolov, K. S., C.; Urlaub, H. 2012. Quantitative mass-spectrometry based proteomics: an overview. *In:* MARCUS, K. (ed.) *Quantitative methods in proteomics.* 233 Spring Street, New York, NY 10013, USA: Humana Press.

O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem,* 250**,** 4007-21.

Oh, C. K., Murray, L. A. & Molfino, N. A. 2012. Smoking and idiopathic pulmonary fibrosis. *Pulm Med,* 2012**,** 808260.

Oh, G.-S., Shen, A. S., Karna, A., Park, R., So, H.-S. & Yang, S.-H. 2013. Activation of Nad(p)h:quinine oxidoreductase 1 (nqo1) enzyme activity attenuates bleomycin-induced lung inflammation and fibrosis in mice. *American Thoracic Society.* Pennsylvania Convention Center.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics,* 1**,** 376-86.

Pandey, A. & Mann, M. 2000. Proteomics to study genes and genomes. *Nature,* 405**,** 837-846.

Pang, M., Kothapally, J., Mao, H., Tolbert, E., Ponnusamy, M., Chin, Y. E. & Zhuang, S. 2009. Inhibition of histone deacetylase activity attenuates renal fibroblast activation and interstitial fibrosis in obstructive nephropathy. *Am J Physiol Renal Physiol,* 297**,** F996-F1005.

Pang, M. & Zhuang, S. 2010. Histone deacetylase: a potential therapeutic target for fibrotic disorders. *J Pharmacol Exp Ther,* 335**,** 266-72.

Papiris, S. A., Kagouridis, K., Kolilekas, L., Bouros, D. & Manali, E. D. 2014. Idiopathic pulmonary fibrosis acute exacerbations: where are we now? *Expert Rev Respir Med,* 8**,** 271-3.

Pardo, A. & Selman, M. 2012. Role of matrix metaloproteases in idiopathic pulmonary fibrosis. *Fibrogenesis Tissue Repair,* 5**,** S9.

Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C. & Scrivens, J. H. 2009. A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches. *Journal of Proteome Research,* 8**,** 3752-3759.

Paulsson, M. 1992. Basement membrane proteins: structure, assembly, and cellular interactions. *Crit Rev Biochem Mol Biol,* 27**,** 93-127.

Peng, J. M. & Gygi, S. P. 2001. Proteomics: the move to mixtures. *Journal of Mass Spectrometry,* 36**,** 1083-1091.

Pepe, M. S., Etzioni, R., Feng, Z., Potter, J. D., Thompson, M. L., Thornquist, M., Winget, M. & Yasui, Y. 2001. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst,* 93**,** 1054-61.

Peterfi, Z., Donko, A., Orient, A., Sum, A., Prokai, A., Molnar, B., Vereb, Z., Rajnavolgyi, E., Kovacs, K. J., Muller, V., Szabo, A. J. & Geiszt, M. 2009. Peroxidasin is secreted and incorporated into the extracellular matrix of myofibroblasts and fibrotic kidney. *Am J Pathol,* 175**,** 725-35.

Pflieger, D., Chabane, S., Gaillard, O., Bernard, B. A., Ducoroy, P., Rossier, J. & Vinh, J. 2006. Comparative proteomic analysis of extracellular matrix proteins secreted by two types of skin fibroblasts. *Proteomics,* 6**,** 5868-79.

Phillips, R. J., Burdick, M. D., Hong, K., Lutz, M. A., Murray, L. A., Xue, Y. Y., Belperio, J. A., Keane, M. P. & Strieter, R. M. 2004. Circulating fibrocytes traffic to the lungs in response to CXCL12 and mediate fibrosis. *J Clin Invest,* 114**,** 438-46.

Piekarz, R. & Bates, S. 2004. A review of depsipeptide and other histone deacetylase inhibitors in clinical trials. *Curr Pharm Des,* 10**,** 2289-98.

Piekarz, R. L., Frye, R., Turner, M., Wright, J. J., Allen, S. L., Kirschbaum, M. H., Zain, J., Prince, H. M., Leonard, J. P., Geskin, L. J., Reeder, C., Joske, D., Figg, W. D., Gardner, E. R., Steinberg, S. M., Jaffe, E. S., Stetler-Stevenson, M., Lade, S., Fojo, A. T. & Bates, S. E. 2009. Phase II multi-institutional trial of the histone deacetylase inhibitor romidepsin as monotherapy for patients with cutaneous T-cell lymphoma. *J Clin Oncol,* 27**,** 5410-7.

Piekarz, R. L., Robey, R., Sandor, V., Bakke, S., Wilson, W. H., Dahmoush, L., Kingma, D. M., Turner, M. L., Altemus, R. & Bates, S. E. 2001. Inhibitor of histone deacetylation, depsipeptide (FR901228), in the treatment of peripheral and cutaneous T-cell lymphoma: a case report. *Blood,* 98**,** 2865-8.

Plumb, R. S., Johnson, K. A., Rainville, P., Smith, B. W., Wilson, I. D., Castro-Perez, J. M. & Nicholson, J. K. 2006. UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun Mass Spectrom,* 20**,** 1989-94.

Pouyani, T., Ronfard, V., Scott, P. G., Dodd, C. M., Ahmed, A., Gallo, R. L. & Parenteau, N. L. 2009. De novo synthesis of human dermis in vitro in the absence of a three-dimensional scaffold. *In Vitro Cell Dev Biol Anim,* 45**,** 430-41.

Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T. & Scrivens, J. H. 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry,* 261**,** 1-12.

Pugacheva, E. N., Jablonski, S. A., Hartman, T. R., Henske, E. P. & Golemis, E. A. 2007. HEF1-dependent Aurora A activation induces disassembly of the primary cilium. *Cell,* 129**,** 1351-63.

Rafii, R., Juarez, M. M., Albertson, T. E. & Chan, A. L. 2013. A review of current and novel therapies for idiopathic pulmonary fibrosis. *J Thorac Dis,* 5**,** 48-73.

Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Colby, T. V., Cordier, J. F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., Hansell, D. M., Johkoh, T., Kim, D. S., King, T. E., Kondoh, Y., Myers, J., Muller, N. L., Nicholson, A. G., Richeldi, L., Selman, M., Dudden, R. F., Griss, B. S., Protzko, S. L., Schunemann, H. J. & Comm, A. E. J. A. 2011. An Official ATS/ERSARS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *American Journal of Respiratory and Critical Care Medicine,* 183**,** 788-824.

Raghu, G., Freudenberger, T. D., Yang, S., Curtis, J. R., Spada, C., Hayes, J., Sillery, J. K., Pope, C. E. & Pellegrini, C. A. 2006. High prevalence of abnormal acid gastro-oesophageal reflux in idiopathic pulmonary fibrosis. *European Respiratory Journal,* 27**,** 136-142.

Raghu, G., Rochwerg, B., Zhang, Y., Garcia, C. A., Azuma, A., Behr, J., Brozek, J. L., Collard, H. R., Cunningham, W., Homma, S., Johkoh, T., Martinez, F. J., Myers, J., Protzko, S. L., Richeldi, L., Rind, D., Selman, M., Theodore, A., Wells, A. U., Hoogsteden, H., Schunemann, H. J., American Thoracic, S., European Respiratory, s., Japanese Respiratory, S. & Latin American

Thoracic, A. 2015. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *Am J Respir Crit Care Med,* 192**,** e3-19.

Ramanathan, R. 2008. Mass spectrometry in drug metabolism and pharmacokinetics. *In:* RAMANATHAN, R. (ed.). Hoboken, New Jersey: John Wiley & Sons.

Ramos, C., Montano, M., Garcia-Alvarez, J., Ruiz, V., Uhal, B. D., Selman, M. & Pardo, A. 2001. Fibroblasts from idiopathic pulmonary fibrosis and normal lungs differ in growth rate, apoptosis, and tissue inhibitor of metalloproteinases expression. *Am J Respir Cell Mol Biol,* 24**,** 591-8.

Rardin, M. J., Newman, J. C., Held, J. M., Cusack, M. P., Sorensen, D. J., Li, B., Schilling, B., Mooney, S. D., Kahn, C. R., Verdin, E. & Gibson, B. W. 2013. Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. *Proc Natl Acad Sci U S A,* 110**,** 6601-6.

Ravaglia, C., Tomassetti, S., Gurioli, C., Piciucchi, S., Dubini, A., Gurioli, C., Casoni, G. L., Romagnoli, M., Carloni, A., Tantalocco, P., Buccioli, M., Chilosi, M. & Poletti, V. 2014. Features and outcome of familial idiopathic pulmonary fibrosis. *Sarcoidosis Vasc Diffuse Lung Dis,* 31**,** 28-36.

Ricard-Blum, S. 2011. The Collagen Family. *Cold Spring Harbor Perspectives in Biology,* 3.

Richeldi, L. 2015. Idiopathic pulmonary fibrosis: moving forward. *BMC Med,* 13**,** 231.

Richeldi, L., Costabel, U., Selman, M., Kim, D. S., Hansell, D. M., Nicholson, A. G., Brown, K. K., Flaherty, K. R., Noble, P. W., Raghu, G., Brun, M., Gupta, A., Juhel, N., Kluglich, M. & du Bois, R. M. 2011. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med,* 365**,** 1079-87.

Richeldi, L., Cottin, V., Flaherty, K. R., Kolb, M., Inoue, Y., Raghu, G., Taniguchi, H., Hansell, D. M., Nicholson, A. G., Le Maulf, F., Stowasser, S. & Collard, H. R. 2014a. Design of the INPULSIS trials: two phase 3 trials of nintedanib in patients with idiopathic pulmonary fibrosis. *Respir Med,* 108**,** 1023-30.

Richeldi, L., du Bois, R. M., Raghu, G., Azuma, A., Brown, K. K., Costabel, U., Cottin, V., Flaherty, K. R., Hansell, D. M., Inoue, Y., Kim, D. S., Kolb, M., Nicholson, A. G., Noble, P. W., Selman, M., Taniguchi, H., Brun, M., Le Maulf, F., Girard, M., Stowasser, S., Schlenker-Herceg, R., Disse, B., Collard,

H. R. & Investigators, I. T. 2014b. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med,* 370**,** 2071-82.

Roberts, P. J. & Der, C. J. 2007. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene,* 26**,** 3291-310.

Rolfo, C., Raez, L. E., Bronte, G., Santos, E. S., Papadimitriou, K., Buffoni, L., van Meerbeeck, J. P. & Russo, A. 2013. BIBF 1120/ nintedanib : a new triple angiokinase inhibitor-directed therapy in patients with non-small cell lung cancer. *Expert Opin Investig Drugs,* 22**,** 1081-8.

Ropero, S. & Esteller, M. 2007. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol,* 1**,** 19-25.

Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics,* 3**,** 1154-1169.

Roth, G. J., Heckel, A., Colbatzky, F., Handschuh, S., Kley, J., Lehmann-Lintz, T., Lotz, R., Tontsch-Grunt, U., Walter, R. & Hilberg, F. 2009. Design, synthesis, and evaluation of indolinones as triple angiokinase inhibitors and the discovery of a highly specific 6-methoxycarbonyl-substituted indolinone (BIBF 1120). *J Med Chem,* 52**,** 4466-80.

Rottoli, P., Magi, B., Perari, M. G., Liberatori, S., Nikiforakis, N., Bargagli, E., Cianti, R., Bini, L. & Pallini, V. 2005. Cytokine profile and proteome analysis in bronchoalveolar lavage of patients with sarcoidosis, pulmonary fibrosis associated with systemic sclerosis and idiopathic pulmonary fibrosis. *Proteomics,* 5**,** 1423-30.

Saito, T., Kasamatsu, A., Ogawara, K., Miyamoto, I., Saito, K., Iyoda, M., Suzuki, T., Endo-Sakamoto, Y., Shiiba, M., Tanzawa, H. & Uzawa, K. 2015. Semaphorin7A Promotion of Tumoral Growth and Metastasis in Human Oral Cancer by Regulation of G1 Cell Cycle and Matrix Metalloproteases: Possible Contribution to Tumoral Angiogenesis. *PLoS One,* 10**,** e0137923.

Sakai, N. & Tager, A. M. 2013. Fibrosis of two: Epithelial cell-fibroblast interactions in pulmonary fibrosis. *Biochimica Et Biophysica Acta-Molecular Basis of Disease,* 1832**,** 911-921.

Sakamoto, K., Taniguchi, H., Kondoh, Y., Wakai, K., Kimura, T., Kataoka, K., Hashimoto, N., Nishiyama, O. & Hasegawa, Y. 2012. Acute exacerbation of IPF following diagnostic bronchoalveolar lavage procedures. *Respir Med,* 106**,** 436-42.

Saketkoo, L. A., Mittoo, S., Huscher, D., Khanna, D., Dellaripa, P. F., Distler, O., Flaherty, K. R., Frankel, S., Oddis, C. V., Denton, C. P., Fischer, A., Kowal-Bielecka, O. M., LeSage, D., Merkel, P. A., Phillips, K., Pittrow, D., Swigris, J., Antoniou, K., Baughman, R. P., Castelino, F. V., Christmann, R. B., Christopher-Stine, L., Collard, H. R., Cottin, V., Danoff, S., Highland, K. B., Hummers, L., Shah, A. A., Kim, D. S., Lynch, D. A., Miller, F. W., Proudman, S. M., Richeldi, L., Ryu, J. H., Sandorfi, N., Sarver, C., Wells, A. U., Strand, V., Matteson, E. L., Brown, K. K., Seibold, J. R. & Group, C.-I. S. I. 2014. Connective tissue disease related interstitial lung diseases and idiopathic pulmonary fibrosis: provisional core sets of domains and instruments for use in clinical trials. *Thorax,* 69**,** 428-36.

