DNA methylation in lung fibroblasts and its role in pulmonary fibrosis

Ian Matthew Garner

University College London

Division of Medicine

UCL Respiratory

Centre for Inflammation and Tissue Repair

Rayne Institute

University Street

London

This thesis is submitted to University College London for the degree of Doctor of Philosophy

Declaration

I, Ian Matthew Garner, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis,

London, October 2015.

Acknowledgements

I would like to acknowledge the tutelage and direction of my supervisor, Professor Robin McAnulty, who has always supported me and given me excellent guidance and opportunities throughout my PhD studies. His unparalleled scientific knowledge surrounding pulmonary fibrosis was invaluable to me and has further fuelled my hope to become a successful scientist. I would also like to acknowledge my secondary supervisor, Dr. Chris Scotton, who has provided excellent feedback and discussions as well as offering a calming presence when presenting at conferences.

I would like to thank past and present members at the Centre for Inflammation and Tissue Repair, in particular 'Team Robin' members who have provided me with support, laughter, and countless ideas throughout my PhD, notably and in no particular order: Dr. Iona Evans, Dr. Josephine Barnes, Mr. Dave Pearce and Miss. Lizzy Peix.

I would also like to thank the Medical Research Council for funding my PhD and Professor David Abraham and the staff of the Centre for Rheumatology and Connective Tissue Diseases.

Finally, I thank my parents for always supporting me emotionally and financially and for encouraging me throughout my studies and my partner, Amy, who has continuously supported me throughout my PhD studies.

Conferences, abstracts and presentations

BALR 2012 (University of Southampton), 17th International Colloquium on Lung and Airway Fibrosis (ICLAF) 2012 (Italy), BSMB 2013 (University of Oxford), Infection, Immunology and Inflammation 2013 (UCL), Division of medicine 2014 (UCL), 18th ICLAF 2014 (Canada), London Matrix Group 2014 (London), American Thoracic Society International conference 2014, BALR 2015 (University of Bath).

Publications

Evans, I.C., Barnes, J.L., Garner, I.M., Pearce, D.R., Maher, T.M., Shi-wen, X., Renzoni, E.A., Wells, A.U., Denton, C.P., Laurent, G.J., Abraham, D.J. & McAnulty, R.J., 2016. Epigenetic Regulation of Cyclooxygenase-2 by methylation of c8orf4 in Pulmonary Fibrosis. *Clinical Science*, 130(8), 575-86.

Garner, I.M., Evans, I.C., Barnes, J.L., Maher, T.M., Renzoni, E.A., Denton, C.P., Scotton, C.J., Abraham, D.J. & McAnulty, R.J., 2016. DNA methylation and its role in pulmonary fibrosis. (In preparation).

Garner, I. M., Evans, I. C., Maher, T. M., Renzoni, E. A., Denton, C. P., Scotton, , C.J., Abraham, D.J. & McAnulty, R.J., 2014. Genome-wide analysis identifies multiple genes with altered methylation and expression in IPF and SSc lung fibroblasts. *International Journal of Experimental Pathology*, 94 (5), A16.

Garner, I.M., Evans, I.C., Barnes, J.L., Maher, T.M., Renzoni, E.A., Denton, C.P., Scotton, C.J., Abraham, D.J. & McAnulty, R.J., 2014. Hypomethylation Of The TNXB Gene Contributes To Increased Expression And Deposition Of Tenascin-X In Idiopathic Pulmonary Fibrosis. *International Journal of Experimental Pathology*, 94 (5), A15.

Abstract

Altered methylation and subsequent changes in gene expression have been implicated in several fibroses including lung however, the full extent and role of altered DNA methylation in fibrotic lung fibroblasts is unknown. Emerging evidence also suggests gender-specific methylation differences are common in disease and could elucidate why diseases characterised by pulmonary fibrosis including idiopathic pulmonary fibrosis (IPF) and systemic sclerosis (SSc) have a sex-biased prevalence. Using a genome-wide array-based approach, this thesis investigates differentially methylated and expressed genes in fibrotic compared to control lung fibroblasts, gender-specific methylation and expression differences and the effects of modulating DNA methylation using a DNA methyltransferase (DNMT) inhibitor, 5-Aza-2'deoxycytidine (5-Aza). Data show primary human IPF and SSc lung fibroblasts have multiple genes with altered DNA methylation and expression compared to control lung fibroblasts. Multiple biological processes were enriched in these genes, many of which are relevant to fibrosis including, transcriptional regulation, extracellular matrix (ECM) organisation, Wnt signalling and apoptosis. Using siRNA knockdown and collagen gel contraction assays, novel genes including Tenascin-XB (TNXB), which encodes the ECM glycoprotein Tenasicn-X (TNX), were identified as having potential functional significance in the pathogenesis of pulmonary fibrosis. Furthermore, multiple genes including TNXB had altered methylation and expression in IPF compared to SSc lung fibroblasts and may distinguish IPF from other diseases associated with pulmonary fibrosis. Multiple genes were identified with gender-specific differences in methylation and expression in lung fibroblasts. Interestingly, multiple genes with altered methylation in IPF males compared to control males were not the same genes with altered methylation in IPF females compared to control females, which may in part explain why IPF predominates in males. The final chapter of my thesis shows 5-Aza treatment alters the methylation and expression of multiple genes in primary human lung fibroblasts. Strong correlation between changes in methylation and changes in expression were identified suggesting DNA methylation can directly regulate the expression of multiple genes in lung fibroblasts.

Contents

List of figures

List of tables

Abbreviations

5-Aza - 5-Aza-2'-deoxycytidine
CF – Control female
CGI – CpG Island
CM – Control male
CpG – Cytosine-phosphate bond guanine
DNMT – DNA methyltransferase
EMT – Epithelial-mesenchymal transition
FC – Fold change
GO-term – Gene Ontology term
IF – IPF female
ILD – Interstitial lung disease
IM – IPF male
IPF – Idiopathic pulmonary fibrosis
LFs – Lung fibroblasts
NSIP – Non-specific interstitial pneumonia
PFAM – Protein families
siRNA – small interfering RNA
SSc – Systemic sclerosis
SSc-PF – Systemic sclerosis with pulmonary fibrosis
TET – Ten-eleven translocation enzymes
TNoM – Threshold number of misclassifications
TNX - Tenascin
TNXB – Tenascin XB
TSS – Transcription start site
UIP – Usual interstitial pneumonia
WNT – Wingless-Type MMTV Integration Site Family, Member
Δβ – Delta-beta (Change in methylation between two groups)
450k – Illumina Infinium HumanMethylation 450 BeadChip array
27k – Illumina Infinium HumanMethylation 27 BeadChip array
HT12v4 – Illumina HumanHT-12 v4 Expression BeadChip array

Chapter 1. Introduction

1.1. Anatomy and physiology of the normal lungs and airways

The human lung consists of an intricate branching network of airways starting at the trachea which links the lungs to the nasopharyngeal space. The trachea extends downwards from the larynx to enter midway in the thorax and terminates into two main bronchi. The right bronchus is wider than the left and based on their relation to the left and right pulmonary arteries, the right main bronchus is eparterial (above the pulmonary artery) whereas the left main bronchus is hyparterial (below the pulmonary artery). The bronchi transition into bronchioles which then lead to the terminal bronchioles. Terminal bronchioles are the last non-alveoliated airways, beyond which are respiratory bronchioles. Respiratory bronchioles have a diameter of less than 0.5 mm and terminate in acini containing multiple clusters of alveoli surrounded by alveolar walls. These alveolar walls contain capillaries, alveolar epithelium, capillary endothelium and extracellular matrix. This forms the gas-exchange region of the lung which contains approximately 300 million alveoli (Ochs et al, 2004) with a gas-exchange surface area between 40 and 80 m^2. The main cell populations located within this gas-exchange region include type I alveolar epithelial cells (AECIs), which form >90% of the gas-exchange surface area, type II alveolar epithelial cells (AECIIs), alveolar capillary endothelial cells and interstitial cells such as fibroblasts and macrophages (**Figure 1.1.1**) (Crapo et al, 1982).

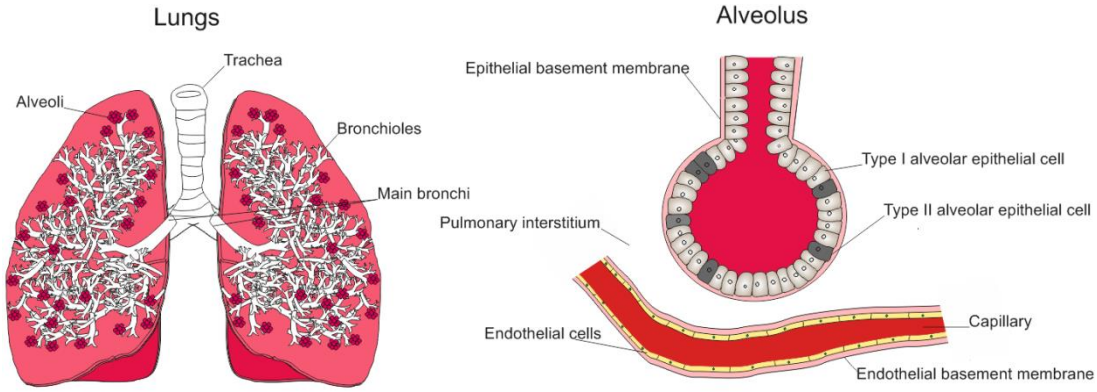


Figure 1.1.1. Anatomy of a normal lung. Right: basic anatomy of the lungs. Left: a normal alveolus. Type I alveolar epithelial cells interface with pulmonary capillaries facilitating efficient gas exchange. Type II alveolar epithelial cells also interface with pulmonary capillaries and are multifunctional cells which secrete pulmonary surfactant (important for pulmonary compliance, reducing surface tension and innate immunity) and multiple molecules involved in innate host defence.

1.1.1. ECM in the normal lung

The ECM is the non-cellular component present within all tissues and organs and is essential for tissue homeostasis and development. The physical, topological and biochemical composition of the ECM is tissue-specific and markedly heterogeneous (Frantz et al, 2010). In the lung the ECM forms the three-

dimensional scaffold of the alveolar walls, composed of a layer of epithelial and endothelial cells, their basement membranes and a thin layer of interstitial space between unfused capillary endothelium and alveolar epithelium basement membranes (West and Mathieu-Costello, 1999) (**Figure 1.1.1.1**). The pulmonary interstitium provides the elastic recoil of the lung and its ECM, which is located between the thin and thick sides of the alveolar septum, is formed by and composed of multiple stromal cells including fibroblasts and multiple macromolecules including, collagens (Bradley et al, 1974), elastin (Rucker and Dubick, 1984, Starcher, 1986, Mecham et al 1991), proteoglycans (Bensadoun et al, 1996) and glycoproteins (Bhattacharyya et al, 1975) (**Figure 1.1.1.1**). In total, approximately 300 proteins, multiple ECM-modifying enzymes and multiple ECM-binding growth factors make up the core matrisome for all ECMs which co-operate to assemble, maintain and remodel the ECM (Hynes and Naba, 2012).

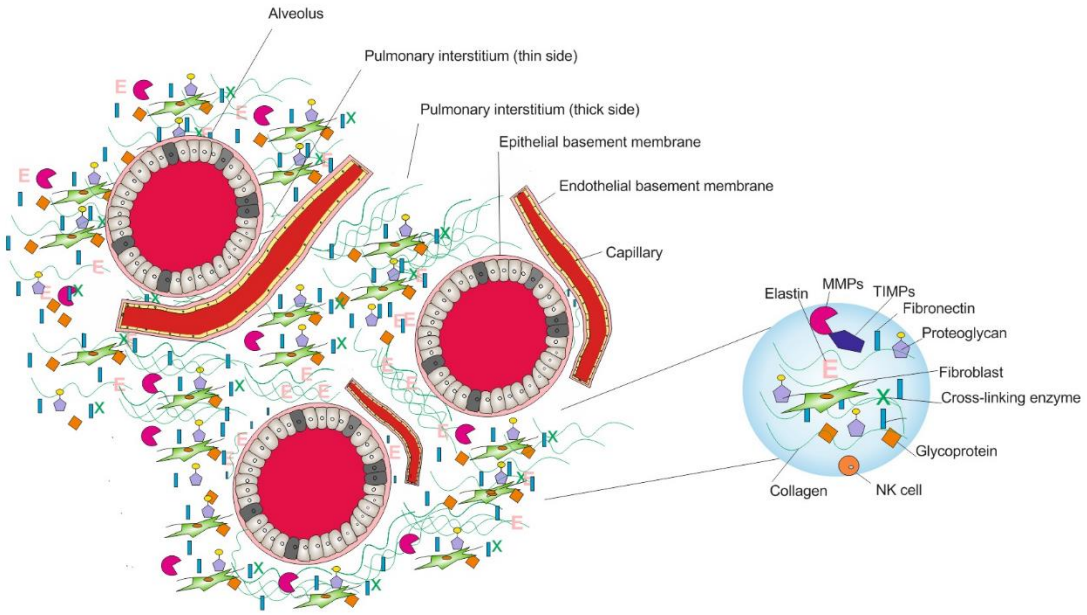


Figure 1.1.1.1. Alveoli and the pulmonary interstitium. The ECM is located between the alveolar epithelial basement membranes and the capillary endothelial basement membranes at the thick sides of the pulmonary interstitium or where gas exchange takes place (thin sides of the pulmonary interstitium). The ECM contains multiple proteins, growth factors and different interstitial cell types including, macrophages, natural killer (NK) cells and fibroblasts.

1.1.2. ECM turnover in the lungs

The ECM is a highly dynamic structure which undergoes rapid rates of synthesis and degradation. Collagens are continually synthesised and degraded throughout life (Laurent, 1982, McAnulty and Laurent, 1987). Peak synthesis occurs during the perinatal stage of development although collagen deposited in the lung increases during mammalian growth, with a forty-fold change in collagen content between birth and adulthood (Mays et al, 1989). Studies on rat and rabbit lungs have shown

lung collagen is subjected to a turnover rate of approximately 10% a day, which suggests small daily changes in ECM turnover can result in large changes to the composition of the ECM over a longer period of time (McAnulty and Laurent, 1987). Within the majority of tissues, collagens are the most abundant proteins found in the ECM. Elastins play a critical role in development of the respiratory system and function in the lung to support the expansion and recoil of the alveoli during breathing. There is minimal turnover of elastin in normal healthy lungs (Shapiro et al, 1990) however, aberrant expression of proteases such as metalloproteinases (MMPs) which can cleave elastin fibres, can result in an increased turnover of elastin and a loss of elasticity (Ashworth et al, 1999). MMPs also play important roles in multiple biological processes including apoptosis and angiogenesis which are essential for normal lung function and in modulating the activity of growth factors/receptors (Pardo and Selman, 2012).

1.1.3. Functions of the ECM in the lung

In the lung, a diverse array of ECM components including collagens, elastin, fibronectin, proteoglycans and glycoproteins support the physical, biochemical and biomechanical functions of the ECM. Under normal conditions these ECM components form a compliant meshwork which allows the lungs to function correctly. The pulmonary ECM provides a physical scaffold for cells, maintains tissue integrity and generates mechanical, tensile and compressive strength (Frantz et al, 2010). In addition to providing structural support, the ECM physically acts as a barrier and anchorage site for cells to adhere to and organise into functional units (Lu et al, 2012). The mechanical properties of the ECM are important for a number of cellular functions including cell proliferation, differentiation, survival and migration (Wells, 2008). The pulmonary ECM is also biologically active and can control cell behaviour by either binding molecules such as growth factors or adhesion molecules or by interacting with cell-surface receptors to generate signal transduction and regulate gene expression (Frantz et al, 2010, Cox and Erler 2011, Kim et al, 2011). Cell-surface receptors belonging to the integrin family are the primary mechanosensors which form key cell-matrix communications and response to physical forces transmitted by surrounding ECM and neighbouring cells (Janoštiak et al, 2014). The ECM also sequesters multiple growth factors and cytokines, thus acting as a local 'reservoir ' that can be used rapidly if needed without the need for de-novo synthesis.

1.1.4. The composition of the main ECM components

1.1.4.1. Collagens

The pulmonary interstitium is mainly comprised of the fibrous collagen type I and together with laminins, gives mechanical strength to the lungs (Frantz et al, 2010). Collagens are the most abundant proteins in the interstitial ECM and provide structural strength to all forms of extracellular matrices. They also regulate cell adhesion, cell migration, support chemotaxis and are important in tissue repair and development (Rozario and DeSimone, 2010). Collagens consist of three polypeptide alpha chains.

Each polypeptide chain contains the amino sequence Gly-x-y where 'X' and 'Y' are predominately proline and hydroxyproline respectively (van der Rest and Garrone, 1991) (**Figure 1.1.4.1.1**). It is via the Gly-x-y sequence that the three polypeptide chains bind via hydrogen bonds, forming triple helices. Collagens are initially synthesised as pro-collagens that require cleavage to become active. To date, 28 collagens have been identified which can be broadly divided into either fibrillar or non-fibrillar (Heino, 2007, Gordon and Hahn, 2010). Fibrillar collagens, of which 11 are known to exist, have structural and mechanical functions and can influence cell behaviour by serving as ligands for receptors and as a reservoir of growth factors (Huxley-Jones et al, 2007). I, II, III, XI are the major fibrillar collagens in ECM, all of which are found within the pulmonary interstitium. Types I and III are the most abundant collagens in the human lungs, totalling > 90% (Bateman et al, 1981, Kirk et al, 1984). Non-fibrillar collagens such as type IV form intricate networks and are essential components of basement membranes together with laminins (Boute et al, 1996, Exposito et al, 2010). Fibroblasts transcribe and secrete the majority of interstitial collagen and can organise collagen fibres by exerting tension on the matrix (De Wever et al, 2008). Covalent cross-linking of collagen fibrils is essential for the normal mechanical properties of ECM, however increased cross-linking causes the stiffness of the ECM to increase. This changes the mechanical properties of the ECM which has been linked to a number of diseases including pulmonary fibrosis (Calderwood et al, 2012, Clarke et al, 2013).

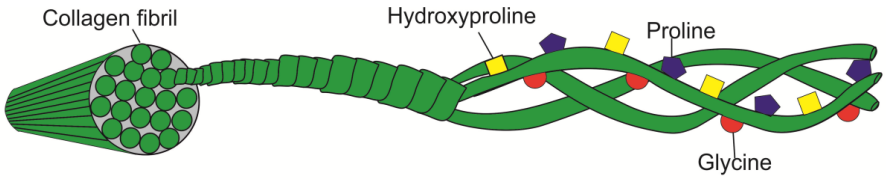


Figure 1.1.4.1.1. Structure of collagen. Collagens are composed of three polypeptide alpha chains each of which contain the Gly-X-Y motif. X and Y can be any amino acid although they are predominately proline and hydroxyproline. Multiple collagen fibres form a collagen fibril.

1.1.4.2. Proteoglycans

Of the ~300 proteins that make up the core matrisome, 36 are proteoglycans (Hynes and Naba, 2012). Proteoglycans are glycosylated proteins that can covalently attach to glycosaminoglycans (GAGs) (Yanagishita, 1993) which play important functional roles in regulating cell adhesion, migration and proliferation (Wight et al, 1992, Järveläinen et al, 2009). The addition of GAGs also gives proteoglycans buffering and hydrating properties (Hynes and Naba, 2012). Proteoglycans such as decorin, aggrecan, versican and perlecan are interspersed among collagen fibrils in different ECMs (Hynes and Naba, 2012) and can directly interact with growth factors and cytokines such as TGFß. Other proteoglycans, which can be membrane bound, also interact with chemical signals and proteases and can function as co-receptors thus playing an important role in regulating their biological activity (Yu and Woessner et al, 2000, Ruiz et al, 2012).

Proteoglycans can be separated into different multiple families (Iozzo and Murdoch, 1996). The two main families are those which encompass proteoglycans containing leucine rich repeats (LRRs) and those that contain link and C-type lectin domains (hyalectans) (Hynes and Naba, 2012). Proteoglycans with LRRs can bind to multiple glycoproteins. For example, the GAG chain of decorin can bind to tenascin-X and mediate its interaction with collagen fibrils (Merline et al, 2009). Hyalectans include versican, aggrecan, brevican and neurocan, all of which can bind to hyaluronic acid (Hynes and Naba, 2012) and play important roles in cell adhesion, migration and apoptosis (Wu et al, 2005).

1.1.4.3. Glycoproteins

Approximately 200 glycoproteins are associated with the matrisome with tenascins, laminins, fibronectins, thrombospondins and fibrillins being some of the most well studied (Hynes and Naba, 2012). Glycoproteins represent a diverse group of proteins that are covalently bound to carbohydrate motifs. They have important functions which include maintaining cell structure, mediating cell adhesion, migration and can have enzymatic and inhibitory activity. Fibronectin is one of the most common protein domains in vertebrates (Alberts, 2002) and alternative splicing of fibronectin results in a variety of isoforms. The type III fibronectin repeat is the most common form which binds to integrins (Carr et al, 1997). Integrins are cell-surface heterodimeric receptors consisting of an alpha and beta glycoprotein subunit (Akiyama, 1996). They link the ECM with the the intracellular cytoskeleton (Plow et al, 2000) and are key mechanosensors, essential in cell signalling and can regulate multiple molecules such as transforming growth factor beta (TGFß) in response to feedback from the ECM (Munger et al, 1999, Shyy and Chien, 2002, Yang et al, 2007).

1.2. The role of fibroblasts in the ECM

The essential role fibroblasts play in maintaining ECM homeostasis by regulating ECM turnover is highlighted by their ability to secrete both MMPs which degrade the ECM and tissue inhibitors of metalloproteinases (TIMPs) which inhibit ECM degradation (Gomez et al, 1997). MMPs/TIMPs are also involved in the regulation of growth factors and cytokine/chemokine activity (Elkington and Friedland, 2006, Wynn, 2007). Furthermore, each fibroblast is capable of secreting more than 5000 molecules of pro-collagen every minute in response to biochemical or mechanical stimuli (McAnulty et al, 1991, Lindahl et al, 2002). Fibroblasts are located in the majority of tissues that are associated with extracellular molecules and constitute the principle cellular component of connective tissue (Tarin and Croft, 1970). They are large, elongated, flat spindle-shaped cells and are characterised by their morphology, ability to adhere to plastic (in vitro) and expression of the intermediate filament associated protein vimentin in the absence of markers associated with other cell lineages (McAnulty, 2007). Multiple ECM proteins including collagens, elastin, proteoglycans, tenascins, laminins and fibronectin, as well as a number of ECM-modifying enzymes including collagen cross-linking enzymes and ECM-degrading enzymes are synthesised by fibroblasts which are essential to numerous biological processes. These include wound healing and repair (Tomasek et al, 2002), regulation of inflammation,

angiogenesis, ECM synthesis and deposition and homeostatic regulation of the ECM (McAnulty 1991, McAnulty, 1995, McAnulty 2002, McAnulty, 2007) (**Figure 1.2.1**).

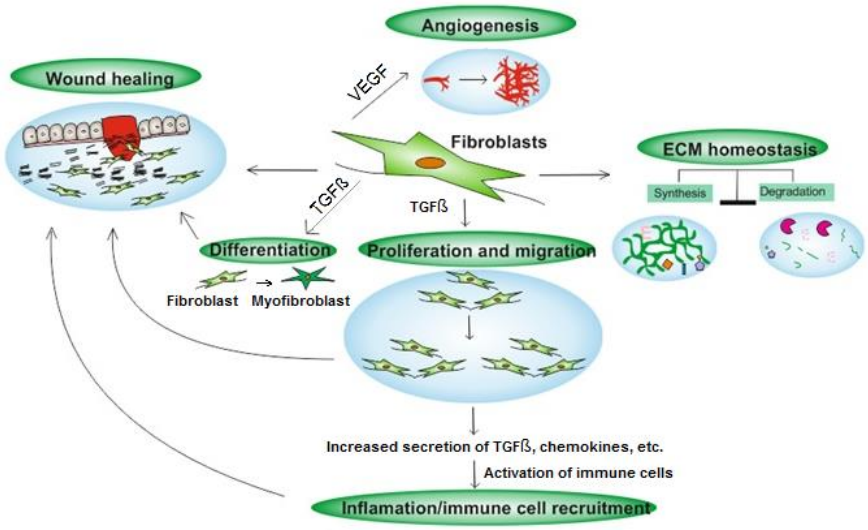


Figure 1.2.1. Role of fibroblasts in the ECM. In would healing fibroblasts proliferate and chemotax to sites of injury to repair the damaged ECM. In inflammation fibroblasts secrete and respond to factors including cytokines and prostaglandins which facilitates immune cell recruitment. In angiogenesis, fibroblasts interact with endothelial cells to induce angiogenesis into tissues which are not accessible to existing blood vessels.

1.2.1. Effect of cytokines on fibroblasts

Multiple cytokines have activating or inhibitory effects on fibroblasts, many of which are produced by fibroblasts themselves (Scotton and Chambers, 2007, Maher et al, 2010). TGFß is the prototypic pro-fibrotic cytokine which is produced by a variety of cells including fibroblasts and acts on fibroblasts and myofibroblasts to induce proliferation, differentiation, cell migration and matrix production (Scotton and Chambers, 2007). Platelet-derived growth factor (PDGF) is another pro-fibrotic cytokine produced by cells including platelets and macrophages which acts similarly to TGFß in context with inducing fibroblast proliferation, differentiation and ECM production (Scotton and Chambers, 2007). Whilst these cytokines are essential for facilitating numerous fibroblast functions (Scotton and Chambers, 2007) , persistent or overexpression can lead to fibrosis as seen in patients with IPF or pulmonary fibrosis associated with systemic sclerosis (SSc-PF) (LeRoy et al, 1989, Yamane et al, 2002). Other pro-fibrotic factors including vascular endothelial growth factor (VEGF), which is overexpressed in SSc, may support fibrosis by causing vascular damage which consequently causes fibroblast activation (Kajihara et al, 2013). In contrast, the anti-fibrotic cytokine prostaglandin E2 (PGE2), which inhibits fibroblast proliferation, collagen production (McAnulty et al, 1997) and apoptosis of alveolar epithelial cells (AECs) (Maher et al, 2010) is decreased in IPF and SSc-PF, highlighting the importance of balancing the correct levels of pro-fibrotic and anti-fibrotic cytokines (**Figure 1.2.1.1**).