Saveliev, S., Bratz, M., Zubarev, R., Szapacs, M., Budamgunta, H. & Urh, M. 2013. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nat Meth,* 10.

Schaefer, C. J., Ruhrmund, D. W., Pan, L., Seiwert, S. D. & Kossen, K. 2011. Antifibrotic activities of pirfenidone in animal models. *Eur Respir Rev,* 20**,** 85-97.

Schmidt, C. & Urlaub, H. 2012. Absolute quantification of proteins using standard peptides and multiple reaction monitoring. *Methods Mol Biol,* 893**,** 249-65.

Schmidt, S. L., Tayob, N., Han, M. K., Zappala, C., Kervitsky, D., Murray, S., Wells, A. U., Brown, K. K., Martinez, F. J. & Flaherty, K. R. 2014. Predicting pulmonary fibrosis disease course from past trends in pulmonary function. *Chest,* 145**,** 579-85.

Schwartz, D. A., Merchant, R. K., Helmers, R. A., Gilbert, S. R., Dayton, C. S. & Hunninghake, G. W. 1991. The Influence of Cigarette-Smoking on Lung-Function in Patients with Idiopathic Pulmonary Fibrosis. *American Review of Respiratory Disease,* 144**,** 504-506.

Scott, J., Johnston, I. & Britton, J. 1990. What Causes Cryptogenic Fibrosing Alveolitis - a Case-Control Study of Environmental Exposure to Dust. *British Medical Journal,* 301**,** 1015-1017.

Seibold, M. A., Wise, A. L., Speer, M. C., Steele, M. P., Brown, K. K., Loyd, J. E., Fingerlin, T. E., Zhang, W., Gudmundsson, G., Groshong, S. D., Evans, C.

M., Garantziotis, S., Adler, K. B., Dickey, B. F., du Bois, R. M., Yang, I. V., Herron, A., Kervitsky, D., Talbert, J. L., Markin, C., Park, J., Crews, A. L., Slifer, S. H., Auerbach, S., Roy, M. G., Lin, J., Hennessy, C. E., Schwarz, M. I. & Schwartz, D. A. 2011. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med,* 364**,** 1503-12.

Selman, M., King, T. E. & Pardo, A. 2001. Idiopathic pulmonary fibrosis: Prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Annals of Internal Medicine,* 134**,** 136-151.

Selman, M., Ruiz, V., Cabrera, S., Segura, L., Ramirez, R., Barrios, R. & Pardo, A. 2000. TIMP-1, -2, -3, and -4 in idiopathic pulmonary fibrosis. A prevailing nondegradative lung microenvironment? *Am J Physiol Lung Cell Mol Physiol,* 279**,** L562-74.

Seymour, S. 2010. *Briefing information of the March 9, 2010 meeting of the Pulmonary-Allergy Drugs Advisory Committee*

(PADAC). In: SEYMOUR, S. (ed.). Beltsville, MD.

Shamis, Y., Hewitt, K. J., Bear, S. E., Alt-Holland, A., Qari, H., Margvelashvilli, M., Knight, E. B., Smith, A. & Garlick, J. A. 2012. iPSC-derived fibroblasts demonstrate augmented production and assembly of extracellular matrix proteins. *In Vitro Cell Dev Biol Anim,* 48**,** 112-22.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research,* 13**,** 2498-2504.

Shi-Wen, X., Rodriguez-Pascual, F., Lamas, S., Holmes, A., Howat, S., Pearson, J. D., Dashwood, M. R., du Bois, R. M., Denton, C. P., Black, C. M., Abraham, D. J. & Leask, A. 2006. Constitutive ALK5-independent c-Jun N-terminal kinase activation contributes to endothelin-1 overexpression in pulmonary fibrosis: evidence of an autocrine endothelin loop operating through the endothelin A and B receptors. *Mol Cell Biol,* 26**,** 5518-27.

Shliaha, P. V., Bond, N. J., Gatto, L. & Lilley, K. S. 2013. Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *J Proteome Res,* 12**,** 2323-39.

Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. C. & Geromanos, S. J. 2006. Absolute quantification of proteins by LCMSE - A virtue of parallel MS acquisition. *Molecular & Cellular Proteomics,* 5**,** 144-156.

Skalli, O., Pelte, M. F., Peclet, M. C., Gabbiani, G., Gugliotta, P., Bussolati, G., Ravazzola, M. & Orci, L. 1989. Alpha-smooth muscle actin, a differentiation marker of smooth muscle cells, is present in microfilamentous bundles of pericytes. *J Histochem Cytochem,* 37**,** 315-21.

Smith, K. T. & Workman, J. L. 2009. Introducing the acetylome. *Nat Biotechnol,* 27**,** 917-9.

Smithmyer, M. E., Sawicki, L. A. & Kloxin, A. M. 2014. Hydrogel scaffolds as in vitro models to study fibroblast activation in wound healing and disease. *Biomater Sci,* 2**,** 634-650.

Soc, A. T. 2000. Idiopathic pulmonary fibrosis: Diagnosis and treatment - International consensus statement. *American Journal of Respiratory and Critical Care Medicine,* 161**,** 646-664.

Sol, E. M., Wagner, S. A., Weinert, B. T., Kumar, A., Kim, H. S., Deng, C. X. & Choudhary, C. 2012. Proteomic Investigations of Lysine Acetylation Identify Diverse Substrates of Mitochondrial Deacetylase Sirt3. *Plos One,* 7.

Song, J. W., Hong, S. B., Lim, C. M., Koh, Y. & Kim, D. S. 2011. Acute exacerbation of idiopathic pulmonary fibrosis: incidence, risk factors and outcome. *European Respiratory Journal,* 37**,** 356-363.

Spange, S., Wagner, T., Heinzel, T. & Kramer, O. H. 2009. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *International Journal of Biochemistry & Cell Biology,* 41**,** 185-198.

Steele, M. P. & Schwartz, D. A. 2013. Molecular Mechanisms in Progressive Idiopathic Pulmonary Fibrosis. *Annual Review of Medicine, Vol 64,* 64**,** 265-276.

Steele, M. P., Speer, M. C., Loyd, J. E., Brown, K. K., Herron, A., Slifer, S. H., Burch, L. H., Wahidi, M. M., Phillips, J. A., 3rd, Sporn, T. A., McAdams, H. P., Schwarz, M. I. & Schwartz, D. A. 2005. Clinical and pathologic features of familial interstitial pneumonia. *Am J Respir Crit Care Med,* 172**,** 1146-52.

Strieter, R. M. & Mehrad, B. 2009. New Mechanisms of Pulmonary Fibrosis. *Chest,* 136**,** 1364-1370.

Struhl, K. 1998. Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev,* 12**,** 599-606.

Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one,* 6**,** e21800.

Takafuta, T., Saeki, M., Fujimoto, T. T., Fujimura, K. & Shapiro, S. S. 2003. A new member of the LIM protein family binds to filamin B and localizes at stress fibers. *J Biol Chem,* 278**,** 12175-81.

Takeda, Y., Tsujino, K., Kijima, T. & Kumanogoh, A. 2014. Efficacy and safety of pirfenidone for idiopathic pulmonary fibrosis. *Patient Prefer Adherence,* 8**,** 361-70.

Tang, G., Minemoto, Y., Dibling, B., Purcell, N. H., Li, Z., Karin, M. & Lin, A. 2001. Inhibition of JNK activation through NF-kappaB target genes. *Nature,* 414**,** 313-7.

Taskar, V. S. & Coultas, D. B. 2006. Is idiopathic pulmonary fibrosis an environmental disease? *Proc Am Thorac Soc,* 3**,** 293-8.

Thabut, G., Mal, H., Castier, Y., Groussard, O., Brugiere, O., Marrash-Chahla, R., Leseche, G. & Fournier, M. 2003. Survival benefit of lung transplantation for patients with idiopathic pulmonary fibrosis. *J Thorac Cardiovasc Surg,* 126**,** 469-75.

Thannickal, V. J. 2010. Aging, antagonistic pleiotropy and fibrotic disease. *International Journal of Biochemistry & Cell Biology,* 42**,** 1398-1400.

Thannickal, V. J., Lee, D. Y., White, E. S., Cui, Z., Larios, J. M., Chacon, R., Horowitz, J. C., Day, R. M. & Thomas, P. E. 2003. Myofibroblast differentiation by transforming growth factor-beta1 is dependent on cell adhesion and integrin signaling via focal adhesion kinase. *J Biol Chem,* 278**,** 12384-9.

ThomsonReuters. 2014. *MetaCore™ Data-mining and pathway analysis* [Online]. Available: http://thomsonreuters.com/metacore/ [Accessed 10/09/2104 2014].

Tibbitt, M. W. & Anseth, K. S. 2009. Hydrogels as extracellular matrix mimics for 3D cell culture. *Biotechnol Bioeng,* 103**,** 655-63.

Tobin, R. W., Pope, C. E., Pellegrini, C. A., Emond, M. J., Sillery, J. & Raghu, G. 1998. Increased prevalence of gastroesophageal reflux in patients with idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine,* 158**,** 1804-1808.

Todd, N. W., Luzina, I. G. & Atamas, S. P. 2012. Molecular and cellular mechanisms of pulmonary fibrosis. *Fibrogenesis Tissue Repair,* 5**,** 11.

Travis, W. D., Costabel, U., Hansell, D. M., King, T. E., Jr., Lynch, D. A., Nicholson, A. G., Ryerson, C. J., Ryu, J. H., Selman, M., Wells, A. U., Behr, J., Bouros, D., Brown, K. K., Colby, T. V., Collard, H. R., Cordeiro, C. R., Cottin, V., Crestani, B., Drent, M., Dudden, R. F., Egan, J., Flaherty, K., Hogaboam, C., Inoue, Y., Johkoh, T., Kim, D. S., Kitaichi, M., Loyd, J., Martinez, F. J., Myers, J., Protzko, S., Raghu, G., Richeldi, L., Sverzellati, N., Swigris, J., Valeyre, D. & Pneumonias, A. E. C. o. I. I. 2013. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med,* 188**,** 733-48.

Trinh, H. V., Grossmann, J., Gehrig, P., Roschitzki, B., Schlapbach, R., Greber, U. F. & Hemmi, S. 2013. iTRAQ-Based and Label-Free Proteomics Approaches for Studies of Human Adenovirus Infections. *Int J Proteomics,* 2013**,** 581862.

Tzouvelekis, A., Bonella, F. & Spagnolo, P. 2015. Update on therapeutic management of idiopathic pulmonary fibrosis. *Ther Clin Risk Manag,* 11**,** 359-70.

Ueda, H., Manda, T., Matsumoto, S., Mukumoto, S., Nishigaki, F., Kawamura, I. & Shimomura, K. 1994. FR901228, a novel antitumor bicyclic depsipeptide produced by Chromobacterium violaceum No. 968. III. Antitumor activities on experimental tumors in mice. *J Antibiot (Tokyo),* 47**,** 315-23.

Usunier, B., Benderitter, M., Tamarat, R. & Chapel, A. 2014. Management of fibrosis: the mesenchymal stromal cells breakthrough. *Stem Cells Int,* 2014**,** 340257.

Valdez, B. C., Brammer, J. E., Li, Y., Murray, D., Liu, Y., Hosing, C., Nieto, Y., Champlin, R. E. & Andersson, B. S. 2015. Romidepsin targets multiple survival signaling pathways in malignant T cells. *Blood Cancer J,* 5**,** e357.

Valeyre, D. 2011. Towards a better diagnosis of idiopathic pulmonary fibrosis. *Eur Respir Rev,* 20**,** 108-13.

van den Blink, B., Wijsenbeek, M. S. & Hoogsteden, H. C. 2010. Serum biomarkers in idiopathic pulmonary fibrosis. *Pulmonary Pharmacology & Therapeutics,* 23**,** 515-520.

van Rijn, A., Paulis, L., te Riet, J., Vasaturo, A., Reinieren-Beeren, I., van der Schaaf, A., Kuipers, A. J., Schulte, L. P., Jongbloets, B. C., Pasterkamp, R. J., Figdor, C. G., van Spriel, A. B. & Buschow, S. I. 2016. Semaphorin 7A Promotes Chemokine-Driven Dendritic Cell Migration. *J Immunol,* 196**,** 459-68.

Vancheri, C. 2012. Idiopathic pulmonary fibrosis: an altered fibroblast proliferation linked to cancer biology. *Proc Am Thorac Soc,* 9**,** 153-7.

Vancheri, C. 2013. Common pathways in idiopathic pulmonary fibrosis and cancer. *Eur Respir Rev,* 22**,** 265-72.

Vancheri, C., Failla, M., Crimi, N. & Raghu, G. 2010. Idiopathic pulmonary fibrosis: a disease with similarities and links to cancer biology. *European Respiratory Journal,* 35**,** 496-504.

VanderMolen, K. M., McCulloch, W., Pearce, C. J. & Oberlies, N. H. 2011. Romidepsin (Istodax, NSC 630176, FR901228, FK228, depsipeptide): a natural product recently approved for cutaneous T-cell lymphoma. *J Antibiot (Tokyo),* 64**,** 525-31.

Vij, R. & Noth, I. 2012. Peripheral blood biomarkers in idiopathic pulmonary fibrosis. *Translational Research,* 159**,** 218-227.

Walther, T. C. & Mann, M. 2010. Mass spectrometry-based proteomics in cell biology. *Journal of Cell Biology,* 190**,** 491-500.