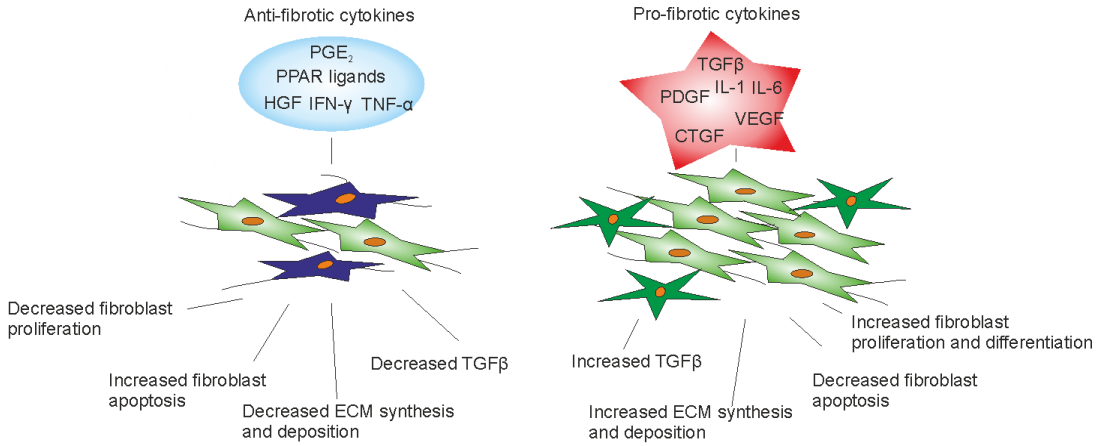


Figure 1.2.1.1. Effect of cytokines on fibroblasts. Anti-fibrotic cytokines including PGE2 can act to decrease fibroblast proliferation and their differentiation into myofibroblasts and increase their apoptosis. Other anti-fibrotic cytokines include: hepatocyte growth factor (HGF), interferon gamma (IFN-γ) and tumor necrosis factor alpha (TNF-α). Pro-fibrotic cytokines including TGFß and VEGF can activate multiple genes and increase fibroblast proliferation and differentiation. Other pro-fibrotic cytokines include: connective tissue growth factor (CTGF), interleukin-1 (IL-1) and interleukin-6 (IL-6).

1.3. ECM in fibrosis

Excessive collagen production and remodelling of the lung architecture is a consequence of altering the balance of fibroblast proliferation, apoptosis and the accumulation and breakdown of the ECM. Under fibrotic conditions such as those found in pulmonary fibrosis, the ECM changes in composition, severely affecting lung function (**Figure 1.3.1**).

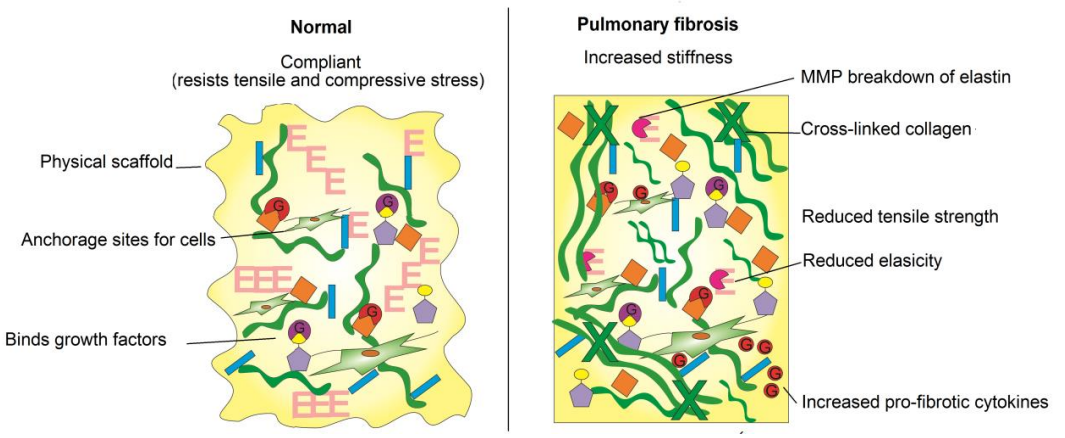


Figure 1.3.1. Normal ECM function and structure compared to fibrotic ECM. The ECM in the normal lung is compliant and has very little elastin turnover. In contrast, a fibrotic lung has a high turnover of elastin, increased cross-linking of collagen and other ECM components, dysregulated MMP/TIMP levels and altered composition of multiple ECM components which leads to an inelastic, non-compliant, stiff matrix.

Type I and III collagens are elevated in IPF patients (Laurent et al, 1988) and dramatic increases in collagen synthesis have been observed in a number of animal models following treatment with bleomycin (Clark et al, 1982, Laurent and McAnulty et al, 1983, Decaris et al, 2014), suggesting that altered turnover of collagen contributes to pulmonary fibrosis. Furthermore, increased collagen cross-linking which results in a stiff matrix has been associated with IPF and SSc (Kadler et al, 2007, Kadler et al, 2008, Olsen et al, 2011, Cox et al, 2013, Ho et al, 2014). Cross-linking is a biochemical process mediated by cross-linking enzymes including transglutaminases (TG) and lysyl-oxidases (LOX) which stabilise and allow correct assembly of ECM components (Kadler et al, 2007, Kadler et al, 2008). Transglutamase 2 (TG2), an enzyme responsible for cross-linking ECM components including collagen, elastin and fibronectin, can alter ECM properties by increasing its resistance to degradation (Olsen et al, 2011). Furthermore, lysyl-oxidase expression is increased after bleomycin-and irradiation-induced pulmonary fibrosis and is strongly associated with areas of fibrosis (Cox et al, 2013). Elastin gene expression is also increased in pulmonary fibrosis (Hoff et al, 1999), after bleomycin treatment in mice (Decaris et al, 2014) and in response to TGFß (Kähäri et al, 1992). Aberrant, disorganised elastin production and increased degradation is also associated with IPF (Laurent and Tetley, 1984, Kristensen et al, 2015). In IPF and SSc-PF, a number of MMPs and TIMPs are aberrantly expressed (**Table 1.3.1**) (Selman et al, 2000, Zuo et al, 2002, Henry et al, 2002, Pardo et al, 2005, Selman et al, 2006, McKeown et al, 2009, Konishi et al, 2009, Zhou et al, 2010). Furthermore, knockout MMP models confer protection against bleomycin-induced pulmonary fibrosis (McKleroy et al, 2013), suggesting the balance of MMP/TIMP expression is important in maintaining ECM function. Multiple glycoproteins also have altered expression in IPF and SSc including, thy-1 cell surface antigen (THY1) (Hagood et al, 2005, Sanders et al, 2008), mucin 1 (MUC1) (Kohno et al, 1993, Yokoyama et al, 2006) and tenascin-C (TNC) (Kuhn and Mason, 1995) suggesting they may be important in the pathogenesis of these diseases.

Tenascins are a family of large glycoproteins found in the ECM and are responsible for a variety of functions including maintaining ECM structure, cell signalling and cell adhesion modifications (Jones and Jones 2000). There are currently four tenascin family members; TNC, tenascin-R, tenascin-W and tenascin-X (TNX). TNC has been linked to IPF and SSc-PF (Kuhn and Mason, 1995, Hisatomi et al, 2009, Brissett et al, 2012, Estany et al, 2014). TNX is encoded by the TNXB gene and is the largest of all tenascins. TNX is believed to play an important role in regulating collagen fibrillogenesis (Minamitani et al, 2004) and can bind to a number of collagens including fibril-associated type XII and XIV collagens (Lethias et al, 2006), elastin (Egging et al, 2007) and decorin (Elefteriou et al, 2001). Deficiency of TNX is associated with hypermobility type Ehlers Danlos syndrome which is characterised by joint hypermobility and skin hyperextensibility (Burch et al, 1997, Mao et al, 2002, Zweers et al, 2003). TNX deficiency is also associated with reduced type I collagen deposition, reduced density of collagen fibrils (Mao et al, 2002), and disruption of the elastic fibre network (Burch et al, 1997). Furthermore, knockout models of TNXB have shown protection against cardiac fibrosis in rat models, mediated by

a decrease in TGFß and an increase in PPARγ (Jing et al, 2011) whereas increased levels of TNX are associated with increased matrix stiffness (Margaron et al, 2010) and fibrous tumours including mesothelioma (Yuan et al, 2009). A list of ECM components which have altered expression in IPF and SSc is shown in **Table 1.3.1.**

ECM component	Disease	Expression	Function
COL1	IPF SSc	Increased Increased	Provides tensile strength. (Laurent et al, 1988).
COL3	IPF SSc	Increased Increased	Provides tensile strength. (Laurent et al, 1988).
Chitinase-3-like protein 1 (CHI3L1)	IPF SSc	Increased Increased	Tissue remodelling and stimulates fibroblast proliferation (Recklies et al, 2002, Nordenbaek et al, 2005, Johansen et al, 2005, Korthagen et al, 2011).
Superoxide dismutase 3 (SOD3)	IPF SSc	Increased Increased	Involved in inflammatory responses (Gao et al, 2008, Laurila et al, 2009, Arcucci et al, 2011).
Secreted protein, acidic, cysteine-rich (SPARC)	IPF SSc	Increased Increased	Suppresses apoptosis of IPF fibroblasts (Zhou et al, 2002, Zhou et al, 2006, Chang et al, 2010).
MUC1	IPF SSc	Increased Increased	Multiple functions including regulation of growth factors, inflammation and apoptosis (Kohno et al, 1993, Yamane et al, 2000, Sato et al, 2000, Ohnishi et al, 2003, Yokoyama et al, 2006, Nath and Mukherjee, 2014).
TG2	IPF	Increased	Cross-links ECM (Olsen et al, 2011).
LOX	IPF	Increased	Cross-links ECM (Cox et al, 2013).
TNC	IPF SSc	Increased Increased	Cell adhesion, fibroblast migration (Trebaul et al, 2007, Brissett et al, 2012).
MMP-1	IPF SSc	Increased Decreased	Processes cytokines, cell migration, cell growth, role unknown in IPF and SSc (Pardo et al, 2012, Herrera et al, 2013).
MMP-2	IPF	Increased	Degrades matrix and non-matrix components (Pardo and Selman, 2006).
MMP-3	IPF	Increased	May contribute to epithelial-mesenchymal transition (EMT) (Radisky et al, 2010, Yamashita et al, 2011) and endostatin release promoting alveolar epithelial cell apoptosis (Richter et al, 2009).
MMP-7	IPF SSc	Increased Increased	Pro-fibrotic, may promote fibrosis via activation of latent TGFß (Pardo et al, 2005, Rosas et al, 2008).
MMP-8	IPF	Increased	Migration of fibrocytes (Moeller et al, 2009, García-de-Alba et al, 2010).
MMP-9	IPF	Increased	Regulates tissue turnover (Lemjabbar et al, 1999, Selman et al, 2000, Suga et al, 2000).
MMP-10	IPF SSc	Increased Increased	Unknown.
MMP-12	IPF SSc	Unknown Increased	May promote fas-induced pulmonary fibrosis (Matute-Bello et al, 2007). Cleaves urokinase-type plasminogen activator receptor causing impaired endothelial cell proliferation, migration and angiogenesis (D'Alessio et al, 2004).
TIMP-1,2,3 and 4	IPF SSc	Increased Increased	Regulates MMP activity and modulates cell functions including apoptosis and cell differentiation (Ramos et al, 2000, Selman et al, 2000, Ries et al, 2014).

Table 1.3.1. ECM Components which have altered expression in IPF and SSc. Multiple collagens, cross-linking enzymes, glycoproteins, metalloproteinases and tissue inhibitors of metalloproteinases have altered expression in IPF and/or SSc.

1.4. Pulmonary fibrosis

Pulmonary fibrosis (PF) is a pathological condition originating from aberrant repair mechanisms in response to acute or chronic lung injury. PF is characterised by progressive scarring of the lungs and excessive ECM deposition in the pulmonary interstitium. The pulmonary interstitium comprises alveolar septa surrounding the bronchial and vascular spaces which are critical areas within the lung where fibrosis exerts its devastating effects. Effacement of this intricate lung architecture by excessive production of ECM, mainly collagens, results in scarring and ultimately leads to decreased lung function, respiratory failure and subsequent death.

Hamman and Rich are seen as the first people to describe pulmonary fibrosis although it has been argued that previous papers in Germany described a number of fibrotic lung diseases before them (Homolka, 1987). In the early 20th century pulmonary fibrosis was described in four patients who displayed rapidly progressive diffuse interstitial fibrosis of the lungs (Hamman and Rich, 1935, Hamman and Rich, 1944). It was later found that not all patients deteriorated as rapidly as Hammond and Rich had first described. This led to distinct types of pulmonary fibroses being described such as diffuse fibrosing alveolitis and idiopathic pulmonary fibrosis (Scadding et al, 1960, Scadding et al, 1967, Crystal et al, 1976).

PF epitomises the end stages of many interstitial lung diseases (ILDs) which describe a diverse heterogeneous group of over 200 pulmonary disorders (Raghu and Brown, 2004). Similar radiological and clinical manifestations are common in ILDs although their aetiologies and pathophysiologies are distinct. ILD and PF often occur as secondary effects of other diseases including rheumatoid arthritis (RA) (Shiel and Prete, 1984), systemic lupus erythematosus (Eisenberg et al, 1973) and SSc (McCarthy et al, 1988, Wells et al, 1994) or from exposure to substances such as asbestos or silica dust (Khalil et al, 2007).

The vast majority of ILDs have no known etiology and are sub-categorised as idiopathic interstitial pneumonias (IIPs), the most prevalent and deleterious being IPF (Amercian Thoracic Society/European Respiratory Society (ATS/ERS), 2002, Travis et al, 2013). Although sufficiently different to warrant designation as separate diseases, many IIPs share similar features. In 2002 the ATS/ERS categorised IIPs into seven groups based on histological, radiological and clinical characteristics (ATS/ERS 2002). Since then the classification of IIPs has been updated (**Figure 1.4.1**) (Travis et al, 2013). This revision of IIP classification includes, recognition that idiopathic NSIP is a distinct clinical entity, sub-divisions of IIPs into major, rare and unclassifiable and grouping major IIPs into either chronic fibrosing, smoking-related or acute/subacute (Travis et al, 2013).

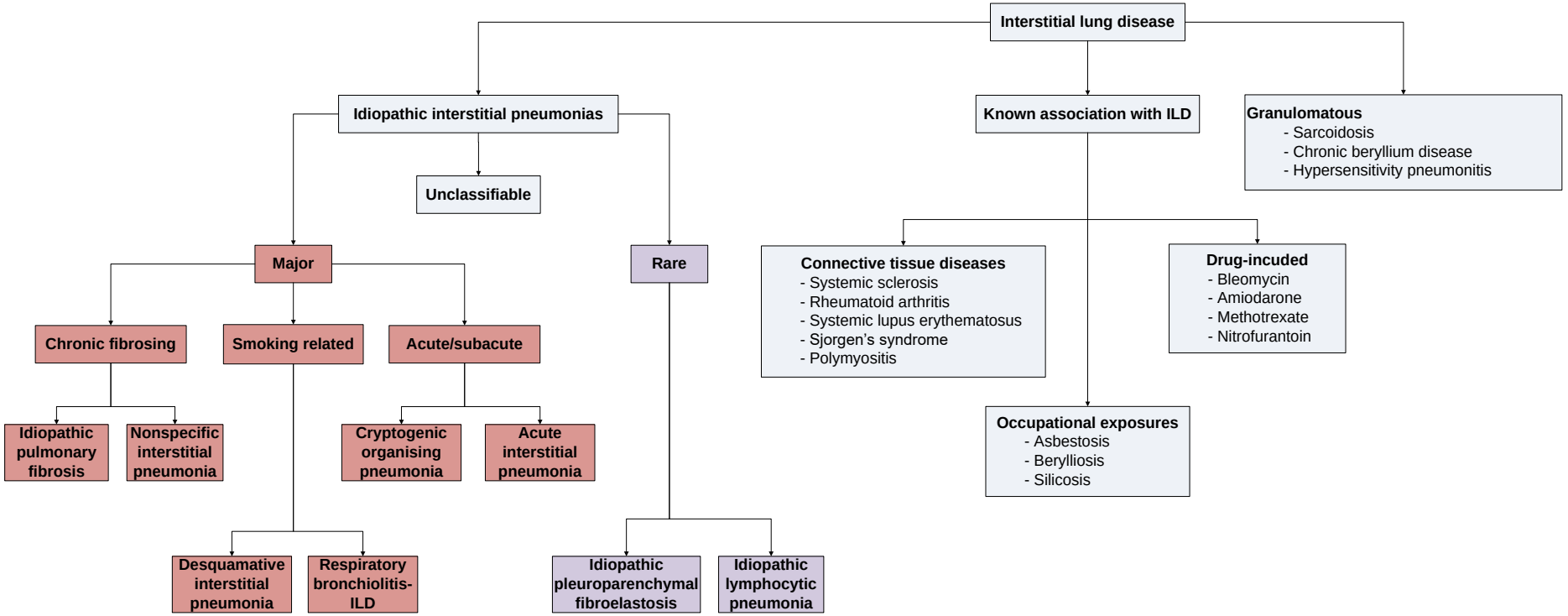


Figure 1.4.1. Classification of ILDs including the revised ATS/ERS classification of idiopathic interstitial pneumonias. (Adapted from Zibrak and Price, 2014).

1.4.1. IPF

IPF is a chronic, progressive fibrotic lung disease which is largely unresponsive to any therapy and remains the most common and deleterious of all IIPs, accounting for 62% of all diagnosed IIP cases (Bjoraker et al, 1998). Manifestations of IPF include, breathlessness on exertion, dry cough and increasing dyspnoea, bibasilar inspiratory crackles and 50% of IPF patients develop thickening of their fingertips, termed 'finger clubbing' (Raghu et al, 2011). IPF is clinically characterised by progressive dyspnoea, dry cough, presence of sub-pleural honeycombing upon high-resolution computed tomography (HRCT) scanning and absence of any known causes of other ILDs (**Table 1.4.1.1**). IPF is also associated with a histological pattern of usual interstitial pneumonia (UIP) (**Figure 1.4.1.1**), although it is not uncommon for UIP and non-specific interstitial pneumonia (NSIP) to exist in different regions of IPF lung biopsies (Monaghan et al, 2004). UIP has the worst prognosis of all IIPs whereas NSIP has a slightly better prognosis and is seen more often in PF associated with connective tissue diseases such as SSc and RA (Wells et al, 1997, Nicholson and Wells, 2001, Bouros et al, 2002). UIP is characterised by abnormal proliferation of mesenchymal cells, honeycombing, co-existence of fibrotic regions with histologically normal lung, distortion of normal lung architecture and presence of fibroblastic foci. Fibroblastic foci are seen as a pathological hallmark characteristic of UIP thought to represent active sites of ongoing fibrogenesis where fibroblasts/myofibroblasts accumulate and contribute to excessive ECM production and collagen deposition (Kuhn and McDonald, 1991).

Major Criteria
Exclusion of other known causes of ILD such as certain drug toxicities, environmental exposures, and connective tissue diseases. Abnormal pulmonary function studies that include evidence of restriction (reduced vital capacity (VC), often with an increased forced expiratory volume in one second to forced vital capacity ratio (FEV1/FVC)) and impaired gas exchange (increased $P(A\text{-}a)O_2$, decreased P_aO_2 with rest or exercise or decreased diffusing capacity of the lungs for carbon monoxide (DLCO). Bibasilar reticular abnormalities with minimal ground glass opacities on HRCT scans. Transbronchial lung biopsy or bronchial alveolar lavage showing no features to support an alternative diagnosis.
Minor criteria
Age >50 years old. Insidious onset of otherwise unexplained dyspnoea on exertion. Duration of illness >3 months. Bibasilar, inspiratory crackles (dry type in quality).

Table 1.4.1.1. Clinical criteria required for IPF diagnosis. VC; vital capacity, FEV1/FVC; forced expiratory volume in 1 second/forced vital capacity, $P(A\text{-}a)O_2$; the alveolar-arterial oxygen difference, P_aO_2; partial pressure of oxygen in the arterial blood. (Taken and adapted from ATS, 2002).

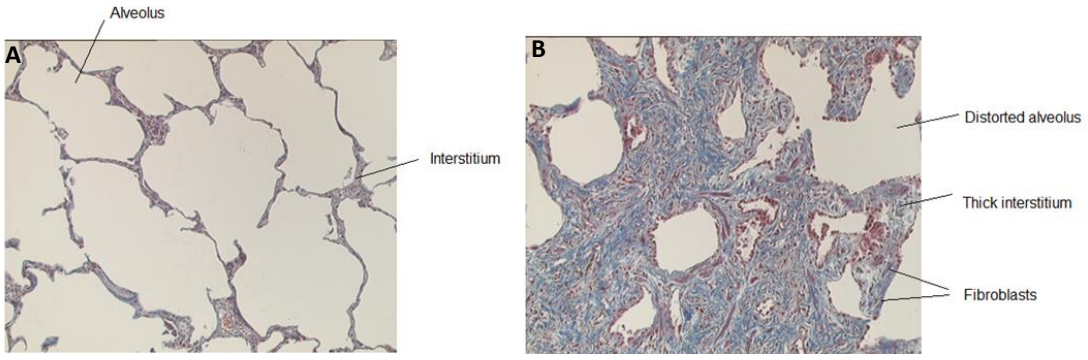


Figure 1.4.1.1. The histology of normal and IPF lung tissue. A) Normal lung tissue. Alveoli are separated by a thin interstitium conducive to lung function. B) IPF lung tissue. Increased collagen (blue staining) content in the ECM causes a thick interstitium and results in the distortion of the normal lung architecture. Brown staining shows cell nuclei.

1.4.2. SSc

SSc is a complex, autoimmune, connective tissue disease, characterised by vascular abnormalities, excessive collagen production and extensive fibrosis of the skin and internal organs (Varga and Abraham, 2007). SSc vasculopathy consists of fibrointimal proliferation of small vessels and Raynaud's phenomenon, which describes cold-induced or stress-induced excessive reduction in blood flow to the extremities, due to vasospasm of small blood vessels (Lewis, 1929, Kahaleh, 2009. ILD occurs in approximately 75% of SSc patients (Bussone and Mouthon, 2011, Tan et al, 2011), including pulmonary arterial hypertension (PAH) and PF which are the leading causes of mortality (Steen and Medsger, 2007).

The exact etiology of SSc is unknown although there is strong evidence that genetic factors, epigenetics and environmental exposures all play an important role (Varga and Abraham, 2007). Inflammation, diffuse endothelial damage, immune abnormalities, microvascular alterations and interstitial and perivascular fibrosis are all characteristic of SSc (Abraham and Varga, 2005, Varga and Abraham, 2007). Depending on the pattern of skin involvement (Varga and Abraham, 2007), SSc can be classified as either diffuse cutaneous SSc (dcSSc) or limited cutaneous SSc (lcSSc). Patients with lcSSc outnumber those with dcSSc by 4.7:1 (Allcock et al, 2004). This classification also in part reflects the severity of SSc. DcSSc is associated with rapid, progressive skin fibrosis and systemic organ involvement (LeRoy et al, 1988, Steen and Medsger 2000, LeRoy and Medsger, 2001). Approximately 20-30% of patients with dcSSc also have the presence of anti-topoisomerase antibodies (ATA) (Steen et al, 1988, Koenig et al, 2008). ATA antibodies are associated with increased mortality and risk of developing pulmonary fibrosis (Steen et al, 1988, Koenig et al, 2008). In contrast, patients with lcSSc tend to have a relatively better prognosis than dcSSc patients, unless PAH develops as a secondary complication (Steen and Medsger, 2003). LcSSc is characterised by slow progressive skin changes with varying degrees of internal organ involvement (Varga and Abraham, 2005). Approximately 70-80% of patients with lcSSc have anti-centromere antibodies (ACA) which increases the risk of developing PAH

(Nihtyanova and Denton, 2010). Both ATA and ACA auto antibodies are useful for diagnosing and classifying SSc patients as although both dcSSc and lcSSc can have ATA antibodies, it is very rare that dcSSc patients have ACA antibodies (Spencer-Green et al, 1997).

1.5. Epidemiology of IPF and SSc

IPF typically affects adults over the age of 40 (Raghu et al, 2011) with a tendency to affect males more than females (Iwai et al, 1994, Scott et al, 1990, Coultas et al, 1994, Mannino et al, 1996, Johnston et al, 1997, Gribbin et al, 2006, Han et al, 2008, Nalysnyk et al, 2012) and whilst still classed as a rare disease, the incidence rates are increasing (Raghu et al, 2006, Gribben et al, 2006, Olson et al, 2007, Navaratnam et al, 2011, Nalysnky et al, 2012), although this may be due to improvements in the earlier and more accurate diagnosis of IPF (Hutchingson et al, 2014). Approximately 66% of IPF patients are older than 60, with a mean age of diagnosis at 66 years (Navaratnam et al, 2011). Age of diagnosis is also a predictor of survival rate, as increased age is strongly associated with an increased risk of death (Raghu et al, 2011, Navaratnam et al, 2011).

IPF predominates in males (Iwai et al, 1994, Scott et al, 1990, Coultas et al, 1994, Mannino et al, 1996, Johnston et al, 1997, Gribbin et al, 2006, Han et al, 2008, Nalysnyk et al, 2012) and may also progress faster in males resulting in decreased survival rates (Gribbin et al, 2006, Han et al, 2008). In contrast, it has been shown the mortality rate amongst women is increasing at a faster rate than in men (Olson et al, 2007), although this may potentially be due to changes in working environments or smoking habits (Ley and Collard, 2013). It has also been shown that race and ethnicity may play a role in susceptibility towards developing IPF although different studies have produced conflicting results (Olson et al, 2007, Swigris et al, 2012). Recent studies show greater and increasing mortality in white populations compared to other racial and ethnic groups (Olson 2007, Swigris et al, 2012). In contrast, a previous study has shown black and Hispanic populations have increased mortality compared to white populations (Lederer et al, 2006).

IPF affects more than 15,000 people in the UK and more than 128,000 people in the USA (Raghu et al, 2006, Navaratnam et al, 2011) with mean survival rates from the time of initial diagnosis dismal, at only 2-3 years (Gribben et al, 2006). In Europe, IPF prevalence ranges from 1.25-23.4 cases per 100,000 population (Nalysnyk et al, 2012). In the UK alone an estimated 5000 people suffering from IPF will die each year, making IPF a substantial cause of mortality in the UK and a more prominent killer than most cancers including leukaemia and kidney cancer (Navaratnam et al, 2011). Currently it is estimated that between 28,000-65,000 people in Europe and 13,000-17,000 people in the USA will die from IPF (Hutchinson et al, 2014, Hutchinson et al, 2015).

SSc typically affects people aged over 45 (Chifflot et al, 2008) and predominates in women (Mayes et al, 2003). The prevalence of SSc varies dramatically according to geographical location and the criteria used to diagnose SSc (Mayes et al, 2003, Chifflot et al, 2008). In European countries, such as France

(Le Guern et al, 2004) and Spain (Arias-Nuñezet al, 2008), the prevalence rate of SSc is estimated at 158 and 277 cases per million, respectively. The prevalence of SSc in the West Midlands of the UK is estimated at 31 cases per million (Silman et al, 1988), whereas in Newcastle, UK, the prevalence is estimated at 88 cases per million (Allock et al, 2000). In different parts of America including South Carolina (Maricq et al, 1989), Michigan (Mayes et al, 2003) and Oklahoma (Arnett et al, 1996) the prevalence is estimated at 286, 276 and 658 cases per a million, respectively. In general, SSc prevalence is estimated at 50-300 cases per million (Chifflot et al, 2008). Approximately 6000 people in the UK suffer from SSc whereas in America, approximately 100,000 people suffer from SSc (Mayes et al, 2003).