Wang, Z., Chen, C., Finger, S. N., Kwajah, S., Jung, M., Schwarz, H., Swanson, N., Lareu, F. F. & Raghunath, M. 2009. Suberoylanilide hydroxamic acid: a potential epigenetic therapeutic agent for lung fibrosis? *Eur Respir J,* 34**,** 145-55.

Wani, P., Teleb, M., Wardi, M. & Salameh, H. J. 2015. Pirfenidone and Nintedanib for idiopathic pulmonary fibrosis: A new drug therapy. *Medical Science Review,* 2**,** 41-48.

Wattiez, R., Hermans, C., Cruyt, C., Bernard, A. & Falmagne, P. 2000. Human bronchoalveolar lavage fluid protein two-dimensional database: study of interstitial lung diseases. *Electrophoresis,* 21**,** 2703-12.

Whitmarsh, A. J. & Davis, R. J. 1996. Transcription factor AP-1 regulation by mitogen-activated protein kinase signal transduction pathways. *J Mol Med (Berl),* 74**,** 589-607.

Whittaker, S. J., Demierre, M. F., Kim, E. J., Rook, A. H., Lerner, A., Duvic, M., Scarisbrick, J., Reddy, S., Robak, T., Becker, J. C., Samtsov, A., McCulloch, W. & Kim, Y. H. 2010. Final results from a multicenter, international, pivotal study of romidepsin in refractory cutaneous T-cell lymphoma. *J Clin Oncol,* 28**,** 4485-91.

Wight, T. N. & Potter-Perigo, S. 2011. The extracellular matrix: an active or passive player in fibrosis? *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 301**,** G950-G955.

Willis, B. C., Liebler, J. M., Luby-Phelps, K., Nicholson, A. G., Crandall, E. D., du Bois, R. M. & Borok, Z. 2005. Induction of epithelial-mesenchymal transition in alveolar epithelial cells by transforming growth factor-beta1: potential role in idiopathic pulmonary fibrosis. *Am J Pathol,* 166**,** 1321-32.

Wilm, M. 2009. Quantitative proteomics in biological research. *Proteomics,* 9**,** 4590-4605.

Wilm, M. 2011. Principles of Electrospray Ionization. *Molecular & Cellular Proteomics,* 10.

Wollin, L., Wex, E., Pautsch, A., Schnapp, G., Hostettler, K. E., Stowasser, S. & Kolb, M. 2015. Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. *Eur Respir J,* 45**,** 1434-45.

Woodcock, H. V., Molyneaux, P. L. & Maher, T. M. 2013. Reducing lung function decline in patients with idiopathic pulmonary fibrosis: potential of nintedanib. *Drug Des Devel Ther,* 7**,** 503-10.

Wullaert, A., Heyninck, K. & Beyaert, R. 2006. Mechanisms of crosstalk between TNF-induced NF-kappaB and JNK activation in hepatocytes. *Biochem Pharmacol,* 72**,** 1090-101.

Wynn, T. A. & Ramalingam, T. R. 2012. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nat Med,* 18**,** 1028-40.

Xaubet, A., Serrano-Mollar, A. & Ancochea, J. 2014. Pirfenidone for the treatment of idiopathic pulmonary fibrosis. *Expert Opinion on Pharmacotherapy,* 15**,** 275-281.

Zhang, H. Y. & Phan, S. H. 1999. Inhibition of myofibroblast apoptosis by transforming growth factor beta(1). *American Journal of Respiratory Cell and Molecular Biology,* 21**,** 658-665.

Zhang, K., Zheng, S. Z., Yang, J. S., Chen, Y. & Cheng, Z. Y. 2013a. Comprehensive Profiling of Protein Lysine Acetylation in Escherichia coli. *Journal of Proteome Research,* 12**,** 844-851.

Zhang, X. Y., Liu, H., Hock, T., Thannickal, V. J. & Sanders, Y. Y. 2013b. Histone Deacetylase Inhibition Downregulates Collagen 3A1 in Fibrotic Lung Fibroblasts. *International Journal of Molecular Sciences,* 14**,** 19605-19617.

Zhang, Y., Feng, X. H. & Derynck, R. 1998. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature,* 394**,** 909-913.

Zhang, Y. & Kaminski, N. 2012. Biomarkers in idiopathic pulmonary fibrosis. *Current opinion in pulmonary medicine,* 18**,** 441-446.

Zhang, Y., Noth, I., Garcia, J. G. & Kaminski, N. 2011. A variant in the promoter of MUC5B and idiopathic pulmonary fibrosis. *N Engl J Med,* 364**,** 1576-7.

Zhang, Y. E. 2009. Non-Smad pathways in TGF-beta signaling. *Cell Res,* 19**,** 128-39.

Zuo, F., Kaminski, N., Eugui, E., Allard, J., Yakhini, Z., Ben-Dor, A., Lollini, L., Morris, D., Kim, Y., DeLustro, B., Sheppard, D., Pardo, A., Selman, M. & Heller, R. A. 2002. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A,* 99**,** 6292-7.

Appendices

Appendix A - Appendix for Chapter 2

Appendix A.1 Data normalisation

The data was normalised using a "top 100" in order to account for differences in LC loading. Below is an example dataset containing expression values (the number of femtomoles measured) for five proteins (A-E) across 6 samples from an experiment; 3 control replicates and 3 test replicates. Proteins A, B and C were identified in all six samples; D and E contain missing values.

Raw data:	Protein	Control 1	Control 2	Control 3	Test 1	Test 2	Test 3
	A	364.0582	7576.885	519.7226	902.2421	4495.058	805.2645
	B	537.731	14304.32	789.5204	1361.345	6181.771	1349.161
	C	711.7483	34786.18	1111.084	2123.86	8805.721	2384.775
	D		8175.074	1052.094	2347.911		1038.953
	E	289.8687	2735.387	948.7108		1030.17	112.2985

Per replicate, each expression value in the dataset was divided by the sum of the femtomoles measured just from proteins A-C as these proteins, containing no missing data, were considered the most stable and abundant and thus the most reliable to use for normalisation.

Sum of A-C:		1613.538	56667.39	2420.327	4387.447	19482.55	4539.201
Each value divided by the sum:	A	0.225627	0.133708	0.214732	0.205642	0.230722	0.177402
	B	0.333262	0.252426	0.326204	0.310282	0.317298	0.297224
	C	0.44111	0.613866	0.459064	0.484076	0.45198	0.525373
	D		0.144264	0.434691	0.535143		0.228885
	E	0.179648	0.048271	0.391976		0.052877	0.02474

Each expression value was multiplied by 1,000 to produce simpler values to manipulate. The data was then logged to base 2.

Each value multiplied by 1,000:	A	225.6274	133.708	214.7324	205.6417	230.7223	177.4023
	B	333.2622	252.426	326.204	310.2818	317.2978	297.2244
	C	441.1105	613.866	459.0636	484.0765	451.9799	525.3734
	D		144.2642	434.6908	535.1429		228.8846
	E	179.6479	48.27092	391.9763		52.87655	24.73971

Each value logged to base 2:	A	7.817798	7.062942	7.746396	7.683989	7.850013	7.470881
	B	8.380514	7.979717	8.349631	8.277435	8.309694	8.215409
	C	8.784996	9.26178	8.84255	8.919091	8.820115	9.037199
	D		7.172569	8.763846	9.06378		7.838476
	E	7.489029	5.593083	8.614623		5.724556	4.628757

Appendix B - Appendix for Chapter 3

Appendix B.1 List of differentially expressed proteins between TGF-β_1-treated fibroblasts and control from the intracellular proteome analysis.

Differentially expressed proteins between TGF-β_1-treated and control fibroblasts

Protein accession	**Gene symbol**	**Fold-change after treatment compared to control**	**T-Test *p*-value**
O00231	PSMD11	-1.18	0.0004
P11216	PYGB	-1.18	0.0007
Q96RQ1	ERGIC2	-2.90	0.0010
P07197	NEFM	-2.60	0.0013
Q13200	PSMD2	1.25	0.0043
Q9H7Z7	PTGES2	1.41	0.0067
P78357	CNTNAP1	-2.03	0.0105
Q15257	PPP2R4	-4.41	0.0108
Q9BW27	NUP85	1.88	0.0110
P84077	ARF1	-1.81	0.0115
P23396	RPS3	1.21	0.0120
Q96BM9	ARL8A	4.88	0.0127
Q08209	PPP3CA	-1.76	0.0142
P48449	LSS	1.71	0.0151
Q12907	LMAN2	1.47	0.0161
Q6IAA8	LAMTOR1	1.47	0.0169
Q08170	SRSF4	-5.53	0.0169
Q9UJX4	ANAPC5	2.19	0.0181
Q15599	SLC9A3R2	-1.32	0.0182
P62191	PSMC1	-1.17	0.0184
Q7L576	CYFIP1	-1.22	0.0185
Q13619	CUL4A	1.41	0.0192
O75874	IDH1	-1.61	0.0194
Q86YZ3	HRNR	-1.44	0.0197
O75369	FLNB	1.40	0.0198
Q7KZF4	SND1	-1.55	0.0212

Q08379	GOLGA2	1.24	0.0212
Q9HCE1	MOV10	-1.72	0.0223
O75762	TRPA1	-3.02	0.0224
P14625	HSP90B1	1.26	0.0225
P50552	VASP	1.14	0.0226
Q16222	UAP1	1.32	0.0227
P01116	KRAS	2.20	0.0252
P17931	LGALS3	-1.31	0.0257
Q96T37	RBM15	1.29	0.0263
P35222	CTNNB1	1.66	0.0266
Q12769	NUP160	1.31	0.0283
Q96TA1	FAM129B	1.18	0.0287
P07384	CAPN1	-1.39	0.0295
P08238	HSP90AB1	-1.96	0.0298
P35221	CTNNA1	2.03	0.0314
Q99497	PARK7	5.99	0.0319
P50995	ANXA11	-1.31	0.0324
O75390	CS	1.10	0.0324
P30153	PPP2R1A	1.20	0.0326
Q13671	RIN1	-2.15	0.0337
Q16881	TXNRD1	-1.33	0.0339
Q8N752	CSNK1A1L	-1.47	0.0361
O15460	P4HA2	1.44	0.0366
P42704	LRPPRC	1.49	0.0384
P14136	GFAP	-2.65	0.0387
P14618	PKM	-1.29	0.0394
Q8TDD1	DDX54	1.29	0.0394
Q01518	CAP1	1.80	0.0399
Q71U36	TUBA1A	36.07	0.0411
Q15435	PPP1R7	1.96	0.0441
Q15813	TBCE	1.65	0.0450
P49588	AARS	1.43	0.0456
Q9UBF8	PI4KB	2.23	0.0461
P07942	LAMB1	-1.41	0.0469
P82650	MRPS22	1.31	0.0483
P14678	SNRPB	2.02	0.0488
Q969V3	NCLN	-1.50	0.0494

Appendix B.2 List of proteins expressed only in TGF-β_1-treated fibroblasts and only in control in cellular proteome data.

Proteins in TGF-β_1-treated fibroblasts only			**Protein in control fibroblasts only**		
Protein accession	**Gene symbol**	**Log2 expression**	**Protein accession**	**Gene symbol**	**Log2 expression**
O00401	WASL	5.88	Q9ULX3	NOB1	-2.49
Q8WXX0	DNAH7	5.36	O43790	KRT86	-3.07
P48723	HSPA13	4.94	Q9NUP9	LIN7C	-3.29
Q9NY61	AATF	4.75	P26022	PTX3	-3.38
Q9Y2H6	FNDC3A	4.71	Q14197	ICT1	-3.47
Q9UHN6	TMEM2	4.61	O00139	KIF2A	-3.48
Q2VIQ3	KIF4B	4.53	P48741	HSPA7	-4.12
Q9HCC0	MCCC2	4.42			
P28340	POLD1	4.28			
P02458	COL2A1	4.28			
Q9Y5V3	MAGED1	4.28			
Q9Y5A9	YTHDF2	4.24			
Q6UB35	MTHFD1L	4.19			
O60306	AQR	4.10			
Q8N5N7	MRPL50	4.07			
Q15155	NOMO1	4.07			
Q02241	KIF23	4.05			
Q92542	NCSTN	4.03			
Q8TD19	NEK9	4.01			
Q9BQ70	TCF25	3.86			
Q92947	GCDH	3.84			
Q8WXX5	DNAJC9	3.74			
Q92925	SMARCD2	3.73			
O75494	SRSF10	3.72			
Q53GS7	GLE1	3.64			
P26440	IVD	3.59			
Q8NEM2	SHCBP1	3.54			
Q6P9B9	INTS5	3.51			
Q8TDX7	NEK7	3.45			
P18858	LIG1	3.40			
Q00403	GTF2B	3.38			
Q9BRA0	NAA38	3.38			
Q9NVU7	SDAD1	3.37			
Q9BYJ9	YTHDF1	3.35			
Q92620	DHX38	3.34			
Q9NRX4	PHPT1	3.33			
Q13185	CBX3	3.28			
Q9Y2R9	MRPS7	3.23			
Q8N0X7	SPG20	3.21			

Q9H330	TMEM245	3.21	
Q9UKG1	APPL1	3.13	
P29279	CTGF	3.11	
Q96BP3	PPWD1	3.07	
Q96EK9	KTI12	3.03	
Q9NVP2	ASF1B	3.01	
Q13636	RAB31	3.00	
O75818	RPP40	2.98	
Q99081	TCF12	2.96	
Q8WUX9	CHMP7	2.94	
Q9BTE7	DCUN1D5	2.84	
P60510	PPP4C	2.82	
Q9NXH9	TRMT1	2.73	
Q9UKU9	ANGPTL2	2.66	
O15013	ARHGEF10	2.57	
Q7Z4W1	DCXR	2.50	
O94768	STK17B	2.49	
P49366	DHPS	2.40	
Q9BV44	THUMPD3	2.22	
Q9UJC5	SH3BGRL2	2.11	
Q9BW83	IFT27	2.10	
Q9NQ50	MRPL40	2.01	
Q9Y343	SNX24	1.96	
P15927	RPA2	1.95	
Q8NGI1	OR56B2P	1.90	
Q9NZA1	CLIC5	1.74	
Q9NR22	PRMT8	1.54	
P48454	PPP3CC	1.32	

Appendix B.3 List of differentially expressed proteins between TGF-β_1-treated fibroblasts and control from the secretome analysis.