1.6. The pathogenesis of IPF and SSc

Intensive research into the etiology of IPF and SSc has identified multiple exogenous agents, pathways, transcription factors and risk factors that are potentially pathogenic. However, the exact molecular and cellular mechanisms underlying both diseases remain unknown. Lack of efficacy using anti-inflammatory drugs in IPF patients and deleterious effects of immunosuppressive therapy led some researchers to challenge the antecedental view of IPF being a chronic inflammatory lung disease (Selman et al, 2001, Raghu et al, 2006, Maher et al, 2007). IPF is now seen as a multifactorial disease caused by chronic epithelial injury with an impaired wound healing response resulting in fibrosis and inflammation with multiple pathways, cell types, genetic and epigenetic factors and the ECM all pathologically implicated (Selman et al, 2001, Chambers et al, 2003, Kinnula et al, 2005, Thannickal et al, 2006, Maher et al, 2007, O'Donoghue et al, 2012, Parker et al, 2014). SSc is a clinically heterogeneous disease and whilst genetic, epigenetic and environmental factors all influence SSc pathogenesis, the exact causes of SSc remain unknown (Kowal-Bielecka et al, 2009).

Much of the current understanding of IPF and SSc has come from the use of animal models of fibrosis, the murine bleomycin model and its variations being the most prominent. However, the bleomycin model fails to epitomise the hallmark characteristics of IPF (Scotton and Chambers, 2010) and SSc (Lakos et al, 2004) and with so many different variations of the bleomycin model, including route of bleomycin administration, dosing and measurements of outcome, it makes any generated results difficult to compare in the literature (Scotton and Chambers, 2010). Although the bleomycin model and other animal models of fibrosis fail to fully recapitulate IPF and SSc phenotypes they have provided researchers with an insight into possible mechanisms underlying IPF and in other diseases where PF often occurs and with careful use can prove a valuable resource.

The majority of people suffering from IPF experience a gradual worsening of their symptoms however, it has been estimated annually that 5-10% of IPF sufferers experience unexplainable rapid deteriorations, defined as acute exacerbations (Song et al, 2011, Collard et al, 2007). Differences in gene expression patterns between stable IPF and IPF with acute exacerbations has recently been reported, highlighting the protein cyclin–A2 (CCNA2) which is essential for controlling the cell cycle,

as being upregulated in IPF with acute exacerbations (Konishi et al, 2009). Recent work by Wootton et al suggests viral infections do not play a key role in acute exacerbations in IPF as in the majority of cases no viral infection was detected (Wootton et al, 2011). However, viral infections including Epstein-Barr virus (EBV) (Egan et al, 1995), hepatitis C (Ueda et al, 1992) and adenovirus (Kuwano et al, 1997) have all previously been implicated in IPF pathology. Furthermore, the number of studies associating viral signatures with IPF is growing (Moore and Moore, 2015), suggesting viral infections may play an important role in the pathogenesis of IPF.

Increased presence of fibroblastic foci has proved one of the most reliable predictors of IPF outcome, with increased numbers of fibroblastic foci correlating with disease severity, a worsened prognosis and decreased survival time (King et al, 2001, Nicholson et al, 2002, Barlo et al, 2010). Proliferation of fibroblasts, their differentiation and secretion of ECM components are normal primary responses to tissue damage (Tomasek et al, 2002) however, progressive, increased ECM deposition and altered ECM turnover rates are hallmark characteristics of many pulmonary diseases including IPF (Laurent et al, 2008) and SSc (Sato et al, 2003). In SSc, fibroblasts from lung explants have a constitutively activated, myofibroblast-like phenotype resulting from circulating auto-antibodies, connective tissue growth factor (CTGF) and IL-6 in response to viral infections (Markiewicz et al, 2004, Abraham and Varga, 2005, Varga and Abraham, 2007, Cox and Erler, 2010). The environmental, genetic and epigenetic factors which may play a role in the pathogenesis of IPF and SSc are summarised in **Figure 1.6.1**. The following sections will discuss each of these pathological mechanisms in more detail.

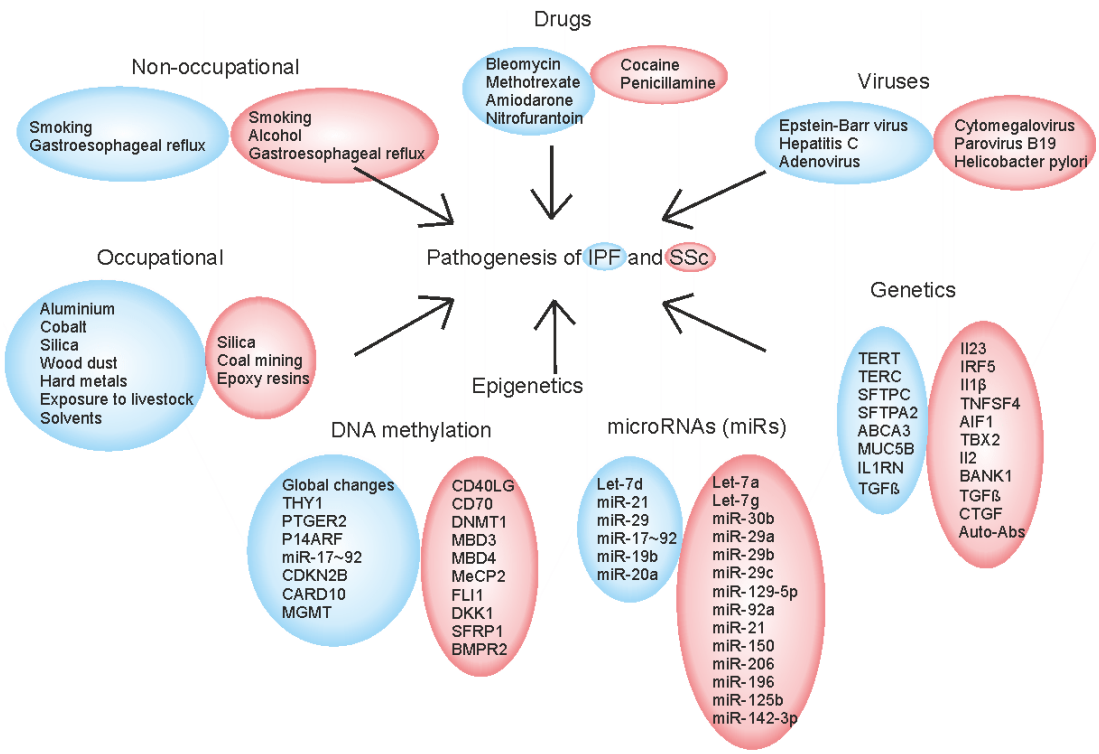


Figure 1.6.1. Possible factors involved in the pathogenesis of IPF and SSc. Blue: factors associated with the pathogenesis of IPF. Red: factors associated with the pathogenesis of SSc.

1.6.1. Altered wound healing in fibrosis

Normal wound healing is a dynamic process achieved by continuous overlapping and highly coordinated phases (Guo and DiPietro 2010). Optimal wound healing in human adults involves rapid haemostasis followed by inflammation, mesenchymal cell differentiation, angiogenesis, re-epithelialisation over the wound surface and suitable synthesis, crosslinking and alignment of collagen to strengthen healing tissue (Gosain and DiPietro, 2004, Guo and DiPietro, 2010). Perturbations to any of these phases can result in impaired wound healing, which can subsequently lead to extensive tissue remodelling and the replacement of functional tissue with permanent scar tissue, as seen in pulmonary fibrosis (Chambers, 2008) (**Figure 1.6.1.1**). Apoptosis plays an important role in the normal wound healing process by removing excessive fibroblasts/myofibroblasts after wound repair (Desmoulière et al, 1997). Myofibroblasts express alpha smooth muscle actin (α-SMA) and in skin models of wound healing they apoptose after resolution, whereas in IPF, fibroblasts and myofibroblasts are more readily activated, secrete exaggerated amounts of ECM proteins and are resistant to apoptosis (Moodley et al, 2004, Thannickal et al, 2006, Maher et al, 2007, Fattman et al, 2008, Maher et al, 2010). In contrast, AECIIs (injury to which is thought to initiate IPF), have increased apoptosis in IPF lungs, with up to 80% shown to have ongoing apoptosis (Korfei et al, 2008).

COX2/PGE2 deficiency has been shown to play an important role in apoptosis in IPF (Maher et al, 2010). Reduced PGE2 causes increased sensitivity of AECs to Fas ligand-induced apoptosis whilst conversely inducing fibroblast resistance to apoptosis (Maher et al, 2010) and to date, remains the only mediator to adequately explain this apoptosis paradox. Increased expression of survivin, a member of the inhibitor of apoptosis protein family, has been linked with promoting apoptosis resistance in IPF fibroblasts (Sisson et al, 2012). IL-6 may also play an essential role in balancing AEC/fibroblast apoptosis as in IPF fibroblasts as IL-6 can induce the anti-apoptotic protein B-cell lymphoma 2 (BCL2) whereas in normal fibroblasts IL-6 can enhance apoptosis by inducing the pro-apoptotic protein, BCL2-associated X protein (BAX) (Moodley et al, 2003). IPF fibroblasts also have increased proliferation and decreased apoptosis when they attach to polymerised collagen (Nho and Hergert, 2014) which can result from increased protein kinase B (AKT) activity via phosphatase and tensin homolog (PTEN) suppression (Nho and Hergert, 2014). In contrast, normal lung fibroblasts have decreased proliferation and increased apoptosis when they attach to polymerised collagen via suppression of the PI3K/AKT pathway as a result of increased PTEN activity (Nho and Hergert, 2014). Furthermore, SPARC, a matricellular protein that regulates tissue repair and wound healing, can increase ß-catenin which subsequently leads to an apoptosis-resistant phenotype in IPF fibroblasts (Chang et al, 2010). Other factors including TGFß (Hagimoto et al, 2002, Murray et al, 2011), found in inflammatory zone 1 (FIZZ1) (Liu et al, 2004, Chung et al, 2007), interleukin-1 beta (IL-1ß) (Zhang et al, 1993) and TNFα (Frankel et al, 2006), have all been linked to regulating apoptosis, yet the exact mechanisms that cause increased AECII apoptosis and increased resistance of fibroblasts apoptosis, remains to be elucidated.

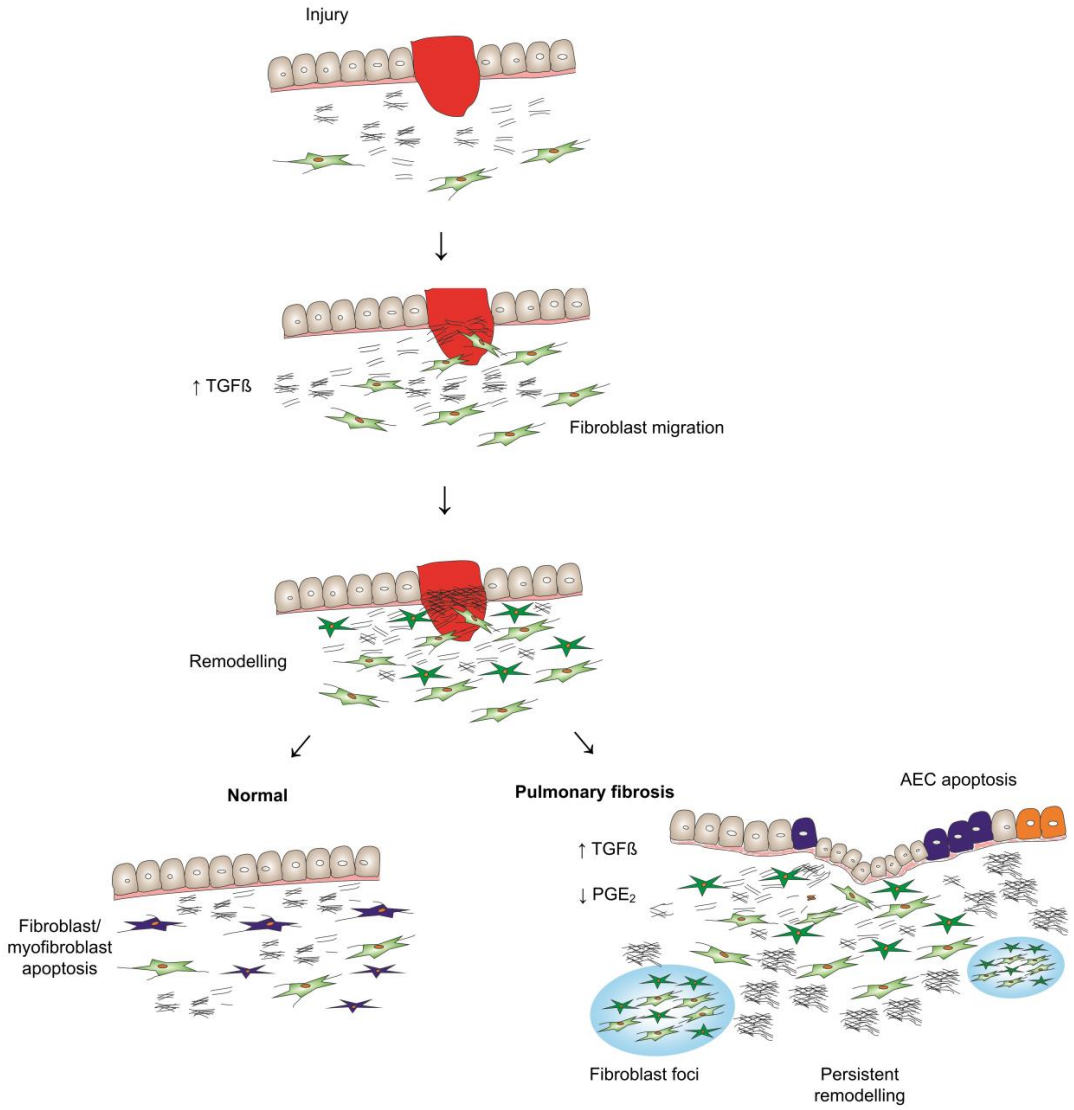


Figure 1.6.1.1. Normal wound healing compared to wound healing in pulmonary fibrosis. After injury to the epithelium, haemostasis occurs resulting in the formation of a fibrin clot. An influx of inflammatory cytokines and growth factors create a pro-fibrotic environment which stimulates fibroblast proliferation and migration towards the provisional matrix. Here, fibroblasts deposit ECM components including collagens and exert forces on the matrix which enables collagen reorganisation. TGFß causes fibroblasts to differentiate into myofibroblasts which together with fibroblasts secrete ECM components and proteases which breakdown ECM. The balance between ECM synthesis and degradation is essential for successful wound repair. Under normal conditions fibroblasts and myofibroblasts apoptose after wound closure however, under pathological conditions such as in pulmonary fibrosis, they persist (are less sensitive to apoptosis) and continue to remodel the matrix. In IPF, fibroblasts and myofibroblasts form fibroblastic foci which represent active areas of fibrosis.

1.6.2. The role of ECM stiffness in IPF and SSc

ECM provides the physical scaffolding for cellular constituents, as well as being involved in the initiation of biochemical and biomechanical events required for tissue homeostasis (Frantz et al, 2010). Daily turnover of ECM in the lung is >10% and small changes over time can result in changes in the ECM composition and fibrosis (McAnulty and Laurent, 1987, McAnulty et al, 1991). The composition of collagens, elastins and proteoglycans which is essential for maintaining normal lung function is perturbed in IPF (Bienkowski and Gotkin, 1995). In normal lungs, collagen accounts for 20% of dry lung weight and is composed of several different types in a location-specific and organised manner (McAnulty and Laurent, 1995). In IPF, collagen deposition is disorganised and both types I and III are associated with significantly altered turnover rates, resulting in abnormal structural remodelling of the lungs (Kirk et al, 1984). Increased production of ECM is central to the pathophysiology of IPF.

Fibroblasts derived from fibrotic diseases such as IPF and SSc have an altered phenotype characterised by increased proliferation and differentiation which subsequently results in exaggerated amounts of ECM components being secreted (Tomasek et al, 2002, Varga and Abraham, 2007). This corresponds with an ECM which is stiff, inelastic and contains an altered composition of ECM components including MMPs/TIMPs, glycoproteins and collagens (**see section 1.3, Table 1.3.1**). Recently it has been shown that the ECM can drive fibroblast phenotype (Marinković et al, 2013, Blaauboer et al, 2013, Parker et al, 2014, Liu et al, 2015). Fibroblasts interact with multiple components of the ECM and interpret signals which control their function and fate (Halliday and Tomasek 1995). Increased matrix stiffness is associated with driving fibroblast activation and myofibroblast differentiation (Liu et al, 2010, Huang et al, 2012) and gene expression can be modulated by the stiffness of the ECM (Booth et al, 2012, Parker et al, 2014). Multiple genes which encode ECM proteins including collagens and laminins are activated in fibroblasts grown on a stiff ECM as found in IPF, regardless of whether they are derived from a healthy or diseased origin (Parker et al, 2014) suggesting mechanical signals strongly influence fibroblast function. This emerging evidence suggests that changes in the composition of the ECM are not just a consequence, but a cause of fibrosis. Current data suggests a positive feedback loop exists between the ECM and fibroblasts whereby the mechanical properties of the ECM can regulate fibroblast gene expression and fibroblast gene expression can regulate the ECM composition (**Figure 1.6.2.1**).

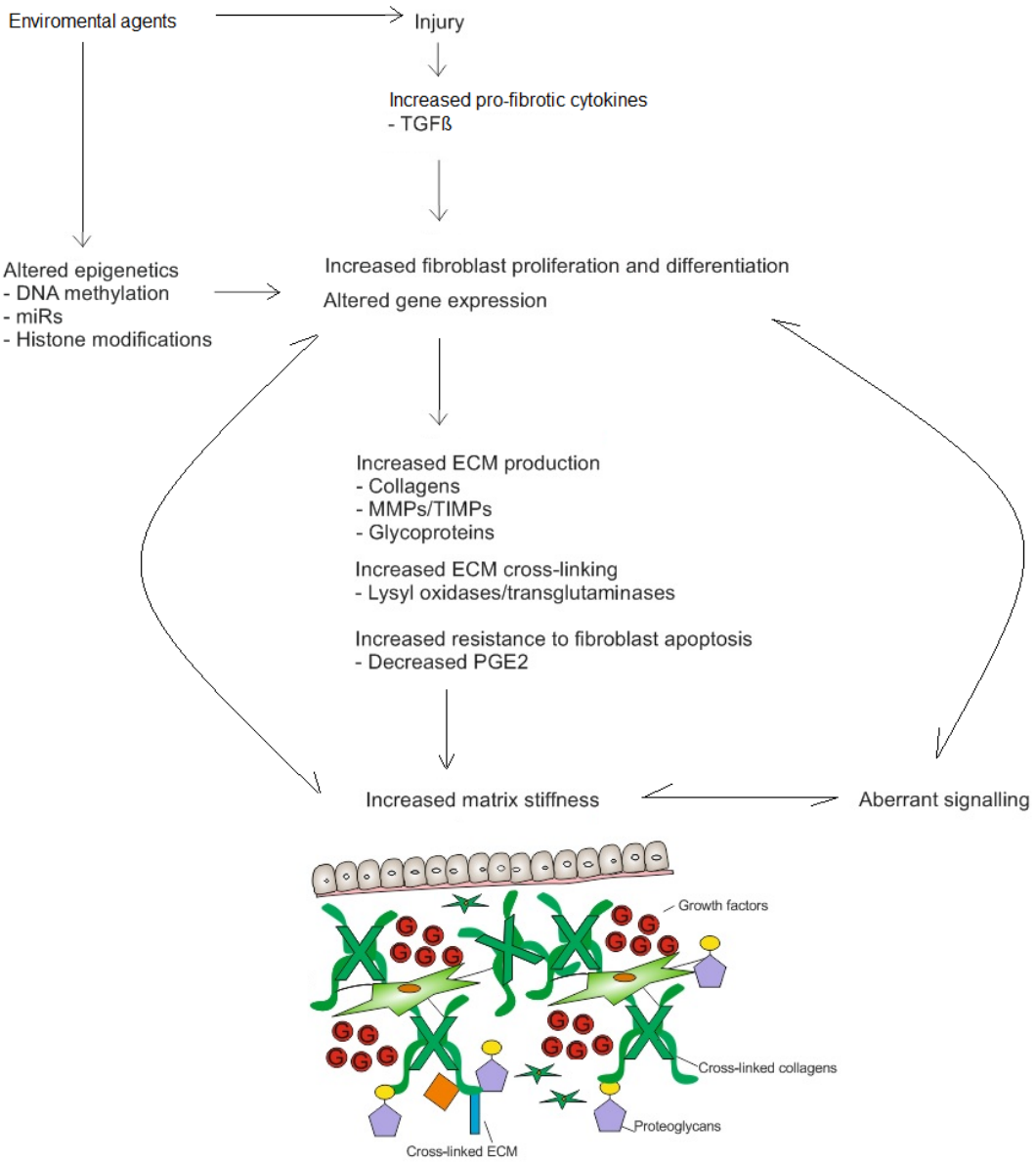


Figure 1.6.2.1. Positive feedback between ECM and fibroblasts. Fibroblasts become activated by potentially multiple pathways and proliferate and differentiate into myofibroblasts, secreting exaggerated amounts of ECM components. This subsequently leads to an altered composition of ECM and changes in the mechanical properties of the ECM. The changes in the mechanical properties of the ECM (increased stiffness) are recognised by integrins (primary mechanosensors) which provide feedback signals to fibroblasts. This signalling further drives fibroblast proliferation/differentiation and may lead to altered gene expression.

1.6.3. Gastroesophageal reflux

Diseases including IPF and SSc-PF have reduced lung compliance due to the altered composition of the ECM. This results in altered intrathoracic pressure which may predispose individuals with these diseases to gastroesophageal reflux (GER) (Raghu et al, 2006). GER occurs when protective mechanisms fail to prevent the reflux and aspiration of acidic and non-acidic products from the gastrointestinal tract. Although no cause and effect relationship between GER and pulmonary fibrosis has been shown, GER is common in IPF patients (Raghu et al, 2006) and approximately 90% of all SSc patients have some degree of gastrointestinal involvement (Turner et al, 1973, Szamosi et al, 2006) which significantly contributes to patient morbidity (Thoua et al, 2010). The pressure gradient between the abdomen and thorax is important in regulating gastroesophageal reflux. However, it is also possible that GER may be a secondary effect of pulmonary fibrosis, rather than a cause, as decreased lung compliance results in increased negative pleural pressure which potentially could result in the reflux of gastric products in the oesophagus (Raghu et al, 2006).

Chronic microaspiration of gastric contents may also cause iterative injury to the alveolar epithelium which is believed to be important in the pathogenesis of both IPF and SSc-PF (Lee et al, 2010). Furthermore, a positive relationship between the degree of acid and non-acid reflux and extent of PF has been shown in SSc (Savarino et al, 2014). Interestingly, acid suppression treatment in IPF patients does not stop GER and more recently acid suppression treatment has been shown to increase non-acidic reflux (Kilduff et al, 2014). The role of non-acid reflux in PF is poorly understood, however a rodent model of chronic aspiration has shown that lung injury can be caused by non-acidic gastric products (Downing et al, 2008). This suggests that increased non-acid reflux could potentially cause lung injury in patients with pulmonary fibrosis. Thus, the recommendation that the majority of patients with IPF and asymptomatic acid reflux should receive acid suppression therapy may be unsuitable, as the increase in non-acidic reflux induced by acid suppression may further cause lung injury (Kilduff et al, 2014). However, the exact role of acidic and non-acidic reflux in PF remains unknown.

1.6.4. The role of environmental factors in IPF

Multiple environmental factors including occupational (Ramage et al, 1988, Scott et al, 1990, Monso et al, 1990, Billings and Howard, 1994, Iwai et al, 1994, Hubbard et al, 1996, Baumgartner et al, 2000), infectious (Ueda et al, 1992, Egan et al, 1995, Kuwano et al, 1997) and non-occupational/non-infectious (Baumgartner et al, 1997) have been linked to ILD, although none have been shown to consistently increase the risk of developing IPF (Figueroa et al, 2010). Furthermore, it is important to note that many of the studies suggesting a role for these factors in IPF were conducted before the 2002 and 2013 IPF criteria guidelines and have subsequently been classified separately from IPF (ATS/ERS, 2002, Travis et al, 2013). Thus the exact cause/causes of IPF still remain unknown. Nonetheless, some of these factors may still play an important role and/or increase the risk of

developing IPF. For example, increased oxidative stress, which has been associated with IPF (MacNee, 2005) can be caused by toxic activation of oxygen species via hard metal and cobalt interactions (Lison et al, 1996, Baumgartner et al, 2000). Increased oxidative stress can also be a consequence of smoking which has been linked to several ILDs and may promote disease progression (MacNee et al, 2005). However, the exact mechanisms by which smoking may contribute to ILD and whether smoking plays an important role in the pathogenesis of IPF remain unknown. The prevalence of tobacco use in IPF patients ranges between 41%-83% (Ryu et al, 2001, Oh et al, 2012) and smoking has been associated with increasing the risk of developing both sporadic IPF and familial PF (Baumgartner et al, 1997, Steele et al, 2005, Taskar and Coultas, 2006). However, perhaps surprisingly, smoking has been associated with increased survival in IPF patients who smoke compared to ex-smokers (King et al, 2001).

Viruses and bacteria have the capacity to cause repetitive lung injury by damaging AECs and by causing apoptosis (Molyneaux and Maher, 2013). Viral infections have previously been implicated in IPF pathology (**Table 1.6.4.1**) however, results from different studies are conflicting (Ueda et al, 1992, Irving et al, 1993, Molyneaux and Maher, 2013). Human herpes viruses (HHV) which includes EBV, CMV and HHV-7 may cause IPF by inducing endoplasmic stress and apoptosis in epithelial cells (Isler et al, 2005, Lawson et al, 2008), processes which have previously been implicated in the pathogenesis of IPF (Maher et al, 2010, Zhong et al, 2011, Tanjore et al, 2012). The role of bacteria is less well known in the pathogenesis of IPF although pathogens including *Haemophilus*, *Streptococcus* and *Pseudomonas* have been reported in broncheoalveolar lavage fluid (BAL) from IPF patients (Richter et al, 2009). Furthermore, progression of IPF has been associated with the presence of *Staphylococcus* and *Streptococcus* in a recent study looking at biomarkers and whether they can predict IPF disease course (Han et al, 2014). Understanding the role of the lung microbiome in the development and progression of IPF will help determine to what extent bacteria play a role in the pathogenesis of IPF. This could subsequently lead to the development of drugs targeting specific bacteria/microbial signatures to help prevent, reverse or inhibit negative features associated with IPF.

Environmental agent	Role/findings in PF	Citations
Wood dust	Accelerates lung function decline.	(Iwai et al, 1994, Hubbard et al, 1996, Scott et al, 1990, Baumgartner, et al 2000, Jacobsen et al, 2008).
Metal dust	Associated with increased incidence and mortality.	(Hubbard et al, 1996, Hubbard et al, 2000).
Silica	Increased risk for developing ILD and potentially, IPF.	(Monso et al, 1990, Taskar and Coultas, 2006).
Sand/stone	Increased risk for developing ILD and potentially IPF.	(Taskar and Coultas, 2006).
Solvents/chemical exposure	Increased TGFß and increased risk for developing ILD and potentially, IPF.	(Billings and Howard, 1994).
Livestock/farming	Increased risk for ILD and potentially, IPF.	(Scott et al, 1990, Iwai et al, 1994, Baumgartner et al, 2000, Gustafson et al, 2007).
Cobalt	Increased risk for developing ILD and potentially IPF.	(Zanelli et al, 1994, Lison et al, 1996).
Aluminium	Increased risk for developing ILD and potentially, IPF.	(Vallyathan et al, 1982, De Vuyst et al, 1986, Jederlinic et al, 1990).
Lead	Increased risk for developing ILD and potentially, IPF.	(Figueroa et al, 1992)
Diesel exhaust	Induces fibrosis, increased number of inflammatory cells in sputa.	(Hyde et al, 1985, Ädelroth et al, 2006).
EBV	Increased in IPF BAL, may cause ER stress and epithelial apoptosis.	(Vergnon et al, 1984, Egan et al, 1995, Stewart et al, 1999, Kelly et al, 2002, Manika et al, 2007, Calabrese et al, 2013),
Hepatitis C	Conflicting data, higher prevalence of Hepatitis C virus markers in IPF.	(Ueda et al, 1992, Arase et al, 2008).
Transfusion transmitted viruses (TTVs)	Found in IPF BAL during acute exacerbations.	(Bando et al, 2008, Wootton et al, 2011).
HHVs	Potential role in initiation and progression of IPF.	(Tang et al, 2003).
Haemophilus	Increased in IPF BAL.	(Richter et al, 2009).
Streptococcus	Increased in IPF BAL.	(Han et al, 2014).
Pseudomonas	Increased in IPF BAL.	(Richter et al, 2009).
Staphylococcus	Increased in IPF.	(Han et al, 2014).

Table 1.6.4.1. Environmental agents potentially involved in the pathogenesis of IPF. Environmental agents including smoking, dust exposure and viruses/bacteria have recently been associated with increased risk of developing IPF however, no single agent to date has been identified as having a causative relationship with IPF.

1.6.5. The role of environmental factors in SSc

Multiple environmental factors have been linked to SSc. These include occupational (Bramwell, 1914, Erasmus, 1957, Rodnan et al, 1967, McCormic et al, 2010), infectious (Vaughan et al, 2000, Lunardi et

al, 2000, Ferri et al, 2002, Lundari et al, 2006) and non-occupational/non-infectious factors (Thompson et al, 2002, Hudson et al, 2011) (**Table 1.6.5.1**). However, due to many of the studies consisting of relatively small sample numbers, it has often been difficult to determine the degree of risk an environmental factor has towards SSc development. Silica dust was one of the first environmental factors linked with SSc via studying workers with occupations involving high silica exposure such as stonemasonry (Bramwell, 1914), gold mining (Erasmus, 1957) and coal mining (Rodnan et al, 1967). A review analysing studies published between 1949-2009 found a significant association between silica exposure and development of SSc (McCormic et al, 2010). Although there were a number of limitations in the study, the strength of the association between silica and SSc appeared stronger in males than females, which may be due to increased exposure to silica in males (McCormic et al, 2010). The exact mechanisms by which silica dust may cause SSc are unknown, however, the adjuvant effect silica displays on antibody production highlights silica's biological plausibility (Uber et al, 1982, Parks et al, 2002). Furthermore, mice develop exacerbated autoimmunity and PF following treatment with crystalline silica (Brown et al, 2004). Solvents (Aryal et al, 2001) and epoxy resins (Yamakage et al, 1980) have also been linked with increased risk of developing SSc.

Multiple bacterial and viral infectious agents have been linked to fibrotic and autoimmune diseases (Leroy, 1996, White et al, 1996, Ferri et al, 2002) and there is mounting evidence suggesting bacterial and viral infectious agents contribute to the development of SSc, potentially via inducing signalling pathways leading to aberrant expression of genes including TGFß (Farina et al, 2014). CMV has been implicated in SSc in several studies (Vaughan et al, 2000, Lunardi et al, 2000, Ferri et al, 2002. Lundari et al, 2006) and may cause apoptosis of endothelial cells which is a characteristic of SSc (Vaughan et al, 2000, Lunardi et al, 2000, Ferri et al, 2002). Furthermore, CMV may also increase apoptosis of macrophages (Pandey and LeRoy, 1998) which play important roles in tissue homeostasis (Mosser and Edwards, 2008) and activation of CTGF which can drive fibroblast activation and increase fibrosis (Inkinen et al, 2005). Another virus, parvovirus B19 (Ferri et al, 2002, Zakrzewska et al, 2009) and the bacteria, *helicobacter pylori* (Yazawa et al, 1998, Kalabay et al, 2002, Radić et al, 2011) have also been implicated in the pathogenesis of SSc.

Non-occupational risk factors such as smoking (Hudson et al, 2011), and alcohol (Thompson et al, 2002) have also been linked to SSc. Smoking causes vascular damage (Powell et al, 1998) thus likely contributes to the vasculopathy seen in SSc. Smoking does not increase the risk of developing SSc but does impact on disease severity (Hudson et al, 2011) and reduces overall survival rates (Hissaria et al, 2011). Overall the impact of a single environmental factor on the development of SSc is likely to be modest, based on large populations exposed to many of the environmental agents associated with SSc, who do not develop SSc (Broen et al, 2014). Nonetheless, multiple environmental agents may contribute to the development of SSc in people who are genetically and/or epigenetically predisposed.

Environmental agent	Role/findings in SSc	Citations
Silica dust	May cause epigenetic modifications.	McCormic et al, 2010.
Vinyl chloride	Associated with increased risk of SSc.	Ostlere et al, 1992, Nietert and Silver, 2000.
Solvents	Associated with increased risk of SSc.	Aryal et al, 2001
EBV	Induces aberrant innate immune response.	Farina et al, 2014.
CMV	Increased CMV antibodies, may cause endothelial apoptosis.	Pandey and LeRoy, 1998, Vaughan et al, 2000, Lunardi et al, 2000, Ferri et al, 2002, Lundari et al, 2006, Namboodiri et al, 2006.
Helicobacter pylori	Increased prevalence, may play a role in oesophageal dysfunction.	Yazawa et al, 1998, Kalabay et al, 2002, Radić et a, 2011.

Table 1.6.5.1. Environmental factors associated with SSc.

1.6.6. The role of genetic factors in IPF

The variability amongst individuals exposed to fibrogenic agents and the development of fibrosis together with evidence from mouse models, where inbred mice have shown different responses to fibrogenic agents has previously led to the belief that genetic factors may play a key role in the pathogenesis of IPF. Although genetic mutations have been associated with IPF (**Table 1.6.6.1**), no single gene has been causally linked to the disease, which may not be surprising due to the heterogeneity often observed between different IPF populations.

IPF generally occurs sporadically however, in rare cases IPF can occur in a familial form (FIPF). This inherited form of IPF is clinically identical to sporadic IPF (Marshall et al, 2000). FIPF studies have provided some of the most compelling evidence that genetic factors play a role in the pathogenesis of IPF (Steele et al, 2005). FIPF is diagnosed when two or more family members have an IIP (Loyd, 2003, Steele et al, 2005) accounting for an estimated 0.5-3.7% of all IPF cases (Lawson and Loyd, 2006), although recent estimates suggest up to 20% of all IPF cases may be familial (Garcia-Sancho et al, 2011). Linkage analysis and candidate gene approaches have identified four genes, telomere reverse transcriptase (TERT) (Armanios et al, 2007, Tsakiri et al, 2007), telomere RNA component (TERC) (Armanios et al, 2007, Tsakiri et al, 2007), surfactant protein C (SFTPC) (Nogee et al, 2001) and surfactant protein A2 (SFTPA2) (Wang et al, 2009) which are associated with FIPF. Mutations in the TERT gene are apparent in 18% of kindreds with FIPF and 3% in sporadic IPF cases (Armanios et al, 2007, Tsakiri et al, 2007, Cronkhite et al, 2008, Diaz de Leon et al, 2010). TERT or TERC mutations which cause telomere shortening, dramatically increase susceptibility to adult-onset IPF (Tsakiri et al, 2007). In sporadic IPF, short telomeres are common in a number of cell types including AECs, lymphocytes and granulocytes (Alder et al, 2008) suggesting telomere dysfunction may frequently occur in fibrotic lung disease. Furthermore, smoking is strongly associated with familial IP (Steele et al, 2005) and can cause shortening of telomere length (Morlá et al, 2006). Telomere length is also influenced by

oxidative stress (Von Zglinicki, 2002) which has been linked with PF (Mastruzzo et al, 2002, Kinnula et al, 2005, Cheresh et al, 2013) and thus, may explain why lung disease is a common phenotype of telomere shortening.

SFTPC and SFTPA2 mutations are known to causes endoplasmic reticulum (ER) stress (Mulugeta et al, 2005, Maitra et al, 2010) which has been strongly associated with a number of diseases including FIPF and sporadic IPF (Lawson et al, 2011, Tanjore et al, 2012, Kropski et al, 2013). Mutations in the gene ATP-binding cassette sub-family A member 3 (ABCA3) have also been identified in children and older patients with chronic ILD (Doan et al, 2008, Young et al, 2008, Campo et al, 2014) and some of these mutations cause an increase in ER stress (Weichert et al, 2011). ER stress has been linked with increased AECII cell death (Mulugeta et al, 2005, Lawson et al, 2008), and generation of a pro-fibrotic AEC phenotype via EMT (Zhong et al, 2011, Tanjore et al, 2011) which may explain how ER stress contributes to PF.

A recent genome-wide linkage study of IPF patients identified a single nucleotide polymorphism (SNP) in the promoter region of the gene mucin 5B (MUC5B) as being associated with the development of FIP and sporadic IPF (Seibold et al, 2011). MUC5B produces the main gel-forming mucin in mucus which functions to provide lubrication, hydrate epithelium and act as a barrier against pathogens and noxious substances (Bansil and Turner, 2006). Therefore, defective production of mucin may enhance susceptibility to environmental cues which may be involved in the pathogenesis of fibrosis. Genome-wide associated studies (GWAS) have confirmed linkage studies and have also identified a number of genes which are associated with increased susceptibility of developing IPF (Mushiroda et al, 2008, Fingerlin et al, 2013, Noth et al, 2013). A number of these genes are involved in host defence, cell-cell adhesion and DNA repair (Fingerlin et al, 2013). Rare mutations in other genes including IL-1RN which encodes the interleukin1 receptor antagonist (IL-1RA) and TGFß1 have also been associated with increasing susceptibility to IPF (Whyte et al, 2000, Hutyrová et al, 2002, Xaubet et al, 2003, Riha et al, 2004, Barlo et al, 2011, Korthagen et al, 2012, Son et al, 2013).

Gene	Role of gene/effect of mutation	Reference
TERT	Telomere shortening	Armanios et al, 2007
TERC	Telomere shortening	Armanios et al, 2007
SFTPC	ER stress	Kropski et al, 2013
SFTPA2	ER stress	Kropski et al, 2013
ABCA3	ER stress	Campo et al, 2014
MUC5B	Defective mucin production	Seibold et al, 2011
IL-1RN	Imbalance of IL1Ra/IL1b	Barlo et al, 2011
TGFB1	May affect disease progression	Xaubet et al, 2003
Toll interacting protein (TOLLIP)	Inhibitory adaptor protein	Noth et al, 2013
Signal peptide peptidase like 2C (SPPL2C)	Unknown	Noth et al, 2013
Desmoplakin (DSP)	Involved in integrity of lung epithelia	Fingerlin et al, 2013
Dipeptidyl-peptidase 9 (DPP9)	Involved in integrity of lung epithelia	Fingerlin et al, 2013
Family with sequence similarity 13 member A (FAM13A)	Signal transduction responsive to hypoxia	Fingerlin et al, 2013
ATPase, class VI, type 11A (ATP11A)	Encodes an integral membrane ATPase	Fingerlin et al, 2013
Oligonucleotide-binding fold containing 1 (OBFC1)	Associated with telomere length	Fingerlin et al, 2013

Table 1.6.6.1. Genetic factors associated with FIPF/IPF.

1.6.7. The role of genetic factors in SSc

Multiple GWAS studies have reported an association of a number of human leukocyte antigen (HLA) and non-HLA genes with SSc susceptibility and/or disease severity (Lambert et al, 2000, Gilchrist et al, 2001, Sato et al, 2004, Gladman et al, 2005, Fonseca et al, 2006, Radstake et al, 2010, Allanore et al, 2011, López-Isac, 2014) (**Table 1.6.7.1**). The low prevalence of SSc means there is a lack of multiplex families and monozygotic twins with the disease (Assassi and Tan, 2005), although the strongest risk factor for developing SSc is a family history of the disease (Englert et al, 1999, Arnett et al, 2001). Compared to the general population, there is a 15-19 fold increase in risk of developing SSc if a sibling has SSc and a 13-15 fold increase in risk of developing SSc for other first-degree relatives (Arnett et al, 2001). Although this appears a large increase in risk it only equates to a 0.026% absolute risk of developing SSc in the general population and just 1.6% in first-degree relatives (Arnett et al, 2001). The prevalence rate of SSc in twins is approximately 4.7% (4.2% in monozygotic; 5.6% in dizygotic) (Feghall-Bostwick et al, 2003). Interestingly only two pairs of twins in this study (one monozygotic and one dizygotic) were concordant for SSc (Feghall-Bostwick et al, 2003). Although many twins in this study were below the typical age on SSc onset, it highlights that genetics are not solely responsible for causing SSc.

Genes	Reference
TGFß	Varga and Pasche, 2009
B-cell scaffold protein with Ankyrin repeats 1 (BANK1)	Dieudé et al, 2009, Rueda et al, 2010
Chromosome 8 open reading frame 13 B-cell signal transducer (C8orf13-BLK)	Gourth et al, 2010, Coustet et al, 2011
TNFAIP3 interacting protein 1 (TNIP1)	Allanore et al, 2011
Psoriasis susceptibility 1 candidate 1 (PSORS1C1)	Allanore et al, 2011
Ras homolog gene family, member B (RHOB)	Allanore et al, 2011
PPARG	López-Isac, 2014
CTGF	Fonseca et al, 2007
IL-1ß	Mattuzzi et al, 2007
Interferon regulatory factor 5 (IRF5)	Dieudé et al, 2009, Ito et al, 2009, Radstake et al, 2010, Sharif et al, 2012
Signal transducer and activator of transcription 4 (STAT4)	Gourh et al, 2009, Tsuchiya et al, 2009, Radstake et al, 2010, Yi et al, 2013
T-box 21 (TBX21)	Gourh et al, 2009
Tumor necrosis factor superfamily member 4 (TNFSF4)	Gourh et al, 2010, Bossini-Castillo et al, 2010
Nitric oxide synthase 3 (NOS3)	Fatini et al, 2006
Monocyte chemoattractant protein-1 (MCP-1)	Karrer et al, 2005
Interleukin 23 receptor (IL-23R)	Agarwal et al, 2009
Interleukin 2 (IL-2)	Mattuzzi et al, 2007

Table 1.6.7.1. Genetic factors associated with SSc.

1.7. Epigenetic mechanisms and their role in pulmonary fibrosis

The word 'epigenetics' was first used to link developmental biology with genetics (Holiday, 2006). Epigenetics is now defined as the study of heritable changes in gene expression that cannot be attributed to changes in the DNA sequence. Types of epigenetic mechanisms include, DNA methylation, histone modifications (acetylation, phosphorylation and methylation) and non-coding RNAs (microRNAs, siRNAs and piRNAs). All these mechanisms function together to modulate chromatin structure and thus regulate gene transcription. Mounting evidence suggests defective epigenetic mechanisms contribute to a variety of complex diseases including cancers (Ngalamika et al, 2012, Jiang et al, 2013), autoimmune diseases (Quintero-Ronderos and Montoya-Ortiz, 2012) and fibrotic diseases (Sanders et al, 2012, Rabinovich et al, 2012, Komatsu et al, 2012, Tampe and Zeisberg, 2013, Zhao et al, 2013, Yang et al, 2014,). Data regarding the role of epigenetic regulatory mechanisms in IPF and SSc including DNA methylation (Sander et al, 2012, Rabinovich et al, 2012, Huang et al, 2014, Yang et al, 2014), microRNAs (Pandit et al, 2010) and histone modifications (Huang et al, 2013) is still in its infancy. However, a number of genes that are potentially regulated epigenetically have been identified and are likely to play a role in the pathogenesis of IPF and SSc.

1.7.1. MicroRNAs and their role in fibrosis

MicroRNAs (miRs) are endogenous, small, non-coding RNAs approximately 21 nucleotides in length. They are evolutionarily conserved and are responsible for mRNA cleavage, transcriptional repression and mRNA destabilisation (Ciechomska et al, 2014). MiRs also play an important role in regulating cell proliferation, differentiation, tissue repair and tissue development (Hwang and Mendell, 2006). There are approximately 1900 miRs currently identified in humans (Kozomara and Griffiths-Jones, 2011), each of which is expressed in a tissue-type and cell-type specific manner and has multiple gene targets (Bartel, 2004, Friedman et al, 2009). In a fibrosis context, several key pro-fibrotic molecules are regulated by multiple miRs (Duisters et al, 2009, Mann et al, 2010, Zhu et al, 2013) as well as multiple ECM proteins (Muth et al, 2010, Maurer et al, 2010, Cushing et al, 2011) and signalling pathways, including TGFß/CTGF (Leask and Abraham, 2006). TGFß is a profibrotic cytokine and a central mediator in fibrosis (Leask and Abraham, 2004). It can induce CTGF, which can in turn feedback to enhance TGFß signalling as well as induce multiple other pro-fibrotic mediators including VEGF, Wnts and integrins (Leask and Abraham, 2006). To date, ten miRs are known to regulate TGFß/CTGF signalling (Vettori et al, 2012) including miR-18a, miR-19a and miR-19b, which have been shown to regulate CTGF in liver and cardiac fibrosis (Kodama et al, 2011, van Almen et al, 2011) and miR-133 and miR-30c which can regulate CTGF in cardiac fibrosis (Duisters et al, 2009). Other studies have shown the miR-200 family may play both direct and indirect roles in fibrogenesis by regulating TGFß2 expression (Wang et al, 2011) and by regulating EMT which could potentially play an important role in fibrosis (Gregory et al, 2008, Wang et al, 2011). However, controversy surrounds the role of EMT in PF as there is evidence suggesting multiple stromal cell populations contribute to PF without evidence for EMT (Rock et al, 2011).

1.7.2. Role of miRs in IPF and SSc

There is limited data regarding the role of miRs in IPF, however a recent study suggested that compared to control lung tissue, 10% of all miRs have significantly altered expression in IPF lung tissue (**Table 1.7.2.1**) (Pandit et al, 2010, Pandit et al, 2011). The miR-17-92 cluster which encodes 6 miRs (miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR-92a-1) has also been associated with IPF (Dakhlallah et al, 2013). MiR-17-92 is down-regulated in IPF lung tissue and miR-19b and miR-20a decrease in proportion to disease severity (Dakhlallah et al, 2013). The down-regulation of these miRs also correlates with increased expression of VEGF, CTGF, granulocyte macrophage colony-stimulating factor (GM-CSF), thrombospondin 1 (TSP-1) and TGFß which have all previously been associated with IPF (Voelkel et al, 2006), (Allen et al, 1999, Pan et al, 2001, Allen et al, 2001, Kono et al, 2011), (Moore et al, 2000), (Ide et al, 2008), (Khalil et al, 1991). Interestingly, the miR-17-92 cluster is methylated and a number of miRs from this cluster appear to be directly regulated by DNA methyltransferase 1 (DNMT1) (Dakhlallah et al, 2013).

miR	Tissue	Expression in IPF	Pro-/anti-fibrotic	Reference
Let-7d	Lung	Down-regulated	Anti	Pandit et al, 2010
miR-21	Lung	Up-regulated	Pro	Liu et al, 2010, Li et al, 2013.
miR-154	Lung fibroblasts	Up-regulated	Pro	Milosevic et al, 2012.
miR-29	Lung, lung fibroblasts	Down-regulated	Anti	Cushing et al, 2011
miR-26a	Lung	Down-regulated	Anti	Liang et al, 2014
miR-17-92 cluster	Lung, lung fibroblasts	Down-regulated	Anti	Dakhlallah et al, 2013
miR-199a-5p	Fibroblasts	Up-regulated	Pro	Lino Cardenas et al, 2013.

Table 1.7.2.1. Dysregulated miRs in IPF. Which tissue the specified miR was studied in, miR expression compared to control groups and whether the miR is associated with being pro- or anti-fibrotic is shown.

Multiple pro-fibrotic and anti-fibrotic miRs are also dysregulated in SSc skin (Li et al, 2012, Zhu et al, 2012) and fibroblasts cultured ex vivo (Honda et al, 2012), compared to healthy controls (**Table 1.7.2.2**). Subsequent studies have shown miRs may play an integral role in the pathogenesis of SSc by modulating multiple fibrosis-related genes including collagens, MMPs and integrins (Broen et al, 2014).

miR	Tissue	Expression in SSc	Pro/anti-fibrotic	Reference
miR-30b	Serum, Skin	Down-regulated	Anti	Tanaka et al, 2013
Let-7a	Serum, Skin fibroblasts	Down-regulated	Anti	Makino et al, 2013
miR-29a/b/c	Skin	Down-regulated	Anti	Bhattacharyya et al, 2013, Ciechomska et al, 2014
Let-7g	Skin	Up-regulated	Unknown	Li et al, 2012
miR-142-3p	Serum	Up-regulated	Unknown	Makino et al, 2012
miR-129-5p	Skin fibroblasts	Down-regulated	Anti	Nakashima et al, 2012
miR-150	Skin fibroblasts	Down-regulated	Anti	Honda et al, 2013
miR-196a	Skin fibroblasts	Down-regulated	Anti	Honda et al, 2012, Honda et al, 2013, Wang et al, 2013, Makino et al, 2013.
miR-145	Skin	Down-regulated	Unknown	Li et al, 2012, Zhu et al, 2012.
miR-125b	Skin	Down-regulated	Unknown	Li et al, 2012, Zhu et al, 2012.
miR-92a	Skin fibroblasts, serum	Up-regulated	Pro	Sing et al, 2012
miR-206	Skin	Down-regulated	Unknown	Li et al, 2012, Zhu et al, 2012.
miR-21	Skin fibroblasts, Skin	Up-regulated	Pro	Zhu et al, 2013
miR-7	Skin fibroblasts, skin	Up-regulated	Anti	Kajihara et al, 2012

Table 1.7.2.2. Dysregulated miRs in SSc. Which tissue the specified miR was studied in, miR expression compared to controls and whether the miR is associated with being pro- or anti-fibrotic is shown.

1.7.3. The role of DNA methylation

DNA methylation is an epigenetic modification essential for maintaining genomic stability, specifying cell fate, genomic imprinting (Li et al, 1993), X-chromosome inactivation and stabilisation (Heard et al, 1997, Sado et al, 2000), protection against retroviruses and transposons (Walsh et al, 1999 and regulating gene expression (Bird, 2002). The first suggestion that DNA methylation could have an important biological function was in 1969 by Griffith and Mahler. In 1975, two independent papers by Riggs and Holliday and Pugh, suggested changes in DNA methylation could directly activate or silence gene expression and could be inherited through somatic cell divisions (Bird, 1978).

Cytosine nucleotides can be methylated at the fifth position on their pyrimidine ring which was first described in 1948 (Hotchkiss, 1948). Methylation occurs via DNA methyltransferase (DNMT) enzymes which use s-adenosyl-L-methionine (SAM) as a methyl donor to covalently attach a methyl group at this position (Adams and Burdon, 1982) (**Figure 1.7.3.1**). Cytosine nucleotides which directly precede guanine nucleotides are known as CpG dinucleotides (where 'p' represents the phosphate bond between cytosine and guanine). DNA methylation occurs almost exclusively in the symmetrical CG context and affects approximately 70-80% of all CpGs (Ehrlich et al, 1982, Bird et al, 1985). Non-CpG methylation can occur in rare circumstances but its function is unknown (Ramsahoye et al, 2000, Haines et al, 2001, Dodge et al, 2002, Ziller et al, 2011).

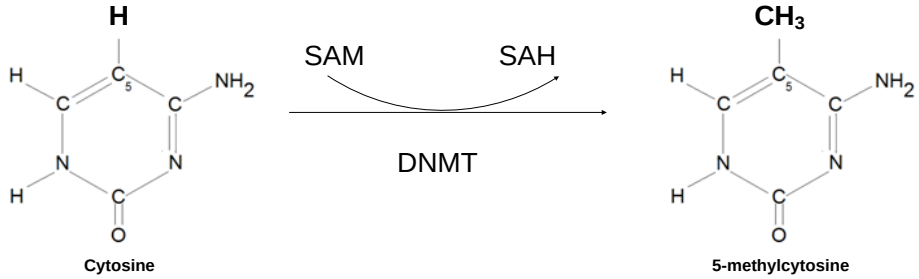


Figure 1.7.3.1. Cytosine conversion to 5-methylcytosine. A methyl group from S-adenosyl-L-methionine (SAM) is transferred to cytosine at the fifth position of its pyrimidine ring via DNA methyltransferase (DNMT) enzymes. This forms 5-methylcytosine and S-adenosyl-L-homocysteine (SAH).