Differentially expressed proteins between TGF-β_1-treated fibroblasts and control			
Protein accession	**Gene symbol**	**Fold-change after treatment compared to control**	**T-Test *p*-value**
P02751	FINC	2.01	0.0022
P07585	PGS2	-1.59	0.0024
Q16270	IBP7	1.58	0.0035
P01034	CYTC	1.32	0.0040
P08253	MMP2	2.15	0.0111
Q15582	BGH3	2.71	0.0132
P151884	LUM	-1.60	0.0353
O00391	QSOX1	-1.20	0.0483

Appendix B.4 List of proteins expressed only from TGF-β_1-treated fibroblasts and only in control in the secretome data.

Proteins from TGF-β_1-treated fibroblasts only			**Proteins from control fibroblasts only**		
Protein accession	**Gene symbol**	**Log2 expression**	**Protein accession**	**Gene symbol**	**Log2 expression**
P01023	A2MG	5.34			
P17936	IBP3	4.62			
P07093	GDN	4.03			
P01024	CO3	3.79			
Q92626	PXDN	3.43			
P08476	INHBA	2.33			

Appendix C - Appendix for Chapter 4

Appendix C.1 List of differentially expressed proteins between FK228-treated fibroblasts and untreated IPF fibroblasts in the IPF +/- FK228 dataset.

Proteins with differential expression between untreated and FK228-treated IPF fibroblasts			
Protein accession	**Protein name**	**Fold-change FK228-treated: untreated**	**T-Test *p*-value**
P09104	Gamma-enolase	10.094	0.029
O00469	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 2	5.807	0.000
Q9BYN0	Sulfiredoxin-1	4.357	0.006
Q9NZU5	LIM and cysteine-rich domains protein 1	3.489	0.024
Q6ZMR3	L-lactate dehydrogenase A-like 6A	3.191	0.026
P48147	Prolyl endopeptidase	3.185	0.007
Q9NYL9	Tropomodulin-3	3.116	0.010
Q9Y6M1	Insulin-like growth factor 2 mRNA-binding protein 2	2.875	0.029
P62140	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	2.829	0.020
O15427	Monocarboxylate transporter 4	2.772	0.013
P12277	Creatine kinase B-type	2.626	0.045
P04179	Superoxide dismutase [Mn]_ mitochondrial	2.530	0.022
P37235	Hippocalcin-like protein 1	2.457	0.042
P36873	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit	2.374	0.031

P05120	Plasminogen activator inhibitor 2	2.294	0.006
Q9Y2D5	A-kinase anchor protein 2	2.276	0.024
Q96AQ6	Pre-B-cell leukemia transcription factor-interacting protein 1	2.228	0.021
P05121	Plasminogen activator inhibitor 1	2.222	0.035
P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex_ mitochondrial	2.092	0.031
P42785	Lysosomal Pro-X carboxypeptidase	1.929	0.012
Q6WCQ1	Myosin phosphatase Rho-interacting protein	1.904	0.031
Q14315	Filamin-C	1.899	0.011
P07355	Annexin A2	1.834	0.021
Q92896	Golgi apparatus protein 1	1.812	0.009
O75369	Filamin-B	1.780	0.031
P61088	Ubiquitin-conjugating enzyme E2 N	1.764	0.016
P05455	Lupus La protein	1.756	0.011
O94905	Erlin-2	1.739	0.008
Q14914	Prostaglandin reductase 1	1.730	0.019
P43686	26S protease regulatory subunit 6B	1.728	0.047
P04899	Guanine nucleotide-binding protein G(i) subunit alpha-2	1.694	0.030
P40939	Trifunctional enzyme subunit alpha_ mitochondrial	1.674	0.022
P07951	Tropomyosin beta chain	1.672	0.037
Q8WX93	Palladin	1.638	0.034
Q05682	Caldesmon	1.628	0.021
Q99614	Tetratricopeptide repeat protein 1	1.619	0.012
P09936	Ubiquitin carboxyl-terminal hydrolase isozyme L1	1.619	0.004
P15559	NAD(P)H dehydrogenase [quinone] 1	1.594	0.038
P06396	Gelsolin	1.586	0.041
P23528	Cofilin-1	1.556	0.021
P46459	Vesicle-fusing ATPase	1.510	0.039
P08238	Heat shock protein HSP 90-beta	1.492	0.034
P11021	78 kDa glucose-regulated protein	1.486	0.022
P60953	Cell division control protein 42 homolog	1.442	0.019
P21980	Protein-glutamine gamma-glutamyltransferase 2	1.435	0.019
P61026	Ras-related protein Rab-10	1.380	0.003
P07237	Protein disulfide-isomerase	1.335	0.019
Q15363	Transmembrane emp24 domain-containing protein 2	1.297	0.030

Q15084	Protein disulfide-isomerase A6	1.288	0.034
Q6IBS0	Twinfilin-2	1.242	0.040
Q14011	Cold-inducible RNA-binding protein	1.227	0.043
O43776	Asparagine--tRNA ligase_ cytoplasmic	1.191	0.023
P04083	Annexin A1	1.180	0.034
P46734	Dual specificity mitogen-activated protein kinase kinase 3	-1.006	0.050
O00231	26S proteasome non-ATPase regulatory subunit 11	-1.100	0.027
P00352	Retinal dehydrogenase 1	-1.117	0.046
P48444	Coatomer subunit delta	-1.172	0.050
P05388	60S acidic ribosomal protein P0	-1.219	0.003
Q99685	Monoglyceride lipase	-1.229	0.047
P14649	Myosin light chain 6B	-1.237	0.029
P78527	DNA-dependent protein kinase catalytic subunit	-1.251	0.045
O00410	Importin-5	-1.317	0.048
Q00688	Peptidyl-prolyl cis-trans isomerase FKBP3	-1.326	0.009
P61978	Heterogeneous nuclear ribonucleoprotein K	-1.326	0.029
Q14974	Importin subunit beta-1	-1.357	0.019
P04844	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	-1.358	0.011
P48556	26S proteasome non-ATPase regulatory subunit 8	-1.382	0.014
O75533	Splicing factor 3B subunit 1	-1.382	0.003
Q00839	Heterogeneous nuclear ribonucleoprotein U	-1.423	0.016
P62258	14-3-3 protein epsilon	-1.427	0.029
P54578	Ubiquitin carboxyl-terminal hydrolase 14	-1.429	0.027
Q9NR12	PDZ and LIM domain protein 7	-1.429	0.011
O95340	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2	-1.454	0.034
P14866	Heterogeneous nuclear ribonucleoprotein L	-1.462	0.003
Q9NPQ8	Synembryn-A	-1.478	0.001
Q9BUF5	Tubulin beta-6 chain	-1.496	0.034
Q99584	Protein S100-A13	-1.529	0.050
P35080	Profilin-2	-1.554	0.039
P62701	40S ribosomal protein S4_ X isoform	-1.572	0.039
O15143	Actin-related protein 2/3 complex subunit 1B	-1.617	0.035
P62280	40S ribosomal protein S11	-1.626	0.032
Q92841	Probable ATP-dependent RNA helicase DDX17	-1.707	0.032
P35237	Serpin B6	-1.709	0.040

O60493	Sorting nexin-3	-1.717	0.009
P46779	60S ribosomal protein L28	-1.736	0.006
P31942	Heterogeneous nuclear ribonucleoprotein H3	-1.742	0.007
P40261	Nicotinamide N-methyltransferase	-1.744	0.022
P35749	Myosin-11	-1.821	0.044
Q13838	Spliceosome RNA helicase DDX39B	-1.830	0.045
P36578	60S ribosomal protein L4	-1.863	0.035
P17844	Probable ATP-dependent RNA helicase DDX5	-1.879	0.031
P43243	Matrin-3	-1.924	0.032
Q13151	Heterogeneous nuclear ribonucleoprotein A0	-1.965	0.025
O43399	Tumor protein D54	-2.086	0.010
P38159	RNA-binding motif protein_ X chromosome	-2.158	0.010
P62879	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	-2.225	0.044
P52788	Spermine synthase	-2.229	0.022
P26583	High mobility group protein B2	-2.241	0.030
P62269	40S ribosomal protein S18	-2.313	0.014
P27635	60S ribosomal protein L10	-2.330	0.019
P08779	Keratin_ type I cytoskeletal 16	-2.369	0.000
Q96AG4	Leucine-rich repeat-containing protein 59	-2.556	0.019
P26447	Protein S100-A4	-2.849	0.044
P09429	High mobility group protein B1	-3.825	0.014
P20336	Ras-related protein Rab-3A	-8.421	0.039

Appendix C.2 List of proteins expressed only in untreated IPF fibroblasts in the IPF +/- FK228 dataset.

Proteins identified in untreated IPF fibroblasts only		
Protein accession	**Protein name**	**Log2 expression**
Q9UHD9	Ubiquilin-2	4.55
Q8TEA8	D-tyrosyl-tRNA(Tyr) deacylase 1	4.50
Q9BVG4	UPF0368 protein Cxorf26	2.78
Q8IYT4	Katanin p60 ATPase-containing subunit A-like 2	5.91
Q9UH62	Armadillo repeat-containing X-linked protein 3	3.35
Q9NRX1	RNA-binding protein PNO1	3.23
P49736	DNA replication licensing factor MCM2	5.18

Q13363	C-terminal-binding protein 1 CTBP1	4.38
P40222	Alpha-taxilin TXLNA	4.43
Q8N684	Cleavage and polyadenylation specificity factor subunit 7	3.89
Q96C36	Pyrroline-5-carboxylate reductase 2	2.67
Q08378	Golgin subfamily A member 3	5.93
Q9HAU4	E3 ubiquitin-protein ligase SMURF2	4.33
Q15042	Rab3 GTPase-activating protein catalytic subunit	3.59
O43617	Trafficking protein particle complex subunit 3	3.94
Q6UB99	Ankyrin repeat domain-containing protein 11	4.72
Q86W28	NACHT_ LRR and PYD domains-containing protein 8	2.55
O15078	Centrosomal protein of 290 kDa	6.08
P31350	Ribonucleoside-diphosphate reductase subunit M2	4.42
P55145	Mesencephalic astrocyte-derived neurotrophic factor	4.83
P52292	Importin subunit alpha-2	5.52
O00541	Pescadillo homolog	4.48
Q9Y2T7	Y-box-binding protein 2	2.77
Q6UVK1	Chondroitin sulfate proteoglycan 4	5.30
P49321	Nuclear autoantigenic sperm protein	5.23
Q14684	Ribosomal RNA processing protein 1 homolog B	3.34
Q9Y2L1	Exosome complex exonuclease RRP44	4.63
Q6P158	Putative ATP-dependent RNA helicase DHX57	4.71
Q9Y5L0	Transportin-3	4.84
P26358	DNA (cytosine-5)-methyltransferase 1	4.91
Q9P258	Protein RCC2	3.57
P42226	Signal transducer and activator of transcription 6	5.10
Q9BQ67	Glutamate-rich WD repeat-containing protein 1	3.99
Q9BWF3	RNA-binding protein 4	3.90
Q8IVM0	Coiled-coil domain-containing protein 50	4.94
P11234	Ras-related protein Ral-B	0.93
Q9UMY4	Sorting nexin-12	3.54
Q9H0X9	Oxysterol-binding protein-related protein 5	3.90
Q9Y314	Nitric oxide synthase-interacting protein	4.89
Q9Y385	Ubiquitin-conjugating enzyme E2 J1	2.67
P50281	Matrix metalloproteinase-14	5.01
Q9NYF8	Bcl-2-associated transcription factor 1	4.74
Q2NL82	Pre-rRNA-processing protein TSR1 homolog	3.79

Q9Y3B4	Pre-mRNA branch site protein p14	2.84
P48960	CD97 antigen	4.98
Q92544	Transmembrane 9 superfamily member 4	4.20
P30530	Tyrosine-protein kinase receptor UFO	4.05
Q86YQ8	Copine-8	1.54
Q9BPX3	Condensin complex subunit 3	4.54
Q8NBX0	Saccharopine dehydrogenase-like oxidoreductase	4.01
P00966	Argininosuccinate synthase	3.40
Q9ULE3	DENN domain-containing protein 2A	3.92
Q6AHZ1	Zinc finger protein 518A	3.92
O95478	Ribosome biogenesis protein NSA2 homolog	2.61
Q7LG56	Ribonucleoside-diphosphate reductase subunit M2 B	2.68
O75526	RNA-binding motif protein_ X-linked-like-2	1.22
P51784	Ubiquitin carboxyl-terminal hydrolase 11	3.40
P24390	ER lumen protein retaining receptor 1	4.42
Q9NVN3	Synembryn-B	1.19
P04818	Thymidylate synthase	4.55
P51451	Tyrosine-protein kinase Blk	3.13
O60573	Eukaryotic translation initiation factor 4E type 2	3.08
P29353	SHC-transforming protein 1	3.20
Q8WZA9	Immunity-related GTPase family Q protein	3.49
Q8NDA8	HEAT repeat-containing protein 7A	4.80
Q9H425	Uncharacterized protein C1orf198	2.54
Q9HCL0	Protocadherin-18	3.41
Q9UNF1	Melanoma-associated antigen D2	4.16
Q5SRE5	Nucleoporin NUP188 homolog	4.13
Q9UBT6	DNA polymerase kappa	3.04
O60502	Bifunctional protein NCOAT	3.94
P46937	Yorkie homolog	5.03
Q96E11	Ribosome-recycling factor_ mitochondrial	4.02
P07948	Tyrosine-protein kinase Lyn	2.27
P48668	Keratin_ type II cytoskeletal 6C	5.22
A6NE52	WD repeat-containing protein KIAA1875	3.07
P24928	DNA-directed RNA polymerase II subunit RPB1	3.74
Q8N9V7	Testis- and ovary-specific PAZ domain-containing protein 1	5.29
Q4W5G0	Tigger transposable element-derived protein 2	3.93

Q5TBB1	Ribonuclease H2 subunit B	6.56
Q9P2K1	Coiled-coil and C2 domain-containing protein 2A	6.33
O75326	Semaphorin-7A	3.94
Q9NP61	ADP-ribosylation factor GTPase-activating protein 3	4.86
Q9NQA5	Transient receptor potential cation channel subfamily V member 5	2.90
P98161	Polycystin-1	4.63
Q9HBG6	Intraflagellar transport protein 122 homolog	5.06
Q8NFJ9	Bardet-Biedl syndrome 1 protein	3.37
Q9UHC1	DNA mismatch repair protein Mlh3	5.74
Q7Z7M9	Polypeptide N-acetylgalactosaminyltransferase 5	4.24
Q9NRM1	Enamelin	3.10
Q9NZN5	Rho guanine nucleotide exchange factor 12	3.92
Q9NX02	NACHT_ LRR and PYD domains-containing protein 2	3.00
Q9Y2W2	WW domain-binding protein 11	3.43
P18858	DNA ligase 1	4.46
O75339	Cartilage intermediate layer protein 1	2.45
Q5VZ89	DENN domain-containing protein 4C	3.98
Q9UPX6	UPF0258 protein KIAA1024	2.22
Q7Z739	YTH domain family protein 3	2.87
Q502W7	Coiled-coil domain-containing protein 38	5.11
A9Z1Z3	Fer-1-like protein 4	2.02
P02461	Collagen alpha-1(III) chain	4.56
Q8WUP2	Filamin-binding LIM protein 1	3.61
Q5T7N2	LINE-1 type transposase domain-containing protein 1	2.58
Q9NSK0	Kinesin light chain 4	2.75

Appendix C.3 List of proteins expressed only in FK228-treated IPF fibroblasts in the IPF +/- FK228 dataset.

Proteins identified in IPF fibroblasts after FK228 treatment only		
Protein accession	**Protein name**	**Log2 expression**
Q9UI42	Carboxypeptidase A4	5.54
O43423	Acidic leucine-rich nuclear phosphoprotein 32 family member C	4.20
Q7L5N7	Lysophosphatidylcholine acyltransferase 2	4.64
O95671	N-acetylserotonin O-methyltransferase-like protein	4.71

Q8NFJ5	Retinoic acid-induced protein 3	5.98
Q7RTV0	PHD finger-like domain-containing protein 5A	2.89
P17900	Ganglioside GM2 activator	3.63

Appendix C.4 List of differentially expressed proteins between FK228-treated fibroblasts and untreated non-IPF fibroblasts in the non-IPF +/- FK228 dataset.

Proteins with differential expression between untreated and FK228-treated non-IPF fibroblasts			
Protein accession	**Protein name**	**Fold-change FK228-treated: untreated**	**T-Test *p*-value**
Q58FF3	Putative endoplasmin-like protein	8.204	0.004
Q6S8J3	POTE ankyrin domain family member E	4.331	0.043
O60282	Kinesin heavy chain isoform 5C	4.114	0.043
P30043	Flavin reductase (NADPH)	3.193	0.034
Q01130	Serine/arginine-rich splicing factor 2	2.276	0.023
Q92597	Protein NDRG1	2.239	0.004
Q9Y2D5	A-kinase anchor protein 2	1.897	0.012
P21926	CD9 antigen	1.763	0.009
O75369	Filamin-B	1.759	0.002
O00151	PDZ and LIM domain protein 1	1.748	0.018
Q96HC4	PDZ and LIM domain protein 5	1.736	0.013
Q14914	Prostaglandin reductase 1	1.636	0.011
Q9Y295	Developmentally-regulated GTP-binding protein 1	1.631	0.021
P53999	Activated RNA polymerase II transcriptional coactivator p15	1.618	0.034
P13674	Prolyl 4-hydroxylase subunit alpha-1	1.590	0.030
P28288	ATP-binding cassette sub-family D member 3	1.550	0.011
O95394	Phosphoacetylglucosamine mutase	1.483	0.019
P38606	V-type proton ATPase catalytic subunit A	1.471	0.027
O94851	Protein-methionine sulfoxide oxidase MICAL2	1.442	0.042
P53004	Biliverdin reductase A	1.442	0.049
O43795	Unconventional myosin-Ib	1.424	0.010
P51911	Calponin-1	1.419	0.015
O00154	Cytosolic acyl coenzyme A thioester hydrolase	1.383	0.014
P11413	Glucose-6-phosphate 1-dehydrogenase	1.380	0.002
Q96CS3	FAS-associated factor 2	1.374	0.012
P48739	Phosphatidylinositol transfer protein beta isoform	1.326	0.029
P16152	Carbonyl reductase [NADPH] 1	1.326	0.007
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1	1.324	0.048
O43865	Putative adenosylhomocysteinase 2	1.302	0.034
P49257	Protein ERGIC-53	1.284	0.015

P11940	Polyadenylate-binding protein 1	1.275	0.017
Q06830	Peroxiredoxin-1	1.257	0.029
Q14444	Caprin-1	1.171	0.006
Q9BS40	Latexin	-1.081	0.027
Q14697	Neutral alpha-glucosidase AB	-1.087	0.002
Q7L1Q6	Basic leucine zipper and W2 domain-containing protein 1	-1.124	0.033
P53618	Coatomer subunit beta	-1.145	0.003
P55072	Transitional endoplasmic reticulum ATPase	-1.183	0.025
P12956	X-ray repair cross-complementing protein 6	-1.215	0.004
P68036	Ubiquitin-conjugating enzyme E2 L3	-1.218	0.027
Q96QK1	Vacuolar protein sorting-associated protein 35	-1.219	0.013
Q00839	Heterogeneous nuclear ribonucleoprotein U	-1.234	0.005
Q13895	Bystin	-1.243	0.019
O60664	Perilipin-3	-1.257	0.021
O00303	Eukaryotic translation initiation factor 3 subunit F	-1.259	0.018
O14979	Heterogeneous nuclear ribonucleoprotein D-like	-1.260	0.008
O14818	Proteasome subunit alpha type-7	-1.279	0.012
P46821	Microtubule-associated protein 1B	-1.291	0.049
Q9H7B2	Ribosome production factor 2 homolog RPF2 PE=1 SV=2	-1.313	0.003
Q15366	Poly(rC)-binding protein 2	-1.323	0.007
Q9UN86	Ras GTPase-activating protein-binding protein 2	-1.336	0.019
P00491	Purine nucleoside phosphorylase	-1.393	0.027
Q9UKD2	mRNA turnover protein 4 homolog	-1.404	0.023
Q01085	Nucleolysin TIAR	-1.421	0.041
Q9BUF5	Tubulin beta-6 chain	-1.448	0.001
Q07960	Rho GTPase-activating protein 1	-1.473	0.006
Q5SSJ5	Heterochromatin protein 1-binding protein 3	-1.508	0.000
O75521	Enoyl-CoA delta isomerase 2_ mitochondrial	-1.540	0.004
P40123	Adenylyl cyclase-associated protein 2	-1.562	0.011
Q9UG63	ATP-binding cassette sub-family F member 2	-1.572	0.009
Q9NPQ8	Synembryn-A	-1.611	0.004
Q96FQ6	Protein S100-A16	-1.634	0.032
Q13310	Polyadenylate-binding protein 4	-1.669	0.006
Q9Y2S7	Polymerase delta-interacting protein 2	-1.750	0.003
P19338	Nucleolin	-1.759	0.038
Q96C86	m7GpppX diphosphatase	-1.826	0.042
P35637	RNA-binding protein FUS	-1.887	0.037
Q99729	Heterogeneous nuclear ribonucleoprotein A/B	-1.894	0.022
P43243	Matrin-3	-1.940	0.004

Q93062	RNA-binding protein with multiple splicing	-1.952	0.015
O15042	U2 snRNP-associated SURP motif-containing protein	-1.966	0.049
Q10567	AP-1 complex subunit beta-1	-1.993	0.023
Q14980	Nuclear mitotic apparatus protein 1	-1.996	0.005
P20700	Lamin-B1	-1.998	0.033
P23919	Thymidylate kinase	-2.192	0.050
O43684	Mitotic checkpoint protein BUB3	-2.221	0.041
P22087	rRNA 2'-O-methyltransferase fibrillarin	-2.291	0.020
Q9UHD9	Ubiquilin-2	-2.480	0.010
P48509	CD151 antigen	-2.554	0.008
P07205	Phosphoglycerate kinase 2	-2.717	0.014
Q13445	Transmembrane emp24 domain-containing protein 1	-2.946	0.009
P55145	Mesencephalic astrocyte-derived neurotrophic factor	-3.006	0.046
Q09028	Histone-binding protein RBBP4	-4.166	0.028
Q9BV57	1_2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	-4.274	0.015

Appendix C.5 List of proteins expressed only in untreated IPF fibroblasts in the non-IPF +/- FK228 dataset.

Proteins identified in untreated non-IPF fibroblasts only		
Protein accession	**Protein name**	**Log2 expression**
A8MTJ3	Guanine nucleotide-binding protein G(t) subunit alpha-	2.54
O76074	cGMP-specific 3'_5'-cyclic phosphodiesterase	4.47
Q5T653	39S ribosomal protein L2_ mitochondrial	3.79
P04818	Thymidylate synthase	5.19
P48730	Casein kinase I isoform delta	2.51
Q6UN15	Pre-mRNA 3'-end-processing factor	3.92
Q86UU1	Pleckstrin homology-like domain family B member 1	4.87
Q9BZE9	Tether containing UBX domain for GLUT4	3.95
Q9H6Y2	WD repeat-containing protein 55	4.24
P51531	Probable global transcription activator SNF2L2	3.69
P51532	Transcription activator BRG1	5.13
Q8NAV1	Pre-mRNA-splicing factor 38A	4.82
P49247	Ribose-5-phosphate isomerase	3.17
Q99811	Paired mesoderm homeobox protein	4.39

Appendix C.6 List of proteins expressed only in FK228-treated non-IPF fibroblasts in the non-IPF +/- FK228 dataset.

Proteins identified in FK228-treated non-IPF fibroblasts only		
Protein accession	**Protein name**	**Log2 expression**
Q96DA2	Ras-related protein Rab-39B	2.66
O75347	Tubulin-specific chaperone A	4.57
Q9BQS8	FYVE and coiled-coil domain-containing protein 1	6.14
Q13637	Ras-related protein Rab-32	3.73
Q92572	AP-3 complex subunit sigma-1	3.93
Q9UI42	Carboxypeptidase A4	5.25
Q8N4C6	Ninein	3.40
Q9BW83	Intraflagellar transport protein 27 homolog	3.63
Q13642	Four and a half LIM domains protein 1	3.88
Q9UGI8	Testin	3.66
Q8NFJ5	Retinoic acid-induced protein 3	5.12
O14966	Ras-related protein Rab-7L1	2.73
Q9H0Q0	Protein FAM49A	2.22
Q9Y244	Proteasome maturation protein	3.33
O94832	Unconventional myosin-Id	3.79
O94901	SUN domain-containing protein 1	5.26
Q9BYX2	TBC1 domain family member 2A	4.93

Appendix D - Appendix for Chapter 5

Appendix D.1 List of differentially expressed proteins between fibroblasts grown for 3 and 6 weeks in the 3D model.

Proteins with differential expression between 3 and 6 weeks of culture in the 3D model			
Protein accession	**Protein name**	**Fold-change 6:3 weeks**	**T-Test *p*-value**
P13796	Plastin-2	17.045	0.026
P05386	60S acidic ribosomal protein P1	3.757	0.012
P19105	Myosin regulatory light chain 12A	2.935	0.049
Q96QK1	Vacuolar protein sorting-associated protein 35	2.017	0.022
P07339	Cathepsin D	1.867	0.046
P18206	Vinculin	1.708	0.045
Q15582	Transforming growth factor-beta-induced protein ig-h3	1.554	0.015
Q02878	60S ribosomal protein L6	1.513	0.033
Q9Y490	Talin-1	1.505	0.008
P55072	Transitional endoplasmic reticulum ATPase	-1.141	0.038
P63244	Guanine nucleotide-binding protein subunit beta-2-like 1	-1.394	0.047

Q15436	Protein transport protein Sec23A	-1.400	0.037
Q99873	Protein arginine N-methyltransferase 1	-1.465	0.002
P49748	Very long-chain specific acyl-CoA dehydrogenase_ mitochondrial	-1.471	0.048
P42785	Lysosomal Pro-X carboxypeptidase	-1.480	0.002
Q9BSJ8	Extended synaptotagmin-1	-1.511	0.002
Q15084	Protein disulfide-isomerase A6	-1.691	0.020
P62424	60S ribosomal protein L7a	-1.890	0.033
Q01995	Transgelin	-1.898	0.034
Q13162	Peroxiredoxin-4	-2.042	0.000
P51149	Ras-related protein Rab-7a	-2.134	0.045
P48047	ATP synthase subunit O_ mitochondrial	-2.406	0.028
Q01105	Protein SET	-2.451	0.009
P00367	Glutamate dehydrogenase 1_ mitochondrial	-2.581	0.038
P15559	NAD(P)H dehydrogenase [quinone] 1	-2.886	0.049
Q7Z3Y7	Keratin_ type I cytoskeletal 28	-5.381	0.047

Appendix D.2 List of proteins expressed only in IPF fibroblasts cultured for 3 weeks in the 3D model.