Increased methylation is strongly associated with the formation of heterochromatin and transcriptional silencing (Keshet et al, 1986, Reik et al, 2001). Methylated cytosine, referred to as 5-methylcytosine (5-mC), is found in approximately 1.5% of genomic DNA (Lister et al, 2009) and acts as a ligand for methyl-binding proteins (MBPs). There are currently 15 defined MBPs all of which contain a methyl-binding domain that exclusively binds to methylated CpGs (mCpGs) (Parry and Clarke 2011). The binding of MBPs to mCpGs causes recruitment of transcriptional repressors, HDACs and chromatin-modifying complexes which induce the formation of heterochromatin, a tightly packed

form of chromatin. These mechanisms cause transcription to be repressed resulting in gene silencing (Fuks, 2005) (**Figure 1.7.3.2**). CpG frequency in the human genome occurs less frequently than expected by random chance as methylcytosine has the ability to spontaneously deaminate to thymine (Bird, 1980). However, small genomic regions (approximately 1kb long) which contain a multitude of CpG dinucleotides in GC-rich regions exist and are predominantly unmethylated (Antequera and Bird, 1993). These regions are termed CpG islands (CGIs), and are located in approximately 70% of all gene promoters (Saxonov et al, 2006). Although promoter-associated CGIs tend to be unmethylated, in specific tissues or during development, a differential methylated state may be adopted (Li, 2002, Song et al, 2005).

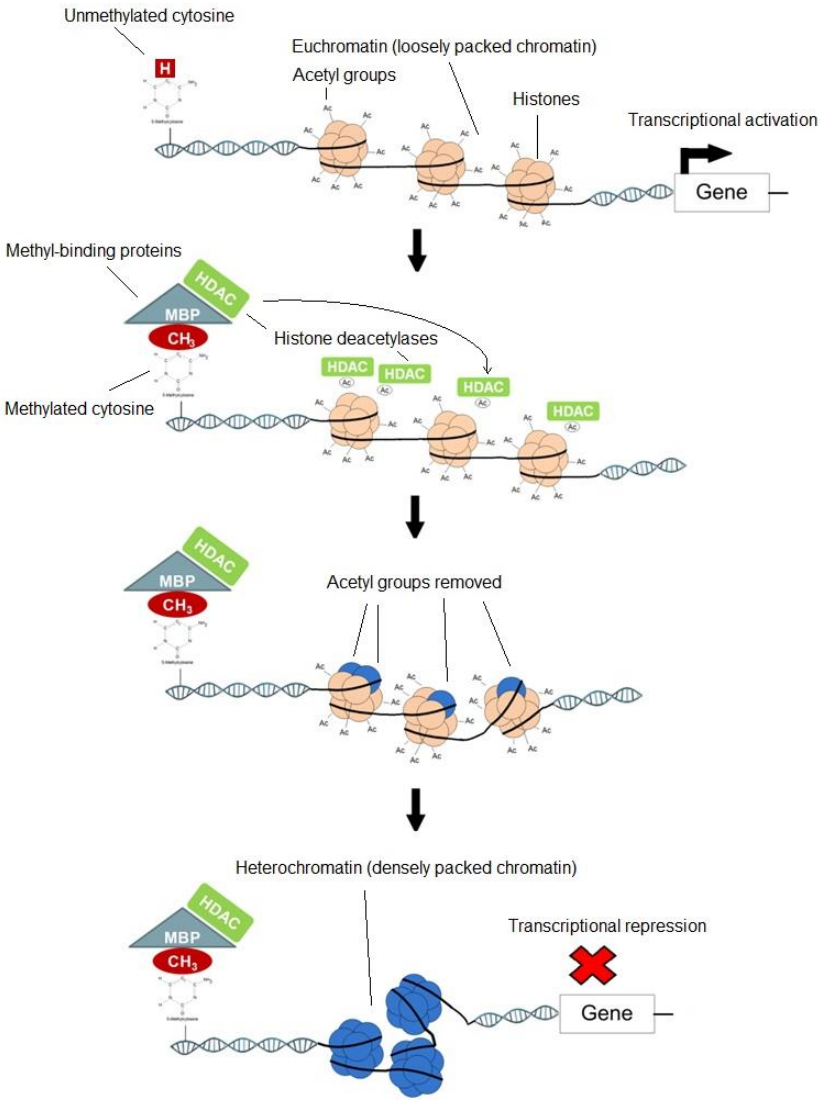


Figure 1.7.3.2. Diagram of how methylation reduces gene expression. Unmethylated cytosine and acetylated histones are associated with transcriptional activity. Acetylation of histones allows chromatin to be loosely packed. Once cytosine becomes methylated, methyl-binding proteins (MBPs) can bind and recruit histone deacetylases (HDACs). HDACs remove the acetyl groups which induces the formation of densely packed chromatin (heterochromatin). This physically blocks transcription factors binding to receptors and is most commonly associated with repressed gene transcription.

1.7.4. Location of CpG methylation

CGIs usually remain unmethylated to facilitate transcription, whereas increased methylation within some CGIs can block transcription (Deaton and Bird, 2011). Increased methylation of the gene body is believed to block aberrant transcription from within the gene, thus functioning to avoid the production of truncated forms of the protein although multiple studies have shown that for some genes, intragenic methylation correlates with increased transcription (Zilberman et al, 2007, Rauch et al, 2009, Kulis et al, 2013). Research into the role of DNA methylation in regulating gene expression has predominantly focused on CpG methylation with emphasis on CGIs located in promoter regions of genes. CGI methylation in promoter regions has been strongly linked to gene silencing in numerous pathologies (Esteller and Herman, 2001, Egger et al, 2004, Robertson, 2005), although the role of methylation in CGIs distal to promoter regions and its ability to regulate transcription is poorly understood (Illingworth et al, 2010).

CGIs that are located remotely from defined promoter regions have been coined 'orphan CGIs' and are more readily methylated than CGIs within promoter regions (Illingworth et al, 2010). Orphan CGIs have the characteristics of functional promoters suggesting they can regulate gene expression and may be associated with novel transcripts that have a regulatory role (Illingworth et al, 2010). DNA methylation outside CGIs could be equally as important in regulating transcriptional activity (Irizarry et al, 2009). Irizarry et al recently demonstrated the majority (76%) of methylation in human colon cancer cells occurred at areas a short distance away from CGIs, coined 'CpG shores' (Irizarry et al, 2009). These regions exist up to 2kb away from a CGI and have been shown to strongly correlate with gene expression (Irizarry et al, 2009). Flanking CpG island shores are CpG shelves which occur 2kb-4kb away from CGIs (Irizarry et al, 2009). CpGs located further than 4kb away from a CGI are said to be located in open sea regions (**Figure 1.7.4.1**)**.** To what effect DNA methylation has on gene expression at CGIs, shores, shelves and open sea regions in IPF and other fibrotic diseases remains unknown.

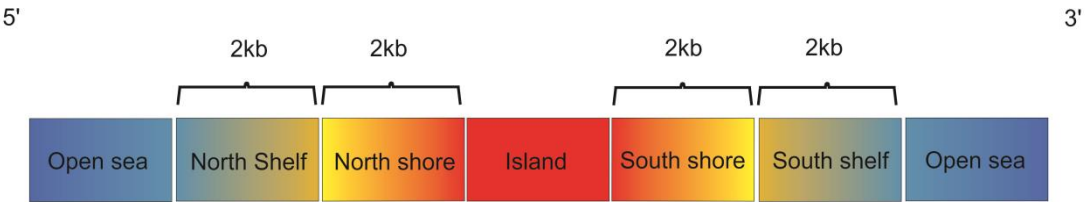


Figure 1.7.4.1. The different regions of a gene in which CpG methylation can occur. CpG **s**hore regions (up to 2kb away) flank CpG islands. CpG shelf regions (up to 2kb away) flank CpG shores. Open sea regions are further than 4kb away from CpG islands.

1.7.5. The role of DNA methyltransferases

As previously mentioned, DNMTs catalyse the formation of 5-mC by transferring a methyl group from S-adenosyl-L-methionine to the C5 position of cytosine residues. DNMT1, 2, 3A, 3B and 3L are the main DNMTs that belong to the DNMT family, although only DNMT1, DNMT3A and DNMT3B have methyltransferase activity. The first DNMT to be biochemically characterised was DNMT1 and is considered the primary maintenance DNA methyltransferase. The primary function of DNMT1 is to methylate newly synthesised DNA during the S phase of the cell cycle (Bird, 2007). DNMT1 recognises and has preference for hemimethylated DNA (methylated on one strand) rather than unmethylated DNA which stops previously unmethylated CpGs becoming methylated during DNA replication, thus maintaining methylation patterns during DNA replication (Okano et al, 1999, Bestor, 2000, Jones and Baylin, 2007). DNMT1 was originally believed to be the sole DNMT responsible for maintenance methylation, however, more recently it has been shown that DNMT3A and DNMT3B also play an important role (Liang et al, 2002, Chen et al, 2003, Jeong et al, 2009). DNMT1 may also act as a de novo methylase at sites of homologous recombination repair (Cuozzo et al, 2007) and can interact with HDACs (Fuks et al, 2000, Rountree et al, 2000, Fuks et al, 2001), other DNMTs, MBPs (Kimura and Shiota, 2003) and genes including ubiquiting-like with PHD and ring finger domains 1 (UHRF1) (Qian et al, 2008), proliferating cell nuclear antigen (PCNA) (Chuang et al, 1997) and DNA methyltransferase 1-associated protein 1 (DMAP1) (Rountree et al, 2000).

Like DNMT1, DNMT3A and DNMT3B are essential for embryonic development (Okano et al, 1999). DNMT3A and DNMT3B are the catalytically active DNMTs in the DNMT3 family. DNMT3L is catalytically inactive due to the mutation of specific catalytic residues (Pacaud et al, 2014). However, DNMT3L is an important co-factor for DNMT3A and DNMT3B which enhances de novo methylation and has recently be shown to interact with multiple transcription factors (Pacaud et al, 2014). The interaction of DNMT3L with specific transcription factors allows DNMT3A and DNMT3B to interact at specific sites on these genes, which are otherwise inaccessible (Pacaud et al, 2014). DNMT3A and DNMT3B are considered de novo DNMTs that are essential for the establishment of DNA methylation patterns during mammalian development and in germ cells (Okano et al, 1999). They do not show any preference for hemimethylated or unmethylated DNA (Okano et al 1999) and whilst both are de novo methylases, they also have important functions in maintenance methylation (Liang et al, 2002, Chen et al, 2003, Jeong et al, 2009). DNMT2 (renamed TRDMT1) plays an important role in catalysing cytosine methylation in RNA substrates (Goll et al, 2006), methylation of tRNAs (Schaefer et al, 2010) and RNA-mediated epigenetic heredity (Kiani et al, 2013). However, the exact biological functions of DNMT2/TRDMT1 remain poorly understood (**Figure 1.7.5.1**).

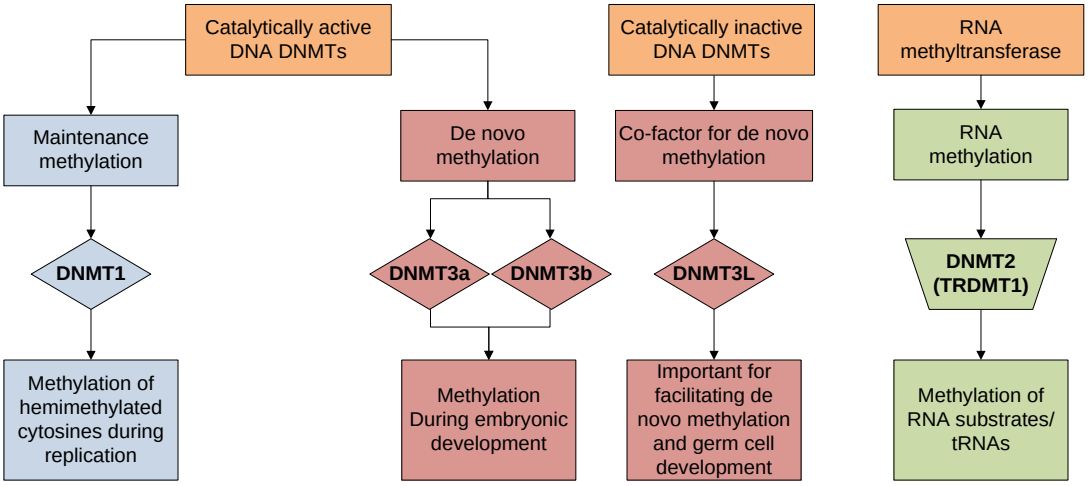


Figure 1.7.5.1. The role of different DNMTs. DNMT1 acts primarily to maintain genomic methylation patterns although it can also act as a de novo methyltransferase. DNMT3A and DNMT3B are primarily de novo methyltransferases but can also play a role in maintenance of DNA methylation. DNMT3L is a co-factor for DNMT3A and DNMT3B and can direct both DNMT3A and DNMT3B to specific sites in multiple transcription factors. DNMT2/TRDMT1 catalyses cytosine methylation in RNA substrates and tRNAs.

1.7.6. DNA demethylation

Although DNA methylation is relatively stable, as compared to histone marks, loss of methylation and DNA demethylation have been observed in a number of different biological contexts (Kohli and Zhang, 2013). DNA demethylation can be active or passive. Passive DNA demethylation requires DNA replication and occurs by successive rounds of DNA replication in the absence of the required DNA methylation machinery such as DNMTs or SAM (Kohli and Zhang, 2013). Active demethylation does not require DNA replication to remove 5-mC. Instead, enzymes belonging to the ten-eleven translocation (TET) family oxidise 5-mC to form 5-hydroxymethylcytosine (5-hmC) (Tahiliani et al, 2009, Ito et al, 2010). 5-hmC represents a key intermediate in active demethylation and can be further oxidised to form 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (Ito et al, 2011, He et al, 2011). These oxidised bases are removed by thymine DNA glycosylase-triggered base excision repair (TDG-BER) to reform unmethylated cytosine (Zhu, 2009, He et al, 2011, Kohli and Zhang, 2013) (**Figure 1.7.6.1**). Interestingly, deletion of TDG does not affect demethylation in zygotes, suggesting other mechanisms exist for removing oxidised substituents (Guo et al, 2014). It has been suggested that DNMTs could theoretically function in demethylation by promoting the removal of oxidised substituents at the 5-C position of the pyrimidine ring although this has not been experimentally confirmed (Kohli and Zhang, 2013).

Nucleotide-based DNMT inhibitors such as 5-Aza-2'-deoxycytidine (5-Aza) can also act to cause demethylation in a DNA replication-dependant manner. 5-Aza is a chemical analogue of cytosine and can incorporate itself into DNA and bind to DNA methyltransferase enzymes which inhibits their activity (Christman, 2002). It has also been suggested that HDAC inhibitors can indirectly reverse CpG methylation by inhibiting ERK signalling which causes down-regulation of DNMT1 (Sarkar et al, 2011). Cytosine methylation is therefore reversible by either biochemical or biological manipulation and represents an attractive and exciting prospect for treating diseases in which aberrant DNA methylation plays an important pathological role.

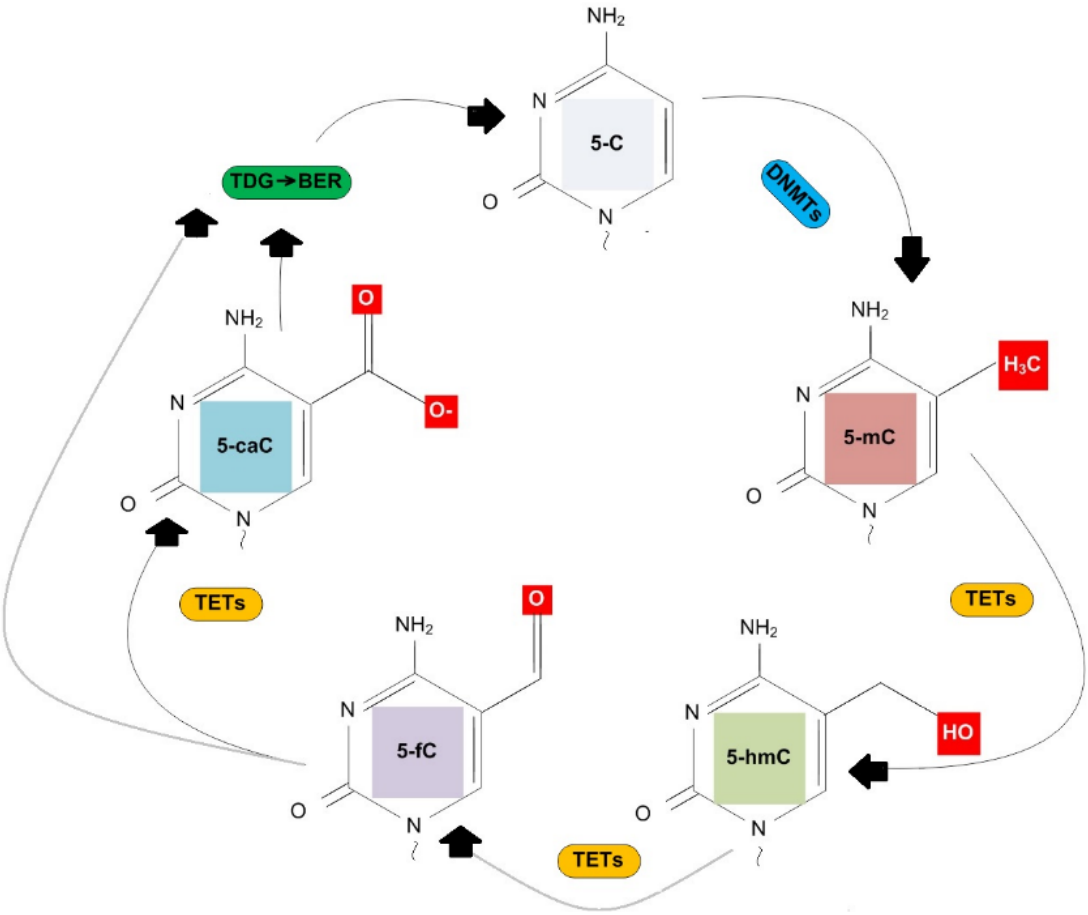


Figure 1.7.6.1. Active demethylation by TET enzymes TDG-triggered base excision repair. Cytosine (5-C) is converted to 5-methylcytosine (5-mC) by DNMTs. 5-mC is then converted to 5-hydroxymethylcytosine (5-hmC) by TET enzymes. 5-hmC can be converted by iterative oxidation to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) which can both be converted back to unmethylated cytosine by thymine DNA glycosylase-triggered base excision repair (TDG-BER). (Redrawn and adapted from Kohil and Zhang, 2013).

1.7.7. Modulation of DNA methylation

Factors including drugs (Christman, 2002, Nielsen et al, 2009), alcohol (Garro et al, 1991, Tao et al, 2011, Philibert et al, 2013), nutrition (Dominguez-Salas et al, 2014), age (Teschdorff et al, 2010, Bell et al 2012, Jung and Pfeifer, 2015), gender (Boks et al, 2009, Zhang et al, 2011, Tapp et al, 2013), smoking (Breitling et al, 2011, Lee and Pausova, 2013) and stress (Roth et al, 2009) can all influence and modulate methylation and as research continues to advance, this list is likely to become much larger (**Figure 1.7.7.1**). Alcohol can induce DNA methylation changes in sperm (Ouko et al, 2009), embryos (Garro et al, 1991, Wolff et al, 1998) and the brain (Otero et al, 2012) and can also disrupt enzymes (such as DNMTs) involved in methionine metabolism (Barak et al, 1987, Garro et al, 1991, Barak et al, 1993, Halsted et al, 1996, Lu et al, 2000). Methionine is a precursor to SAM which acts as a methyl donor in multiple reactions and plays an essential role in DNA methylation, thus changes to methionine metabolism can directly affect SAM-dependent methylation reactions. Alcohol may reduce absorption of folate which is also a key component of the methionine-homocysteine cycle (Halsted et al, 2002). There is strong evidence that folic acid is important in epigenetic programming (Steegers-Theunissen et al, 2009, Akchiche et al, 2012, Guéant et al, 2013) and maternal nutrition at conception has been associated with epigenetic alterations (Dominguez-Salas et al, 2014).

Cigarette smoke can modulate DNMT1 expression (Lee and Pausova, 2013) and induce hypoxia, which in turn leads to hypoxia-induced factor 1A (HIF-1A)-dependent up-regulation of methionine adenosyltransferase 2A (Suter et al, 2011). This enzyme is involved in the synthesis of SAM (Liu et al, 2011), thus smoking reduces methyl group availability. Cigarette smoke also contains carcinogens such as arsenic, chromium and nitrosamines which cause double-stranded breaks in DNA (Huang et al, 2012). DNMT1 is recruited to these repair sites (Mortusewicz et al, 2005) and can methylate adjacent CpGs (Cuozzo et al, 2007). A number of studies have identified smoke-related changes in DNA methylation (Breitling et al, 2011, Monick et al, 2012, Shenker et al, 2012, Sun et al, 2013, Besingi and Johansson, 2014, Dogan et al, 2014, Harlid et al, 2014, Elliott et al, 2014, Guida et al, 2015) and studies on women who smoke during pregnancy have shown genome-wide (Suter et al, 2011) and gene-specific (Maccani et al, 2013) changes in methylation. Furthermore, a recent study has shown that some CpGs can revert back to methylation states typical of non-smokers after smoking cessation, whereas other CpG sites remain differentially methylated even after more than 35 years smoking cessation (Guida et al, 2015).

Aging has been associated with altered methylation in multiple species (Vanyushin et al, 1973, Wilson et al, 1987, Richardson, 2003) and multiple studies have shown genes and genomic regions which either get hypermethylated or hypomethylated with age (Teschdorff et al, 2010, Bell et al, 2012). Global hypomethylation has been reported in old age (Fuke et al, 2004, Fraga and Esteller, 2007, Heyn et al, 2012) although CpG islands tend to gain methylation with age (Calvanese et al, 2009). It has also been shown that some CpGs have a linear change in methylation with age (Hannum et al, 2013),

however, it is unknown how epigenetic modifications including DNA methylation are regulated during aging.

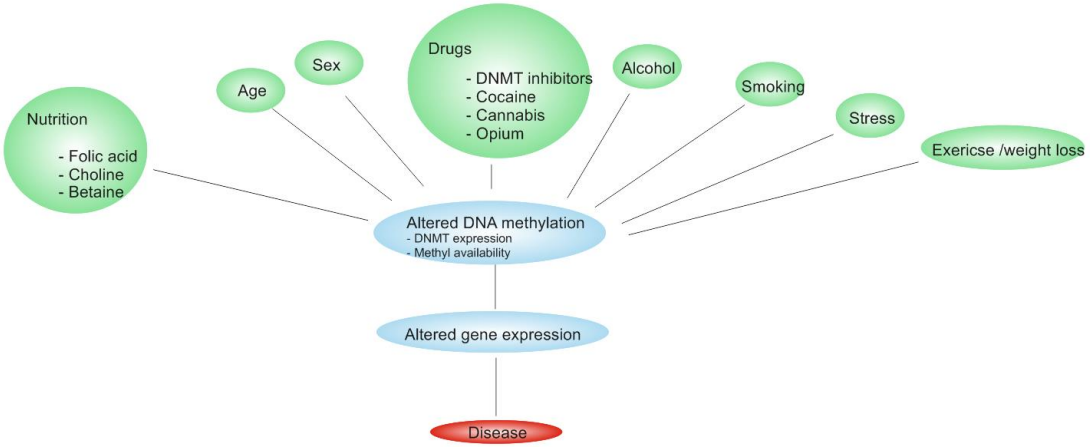


Figure 1.7.7.1. Factors which modulate DNA methylation. DNA can be influenced by multiple factors including, diet, lifestyle, drugs, stress, age and sex.

1.7.8. Differences in male and female methylation

As previously mentioned, DNA methylation plays an important role in genomic imprinting and X-chromosome inactivation. Genomic imprinting is a mechanism which determines the expression of a gene based on its parent of origin (Sharp et al, 2011). Either the paternal or maternal allele is hypermethylated which leads to monoallelic expression (Sharp et al, 2011). The epigenetic marks of imprinted genes are established in the male or female germline and maintained after fertilisation and during development (Kelsey et al, 2007, Ideraabdullah et al, 2008). Genomic imprinting is essential for multiple processes including foetal development and somatic differentiation (Reik et al, 2003).

X-chromosome inactivation is a mechanism that equalises the expression of sex-linked genes between males and females (Lyon, 1962) and results in silencing the majority of genes on one of the two X-chromosomes in each somatic cell of females (Carrel and Willard, 2005, Sharp et al, 2011). It has been shown that methylation profiles of human active and inactive X-chromosomes exist between males and females (Sharp et al, 2011). Furthermore, the majority of CpG sites on X-chromosomes in primates such as great apes show increased methylation in females compared to males which is consistent with the role of methylation in X-inactivation (Hernando-Herraez et al, 2013).

In the context of healthy individuals, it has previously been shown that male and female cells have different methylation patterns (Sarter et al, 2005, Liu et al, 2010, Hall et al, 2014), however, the effects of sex on genome-wide DNA methylation is poorly understood. Liu et al, (2010) studied the effects of sex on genome-wide methylation using saliva from humans and showed that the influence of sex on

methylation was CpG site-specific (Liu et al, 2010). Xu et al, showed sex-specific methylome profiles in the human prefrontal cortex (Xu et al, 2014). For some genes sex has been shown to be as strong an indicator of methylation as age (Sartar et al, 2005). It has also been shown that age as well as sex affects DNMT3B expression in human liver samples (Xiao et al, 2008). Females have significantly higher expression of DNMT3B which could in turn, influence global DNA methylation levels (Xiao et al, 2008). During early human embryo development, genome-wide methylation in male pronuclei is decreased compared to female pronuclei, at the end of the zygotic stage (Guo et al, 2014). Furthermore, human male foetuses show decreased methylation of multiple immune response genes compared to females (Flanagan, 2014). In mice, primordial germ cells have different methylation patterns in males and females during development (Kobayashi et al, 2013). Furthermore, multiple genes in mouse hybrid strains have altered methylation between males and females (Orozco et al, 2014).

In a disease context, it is often the case that one sex has a greater predisposition or tendency to be affected by a specific disease such as in IPF, which affects males more than females (Scott et al, 1990, Iwai et al, 1994, Coultas et al, 1994, Mannino et al, 1996, Johnston et al, 1997, Gribbin et al, 2006, Han et al, 2008, Nalysnyk et al, 2012) and SSc, which affects more females than males (Varga and Abraham, 2007). In many cases gene expression profiles are different between healthy and diseased patients. Genetic factors alone cannot explain these differences suggesting a role for epigenetic mechanisms such as DNA methylation, which has been shown in diseases such as familial breast cancer (Pinto et al, 2013) however, there are very few studies highlighting the role of methylation in sex-biased diseases.