Proteins identified only in fibroblasts cultured for 3 weeks in the 3D model		
Protein accession	**Protein name**	**Log2 expression**
Q96I99	Succinyl-CoA ligase [GDP-forming] subunit beta_ mitochondrial	5.248
Q53H12	Acylglycerol kinase_ mitochondrial	6.172
P32189	Glycerol kinase	5.067
Q16836	Hydroxyacyl-coenzyme A dehydrogenase_ mitochondrial	4.820
Q9UPT5	Exocyst complex component 7	6.042
P48735	Isocitrate dehydrogenase [NADP]_ mitochondrial	5.303
Q02218	2-oxoglutarate dehydrogenase_ mitochondrial	6.217
Q16643	Drebrin	5.590
P80723	Brain acid soluble protein 1	4.899
P25325	3-mercaptopyruvate sulfurtransferase	3.418
Q9H8Y8	Golgi reassembly-stacking protein 2	5.545
Q9Y6C9	Mitochondrial carrier homolog 2	3.755
Q14247	Src substrate cortactin	6.648
Q3SY69	Mitochondrial 10-formyltetrahydrofolate dehydrogenase	5.201
Q14847	LIM and SH3 domain protein 1	5.283
Q15393	Splicing factor 3B subunit 3	5.611
P36776	Lon protease homolog_ mitochondrial	5.477
Q92747	Actin-related protein 2/3 complex subunit 1A	5.881
Q15717	ELAV-like protein 1	4.573
P37235	Hippocalcin-like protein 1	3.317
P05109	Protein S100-A8	3.643
P11177	Pyruvate dehydrogenase E1 component subunit beta_ mitochondrial	5.061

Q8IVL6	Prolyl 3-hydroxylase 3	4.926
Q9UPY8	Microtubule-associated protein RP/EB family member 3	2.054
Q16629	Serine/arginine-rich splicing factor 7	4.712
Q9BS40	Latexin	6.306
Q7Z4N8	Prolyl 4-hydroxylase subunit alpha-3	4.908
P13798	Acylamino-acid-releasing enzyme	4.901
Q07955	Serine/arginine-rich splicing factor 1	4.482
Q96DZ1	Endoplasmic reticulum lectin 1	4.405
P09601	Heme oxygenase 1	4.498
Q2VIR3	Putative eukaryotic translation initiation factor 2 subunit 3-like protein	0.535
Q9Y285	Phenylalanine--tRNA ligase alpha subunit	5.242
Q06136	3-ketodihydrosphingosine reductase	4.387
Q8NBJ7	Sulfatase-modifying factor 2	4.088
P84103	Serine/arginine-rich splicing factor 3	5.038
Q9NUU7	ATP-dependent RNA helicase DDX19A	3.208
Q9UL25	Ras-related protein Rab-21	4.340
Q13409	Cytoplasmic dynein 1 intermediate chain 2	6.012
P61201	COP9 signalosome complex subunit 2	5.784
P45877	Peptidyl-prolyl cis-trans isomerase C	5.445
Q9Y4Z0	U6 snRNA-associated Sm-like protein LSm4	5.311
O75822	Eukaryotic translation initiation factor 3 subunit J	3.735
Q9UBS4	DnaJ homolog subfamily B member 11	5.095
O60256	Phosphoribosyl pyrophosphate synthase-associated protein 2	3.886
Q6YN16	Hydroxysteroid dehydrogenase-like protein 2	4.443
Q8WWX9	Selenoprotein M	4.827
Q6P2Q9	Pre-mRNA-processing-splicing factor 8	4.129
Q99653	Calcineurin B homologous protein 1	3.817
Q9Y6Y8	SEC23-interacting protein	5.004
P17302	Gap junction alpha-1 protein	5.403
P60891	Ribose-phosphate pyrophosphokinase 1	3.672
P02753	Retinol-binding protein 4	4.215
Q96IJ6	Mannose-1-phosphate guanyltransferase alpha	4.492
Q9BVG4	UPF0368 protein Cxorf26	3.741
P62273	40S ribosomal protein S29	6.185
Q9UNW1	Multiple inositol polyphosphate phosphatase 1	4.889
Q99487	Platelet-activating factor acetylhydrolase 2_ cytoplasmic	4.799
Q86W92	Liprin-beta-1	3.194
P20648	Potassium-transporting ATPase alpha chain 1	2.338
Q9NPQ8	Synembryn-A	3.384
Q9BRK5	45 kDa calcium-binding protein	4.094
Q9H0R4	Haloacid dehalogenase-like hydrolase domain-containing protein 2	4.770
P30504	HLA class I histocompatibility antigen_ Cw-4 alpha chain	2.157
O00763	Acetyl-CoA carboxylase 2	6.104
Q6ZW49	PAX-interacting protein 1	3.059
Q13151	Heterogeneous nuclear ribonucleoprotein A0	4.536
Q6NUQ4	Transmembrane protein 214	5.532
Q9BTZ2	Dehydrogenase/reductase SDR family member 4	4.014

Q01844	RNA-binding protein EWS	5.177
A1XBS5	Protein FAM92A1	-0.188
Q08AM6	Protein VAC14 homolog	4.925

Appendix D.3 List of proteins expressed only in IPF fibroblasts cultured for 6 weeks in the 3D model.

Proteins identified only in fibroblasts cultured for 6 weeks in the 3D model		
Protein accession	**Protein name**	**Log2 expression**
Q9UDY4	DnaJ homolog subfamily B member 4	5.682
P68543	UBX domain-containing protein 2A	4.870
Q6NUP7	Serine/threonine-protein phosphatase 4 regulatory subunit 4	6.132

Appendix D.4 List of identified ECM proteins and their expression values in the 3D model and monolayer cultures.

		3 weeks 3D system expression values			6 weeks 3D system expression values			Monolayer culture system expression values
Protein accession	**Protein name**	**Replicate 1**	**Replicate 2**	**Replicate 3**	**Replicate 1**	**Replicate 2**	**Replicate 3**	**Replicate 1**
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein		7.420	7.470		7.208	7.546	
P21810	Biglycan		3.838	4.169		3.726	4.783	
Q05682	Caldesmon	7.704	7.456	6.958	7.813	6.818	6.513	7.906
P02452	Collagen alpha-1(I) chain		7.634	6.353		5.936	5.764	8.060
P02461	Collagen alpha-1(III) chain		6.059	5.564		6.059		5.245
P12109	Collagen alpha-1(VI) chain	9.226	9.374	8.881	9.689	9.415	9.477	6.333
Q02388	Collagen alpha-1(VII) chain	8.145	5.410			4.665	7.373	
P27658	Collagen alpha-1(VIII) chain		5.530	5.605		4.828	4.816	
Q99715	Collagen alpha-1(XII) chain	6.695	6.505	6.821	7.391	6.506	7.196	7.616
P39059	Collagen alpha-1(XV) chain		4.888			5.048		
P39060	Collagen alpha-1(XVIII) chain					6.052		
Q86Y22	Collagen alpha-1(XXIII) chain	3.338						
P08123	Collagen alpha-2(I) chain	7.710	6.423	5.187		5.059	5.271	6.795
P08572	Collagen alpha-2(IV) chain	5.832	6.370	5.997		5.899	5.993	
P05997	Collagen alpha-2(V) chain		5.282	5.029		7.999		
P12110	Collagen alpha-2(VI) chain	8.810	8.193	8.794	9.422	9.246	8.718	6.802
P12111	Collagen alpha-3(VI) chain	9.701	9.326	9.935	9.670	10.033	10.469	7.674
A8TX70	Collagen alpha-5(VI) chain					4.336	7.811	
Q9Y5P4	Collagen type IV alpha-3-binding protein	4.284						
P07585	Decorin	3.843	5.581	5.484		4.311		3.234
Q16610	Extracellular matrix protein 1							4.242
P35555	Fibrillin-1	5.019	4.392	6.481		3.650		
P02751	Fibronectin	6.671	8.004	7.755	6.440	8.101	7.751	8.391
P07942	Laminin subunit beta-1		6.033	5.643				4.886
P55268	Laminin subunit beta-2							4.756
P51884	Lumican		3.625					
P50281	Matrix metalloproteinase-14		4.438	4.444		3.425		4.785
Q9ULZ9	Matrix metalloproteinase-17				5.662			
O60882	Matrix metalloproteinase-20	4.965						
P01033	Metalloproteinase inhibitor 1		4.991	4.890			4.369	3.765
P35625	Metalloproteinase inhibitor 3							6.163
Q14112	Nidogen-2		6.197	5.268		6.260		
Q15063	Periostin		6.352					
Q8NBJ5	Procollagen galactosyltransferase 1	5.279	5.792	4.903		4.075	3.555	6.044
Q02809	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 1	8.041	8.148	7.115	7.751	7.909	7.613	6.799
O00469	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 2	9.298	8.718	8.038	8.848	8.827	8.627	4.729
O60568	Procollagen-lysine_2-oxoglutarate 5-dioxygenase 3	6.386	6.800	6.072		5.866	5.230	5.899
Q32P28	Prolyl 3-hydroxylase 1	6.528	6.321	6.419		6.603	5.833	6.777
Q8IVL6	Prolyl 3-hydroxylase 3	4.985	4.901	4.892				5.346
P13674	Prolyl 4-hydroxylase subunit alpha-1	7.800	7.471	7.152	6.609	7.572	7.572	6.299
O15460	Prolyl 4-hydroxylase subunit alpha-2	7.529	6.861	6.364	7.354	7.372	7.405	6.556
Q7Z4N8	Prolyl 4-hydroxylase subunit alpha-3	4.368	5.227	5.130				2.771
P24821	Tenascin	7.557	7.951	8.000	8.014	8.371	7.772	
P07996	Thrombospondin-1	5.448	6.128	6.550		6.957	6.171	5.967
p35442	Thrombospondin-2	4.248		3.262				
Q15582	Transforming growth factor-beta-induced protein ig-h3	9.087	8.893	8.672	9.427	9.432	9.701	3.733
P04004	Vitronectin	4.451						

Appendix D.5 List of proteins with differential expression between untreated IPF fibroblasts and IPF fibroblasts treated for 3 weeks with FK228 in the 3D model (6 weeks culture).

Proteins with differential expression between untreated and FK228-treated fibroblasts in the 3D model (chronic)			
Protein accession	**Protein name**	**Fold-change FK228-treated: untreated**	**T-Test *p*-value**
P15559	NAD(P)H dehydrogenase [quinone] 1	5.329	0.020
P54652	Heat shock-related 70 kDa protein 2	5.159	0.018
P35232	Prohibitin	3.912	0.005
P21912	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit_ mitochondrial	3.732	0.035
P05091	Aldehyde dehydrogenase_ mitochondrial	3.202	0.019
P00367	Glutamate dehydrogenase 1_ mitochondrial	2.893	0.020
P45880	Voltage-dependent anion-selective channel protein 2	2.838	0.032
Q15008	26S proteasome non-ATPase regulatory subunit 6	2.639	0.009
P42785	Lysosomal Pro-X carboxypeptidase	2.575	0.008
P12277	Creatine kinase B-type	2.468	0.008
P38606	V-type proton ATPase catalytic subunit A	2.305	0.012
Q9HDC9	Adipocyte plasma membrane-associated protein	2.284	0.004
Q9NVD7	Alpha-parvin	2.275	0.019
Q14764	Major vault protein	2.227	0.012
P51659	Peroxisomal multifunctional enzyme type 2	2.185	0.018
P04424	Argininosuccinate lyase	2.148	0.015
Q9BS26	Endoplasmic reticulum resident protein 44	2.108	0.018
P02794	Ferritin heavy chain	1.909	0.006
P06576	ATP synthase subunit beta_ mitochondrial	1.908	0.041
P55084	Trifunctional enzyme subunit beta_ mitochondria	1.694	0.038
Q9NR31	GTP-binding protein SAR1a	1.686	0.024
O00231	26S proteasome non-ATPase regulatory subunit 11	1.355	0.048
P21333	Filamin-A	-1.382	0.030
O94973	AP-2 complex subunit alpha-2	-1.659	0.003
P30153	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	-1.757	0.022
P18669	Phosphoglycerate mutase 1	-1.761	0.043
P06744	Glucose-6-phosphate isomerase	-1.844	0.023

Q9NZN4	EH domain-containing protein 2	-1.904	0.030
P04632	Calpain small subunit 1	-1.976	0.044
Q96QK1	Vacuolar protein sorting-associated protein 35	-1.988	0.026
P53396	ATP-citrate synthase	-2.087	0.035
P50454	Serpin H1	-2.369	0.003
P05023	Sodium/potassium-transporting ATPase subunit alpha-1	-2.376	0.037
P04406	Glyceraldehyde-3-phosphate dehydrogenase	-2.426	0.007
Q96AY3	Peptidyl-prolyl cis-trans isomerase FKBP10	-2.449	0.023
O15460	Prolyl 4-hydroxylase subunit alpha-2	-3.254	0.020
P34931	Heat shock 70 kDa protein 1-like	-7.702	0.010

Appendix D.6 List of proteins only identified in untreated IPF fibroblasts compared to IPF fibroblasts treated for 3 weeks with FK228 in the 3D model (6 weeks culture).

Proteins only identified in untreated fibroblasts in the 3D model (chronic)		
Protein accession	**Protein name**	**Log2 expression**
Q9NY33	Dipeptidyl peptidase 3	4.698
Q92619	Minor histocompatibility protein HA-1	3.055
Q96JG9	Zinc finger protein 469	5.981
Q96A08	Histone H2B type 1-A	4.894
A4D2P6	Delphilin	3.203
P68543	UBX domain-containing protein 2A	4.870
Q6NUP7	Serine/threonine-protein phosphatase 4 regulatory subunit 4	6.132

Appendix D.7 List of proteins only identified in IPF fibroblasts treated for 3 weeks with FK228 compared to untreated IPF fibroblasts in the 3D model (6 weeks culture).

Proteins only identified in FK228-treated fibroblasts in the 3D model (chronic)		
Protein accession	**Protein name**	**Log2 expression**
Q96I99	Succinyl-CoA ligase [GDP-forming] subunit beta_ mitochondrial	5.351
Q53H12	Acylglycerol kinase_ mitochondria	5.609
P32189	Glycerol kinase	5.020
Q16836	Hydroxyacyl-coenzyme A dehydrogenase_ mitochondrial	5.586
P07099	Epoxide hydrolase 1	6.037
O94925	Glutaminase kidney isoform_ mitochondrial	5.978
O75348	V-type proton ATPase subunit G 1	5.100

Q12840	Kinesin heavy chain isoform 5A	4.853
P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex_ mitochondrial	6.587
Q92747	Actin-related protein 2/3 complex subunit 1A	5.690
Q9Y394	Dehydrogenase/reductase SDR family member 7	4.136
Q13637	Ras-related protein Rab-32	5.038
O94788	Retinal dehydrogenase 2	2.964
O95671	N-acetylserotonin O-methyltransferase-like protein	6.253
P49207	60S ribosomal protein L34	4.499
O15230	Laminin subunit alpha-5	5.573
P16402	Histone H1.3	5.809
Q92743	Serine protease HTRA1	6.187

Appendix D.8 List of proteins with differential expression between untreated IPF fibroblasts and IPF fibroblasts treated for 48 hours with FK228 in the 3D model (3 weeks culture).

Proteins with differential expression between untreated and FK228-treated fibroblasts in the 3D model (acute)			
Protein accession	**Protein name**	**Fold-change FK228-treated: untreated**	**T-Test *p*-value**
P05091	Aldehyde dehydrogenase_ mitochondrial	2.529	0.032
Q06830	Peroxiredoxin-1	1.959	0.004
P15559	NAD(P)H dehydrogenase [quinone] 1	1.867	0.046
Q7KZF4	Staphylococcal nuclease domain-containing protein 1	1.581	0.050
P05556	Integrin beta-1	1.420	0.029
O14773	Tripeptidyl-peptidase	-1.308	0.010
O43776	Asparagine--tRNA ligase_ cytoplasmic	-1.318	0.034
P17858	6-phosphofructokinase_ liver type	-1.319	0.048
Q8NBS9	Thioredoxin domain-containing protein 5	-1.342	0.014
P21399	Cytoplasmic aconitate hydratase	-1.402	0.020
P46940	Ras GTPase-activating-like protein IQGAP1	-1.498	0.006
P10768	S-formylglutathione hydrolase	-1.503	0.047
P27635	60S ribosomal protein L10	-1.756	0.043
P16278	Beta-galactosidase	-1.793	0.043
Q9BSJ8	Extended synaptotagmin-1	-2.083	0.006
Q3ZCM7	Tubulin beta-8 chain	-16.685	0.047
P05787	Keratin_ type II cytoskeletal 8	-26.584	0.007
P17661	Desmin	-34.150	0.014

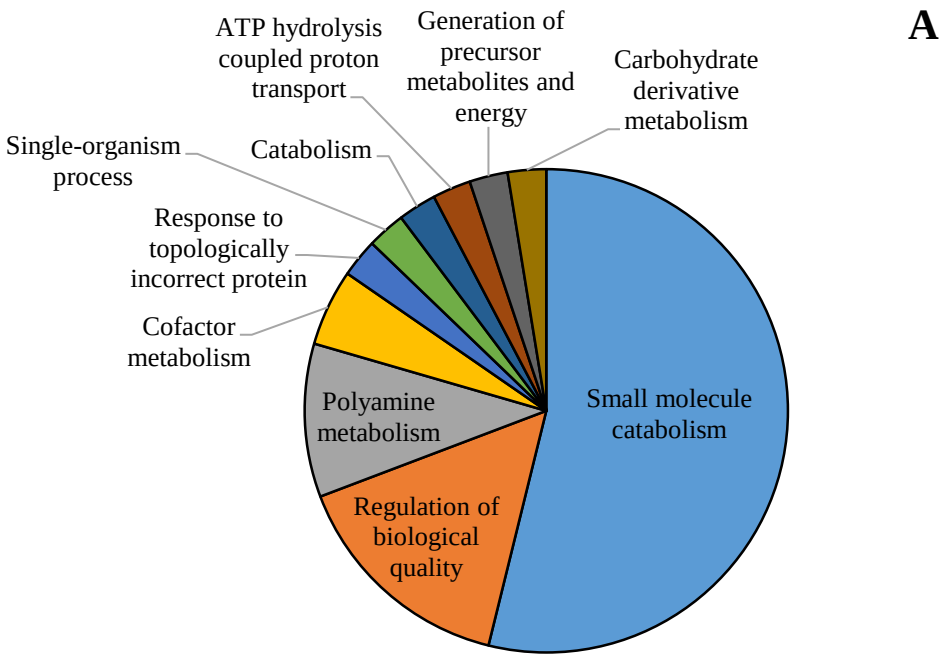
Appendix D.9 List of proteins only identified in untreated IPF fibroblasts compared to IPF fibroblasts treated for 48 hours with FK228 in the 3D model (3 weeks culture).

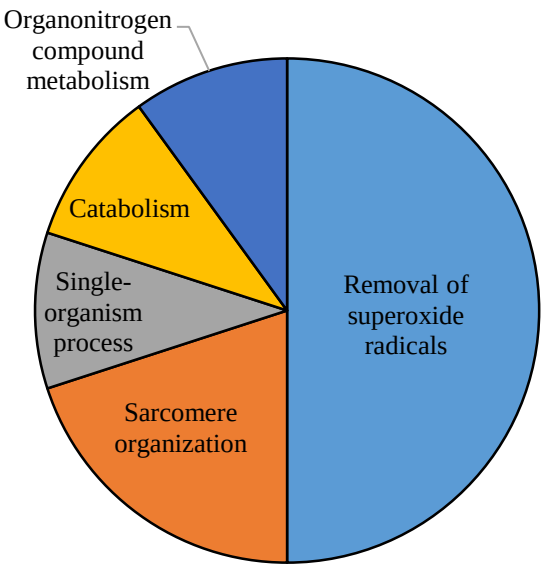
Proteins only identified in untreated fibroblasts in the 3D model (acute)		
Protein accession	**Protein name**	**Log2 expression**
Q99541	Perilipin-2	5.257
A8MTJ3	Guanine nucleotide-binding protein G(t) subunit alpha-3	0.641
Q96M27	Protein PRRC1	5.340
P22087	rRNA 2'-O-methyltransferase fibrillarin	5.414
P61970	Nuclear transport factor 2	6.231
Q9H853	Putative tubulin-like protein alpha-4B	6.92
Q9ULC3	Ras-related protein Rab-23	4.144
Q9NP97	Dynein light chain roadblock-type 1	4.955
P11488	Guanine nucleotide-binding protein G(t) subunit alpha-1	0.394
P06730	Eukaryotic translation initiation factor 4E	5.796
O75165	DnaJ homolog subfamily C member 13	5.597
Q13131	5'-AMP-activated protein kinase catalytic subunit alpha-1	4.805
P02749	Beta-2-glycoprotein 1	5.265
Q9UHD1	Cysteine and histidine-rich domain-containing protein 1	4.771
Q9Y639	Neuroplastin	6.225
Q9Y2A7	Nck-associated protein 1	5.148
Q9H223	EH domain-containing protein 4	3.051
O95249	Golgi SNAP receptor complex member 1	4.711
Q9NZ32	Actin-related protein 10	4.504
Q15629	Translocating chain-associated membrane protein 1	5.364
Q13748	Tubulin alpha-3C/D chain	5.990
P61201	COP9 signalosome complex subunit 2	5.784
P45877	Peptidyl-prolyl cis-trans isomerase C	5.445
Q6YN16	Hydroxysteroid dehydrogenase-like protein 2	4.443
Q8WWX9	Selenoprotein M	4.827
Q6P2Q9	Pre-mRNA-processing-splicing factor 8	4.129
Q99653	Calcineurin B homologous protein 1	3.817
Q14956	Transmembrane glycoprotein NMB	6.238
Q9GZV4	Eukaryotic translation initiation factor 5A-2	6.982
Q5JQF8	Polyadenylate-binding protein 1-like 2	2.180
P62979	Ubiquitin-40S ribosomal protein S27a	8.338
P12429	Annexin A3	1.264
P60891	Ribose-phosphate pyrophosphokinase 1	3.672
Q96IJ6	Mannose-1-phosphate guanyltransferase alpha	4.492
Q99487	Platelet-activating factor acetylhydrolase 2_ cytoplasmic	4.799
P20648	Potassium-transporting ATPase alpha chain 1	2.338
Q969G5	Protein kinase C delta-binding protein	5.302
Q9H0R4	Haloacid dehalogenase-like hydrolase domain-containing protein 2	4.770
O00139	Kinesin-like protein KIF2A	2.446
O00763	Acetyl-CoA carboxylase 2	6.104
Q6ZW49	PAX-interacting protein 1	3.059

Q13151	Heterogeneous nuclear ribonucleoprotein A0	4.536
Q6NUQ4	Transmembrane protein 214	5.531
Q9BTZ2	Dehydrogenase/reductase SDR family member 4	4.014
Q01844	RNA-binding protein EWS	5.177
A1XBS5	Protein FAM92A1	-0.188
Q08AM6	Protein VAC14 homolog	4.925

Appendix D.10 Gene ontology enrichment analysis of proteins with significantly different expression levels (p<0.05) between IPF fibroblasts grown in the 3D in vitro culture system for A.) 6 weeks, with or without 3 weeks of FK228 treatment and B.) 3 weeks, with or without 48 hours of FK228 treatment.

Individual GO terms were combined into related terms using REVIGO that are reflected in the pie chart. Size of the terms is adjusted (log10 p-value) to reflect the p-values.





B

Appendix E - Appendix for Chapter 6

Appendix E.1 Results using Choudhary method (original)

Protein Accession	Protein Name	Site of acetylation	Sequence	b/y ions
A5A3E0/ Q6S8J3	POTE ankyrin domain family member F/E	K(1)	KPFGLRSK	y7y8
Q5VTE0/ P68104	Putative elongation factor 1-alpha-like 3/Elongation factor 1-alpha 1	K(4), K(*)	SAQKAQKAK	b3b5b6b8b 9*y3y8* b3°b5y4y5
Q02878	60S ribosomal protein L6	K(8)	FVIATSTKIDI SNVKIPK	b3b5°b14°b 17y9°
P84098	60S ribosomal protein L19	K(9), K(4)	EEIIKTLSK TLSK	b3y4y9 b4°y4
Q13748/ Q71U36 Q6PEY2/ Q9BQE3/ P68363	Tubulin alpha-1A/3C/D/1B/1C/3E chain	K(10)	DVNAAIATIK	b6b7y10

Appendix E.2 Results using Choudhary method modification 1

Protein Accession	Protein Name	Site of acetylation	Sequence	b/y ions
Q5VTE0/ P68104	Putative elongation factor 1-alpha-like 3/Elongation factor 1-alpha 1/2	K(7), K(14)	RYEEIVKEVSTYIK	b9y5y5°
Q5VTE0/ P68105/ Q05639	Putative elongation factor 1-alpha-like 3/Elongation factor 1-alpha 2	K(11)	GSFKYAWVLDK	b8y1y3y5y6y10°
Q05639	Elongation factor 1-alpha 2	K(*)	VETGVLKPGMVVTFAPVNVTTEVKSVEMHHEALSEALPGDNVGFNVK	b21b29°y10y13°y14

Appendix E.3 Results using Choudhary method modification 2

Protein Accession	Protein Name	Site of acetylation	Sequence	b/y ions
Q04721	Neurogenic locus notch homolog protein 2	K(35), K(22)	CEMDINECHSDPCQNDATCLDKIGGFTCLCMPGFK	b5b7*b8b29y6y13y14°b18b24*b29y11y22°y23
		K(14)	CQTDMNECLSEPCKNGGTCSDYVNSYTCK	b21*y13
		K(14), K(42)	CVNTDGAFHCECLKGYAGPRCEMDINECHSDPCQNDATCLDK	b12b12*b15°b25*y15°y17
		K(23)	DGYEPCVNEGMCVTYHNGTGYCKCPEGFLGEYCQHR	b9*b12°b13b29°b30*b33°b34y14y35y36