1.7.9. Role of DNA methylation in disease

Defective DNA methylation in mammals is embryonic lethal highlighting its biological importance (Li et al, 1992, Okano et al, 1999). Mounting evidence suggests that DNA methylation has an important regulatory role in a growing number of diseases. In 1983 it was first demonstrated that genomes of cancer cells were hypomethylated compared to healthy cells (Feinberg and Vogelstein, 1983, Feinberg and Tycko, 2004). Aberrant DNA methylation is now associated with multiple different cancers (Ting et al, 2006) including ovarian (Ahluwalia et al, 2001) colon (Nakamura and Takenaga, 1998) and lung (Esteller et al, 2001, Zöchbauer-Müller et al, 2002). Aberrant DNA methylation has also been linked to a number of neurodegenerative diseases including Alzheimer's disease (Siegmund et al, 2007, Mastroeni et al, 2009, Sung et al, 2011, Chouliaras et al, 2013), Parkinson's disease (Jowaed et al, 2010, Matsumoto et al, 2010) and Huntington's disease (Thomas et al, 2013) and autoimmune diseases including systemic lupus erythematosus (Lu et al, 2002, Lu et al, 2007, Javierre et al, 2010), RA (Kim et al, 1996, Fu et al, 2011), multiple sclerosis (Mastronardi et al, 2007) and SSc (Wang et al, 2006, Jüngel et al, 2011) and fibrotic diseases including liver fibrosis (Komatsu et al, 2012), kidney fibrosis (Bechtel et al, 2010) and IPF (Sanders et al, 2012, Rabinovich et al, 2012, Huang et al, 2014, Yang et al, 2014).

Aberrant expression of DNMTs are also associated with disease. Loss of imprinting and aberrant X-chromosome inactivation are common manifestations with mutations or loss of expression of DNMT1 (Li et al, 1993). Neurodegenerative diseases (Desplats et al, 2011, Klein et al, 2011, Winklemann et al, 2012), cancers (Vogelstein et al, 2013) and a number of genetic disorders have also been associated with aberrant expression of DNMT1 (Robertson, 2005, Feinberg, 2007). Mutations in the DNMT3B gene cause a rare autosomal disease called immunodeficiency, centromere instability, facial abnormalities syndrome (Okano et al, 1999, Hansen et al, 1999). Mutations in DNMT3B has also been associated with hypomethylation of multiple X-linked genes (Li, 2002) and increased expression of DNMT3B has been linked to ischemic heart disease (Watson et al, 2014). Increased expression of both de novo DNMTs (DNMT3A and DNMT3B) has also been shown in IPF lung tissue (Sanders et al, 2012) which may have an effect on expression and methylation patterns in IPF, however they do not appear to affect global DNA methylation levels (Sanders et al, 2012).

1.7.10. The role of DNA methylation in IPF

The role of DNA methylation in the pathogenesis of IPF is poorly understood. Currently, three studies exist looking at global DNA methylation in human lung tissue (Sanders et al, 2012, Rabinovich et al, 2012, Yang et al, 2014). These studies have shown multiple genes in IPF with altered methylation, some of which inversely correlate with gene expression (Sanders et al, 2012 and Rabinovich et al, 2012, Yang et al, 2014). However, lung tissue contains multiple distinct cell types, thus making it impossible to determine cell-specific changes in methylation. The use of lung tissue could also potentially cause over or under estimation of methylation changes due to the multiple different cell types. Only one study exists that looks at global DNA methylation levels in fibroblasts (Huang et al, 2014). This study identified multiple genes that have altered methylation in IPF lung fibroblasts but was unable to accurately determine what effect the observed changes in DNA methylation had on gene expression as they did not use expression data from the same cell lines (Huang et al, 2014). Other studies have focused on gene-specific DNA methylation in IPF lung fibroblasts and have provided evidence that DNA methylation in promoter regions of genes can directly regulate their expression. These genes include THY-1 (Sanders et al, 2008, Robinson et al, 2012), Prostaglandin E receptor 2 (PTGER2) (Huang et al, 2010), P14 alternative reading frame (P14ARF) (Cisneros et al, 2012) and miR-17-92 (Dakhlallah et al, 2013) all of which have hypermethylation which corresponds with reduced gene expression (**Table 1.7.10.1**).

Gene	Tissue	Methylation	Expression	Reference
THY1	Lung fibroblasts	Hypermethylated	Decreased	Sanders et al, (2008)
P14ARF	Lung fibroblasts	Hypermethylated	Decreased	Cisneros et al, (2012)
PTGER2	Lung fibroblasts	Hypermethylated	Decreased	Huang et al, (2010)
miR-17-92	Lung tissue Lung fibroblasts	Hypermethylated	Decreased	Dakhlallah et al, (2013)
Multiple	Lung fibroblasts	Hyper and hypo methylation	Increased and decreased	Huang et al, (2014)
Cyclin-dependent kinase 4 inhibitor B (CDKN2B)	Lung fibroblasts	Hypermethylated	Decreased	Huang et al, (2014)
Caspase recruitment domain family, member 10 (CARD10)	Lung fibroblasts	Hypermethylated	Decreased	Huang et al, (2014)
O-6-methylguanine-DNA methyltransferase (MGMT)	Lung fibroblasts	Hypomethylated	Increased	Huang et al, (2014)
Multiple	Lung tissue	Hyper and hypo methylation	Increased and decreased	Sander et al, (2012), Rabinovich et al, (2012), Yang et al, (2014)

Table 1.7.10.1. Summary of DNA methylation alterations in IPF. The name of the genes studied, the source of tissue/cell used, and the methylation/expression state compared with controls is shown.

1.7.11. The role of DNA methylation in SSc

In peripheral blood mononuclear cells derived from SSc patients, DNA methylation patterns of genes on the X chromosome are different in monozygotic twins which has been proposed to affect susceptibility to SSc (Selmi et al, 2012). Global DNA methylation levels are decreased in SSc CD4+ T cells and a number genes have increased expression (Lei et al, 2009, Lian et al, 2012, Jiang et al, 2012) (**Table 1.7.11.1**). In contrast, skin fibroblasts show global hypermethylation and decreased expression of DNMT1, Methyl-CpG binding domain protein 1 (MBD1) and Methyl-CpG binding protein 2 (MeCP2) (Wang et al, 2006). Furthermore, FLI1 (Wang et al, 2006) and two Wnt antagonists: Dickkopf-related protein 1 (DKK1) and Secreted frizzled-related protein 1 (SFRP1) (Dees et al, 2013) are hypermethylated and correlate with decreased expression. Treatment with a DNMT inhibitor (5-Aza) restores expression of DKK1 and SFRP1, inhibits Wnt signalling and ameliorates experimental fibrosis (Dees et al, 2014) suggesting DNA methylation plays an important role in Wnt signalling in SSc. Bone morphogenetic protein receptor type II (BMPR2), a gene with important functions in vascular cell proliferation and apoptosis has been shown to be hypermethylated and have decreased expression in microvascular endothelial cells (MVECs) (Wang et al, 2013). This could explain the increased apoptosis of MVECs in SSc which is seen in all affected organs (Sgonc et al, 1996). In the context of pulmonary fibrosis associated with SSc, to the author's knowledge, there are no studies examining DNA methylation from SSc-PF lung tissue or lung fibroblasts at a gene-specific or global level.

Gene	Tissue	Methylation	Expression	Reference
Cluster of differentiation 40 ligand (CD40LG)	CD4+ T-cells	Hypomethylated	Increased	Lian et al, (2012)
Cluster of differentiation 70 (CD70)	CD4+ T-cells	Hypomethylated	Increased	Jiang et al, (2012)
DNMT1	CD4+ T-cells	Hypomethylated	Increased	Lei et al, (2009)
MBD3	CD4+ T-cells	Hypomethylated	Increased	Lei et al, (2009)
MBD4	CD4+ T-cells	Hypomethylated	Increased	Lei et al, (2009)
DNMT1	Skin fibroblasts	Hypermethylated	Decreased	Wang et al, (2006)
MBD1	Skin fibroblasts	Hypermethylated	Decreased	Wang et al, (2006)
MeCP2	Skin fibroblasts	Hypermethylated	Decreased	Wang et al, (2006)
FLI1	Skin fibroblasts	Hypermethylated	Decreased	Wang et al, (2006)
DKK1	Skin fibroblasts PBMCs	Hypermethylated	Decreased	Dees et al, (2013)
SFRP1	Skin fibroblasts PBMCs	Hypermethylated	Decreased	Dees et al, (2013)
BMPR2	MVECs	Hypermethylated	Decreased	Wang et al, (2013)

Table 1.7.11.1. Summary of DNA methylation alterations in SSc. The name of the genes studied, the source of tissue/cell used, and the methylation/expression state compared with controls is shown.

1.8. IPF therapeutic options

The lack of an effective therapy in IPF is a consequence of not knowing the etiology of IPF. Previous studies have tested numerous drugs aimed towards treating IPF, none of which have elicited a consistent, beneficial response in IPF patients. These unsuccessful treatments include anti-inflammatory agents (Nagai et al, 1999, Flaherty et al, 2001, Richeldi et al, 2003, Collard et al, 2004), anti-coagulants (Noth et al, 2012), endothelin receptor antagonists (King et al, 2008, Jackson et al, 2010, King et al, 2011, Costabel et al, 2011) and anti-fibrotics (King et al, 2009). A number of drug trials have reported no beneficial outcome, worsening of disease or have even been stopped due to increased mortality (Raghu et al, 2011, Raghu et al, 2012).

Currently, pirfenidone is the only drug approved for the treatment of IPF and has acceptable tolerability in clinical trials (Noble et al, 2011). Pirfenidone has anti-inflammatory and antioxidant properties that inhibit TGFß in vitro (Walter et al, 2006) and in animal models of fibrosis, pirfenidone has been shown to inhibit TGFß mRNA and TGFß-stimulated collagen production (Di Sario et al, 2002, Oku et al, 2008). In IPF patients, pirfenidone is associated with reduced fibroblast proliferation (Schaefer et al, 2011), improved progression-free survival (Spagnolo et al, 2010) and a reduction in acute exacerbations of IPF (Azuma et al, 2005), although side-effects including dermatological, gastrointestinal and neurological symptoms are common (Noble et al, 2011, Cottin, 2013, Jenkins et al, 2013). Nintedanib, a small molecule tyrosine kinase inhibitor which targets growth factor receptors (PDGF, VEGF and FGF) associated with IPF (Hilberg et al, 2008, Richeldi et al, 2011) was recently shown to significantly reduce the annual decline in forced vital capacity (FVC) in IPF patients enlisted on a phase III trial (INPULSIS) compared to placebo (Richeldi et al, 2014). Pirfenidone is the first drug to offer some benefits to patients with IPF and nintedanib may soon be available however, with the exception of lung transplants, no treatment to date exists that can reverse or prevent IPF.

1.8.1. SSc-PF therapeutic options

The majority of SSc-PF cases are characterised by a NSIP histopathological pattern (King, 2005) which is associated with a better prognosis compared to UIP, the characteristic histopathological pattern of IPF (Monaghan et al, 2004). However, as with IPF, the treatment of SSc-PF is not well established. A number of drugs including corticosteroids, anti-fibrosing agents and immunosuppressive agents have been proposed for treating SSc-PF (Bussone and Mouthon, 2011). Corticosteroids, which have anti-inflammatory, immunosuppressive and anti-fibrosing properties, are recommended (at a low dose) for SSc patients with severe or worsening interstitial lung disease (Wells and Hirani, 2008), however the efficacy of corticosteroids in SSc-PF are unknown and high doses are associated with scleroderma renal crisis (Trang et al, 2012). Immunosuppressive agents such as cyclophosphamide which suppress lymphokine production are widely used in the treatment of SSc-PF however, multiple studies have shown no long-term benefits of cyclophosphamide compared to placebos (Hoyles et al, 2006, Tashkin et al, 2007, Nannini et al, 2008). That said, cyclophosphamide combined with azathioprine (another immunosuppressive drug) may be beneficial by stabilising or improving lung function in worsening cases of SSc-PF (Paone et al, 2007, Bérezné et al, 2008). Anti-fibrotic agents including D-Penicilliamine, IFN-γ and IFNα have all failed to provide a beneficial effect in patients with SSc (Grassegger et al, 1998, Black et al, 1999, Clements et al, 1999), thus no anti-fibrotic therapy currently exists. Two recent studies by Udwadia et al and Miura et al examining 1 and 5 patients with SSc-PF (respectively) suggest pirfenidone may provide some beneficial effects including improved lung function and increased vital capacity (Miura et al, 2014, Udwadia et al, 2015). Currently a phase II trial (LOTUSS) is examining the safety and tolerability of pirfenidone in patients with SSc-PF and studies examining other promising drugs including imatinib, dazatinib, rituximab and mycophenolate mofeti are ongoing, however to date, there is no effective 'gold standard' treatment currently recommended for patients with SSc-PF (Bussone and Mouthon et al, 2011).

1.9. Summary and Hypothesis

This thesis will investigate the role of DNA methylation in fibrotic lung fibroblasts derived from patients with either IPF or SSc-PF. PF often occurs in patients with SSc and although SSc-PF and IPF are distinct diseases, they share some overlapping clinical similarities including reduced FVC and dypsnea on exertion (Herzog et al, 2014). PF is a pathological condition originating from aberrant repair mechanisms in response to acute or chronic lung injury, characterised by progressive scarring of the lungs and excessive ECM deposition in the pulmonary interstitium and often occurs at the end-stages of many interstitial lung diseases. Fibroblasts and myofibroblasts are the key effector cells of fibrogenesis in which multiple genes have been identified as being aberrantly expressed. DNA methylation is one of many epigenetic mechanisms that play a key role in gene regulation.

Recent evidence has identified a number of genes in IPF lung fibroblasts and SSc fibroblasts, as having altered methylation which correlate to gene expression, however the full extent and role of DNA methylation in IPF and SSc-PF remains unknown. This is the first study to date to examine genome-wide methylation in lung fibroblasts derived from SSc patients and the first study to examine the effects of DNA methylation of gene expression in the same lung fibroblasts derived from IPF patients.

A number of genes have altered methylation which correspond to altered gene expression in IPF lung fibroblasts (Sanders et al, 2008) and many of these encode glycoproteins. Genes of particular interest are the tenascins which have important functions during wound healing, however, their role in fibrosis and wound healing and how they are regulated is poorly understood. TNC has recently been shown to have increased expression in IPF (Estany et al, 2014) and SSc-PF (Brissett et al, 2012, Inoue et al, 2013) and is upregulated in response to TGFß (Estany et al, 2014). Furthermore aberrant expression of TNXB (the gene which encodes the protein TNX) plays an important pathogenic role in diseases such as cardiac fibrosis (Jing et al, 2011), Ehlers-Danlos syndrome (Burch et al, 1997, Mao et al, 2002, Zweers et al, 2003) and mesothelioma (Yuan et al, 2009). TNX can also associate with different collagen types and has the potential to activate latent TGFß (Alcaraz et al, 2014), the pro-fibrotic cytokine universally linked with fibrosis (Leask and Abraham, 2004). Furthermore, TNXB expression has been linked with DNA methylation, therefore the regulation and role of TNX in IPF and other fibrotic diseases including SSc-PF warrant further investigation.

This thesis will address the following hypothesis:

Aberrant DNA methylation in fibrotic lung fibroblasts derived from patient with IPF or SSc-PF is responsible for altering the expression of multiple genes which may contribute to the development of pulmonary fibrosis.

The aims of this thesis are to:

- Identify genes with altered methylation/expression in IPF and SSc compared to control lung fibroblasts.

- Identify overlapping and distinct genes between IPF and SSc lung fibroblasts which have altered methylation/expression compared to control lung fibroblasts.

- Correlate DNA methylation and gene expression data to identify genes potentially regulated directly by methylation.

- Identified biological processes enriched in genes with altered methylation/expression in IPF and SSc compared to control lung fibroblasts.

- Determine whether sex has an influence on methylation and/or expression of genes.

- Determine the effect of DNMT inhibition, using the demethylating agent, 5-Aza-2'-deoxycytidine, on methylation/expression of genes in control, IPF and SSc lung fibroblasts.

- Elucidate the role of DNA methylation in TNXB gene regulation and the role of TNX in IPF.

Chapter 2. Materials and methods

2.1. General plastic-and glass-ware and chemicals

All chemicals used were of analytical grade. Water used for the preparation of buffers was distilled and deionised using a Millipore water purification system (Millipore Ltd, UK). Plastic-and glass-ware included sterile tissue culture flasks, plates, disposable pipettes (Nunc, Denmark), Falcon tubes (Sigma, UK) and Poly-L-Lysine slides (VWR International, USA). Trypsin-EDTA (Fisher Scientific, USA), penicillin, streptomycin, foetal calf serum, Dulbecco's modified Eagle's medium (Life Technologies, UK), amphotericin B (Sigma, UK), NF-H_2O (Ambion, USA), BSA (Bioline, USA), Agarose (Merick, Germany), 96-well plates (Thermo Scientific, USA), 0.2ml PCR tubes (Starlab, UK), eppendorfs (Sarstedt, Germany), RNaseZAP (Sigma, UK), TRIzol (Invitrogen, UK), SYBR (Eurogentec, Belgium), q-Script cDNA supermix (Quanta Biosciences, USA), EZ-DNA methylation-Gold kit (Zymo Research, USA), mineral oil, (Sigma, UK), 10x PCR buffer (Applied Biosystems, USA), magnesium chloride (Applied Biosystems, USA), QIA purification kit, (Qiagen, Germany), Opti-MEM (Life Technologies, UK), ethanol (VWR International, France), chloroform (Sigma, UK).

2.2. Reagents, inhibitors and antibodies

INTERFERin (Polyplus Transfection, USA), HEPES buffer (Sigma, USA), saponin (Sigma, USA), 5-Aza-2'-deoxycytidine (5-Aza) (Sigma, UK), polyclonal rabbit anti-TNXB, (ab111270) (AbCam, UK), goat-anti-rabbit IgG HRP, (P0448) (Dako, Denmark), TNXB, (AM16708) and non-targeting control siRNA, (AM4390846) (Ambion, USA), DNA HyperLadder (Bioline, USA), rat tail type I collagen, (A1048301) (Invitrogen, UK).

2.3. Tissue culture

2.3.1. Isolation of lung fibroblasts

Fibrotic primary human lung fibroblast cell lines were derived from either IPF or SSc-PF lung explants. The SSc-PF cell lines were a kind gift from Professor David Abraham at the Centre for Rheumatology and Connective Tissue Diseases, UCL, UK, and are referred to throughtout the methods and results sections of this thesis as ‘SSc lung fibroblasts’. Primary human control lung fibroblasts were derived from histologically normal areas of lung parenchyma distal to tumour mass in lung cancer patients. All tissues were obtained with patient consent and approved for use by the relevant research ethics committee. Sections ($1mm^3$) cut from lung biopsies were placed ~1cm apart on cell culture Petri dishes with 2ml of Dulbecco's modified Eagle's medium (Life Technologies, UK) supplemented with 10% (v/v) foetal calf serum (Life Technologies, UK), penicillin (100U/ml) (Life Technologies, UK), streptomycin (100µg/ml), L-glutamine (4mM) (Life Technologies, UK) and 2.5µg/ml amphotericin B (Sigma, UK). Cells were incubated at 37°C in a humidified atmosphere of air containing 10% CO_2. Once cells had attached, a further 8ml of culture medium was added (drop wise). Culture medium was replaced with fresh medium one day after isolation and 200µl of amphotericin B was added (drop-wise) every 48

hours. Medium was changed every seven days. Static cultures of cells were maintained in T175cm^2 plastic culture flasks at 37^{o}C in a humidified atmosphere of air containing 10% CO_2. Cells were passaged using trypsin-EDTA upon reaching 80% confluence and characterised by their morphology, their expression of several markers including α-smooth muscle actin and vimentin and their ability to secrete collagens in the presence of TGFß.

2.3.2. Cell culture

For all studies, IPF, SSc and control lung fibroblasts were used at passage ≤8 to ensure cells did not undergo senescence. Cells were routinely plated in T175cm^2 tissue culture flasks in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) foetal calf serum, penicillin (100U/ml), streptomycin (100µg/ml) and L-glutamine (4mM). Cell cultures were maintained in T175cm^2 plastic culture flasks at 37^{o}C in a humidified atmosphere of air containing 10% CO_2. Upon reaching 80% confluence cells were split. Medium was removed and cells were washed twice to remove any residual FCS. 2ml of trypsin-EDTA was added to each T175cm^2 flask, ensuring the whole flask's surface area was covered and incubated at 37^{o}C for approximately 2 minutes. Supplemented DMEM (23ml) was then added to each flask to neutralise the trypsin-EDTA and transferred to a sterile 50ml falcon tube. Cells were gently centrifuged at 200g for 5minutes. The supernatant was removed, leaving the cell pellet. Cells were reconstituted in 10ml of supplemented DMEM and counted using a Scepter^{TM} 2.0 handheld cell counter, (C85360) with 60µm tips (PHCC60050) (Millipore, UK). Cells were then split into new T175cm^2 flasks.

For Illumina microarray studies, 500,000 lung fibroblasts were counted using a Scepter^{TM} 2.0 handheld cell counter, (C85360) and seeded in T175cm^2 flasks at 10% confluence 24 hours before treatment with or without 5-Aza-2'-deoxycytidine (5-Aza) (Sigma, UK). 1µM of 5-Aza was added to the culture medium and replaced at 24 hour intervals until the cells were confluent (≥ 1 week, to ensure a minimum of 3 population doublings). Cells were then harvested for total RNA (TRIzol) or DNA (Nucleon) extraction. This was conducted by Dr. I.C. Evans. The number, gender and the mean age ± the standard deviation of the lung fibroblasts analysed using Illumina microarrays were as follows: IPF, n=5, aged 66.6 ±8 years, 2 male; SSc, n=7, aged 51.7 ±3.7 years, 1 male; and control, n=6, aged 58.3 ±14.5 years, 2 male (**Table 2.3.2.1**).

Fibroblast cell lines	Gender	Age
Non-fibrotic controls		
C1	F	69
C2	F	73
C3	F	33
C4	M	55
C5	M	66
C6	F	54
IPF		
IPF1	F	61
IPF2	M	58
IPF3	M	77
IPF4	F	73
IPF5	F	64
SSc-PF		
SSc1	M	47
SSc2	F	56
SSc3	F	49
SSc4	F	53
SSc5	F	56
SSc6	F	48
SSc7	F	53

Table 2.3.2.1. Demographic data of lung fibroblast cell lines. This table shows the gender and age of all lung fibroblasts used in the methylation and expression microarray studies. IPF, n=5, aged 66.6 ±8 years, 2 male; SSc, n=7, aged 51.7 ±3.7 years, 1 male; and control, n=6, aged 58.3 ±14.5 years, 2 male.

2.4. RNA extraction

All equipment was cleaned with RNaseZAP (Sigma, USA) prior to use and nuclease-free pipette tips were used (Continental Lab Products, UK) to minimise RNA degradation. TRIzol (Invitrogen, UK) was used to isolate RNA from primary lung fibroblasts. TRIzol is a monophasic solution of phenol and guanidinium isothiocyanate which solubilises biological material and denatures protein whilst maintaining RNA integrity (Rio et al, 2010). To isolate RNA, TRIzol was added to primary lung fibroblast cell cultures and incubated for 5 minutes at room temperature to ensure complete dissociation of nucleo-protein complexes. Chloroform (200µl per/ml of TRIzol reagent) was added to each sample and vortexed for 10 seconds followed by an incubation period of 5 minutes at room temperature to allow separation of the upper aqueous phase and a lower organic layer. Samples were then centrifuged at 16000g for 15 minutes at 4^{o}C. The upper clear layer containing approximately 250µl RNA was pipetted into a new tube. An equal amount of 2-propanol (250µl) was added to each sample and incubated for 10 minutes at room temperature to precipitate the RNA. Samples were then centrifuged at 16000g for 15 minutes at 4^{o}C. Supernatant was discarded leaving the pellet. Ethanol 80%/1ml (VWR Chemical, UK) was added to each sample, vortexed for 10 seconds and centrifuged at 16000g for 10 minutes at 4^{o}C. Supernatant was removed leaving the pellet which was left to air dry.

Once dried, samples were re-suspended in 12.5µl of nuclease-free H_2O (NF-H_2O) (Ambion, USA). 10X buffer (1.5µl) and 1µl DNase-1 enzyme (Ambion, USA) was added to each tube, vortexed for 10 seconds, pulse spun and incubated for 20 minutes at 37^oC. Inactivation buffer (2µl) was then added to each tube and incubated for 5 minutes at room temperature. Samples were centrifuged at 16000g for 3 minutes and the supernatant placed into a new tube.

The extracted RNA was quantified by measuring the 260nm wavelength absorbance reading (A260), using a NanoDrop Bioanalyzer ND1000 (NanoDrop, UK). The A260 reading represents the wavelength of maximum absorption of light by RNA and a reading of 1.0 is equivalent to 40µg/ml. The purity of RNA was determined by measuring the A260/A280 ratios where a ratio of ~2.0 is considered pure. As RNA has a maximum absorption of A260, any absorption at the 280nm wavelength could indicate the presence of non-RNA material such as protein. RNA with a 260/280 ratio between 1.8-2.2 was considered acceptable for use in experiments. For studies utilising the Illumina HT12v4 BeadChip expression microarray, 1µg of RNA was sent for analysis.

2.5. DNA extraction

A Nucleon blood and cell culture (BACC2) DNA extraction kit (Amersham, UK) was used to harvest DNA from the same primary lung fibroblast cell cultures used for the gene expression microarray. Following the manufacturer's protocol, lung fibroblasts were trypsinised, counted using a Scepter^{TM} 2.0 handheld cell counter, (C85360) with 60µm tips (PHCC60050) (Millipore, UK) and centrifuged in 15ml falcon tubes at 4^OC for 5 minutes. For 1x 10^6 cells, 250µl of reagent A was added for 5 minutes followed by centrifuging at 1300g for 5 minutes. Supernatant was discarded and 500µl of reagent B was added to each sample and vortexed gently. To remove proteins, 125µl of sodium perchlorate solution was added to each tube and inverted several times. DNA was extracted by adding 500µl of chloroform (Sigma, UK) followed by inverting the tube several times. Nucleon resin (75µl) was then added (which covalently binds proteins and traps proteinaceous material allowing the recovery of high quality DNA) followed by centrifuging at 1300g for 3 minutes. DNA was precipitated by transferring the upper phase into a new tube and adding ~ 2x the volume of ethanol to each sample. The DNA was centrifuged at 4000g for 5 minutes and supernatant discarded, followed by washing the DNA in 2ml of 70% ethanol and re-centrifuging. The DNA pellet was then air-dried and re-dissolved in NF-H_20 (Ambion, USA).

The quantity of the DNA was determined using a Quant-iT^{TM} Picogreen kit (Invitrogen, UK). Deionised water pre-treated with 0.1% diethyl pyrocarbonate (DEPC) for 12 hours at 37^oC was autoclaved for 15 minutes and used to dilute a 20X Tris-EDTA (TE) stock solution to form a 1XTE solution. The Quant-iT^{TM} Picogreen kit was warmed to room temperature and diluted 1:200 in the 1XTE buffer. A standard curve was prepared by diluting 100µg of the lambda DNA standard 50-fold in 1XTE to make a 2µg/ml working solution. DNA samples were diluted in the picogreen solution and incubated at room temperature for 5 minutes. Fluorescent readings were measured using a florescent plate reader and

compared to the standard curve to determine their concentrations. DNA samples with an A260/280 ratio of 1.8-2.1 were considered acceptable for experimental use. This was conducted by Dr. I.C Evans.