		K(23)	DGYEPCVNEGMC VTYHNGTGYCKC PEGFLGEYCQHRD PCEK	b14°b37*y34
		K(24), K(38)	GADCTEDVDECA MANSNPCEHAGK CVNTDGAFHCEC LK	b12b16b16°b18* b20°b22*y7y12 y12°y25y26y27* y35°b6°y7y13°y28
		K(22)	GALCDTNPLNGQ YICTCPQGYKGAD CTEDVDECAMAN SNPCEHAGK	b7*b8b21°b23° y7°y9*y17°y30* y33*
		K(1)	KGATCINGVNGFR CICPEGPHHPSCYS QVNECLSNPCIHG NCTGGLSGYK	b8°y7°y16°
		K(36)	MEVNETQYNEMF GMVLAPAEGTHP GIAPQSRPPEGKHI TTPR	y20
		K(*)	NKGTCVQKK	b4y5y7°
		K(7)	SRRPSAK	b3b4b4°
		K(*)	TQLLYLLAVAVVI ILFIILLGVIMAKR K	b10b10°b15*b25
Q04721/ Q7Z3S9	Neurogenic locus notch homolog protein 2/Notch homolog 2 N-terminal-like protein	K(14)	DTYECTCQVGFTG KECQWTDACLSH PCANGSTCTTVAN QFSCK	b40y26*
P04264	Keratin_ type II cytoskeletal 1	K(*)	DAKNKLNDLEDA LQQAK	b12b14b16°y7*
Q9BYS1	Keratin-associated protein 1-5	N-TERM(1)	MTCCQTSFCGYPS FSISGTCGSSCCQP SCCETSCCQPR	b5*b19y7y30y38

Appendix E.4 Results using Choudhary method modification 3 (Control)

Protein Accession	Protein Name	Site of acetylation	Sequence	b/y ions
Q3LI64	Keratin-associated protein 6-1	N-TERM(1)	CGSYYGNYYGTP GYGFCGYGGLGY GYGGLGCGYGSC CGCGFR	b6b11b11°b19 *b31*b32y4y2 3y33y34°y37
P80294	Metallothionein-1H	N-TERM(1)	DPNCSCEAGGSCA CAGSCK	b5°b18
		K(*)	MDPNCSCEAGGS CACAGSCKCK	b4b6y3
		N-TERM(1), K(20)	MDPNCSCEAGGS CACAGSCKCKK	b5b13b17y9°
Q9UJU5	Forkhead box protein D3	K(22)	EAAGAGAGPGGD VGAPEADGCK	b17
		K(29)	GGVGGEEGGASG GGPGAGSGSAGG LAPSKPK	b13°b29y5°y2 6y30
Q14568	Putative heat shock protein HSP 90-alpha A2	K(4)	ECDKEVSDDETEE KEDKEEEK	y10°y11
		K(13),K(17)	EVSDDETEEKEDK EEEK	b10°b15y9
		K(10), K(13)	EVSDDETEEKEDK EEEK	b10b15°y6y11
		K(41)	MPEETQTQDQPM EEEEVETFAFQAEI AQLMSLIINTFYSN K	b10°y7°y8y12 °y14
Q8NGT7	Olfactory receptor 2A12	K(1)	KILSLFYSLFNPIL NPLIYSLRNAEVK	b10°
Q9BYS1	Keratin-associated protein 1-5	N-TERM(1)	TCCQTSFCGYPSF SISGTCGSSCCQPS CCETSCCQPR	y10y29°
Q13424	Alpha-1-syntrophin	K(13)	TIVFIIHSFLSAKVT	b3b9

A6NN90	Uncharacterized protein C2orf81	K(52)	VRPPTVPVPQVDI VPGRLSEAEMALT ALEEGEDVVGDIL ADLLARVMDSAF K	b30°b43°y30°

Appendix E.5 Results using Choudhary method modification 3 (+FK228)

Protein Accession	**Protein Name**	**Site of acetylation**	**Sequence**	**b/y ions**
P58173	Olfactory receptor 2B6	K(40)	LDTKLHTPMYF FLTNLSLLDLC YTTCTVPQML VNLCSIRK	b33b39y8°y12° y21
Q96HV5	Transmembrane protein 41A	K(16)	LLAIAMVALIP GTLIK	b10y6°
O95873	Uncharacterized protein C6orf47	K(26), K(28)	VDSSSENSGSD WDSAPETMED VGHPKTK	b7b18°y3°y4°y15 y16°y25*
Q8NBI3	Draxin	K(21)	GLNNKCFDDC MCVEGLRCYA K	b9°b11°b15b15* b18°y16
Q495W5	Alpha-(1_3)-fucosyltransferase 11	K(11)	FWDYLHEIFMK RQHL	b3y4*y6*y11°
Q15061	WD repeat-containing protein 43	K(1)	KPLTSNCTIQIA TPGKGKK	b3b9°b13*y8y12° y15
		K(19)	LVYEEESSEEES DDEIADKDSED NWDEDEEESES EK	b26*b28*y4°
		K(14)	DEDVEEEDEDA EGKDEENGED RDTASEKELNG	b12°b34°y35*

			DSDLDPENESEEE	
		K(*)	LSHLHGKLILLITQVTASEKTK	b4y6y9y17°
		K(38)	AKLVYEEESSEEESDDEIADKDSEDNWDEDEEESESEK	b9b22b22°y14y17
		K(7), K(20)	LSHLHGKLILLITQVTASEK	b3°b7b12y2
		K(31)	DSEDNWDEDEEESESEKDEDVEEEDEDAEGKDEENGEDR	b9°b13°b14y14y14°y15
P27797	Calreticulin	K(15), K(25)	KEEEEAEDKEDDEDKDEDEEDEEDK	b23°
		K(10)	DEDEEDEEDKEEDEEEDVPGQAK	y21°
P08118	Beta-microseminoprotein	K(44)	CMDLKGNKHPINSEWQTDNCETCTCYETEISCCTLVSTPVGYDK	b5°b9°b24°b28°b41°
P42127	Agouti-signaling protein	K(4)	NSCKPPAPACCDPCASCQCR	b5y4y11°y17°
		K(4)	NSCKPPAPACCDPCASCQCRFFR	b14*y16
P25713	Metallothionein-3	N-TERM(1)	MDPETCPCPSGGSCTCADSCKCEGCKCTSCK	b21°y2y15
		K(21)	MDPETCPCPSGGSCTCADSCK	b19°y1

O43719	HIV Tat-specific factor 1	K(39)	MQELYGDGKD GDTQTDAGGE PDSLGQQPTDT PYEWDLDK	b2b3°b13°b16* b34y7y7°y18*y 19°y22 y28y30y32*y36 y39
		K(27)	EFDEDSDEKEE EEDTYEKVFDD ESDEK	b7b12°b27y8y9 ° y12y17°
		K(*)	FEKTEDGGEFE EGASENNAKES SPEK	y10*y13
		K(16), K(26)	EDEDADGKEV EDADEKLFEDD DSNEK	b19°y17*y18
		K(7), K(11)	LHVEVAKFQL K	b4b9y3*y5*y6y 7
		K(10)	NDCEENGLAK ESEDDLNK	b11b18°y5*y6* y13°
		K(*)	DGDTQTDAGG EPDSLGQQPTD TPYEWDLDKK	b4b5°b18*b29*
		K(8), K(16), K(26)	EDEDADGKEV EDADEKLFEDD DSNEK	b10°b11°y5°y7* y23°
		K(*), K(*)	ESEDDLNKESE EEVGPTKESEE DDSEKESDEDC SEK	b7*b8*y4°
		K(29), K(30)	HFSEHPSTSKM NAQETATGMA FEEPIDEKK	b10b14°b17b23 *b26°y6°y13y1 9*y29 y29*
		K(23)	TEDGGEFEEGA SENNAKESSPE K	b8b12°b17
		K(*)	LFEESDDKEDE	y4y6

		DADGKEVEDA DEK	
	K(*), K(*)	LFEESDDKEDE DADGKEVEDA DEK	b10°b15y14y14 °
	K(18)	LHVEVAKFQL KGEYDASK	b4b8b9°b14b16 y7 y15

Appendix E.6 Results using Li method

Protein Accession	**Protein Name**	**Site of acetylation**	**Sequence**	**b/y ions**
P82933	28S ribosomal protein S9_ mitochondrial	K(*)	QLIEPVQYDEQG MAFSKSEGK	b4b5*b6*y1 1
P39023	60S ribosomal protein L3	K(*)	KWQDEDGKK	b4°b6b9y5°
P11021	78 kDa glucose-regulated protein	K(1),K(4) K(4)	KTGKDVR QATKDAGTIAGL NVMR	b3b4b5°b6 b4b4*b5*
O43707	Alpha-actinin-4	K(11)	RDHALLEEQSK	b4b9*y5y6
O95817	BAG family molecular chaperone regulator 3	K(*)	KEVDSKPVSQKP PPPSEK	b2°y3y5
P27797	Calreticulin	K(8)	QDEEQRLKEEEE DK	b2b3b3*y5y 5°y6 y14
O43852	Calumenin	K(8)	DGDLIATK	b5°y2°y3
P23528	Cofilin-1	K(12)	HELQANCYEEV KDR	b14y5y6y14
P23528/ P60981	Cofilin-1/Destrin	K(9)	SKMIYASSKDAI K	b5b12b13°y4 °y9
P14625	Endoplasmin	K(10)	EFEPLLNWMK	b4y4y5y6
P21333	Filamin-A	K(2), K(8)	VKETADFK	b3b4b6°y3°

P09382	Galectin-1	N-TERM(1)	ACGLVASNLNL KPGECLR	b4b9y6y13y 14y16 y18
P04406	Glyceraldehyde-3-phosphate dehydrogenase	K(4) K(8)	KVVKQASEGPL K TVDGPSGKLWR	b10y9 b2b9°y4y5y6 y7y8 y9
P08107	Heat shock 70 kDa protein 1A/1B	K(4) K(1)	GETKAFYPEEISS MVLTK KFGDPVVQSDM K	b5b6°b7 b6b10y1y3y 6°y7*
P08238/ Q58FF7	Heat shock protein HSP 90-beta	K(11)	EISDDEAEEEK	b4y3y5
Q9H910	Hematological and neurological expressed 1-like protein	K(11)	SIPAGAEPGEKG SAR	b4y8y13
P20700	Lamin-B1	K(7)	QIEYEYK	b3°b4°
P27816	Microtubule-associated protein 4	K(5) K(4), K(11) K(1)	ASPSKPASAPAS R SGSKSTQTVAK KPESNAVTK	b2b5°y4 y3y4y6*y6° b1b2b7*y8y 8°
P35579	Myosin-9	K(15) K(17) K(13) K(11) K(22)	ADFCIIHYAGKV DYK KKMEDSVGCLE TAEEVK QFRTEMEDLMSS K RALEQQVEEMK TEMEDLMSSKD DVGKSVHELEK	y3°y12 b5b5°b11b15 y3 °y7° b5y7° b3b4y11 b1°b6b7°y5y 6°
P12036	Neurofilament heavy polypeptide	K(1) K(*)	KGGAGGTR SSSTDQKDSKPP EKATEDK	b2b5y3 b10y4y4°

		K(14) K(7)		y4y7 b7*b9b10
P07197	Neurofilament medium polypeptide	K(8)	EEVEQETKEK	b5°y3y6*y7y9y9°
Q6S8J3/A5A3E0/P0CG38/P0CG39	POTE ankyrin domain family member E/F/I/J	K(10)	MNSELSLSCK	b5b5*b8b8*y8y8°
P02545	Prelamin-A/C	K(5), K(12)	LQTMKEELDFQK	b3*b7b8b12 y2*y3 y3*y5y6*
P07737	Profilin-1	K(10)	EGVHGGLINK	b2b3b5b9*
P07237	Protein disulfide-isomerase	K(10)	NFEDVAFDEK	b5*b7y3°y5y5°y7°
P30101	Protein disulfide-isomerase A3	K(3)	NAKGSNYWR	b4b7y9
A6NMY6/P07355	Putative annexin A2-like protein/Annexin A2	K(18)	LSLEGDHSTPPSAYGSVK	b12y4y8
Q58FF3/P14625	Putative endoplasmin-like protein/Endoplasmin	K(14)	YNDTFWKEFGTNIK	b6y7
P08559	Pyruvate dehydrogenase E1 component subunit alpha_ somatic form_ mitochondrial	K(8)	SDPIMLLKDR	b3°y4y5y6
P14618	Pyruvate kinase isozymes M1/M2	K(10)	ITLDNAYMEK	b5*b8°y3y7y8y10
P50395/P31150	Rab GDP dissociation inhibitor beta/alpha	K(9)	DWNVDLIPK	b4b7°
P29508	Serpin B3	K(1)	KINSWVESQTNEK	b7y2°y4y6y7y13
Q14247	Src substrate cortactin	K(8)	TEKHASQK	b4b5y3*

P38646	Stress-70 protein_ mitochondrial	K(12)	QAASSLQQASLK LFEMAYK	b4b7°b9y6y1 9
P37837	Transaldolase	K(9)	AAQASDLEKIHL DEK	b2b3b9*b15 y2y2°y4y5y7 y8y9y11y12 y13*y15
P60174	Triosephosphate isomerase	K(14), K(18)	VVFEQTKVIADN VKDWSK	b4b5b6°y5
O43399	Tumor protein D54	K(6)	NSATFKSFEDR	y3y8°
Q9H7Y2	Uncharacterized protein FLJ14100	K(19)	SQREVQTVCTG VSGDQAQK	b7b9y5*y5°y 7y8*y9*