2.5.1. Bisulfite conversion of genomic DNA

For the Illumina Infinium HumanMethylation450 BeadChip microarray, 500ng of each genomic DNA sample was sent to Cambridge Genomic Services (CGS, UK) to be bisulfite converted using an EZ DNA Methylation-Gold™ Kit (Zymo Research, USA). For validation studies, the same DNA was bisulfite converted in-house. The manufacturer's protocol was followed which essentially involved adding a CT conversion reagent to the DNA and thermal cycling at 98°C for 10 minutes and then at 64°C for 2.5 hours, followed by 4°C until further processing. DNA from the lung fibroblast sample was then added to a binding buffer and centrifuged at ≥ 10000g for 30 seconds. Flow through was discarded. Wash buffer was added and centrifuged at ≥ 10000g for 30 seconds. Samples were incubated in desulphonation buffer at room temperature for 20 minutes and then centrifuged at ≥ 10000g for 30 seconds. Wash buffer was added and centrifuged at >10000g for 30 seconds. Elution buffer was then added and centrifuged at ≥ 10000g for 30 seconds to elute the DNA. Bisulfite converted DNA was analysed by an Illumina Infinium HumanMethylation450 BeadChip microarray and specific regions of the TNXB gene were bisulfite sequenced to validate the array (see **section 2.11**). **Figure 2.5.1.1** shows how bisulfite conversion only affects unmethylated cytosines.

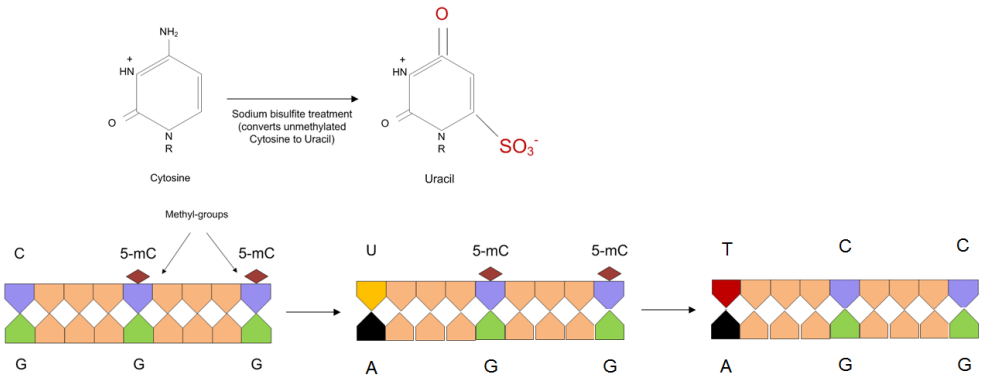


Figure 2.5.1.1. Bisulfite conversion of genomic DNA. Bisulfite conversion results in unmethylated cytosines being converted to uracil. Subsequent PCR amplification results in thymine in the place of where the original unmethylated cytosines were. Methylated cytosine residues remain unaffected.

2.6. Illumina microarrays

The Illumina Infinium HumanMethylation450 BeadChip methylation microarray (Illumina, USA) and the Illumina Infinium HT 12v4 BeadChip expression microarray (Illumina, USA) were used to analyse genome-wide methylation and expression in primary human lung fibroblasts derived from IPF (n=5), SSc (n=7) and control (n=6) lung. Both these microarrays employ Illumina's BeadArray technology which consists of 3µm silica beads, each of which is covered in multiple copies of a probe (a specific 50-mer (50 nucleotides in length) oligonucleotide that acts as the capture sequence). The beads are

uniformly distributed (~5.7µm), randomly assembled into microwells and held in place by Van der Waals force and hydrostatic interaction with the walls of the well.

2.6.1. Illumina Infinium HT 12v4 BeadChip microarray

The Illumina Infinium HT 12v4 BeadChip expression (Illumina, USA) simultaneously profiles 47231 transcripts and known splice variants from the RefSeq database release 38. Oligonucleotide (50-mer) probes on the HT12v4 expression microarray are specific to each gene and immobilised to beads, with approximately 30 beads per a probe on the array. Each probe is designed using a multi-step algorithm to optimise parameters including self-complementarity for hairpin structure prediction, melting temperature for hybridisation uniformity, distance from 3' end of transcript and lack of similarity to other genes. A 29-mer address sequence on each probe helps identify the location of each bead and validates the hybridisation process for each bead on the array. Each nucleotide is biotin-labelled and streptavidin-Cy3 is used to detect for downstream analysis (**Figure 2.6.1.1**).

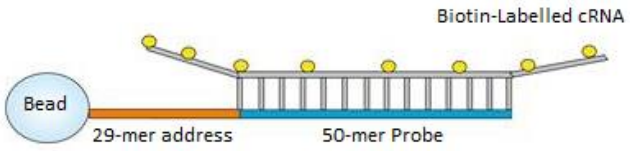


Figure 2.6.1.1. Gene expression profiling Illumina bead design. Thousands of oligomers (one shown for simplicity) are attached to a single bead. The address sequence is used to identify the oligonucleotide after it has been deposited on the array. (Figure adapted and redrawn from http://www.illumina.com/technology/beadarray-technology/direct-hybridization-assay.html).

2.6.2. Illumina Infinium HumanMethylation450 BeadChip microarray

The Illumina Infinium HumanMethylation450 BeadChip microarray (Illumina, USA) is the successor to the Illumina Infinium HumanMethylation 27k BeadChip microarray (Illumina, USA) and interrogates methylation at 482421 CpG sites across the genome, including 90% of the CpGs covered by the 27k array (Bibikova et al, 2011, Sandoval et al, 2011). 99% of all RefSeq genes with an average of 17 CpGs/gene and 96% of all known CpG islands are covered. Furthermore, regions flanking CpGs islands including shore, shelf and open sea are also covered, representing genome-wide coverage. The 450k array uses two types of probe to detect methylation; Infinium type I probes which use two probes per a CpG to detect methylation (one specifically to hybridise to methylated DNA and the other to specifically hybridise to unmethylated DNA) and Infinium type II probes which use just one probe per a CpG locus. Type I probes are both in the same colour channel. The % of methylation (Beta-value) at any given CpG is calculated by comparing the intensities from the two different probes in the same colour channel (beta=M/ (U+M). Type II probes utilise either green or red dye colours to distinguish between methylated or unmethylated CpGs. The beta-value is calculated by comparing the two colours at the same CpG locus (beta= green (M)/ (red (U) + green (M)). Each probe is 50-mer

oligonucleotide, attached to a bead and replicated multiple times on the array. This means the methylation status of each CpG can be analysed thousands of times on a single run (**Figure 2.6.2.1**).

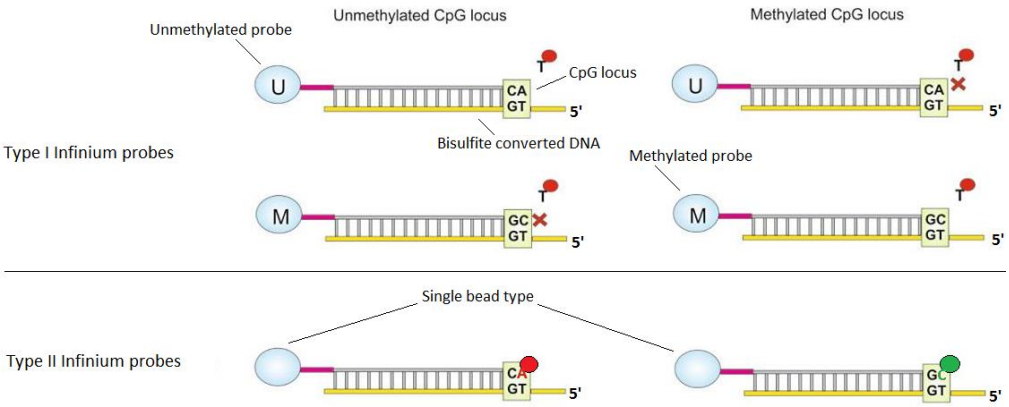


Figure 2.6.2.1. Type I and Type II Infinium probes used in the 450k array. Type I probes use two probes (one unmethylated, one methylated) to detect methylation at a CpG using the same colour channel. Type II probes use one probe to detect methylation at a CpG using either green (for methylated) or red (for unmethylated). (Adapted and redrawn from http://www.illumina.com/technology/beadarray-technology/infinium-methylation-assay.html).

2.6.3. Normalisation of the Illumina Infinium HT 12v4 BeadChip array

Data from the Illumina Infinium HT 12v4 BeadChip expression (Illumina, USA) was normalised and subjected to quality control by Cambridge Genomic Services (CGS, UK). The lumi R package (Du et al, 2008) was used to determine if a gene was significantly ($P<0.01$) expressed above the background as defined by the negative control probes. For any given probe to pass selection, it must have been detected in at least one sample ($P<0.01$ in Lumi). Normalisation of the HT12v4 array was done in two steps: first the data was transformed using variance stabilisation to stabilise the variance of larger intensities and reduce the number of false-positive results (Lin et al, 2008) followed by quantile normalisation to reduce background noise using the R package Lumi (Du et al, 2008). A basic analysis was performed by Cambridge Genomic Services (CGS, UK), using the limma R package (Smyth et al, 2004, which included calculating the fold-changes in expression between different groups, the number of probes detected in each sample and false-discovery P values.

2.6.4. Normalisation of the Illumina Infinium HumanMethylation 450 array

The Illumina Infinium HumanMethylation 450 BeadChip array was normalised and subjected to quality control by Cambridge Genomic Services (CGS, UK). Raw data was obtained using Genome Studio software (Illumina, USA) and processed using the lumi R package to correct for colour bias present on the array due to different dyes used on the array. To correct this bias, Infinium type I and type II probes and both colour channels were separated and smooth quantile normalsation was applied. After this correction, both channels and probe types were combined and quantile normalisation was performed

using the lumi R package (Du et al, 2008). Beta-values were then calculated and probes demonstrating p values >0.01 removed.

2.6.5. Filtering non-specific probes and probes covering a SNP

Several studies have identified non-specific probes on the Illumina Infinium HumanMethylation450 BeadChip microarray as well as more SNPs which cover CpGs in addition to the ones already documented on the array (Price et al, 2013, Chen et al, 2013). I matched a list of all non-specific probes (41937) identified by Price et al was to our data. Price et al followed a protocol developed by Chen et al, 2011 for the 27k array to determine non-specific probes on the 450k array. Essentially, the basic local alignment search tool-like analysis tool (BLAT) (Kent, 2002) was used to align the 50 nucleotide probe sequences to four different versions of the hg19 draft sequence genome. These included a fully unmethylated 'bisulfite-treated' genome (where all Cs were converted to Ts), a fully methylated 'bisulfite-treated' genome (where only non-CpG Cs converted to Ts) and both the above versions on the reverse complement sequence (Price et al, 2013). For a probe to be considered non-specific, the following criteria had to be met: at least 90% identity over the aligned region (calculated by the dividing the number of matching bases by the span of the sequence match), at least 40/50 base pairs matching, no gaps (to avoid the potential of compromising the hybridisation between probes and cross-reactive sequences) and the 50th nucleotide (where the probe hybridises to the target CpG) had to align (Chen et al, 2011, Price et al, 2013). All non-specific probes were removed as they have the potential to hybridise to multiple genomic locations and thus could measure methylation at multiple CpG sites. Any CpG site which was covered by a SNP was also removed to avoid potentially assessing genotype rather than methylation differences. This left 324973 probes on the array.

2.6.6. Statistical analysis of the Illumina Infinium HT 12v4 BeadChip array

Due to a small sample size and heterogeneity that exists between IPF samples (Martinez et al, 2005, Habiel and Hogaboam, 2014, DePianto et al, 2015), very few genes reached statistical significance in the expression analysis using a false-discovery rate P value of <0.05. To minimise the number of false negative and false-positive results and in order not to miss true positive results, a threshold number of misclassifications (TNoM) score of ≤1 with a p value <0.05 was used. TNoM is a threshold-based method that separates class values (ie. control compared to IPF) to determine if a gene's expression is above or below a given threshold. This was calculated by listing each cell lines' expression value in ascending order and then setting a threshold (**Figure 2.6.6.1**). The TNoM score represents the minimum number of misclassifications possible (Bon-Dor et al, 2000).

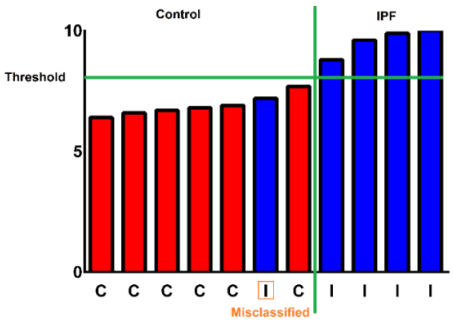


Figure 2.6.6.1. TNoM definition. Control (n=6) and IPF (n=5) gene expression values listed in ascending order. Four IPF values are above the threshold, 1 is below the threshold representing a misclassification. Therefore in the above example, the minimum number of misclassifications = 1.

2.6.7. Statistical analysis of the Illumina Infinium HumanMethylation 450 array

Due to a small sample size and heterogeneity that exists between IPF samples (Martinez et al, 2005, Habiel and Hogaboam, 2014, DePianto et al, 2015), very few genes reached statistical significance in the methylation analyses using an false-discovery rate P value of <0.05. This has previously been shown in another IPF study using the smaller Illumina 27K microarray (Huang et al, 2014). To minimise the number of false negative and false positive results and in order to not miss true positive results, a delta-beta (Δβ) of ≥0.136 (≥13.6%) change in methylation with a non-stringent p value of <0.05 was used as a statistical cut-off based on previous studies utilising Illumina microarrays which show a P<0.05 and 13.6% change in methylation can detect differences with 95% confidence (Bibikova et al, 2009, Lokk et al, 2012).

2.7. Bioinformatics analysis of microarray data

2.7.1. Data visualisation

The R 3.2.0 (x64bit) statistical program and associated packages (including ggplot2 (Wickham, 2009), VennDiagram (Chen and Boutros, 2011), and RColorBrewer) and Microsoft Excel were used for visualising and analysing Illumina microarray data. Using R, I developed scripts which analysed Illumina microarray data saved in comma delimited file format (.csv). After installing and loading in the appropriate R package, .csv files were loaded and read into R. The scripts were then applied to the data and all images exported as .png files unless otherwise stated. Histogram scripts were used to analyse the distribution of genome-wide methylation and expression (**Figure 2.7.1.1**). For methylation and expression distribution, class boundaries (binwidth) were set at 0.05 (5% methylation) and 0.5 respectively. Scatter plots (**Figure 2.7.1.2**), correlation (**Figure 2.7.1.3**) and bar graph (**Figure 2.7.1.4**) scripts were used to determine the distribution of CpG methylation in relation to CpG location, chromosomal location and for correlation between CpG methylation and gene expression. Weighted Venn diagram scripts were used to determine the number of distinct and overlapping genes which had altered methylation/expression (**Figure 2.7.1.5**). Heatmap/cluster scripts were used to determine clustering of cell lines (**Figure 2.7.1.6**).

```
library(ggplot2)
A<-read.csv("filename.csv", as.is=T)
g<-ggplot(A, aes(x=column1))+geom_histogram(binwidth=0.05, color="black", fill="colour")+
xlim(c(0,1.0))+ylim(c(0,100.0))+
facet_wrap(~columnX)
g<-g+labs(x="x-axis title", y="y-axis title")
g
g + theme(text = element_text(size=18),
axis.text.x=element_text(size=14, colour= 'black', angle=90),
axis.text.y=element_text(size=14, colour= 'black'),
axis.title.x = element_text(color="black"),
    axis.title.y = element_text(color="black"),
legend.key=element_rect(fill='NA'),
legend.title=element_blank())
```

Figure 2.7.1.1. Colour-coded script used for the generation of histograms using R 3.2.0. Blue: loads the required package into R console. Green: reads in the file to be used for analysis and gives it a simple label of 'A'. Orange: creates the plot based on the values in 'column 1', sets each histogram border at 0.05 (representing 5% methylation), outlines each bar in black and fills it with a choice of colour and sets a limit on the plot size. In this example xlim is set at '1' (100% methylation with each bar representing 5% methylation intervals) and ylim at 50000 (representing the maximum frequency of CpGs within each bar). Red: the histogram created using 'column1' is further categorised by 'column X' (I.e, frequency of CpGs with altered methylation, based on location). Purple: customises the plot visuals including the x-axis and y-axis labels, their size and their colour. All figures in bold text can be changed without affecting the script.

```
library(ggplot2)
A<-read.csv("filename.csv", as.is=T)
g<-ggplot(A, aes(column1, column2, color=factor(columnX)))+geom_point()+
 xlim(c(0,1.0))+ylim(c(0,1.0))+
  coord_equal()+
scale_color_manual(values=c("colour1", "colour2"))
g<-g+labs(x="x-axis title", y="y-axis title")
g
g + theme(axis.text.x=element_text(size=8, colour= 'black'),
axis.text.y=element_text(size=8, colour= 'black'),
axis.title.x = element_text(color="black"),
    axis.title.y = element_text(color="black"),
legend.key=element_rect(fill='NA'),
legend.title=element_blank())
```

Figure 2.7.1.2. Colour-coded script used for the generation of scatter plots using R 3.2.0. Blue: loads the required package into R console. Green: reads in the file to be used for analysis and gives it a simple label of 'A'. Orange: creates the plot based on the values in 'column 1' and 'column 2'. Defines them to be coloured based on values given in 'column X'. Sets a limit on the plot size (in this example xlim and ylim are set for beta values of 0 (0% methylation) and 1 (100% methylation). Red: sets the colours of 'column 1' and 'column 2'. Purple: customises the plot visuals including the x-axis and y-axis labels, their size and their colour. All figures in bold text can be changed without affecting the script.

```
library(ggplot2)
A<-read.csv("filename.csv", as.is=T)
g<-ggplot(A, aes(gene1expression, gene1methylation))+
geom_point(size=5, aes(color=factor(celline)))+
geom_smooth(method=lm, se=TRUE)+
scale_color_manual(values=c("red", "blue", "orange"))
g<-g+labs(x="x-axis title", y="y-axis title")
g
g + theme(axis.text.x=element_text(size=14, colour= 'black'),
axis.text.y=element_text(size=14, colour= 'black'),
axis.title.x = element_text(size=18, color="black"),
    axis.title.y = element_text(size=18, color="black"),
legend.key=element_rect(fill='NA'),
legend.title=element_blank())
```

Figure 2.7.1.3. Colour-coded script used for the generation of correlation plots using R 3.2.0. Blue: loads the required package into R console. Green: reads in the file to be used for analysis and gives it a simple label of 'A'. Orange: creates the plot based on the values in 'gene1expression' and 'gene1methylation' and colours each point based on 'celline'. Red: creates a regression line and shades regions indicating 95% confidence intervals, colours each 'celline' a different colour. Purple: customises the plot visuals including the x-axis and y-axis labels, their size and their colour. All figures in bold text can be changed without affecting the script.

```
library(RColorBrewer)
A<- read.table("filename.csv", header=T, sep="\t")
myColours <- brewer.pal(x,"colour")
par(las=2)
par(mar=c(5,8,4,2))
bplt<-barplot(A$OB, horiz=TRUE, col=myColours, xlim=c(0,1200), names.arg=c("Island", "N shelf", "N
shore", "Open sea", "S shelf", "S shore"), cex.names=2, cex.axis=2)
text(x= A$OB+80, y= bplt, labels=as.character(A$OB), xpd=TRUE, cex=2)
```

Figure 2.7.1.4. Colour-coded script used for the generation of bar graphs used in the location of CpG methylation using R 3.2.0. Blue: loads the required package into R console. Green: loads the file into a table in R. Orange: 'x' sets the number of colours and 'colour', the colours to be used in the bar graph. Red: sets position of text and margins. Purple: customises the plot visuals including size of text, name and location of labels and positioning. All figures in bold text can be changed without affecting the script.

```
library(VennDiagram)
venn.diagram(x = list(IPF = 1:x, SSc = y:z),
filename = "name.tiff", height = 3000, width = 3000, resolution = 500,
units = "px", compression = "lzw", lwd = 4, fill = c("colour1", "colour2"),
    alpha = 0.75, label.col = "black", cex = 2, fontfamily = "serif", fontface = "bold",
rotation.degree = 180,
    cat.col = c("black"), cat.cex = 2, cat.fontfamily = "serif", cat.fontface = "bold",
    cat.dist = c(0.05, 0.05), cat.pos = c(90, 270), margin = 0.1)
```

Figure 2.7.1.5. Colour-coded script used for the generation of weighted Venn diagrams using R 3.2.0. Blue: loads the required package into R console. Green: creates the Venn diagram based on values entered at x, y and z. Orange: creates the image as a .tiff file and sets the width, height and resolution of image. Red: colours the Venn diagram in based on 'colour1' and 'colour2'. Purple: customises the plot visuals including size of text, font, location of labels and transparency. All figures in bold text can be changed without affecting the script.

```
library(RColorBrewer)
A <- read.table("filename.csv", head=T, row.names=1, sep=",")
as.data.frame(A)
B <- as.matrix(scale(A))
C <-colorRampPalette(c("colour1","colour2", "colour3", "colour4"))(256)
heatmap(B, Colv=F, scale='none', col=C)
```

Figure 2.7.1.6. Colour-coded script used for the generation of heatmap plots using R 3.2.0. Blue: loaded the required package into R console. Green: reads the file into R and converts it into a table. Orange: converts the table into a data frame and then into a matrix. Red: creates the heatmap and manually colours the heatmap base on low (colour1) to high (colour 4) values.

2.7.2. Enrichment and network analysis

The online tool, BioMart (Smedley et al, 2015) in Ensembl (Cunningham et al, 2015) was used to extract data from the Ensembl 81 database in order to map all genes with their respective protein family domains (PFAMs). Using the gene list containing genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF compared to control lung fibroblasts as an example, each gene within the list was compared with the full list of genes in Microsoft Excel 2013 using the Excel formula "=NOT(ISNA(VLOOKUP(X1,$Y:$Y,1,FALSE)))", where 'X' represented the gene of interest within a given list and '$Y:$Y' represented the full list of genes and their corresponding PFAMs. This formula returned a value of 'TRUE' if the gene was identified or 'FALSE' if the gene was not in the gene list. All values returning FALSE were filtered leaving the genes with significantly altered methylation in IPF compared to control lung fibroblasts and the PFAMs they were associated with. The genes were then given different values based on whether they had CpGs with increased, decreased or both increased and decreased methylation compared to control lung fibroblasts for later use in network visualisation. This was done using the Excel formula "=IFERROR(INDEX($X:$X,MATCH(Y,$Z:$Z,0)),"")", where '$X:$X' represents the gene list, 'Y' represents the specific gene, and $Z:$Z' represents the specific value. The domain centric gene ontology (dcGO) database (Fang and Gough, 2013) was used to find biological processes enriched in PFAMs containing genes with significantly altered methylation/expression. A false discovery rate (FDR) threshold p value <0.01 was used as a cut-off. All levels of granularity (highly general, general, specific and highly specific) were included.

The network integration and visualisation analysis tool, Cytoscape v3.2.1 (Shannon et al, 2003) was used to visualise networks of genes which shared specific PFAMs and groups of PFAMs enriched in specific biological processes. Genes and their corresponding PFAMs were imported into Cytoscape as nodes. The list of values indicating the direction of methylation change was imported into cytoscape and assigned to the nodes. VizMapper in Cytoscape was used to colour-code each node based on the direction of methylation. Genes which also had a change in expression were annotated using the import custom graphics option within VizMapper. Genecodis 3 web enrichment tool (Carmona-Saez et al, 2007, Nogales-Cadenas et al, 2009, Tabas-Madrid et al, 2012) was used to determine GO biological processes enriched in genes with significantly altered methylation/expression. The

hypergeometric statistical test was used with the Benjamini and Hochberg FDR p value correction test. Enriched biological processes with a corrected hypergeometric P value <0.05 were considered statistically significant.

2.7.3. Protein-protein interactions and KEGG pathway analysis

The STRING v10 database (www.string-db.org) was used to analyse protein-protein interaction networks and how genes with altered methylation and/or expression potentially interacted with each other (Szklarczyk et al, 2015). Specific gene lists of interest were entered as 'user payload' datasets. Each gene was assigned a hexadecimal colour which represented the direction of change to methylation and/or expression. The analysis was run using the default parameters which were set to confidence score ≥0.400 with all active prediction methods selected.

For visualisation of pathways, KEGG pathways were downloaded as .kgml files from http://www.genome.jp/kegg/pathway.html and read into cytoscape using the Cytoscape plugin, KEGGscape (Nishida et al, 2014). Attribute tables were downloaded into .csv files and read using Excel. These files were then modified to include all genes associated in the pathway with altered methylation and/or expression by using the rest.kegg.jp/list/ function to convert all HSA KEGG IDs to gene symbols. The files were then uploaded as tables and merged to the appropriate cytoscape KEGG pathway. The pathway was then visually enhanced to show genes with altered methylation and/or expression using the VizMapper tool within Cytoscape.

2.8. RT-PCR and real-time qRT-PCR validation of microarray data

2.8.1. cDNA synthesis

To validate the expression microarray, cDNA was prepared by reverse transcription of RNA. All components were placed on ice. RNA (1µg) was made up to a volume of 16µl with NF-H_2O (Ambion, USA) in a 0.2ml PCR tube (Starlabs, UK). A master-mix containing 4µl of q-Script cDNA SuperMix (5X reaction buffer containing optimised concentrations of, MgCl2, dNTPs, primers, RNase inhibitor protein, qScript reverse transcriptase and stabilisers) (Quanta Biosciences, USA) was then added. RT-PCR was performed using a tetrad PTC-225, Peltier Thermal cycler, (Global Medical Instrumentation, USA) with cycle conditions set to: 25^oC for 5 minutes, 42^oC for 30 minutes, 85^oC for 5 minutes and held at 4^oC. After completion of cDNA synthesis the product was diluted 1:4 with NF-H_2O (Ambion, USA) and divided into aliquots which were immediately frozen at -20^oC for future analysis by qPCR.

2.8.2. Primer design

All primers were designed using Primer Blast software. The primer design parameters were set as follows: product size 80-180bp (to ensure efficient amplification), maximum temperature melting difference 1^oC, primer pair separated by at least 1 intron (when possible, to allow differentiation between amplification of cDNA and potential genomic DNA by melting curve analysis), primer size 18-

25 nucleotides long (to provide practical annealing temperatures), primer GC % >50%, melting temperature 58°C-62°C, maximum poly-X 3 (when possible, to avoid inappropriate hybridisation), maximum self and pair complementarity at 4.0 and maximum 3' self and pair complementarity at 2.0 (when possible) and when possible, primers at the 3' end ended with GC (to enhance annealing of the end that is extended). All primers were produced by Invitrogen Life Sciences, UK. Tris-EDTA (100µl) was added to solubilise each primer, followed by vortexing and centrifugation. Primers were then quantified using a NanoDrop Bioanalyzer ND1000 (NanoDrop, UK) and made to 100µM with Tris-EDTA. For optimisation, primers were made to 800nM and 400nM in NF-H_2O (Ambion, USA). A 1:2 cDNA dilution series was performed to check primer efficiency. Primer efficiency was determined by plotting the known concentrations of the dilution series with their corresponding C_t value. The C_t value represents the cycle number at which the fluorescent signal of the reaction crosses the threshold. From the standard curve, the correlation coeffiecient and slope of the log-linear phase of the amplification were calculated. Only primers with an $R^2 \geq 0.9$ and a slope between -3.58 and -3.10 (close to 90%-100% efficiency) were used. 100% efficiency meant the template doubled after each thermal cycle during exponential amplification. A list of primers used in qRT-PCR is shown in **Table 2.8.2.1**.

Gene	Forward primer	Reverse primer
IL8	5'-GGACCACACTGCGCCAACACA-3'	5'TCTCCACAACCCTCTGCACCCA-3'
CADM1	5'-AGTACAGTATAAGCCTCAAGTGC-3'	5'-CCCAAGTTACCATCACAGGC-3'
TNXB	5'-GAGGGAGACTTCCCTGTCCTGCC-3'	5'ACCAGGAGAACCAGGCTGGAGG-3'
PPARγ	5'-AGGCGAGGGCGATCTTGACAG-3'	5'-GATGCGGATCGCCACCTCTTT-3'
EIF1AY	5'-AGCTCTGGGTTTGTGAATAGC-3'	5'-ACTTGTGGCACTGCAATTTGA-3'
MMP10	5'-GCTCGCCCAGTTCCGCCTTT-3'	5'-GCAGGATCACACTTGGCTGGCA-3'
MMP12	5'-CACTTCTTGGGTCTGAAAGTGA-3'	5'-GAGGTGCGTGCATCATCTC-3'
TGFβ	5'-AGAGCAGAAGGAGGACCAGT-3'	5'- CGGAAGTCAGAGAGTGAGGC-3'
DCN	5'-GGAATAATAAGACACGCCCTGA-3'	5'-AGATGCTGCTTTCTCCCTCT-3'
COL6A3	5'-TTGCTCTGCCCTCAGCG-3'	5'-AGCGGTTCACTTGCTATTTCTTT-3'
COL12A1	5'-GCTACCTCTCCCTGTTGCCG-3'	5'-CACTCCATCCCTTCTGCCTCAA-3'
HPRT	5'-TGACACTGGCAAAACAATGCA-3'	5'-GGTCCTTTTCACCAGCAAGCT-3'
YWHAZ	Primerdesign Ltd (reference gene assay)	Primerdesign Ltd (reference gene assay)
EIF4A2	Primerdesign Ltd (reference gene assay)	Primerdesign Ltd (reference gene assay)

Table 2.8.2.1. Primers used for qRT-PCR. Interleukin 8 (IL8), Cell adhesion molecule 1 (CADM1), Tenascin XB (TNXB), Peroxisome proliferator-activated receptor gamma (PPARγ), Eukaryotic translation initiation factor 1A, Y-linked (EIF1AY), Matrix metallopeptidase 10 (MMP10), Matrix metallopeptidase 12 (MMP12), Transforming growth factor beta (TGFß), Decorin (DCN), Collagen, type VI, alpha 3 (COL6A3), Collagen, type XII, alpha 1 (COL12A1), Hypoxanthine-guanine phosophoribosyltransferase (HRPT), Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (YWHAZ), Eukaryotic translation initiation factor 4A, isoform 2 (EIF1AY).

2.8.3. qRT-PCR

All qRT-PCR was performed using an Eppendorf Realplex 4 Mastercycler (Eppendorf, Germany). Master-mixes for 10µl qPCR reactions consisting of 5µl SYBR green mix (Eurogentec, Belgium), 2µl NF-H_2O (Ambion, USA) and 1µl of forward/reverse primers at a final concentration of 800nM were made on ice and 8µl was added to each well of a 96-well plate (Thermo Scientific, USA). cDNA (2µl at a concentration of 12.5ng/µl) was then added to each well, the plate was sealed, vortexed for 30 seconds and centrifuged at 1100g for 2 minutes. The qPCR cycling conditions were as follows: 95^{o}C for 5 minutes (activation of SYBR green), followed by 40 cycles of denaturing at 95^{o}C for 15 seconds, annealing at 62^{o}C for 45 seconds and extension at 72^{o}C for 15 seconds. Melting curve analysis was used to confirm the specificity of the PCR product.

2.8.4. Statistical analysis of qRT-PCR

A geNorm 12 housekeeping reference gene kit (Primerdesign, UK), was used to normalise qPCR data. The geNorm algorithm was applied to determine the best housekeeping reference genes to normalise qRT-PCR data to by ranking each reference gene in order of stability of expression using the geNorm software provided. The geometric mean of three housekeeping genes; Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein, Zeta (YWHAZ), Eukaryotic Translation Initiation Factor 4A2 (EIF4A2) and Hypoxanthine Phosphoribosyltransferase (HPRT) were chosen as they were the most stable across different samples. The fold-changes in mRNA expression were calculated using the standard $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). ΔC_t values were calculated using the formula C_t (gene)-C_t (geometric mean of 3 housekeeping genes). $\Delta\Delta C_t$ values were calculated using the formula ΔC_t (gene) - ΔC_t (average basal gene). Fold-changes were calculated using the standard $2^{-\Delta\Delta C_t}$ methods. Statistical analyses were performed on the ΔC_t values using GraphPad PRISM 6 software (GraphPad Software, USA). A Student's t-test was used to detect significant differences between two different groups. One-way analysis of variance with a Tukey's post hoc or two-way analysis with a Bonferroni post hoc test was used to compare multiple groups. A p value of <0.05 was considered to be statistically significant.

2.9. Histological analysis

2.9.1. Slide preparation

Formalin-fixed, paraffin-embedded sections of lung tissue derived from in-house archival banks obtained with patient consent and ethical approval for use in research, were placed on a cold plate (Tissue Tek III) (Sakura, USA) before 3µm lung sections were cut and adhered to Poly-L-lysine slides (VWR International, USA). All slides were left overnight to dry at room temperature. Slides were then secured in a cassette allowing lung sections to be dewaxed in xylene, rehydrated in decreasing concentrations of ethanol (100%, 70% and 30%) and washed in PBS.

2.9.2. Antigen retrieval

To determine the best method for detecting TNXB, three different antigen unmasking techniques were tested: no antigen retrieval using 1x Tris Buffered Saline (1xTBS), microwaving lung sections in 10mM citrate buffer, pH6.0 (2 x 10 minutes), and incubation with 0.05% saponin (Sigma, USA) at room temperature for 30 minutes. Slides were washed in 1xTBS for 2 x 5 minutes and then incubated at room temperature for 30 minutes with 3% hydrogen peroxide to block endogenous peroxidase. Slides were then washed in 1xTBS for 2 x 5 minutes and incubated for 20 minutes at room temperature with 2-3 drops of horse serum block (ImmPRESS reagent kit) (Vector Laboratories, UK). Different concentrations of primary antibodies were made in 1%BSA/TBS to determine optimum conditions. Slides were incubated at 4^{o}C overnight in a humidified chamber. Slides were washed in 1xTBS (2 x 5 minutes). ImmPRESS reagent anti-rabbit IgG peroxidase (Vector Laboratories) was added to slides and incubated at room temperature for 30 minutes. Slides were washed in 1xTBS (2 x 5 minutes) and incubated at room temperature for 3 minutes with ImmPACT NovaRED peroxidase (Vector Laboratories). Slides were then rinsed with H_2O, counterstained with haematoxylin and sealed with coverslips. Lung tissue sections were then visualised and analysed using a Nanozoomer (Hamamatsu, Japan) and NanoZoomer Digital Pathology Virtual Slide Viewer software (Hamamatsu, Japan). Saponin and no antigen retrieval methods showed similar staining and 5µg/ml of primary antibody was determined suitable to use (**Figure 2.9.2.1**). After optimisation, no antigen retrieval and staining with 5ug/ml antibody were used for all IHC experiments.

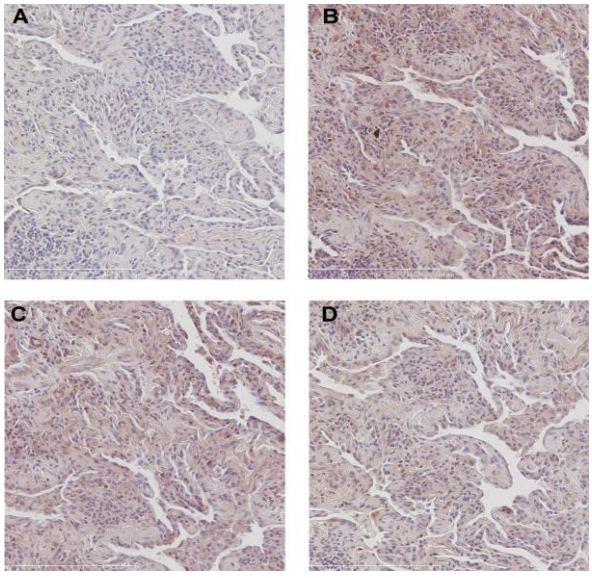


Figure 2.9.2.1. Immunohistochemical staining of TNXB in human lung using no antigen retrieval. Histological sections of human IPF lung specimens stained with: A) 20µg/ml IgG isotype control, B) 20µg/ml TNXB, C) 10µg/ml TNXB, D) 5µg/ml TNXB. At the highest concentration (20µg/ml), TNXB staining was very strong. Very weak staining was observed using an IgG isotype control at a 20µg/ml concentration. To avoid high background staining a 5µg/ml concentration of anti-TNXB was chosen.

2.10. siRNA transfection

IPF and control primary human lung fibroblasts were transfected using INTERFERin (Polyplus Transfection, USA) following the manufacturer's protocol. Ambion siRNA targeting TNXB was reconstituted to 20µM stocks in NF-H20 (Ambion, USA). A range of concentrations (1µM, 5µM 10µM and 20µM) were tested to optimise TNXB knockdown, with 10µM being the most efficient. Cells at passage ≤ 8 were seeded in a 24-well plate at 25,000 cells per well at 30-50% confluence and left overnight. The following day 6 pmoles of TNXB siRNA (Ambion, USA) or negative control siRNA (Ambion, USA) was diluted in 100µl of Opti-MEM (Life Technologies, USA). INTERFERin (3µl) (Polyplus Transfection, USA) was added to the siRNA duplexes and immediately vortexed for 10 seconds. To allow transfection complexes to form between siRNA duplexes and INTERFERin, the mix was incubated for 10 minutes at room temperature. Fresh pre-warmed DMEM (0.5ml) was added to each well with 100µl of transfection mix resulting in siRNA duplexes at 10nM per a well. The 24-well plates were incubated at 37^{o}C for 24 hours and 48 hours. Cells were lysed using TRIzol reagent (Invitrogen, UK). RNA was isolated as previously described and analysed by qRT-PCR.

2.10.1. Collagen gel contraction assays

Collagen gel contraction assays are frequently used to study cell-mediated reorganisation and contraction of ECM (Vernon and Gooden, 2002) and are regarded as an in vitro model of wound contraction. Type I collagen is used as it easily polymerises to form a fibrillary network (Vernon and Gooden, 2002). When fibroblasts are transferred into collagen gels, they are able to reorganise the collagen fibres and subsequently contract the collagen gel (Bell et al, 1979). The collagen gels used to examine lung fibroblast-mediated collagen gel contraction were based on a previous study using fibroblasts (Tingstrom et al, 1992). Plates (24-wells) were coated with 1ml of 2% BSA in PBS overnight at 37^{o}C and then washed PBS (3 x 1 minute) to stop cells attaching. Neutral collagen solution was prepared by mixing DMEM, 0.2 HEPES buffer (pH 8.0) (Sigma, UK), and 3mg/ml rat tail type I collagen, A1048301 (Invitrogen, UK) on ice in a 5:1:4 ratio (by volume), resulting in a collagen gel solution with the final concentration of 1.2mg/ml collagen. Primary human lung fibroblasts derived from IPF and control lung were used at passages ≤8. Cells were cultured on 6-well plates, harvested at confluency and resuspended in DMEM (Life Technologies, UK) at 0.5x10^{6} cells/ml. Primary human lung fibroblasts (100,000 cells) in 200µl DMEM were slowly added to the collagen gel solution. Collagen solution containing lung fibroblasts (1ml) was slowly added into each pre-coated well. Collagen was left for 30mins to polymerise and 1ml of DMEM was added to each well. Using a sterile needle, collagen gels were detached from the side of the well to ensure gels were 'free floating' in media.

For collagen gel assays examining the effects of TNXB knockdown, cells were treated with 10nM TNXB siRNA (Ambion, USA) or 10nM non-targeting negative control siRNA (Ambion, USA) and incubated at 37^{o}C for 48 hours, after which they were trypsinised and resuspended in fresh DMEM before being added to the collagen gels. Photos of collagen gels were taken at every 24 hour time-point for 72

hours. Image J v1.46r (Schneider et al, 2012) was used to measure gel contraction by subtracting the circumference of the collagen gel at a set time-point from the original gel circumference. Statistical analyses were performed using GraphPad PRISM 6 software (GraphPad Software, USA). A Student's t-test was used to detect significant differences between two different groups. One-way analysis of variance with a Tukey's post hoc or 2-way analysis with a Bonferroni post hoc test was used to compare multiple groups. A p value of <0.05 was considered to be statistically significant.

2.11. Bisulfite sequencing of the TNXB gene

2.11.1. Bisulfite primers for TNXB sequencing

DNA was bisulfite converted in-house using an EZ DNA Methylation-Gold™ Kit (Zymo Research, USA) as previously described in section **2.5.1**. TNXB bisulfite primers were designed using Methyl Primer Express v1.0 (Applied Biosystems, USA), Primer Blast (http://blast.ncbi.nlm.nih.gov/) and Primer3 software (Rozen and Skaletsky, 2000). Melting temperatures were increased by increasing the length of primers. For unbiased amplification, no more than 1 CpG dinucleotide was included in a primer. The TNXB bisulfite primers used for the CpG island in exon 3 were; forward primer, 5'-TTTGAGAAGTTTGTTYGGTATATATA-3', reverse primer, 5'-CTAAAACTTACCTCTCCCCTC-3' and exon 10; forward primer, 5'-GGGAAGGTTGGGAGTTAGTAG-3', reverse primer, 5'-ACAACAACAACAAAACCAAC-3'.

2.11.2. RT-PCR and PCR purification

RT-PCR was performed using a tetrad PTC-225, Peltier Thermal cycler (Global Medical Instrumentation, USA). For the TNXB bisulfite primers, a master-mix was prepared on ice, containing all reagents except the bisulfite-converted DNA. The final volume of each PCR reaction was 45µl, containing; 4.5µl of 10xPCR buffer (Applied Biosystems, USA), 3.6µl of magnesium chloride (25mM) (Applied Biosystems, USA), 2.25µl of both forward and reverse primers, 29.55µl of NF-H_2O (Ambion, USA), 0.9µl of dNTP (10mM) (Applied Biosystems, USA) and 0.45µl of Taq (Applied Biosystems, USA). 1.5µl of DNA was then added, vortexed and 20µl of mineral oil (Sigma, USA) was added to prevent any evaporation. PCR cycling conditions were used based on the primer melting temperatures and the PCR product length and were as follows: 94°C for 5 minutes, followed by 10 cycles of 94°C for 20seconds, touchdown from 60°C to 50°C (-1 degree/cycle) for 20 seconds and 72°C for 30 seconds, followed by 35 cycles of 94°C for 30 seconds, 62°C for 30 seconds, 72°C for 7.5 minutes and 25°C for 30 seconds (extention times were longer than normal as uracil decreased the rate of DNA polymerisation) (Darst et al, 2011). Afterwards samples were held at 4°C until storage.

TNXB PCR products were resolved on a 1% agarose gel made with 1xTBE buffer (1g of agarose/100ml 1xTBE), containing 3µl of a 10000X stock solution of GelRed (Biotium, UK). After the gel had set, the gel was covered with 1xTBE buffer. A 5x DNA loading buffer (Bioline, UK) was added to each DNA sample and a molecular weight maker, HyperLadder (Bioline, UK), was then added to the first lane

followed by each DNA sample. The 1% agarose gel was run at 80v for 1.5 hours. DNA bands were observed using a Syngene GeneGenuis imagining system (Syngene, UK). A QIAquick PCR purification kit (Qiagen, Germay) was used to purify the TNXB PCR product. Buffer was added to the sample and centrifuged at ≥ 10000g for 30 seconds to bind DNA. Flow through was discarded and buffer was added to wash DNA. DNA was then eluted by adding 50µl of 10mM Tris-Cl, pH 8.5 and centrifuging at ≥ 10000g for 1 minute.

2.11.3. Bisulfite sequencing analysis

Bisulfite-converted DNA samples were sent to the Wolfson Institute for Biomedical Research (UCL, UK) for DNA sequencing. Template (10µl per a reaction) was provided in 1.5ml Eppendorf tubes at a concentration of ~4ng/µl. Custom primers were provided at a concentration of 5pmole/µl as requested. Analysis of methylation was performed using Sequence Scanner 2 software (Applied Biosystems, USA). The methylation percentage at each CpG site was determined by measuring the area of the C and T peaks using ImageJ software (Abramoff et al, 2004, Erfurth et al, 2006). Data was normalised to area measurements of C peaks from non-methylated cytosine nucleotides.

2.12. General statistical analysis

All analysis for methylation data is presented as the mean values ± the S.E.M unless otherwise stated. All analysis for expression data is presented as the geometric mean ± 95 confidence intervals (CI) unless otherwise stated. Statistical analyses were performed using GraphPad PRISM 6 software (GraphPad Software, USA). A Student's t-test was used to detect significant differences between two different groups. One-way analysis of variance with a Tukey's post hoc or 2-way analysis with a Bonferroni post hoc test was used to compare multiple groups. A p value of <0.05 was considered to be statistically significant.

Chapter 3. DNA methylation in lung fibroblasts and its role in pulmonary fibrosis

3.1. Introduction

Aberrant DNA methylation of multiple cell types has been implicated in a number of different fibrotic diseases including liver fibrosis (Komatsu et al. 2012), kidney fibrosis (Bechtel et al, 2010), IPF (Sanders et al, 2012, Rabinovich et al, 2012) and SSc (Wang et al, 2006). Genes including THY1 (Sanders et al, 2008), PTGER2 (Huang et al, 2010), P14ARF (Cisneros et al, 2012) and miR-17-92 (Dakhlallah et al, 2013) have all previously been identified as having altered methylation in IPF compared to control lung fibroblasts. Multiple genes including DNMT1, MBD1, MeCP2, FLI1 (Wang et al, 2006), DKK1 and SFRP1 (Dees et al, 2013) have altered methylation in SSc compared to control skin fibroblasts however, there have been no studies that examine methylation of lung fibroblasts in SSc-PF patients. Throughout the following result sections, all figures and text with 'SSc' refer to patients with SSc-PF as stated in **section 2.3.1**.

Approximately 70% of all genes have CpG islands (CGIs) in their promoter regions (Saxonov et al, 2006) and aberrant methylation of CpGs located in CGIs has previously been linked to numerous diseases including IPF and SSc (Sanders et al, 2008, Wang et al, 2006). However, DNA methylation outside CGIs could be equally as important in regulating transcriptional activity (Irizarry et al, 2009). CpG methylation in shore regions (which flank CpG islands and can be located up to 2kb away) and shelf regions (which flank shore regions and can be up to 4kb away from the CpG island), may play an important role in disease but the role of methylation in these locations remains poorly understood (Irizarry et al, 2009). CpGs located beyond shelf regions are denoted as being in 'open sea' regions and their importance also remains poorly understood.

The primary aims of the following experiments were to determine:

- Methylation and expression profiles of primary human lung fibroblasts derived from IPF and SSc-PF patients compared with control lung fibroblasts.
- The number and location of CpGs with altered methylation in IPF and SSc-PF compared to control lung fibroblasts and compared to each other.
- Overlapping and distinct CpGs with altered methylation in IPF and SSc-PF compared to control lung fibroblasts.
- Biological processes enriched in differentially methylated/expressed genes.
- How differentially methylated/expressed genes could potentially interact with each other in pathways relevant to fibrosis.

3.2. Genome-wide distribution of DNA methylation in lung fibroblasts

After filtering out non-specific probes and probes which covered a SNP, 324973 CpGs remained on the array. The average methylation value (ß value) of each CpG in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts identified a bimodal distribution of CpG methylation on autosomes, with the highest frequency of CpGs having low methylation (0-15%) and high methylation (80-90%) (**Figure 3.2.1**). The X-Chromosome also had a bimodal distribution but CpGs were more partially methylated with the highest frequencies of CpGs having 25-35% and 75-80% methylation (**Figure 3.2.1**).

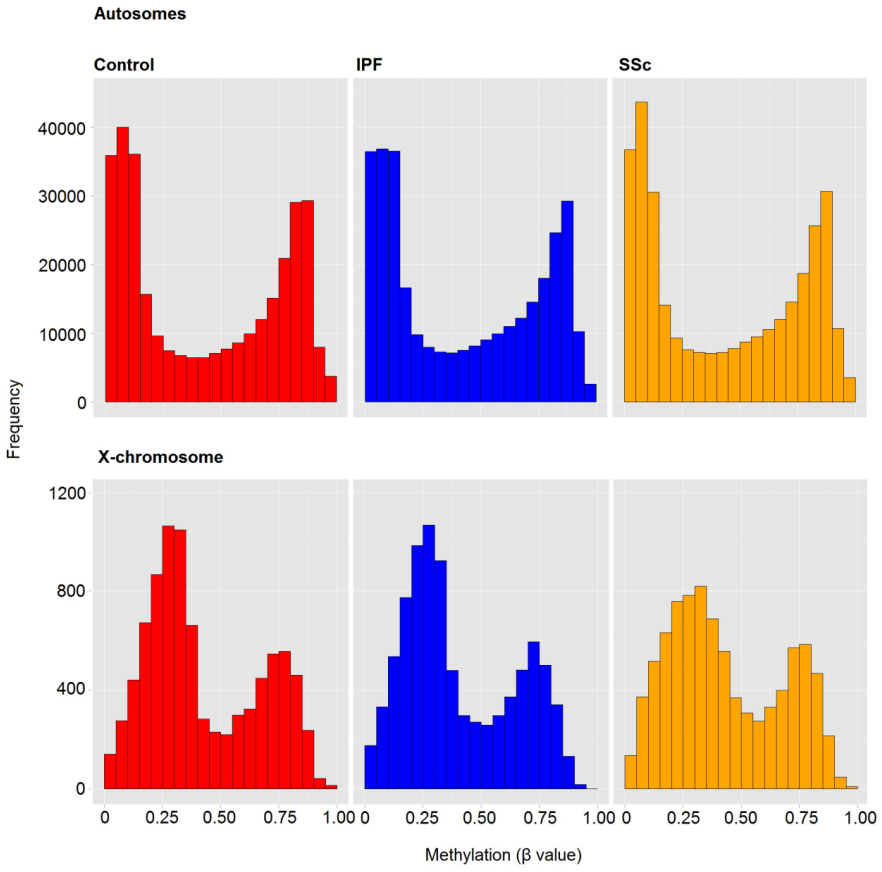


Figure 3.2.1. Distribution of CpG methylation across the genome in autosomes and the X-chromosome. Each bar represents the number of CpGs within each boundary. The β value represents the level of methylation (0= unmethylated, 1= methylated). Red: Control (n=6), blue: IPF (n=5) and yellow: SSc (n=7) lung fibroblasts.

A high frequency of CpGs with low (0-15%) methylation were located in CGIs within 1.5kb of their corresponding genes transcription start site (TSS) (**Figure 3.2.2**). This confirms previous data which showed hypomethylated CpG sites are more commonly found in CpG islands and within 1.5kb of the TSS of a gene (Jones, 2012, Wagner et al, 2014). In contrast, CpGs located in islands further than 1.5kb away from their corresponding genes TSS showed a bimodal pattern of methylation with the highest frequencies of CpGs occurring at 0-15% and 80-85% methylation (**Figure 3.2.2**).

CpGs in north and south shore regions within 1.5kb of their corresponding genes TSS had higher frequencies of CpGs with low methylation compared to CpGs in north and south shore regions further than 1.5kb away from their corresponding genes TSS (**Figure 3.2.3**). In north and south shelves, CpGs within and further than 1.5kb of their genes corresponding TSS had the highest frequency of CpGs with high (85-90%) methylation. CpGs in open sea regions within 1.5kb of their corresponding genes TSS had a bimodal distribution of CpG methylation, with the highest frequencies of CpGs having 10-15% and 85-90% methylation. The majority of CpGs in open sea regions further than 1.5kb away from their corresponding genes TSS had high (85-90%) methylation (**Figure 3.2.3).**

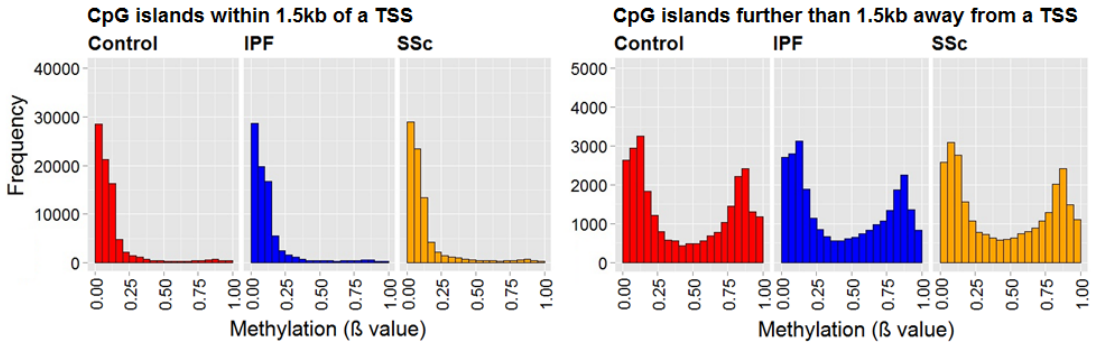


Figure 3.2.2. Distribution of CpG methylation across the genome in CpG islands in relation to the distance from the corresponding gene's transcription start site (TSS). Each bar represents the number of CpGs within each boundary. The β value represents the level of methylation (0= unmethylated, 1= methylated). Red: Control (n=6), blue: IPF (n=5) and yellow: SSc (n=7) lung fibroblasts.

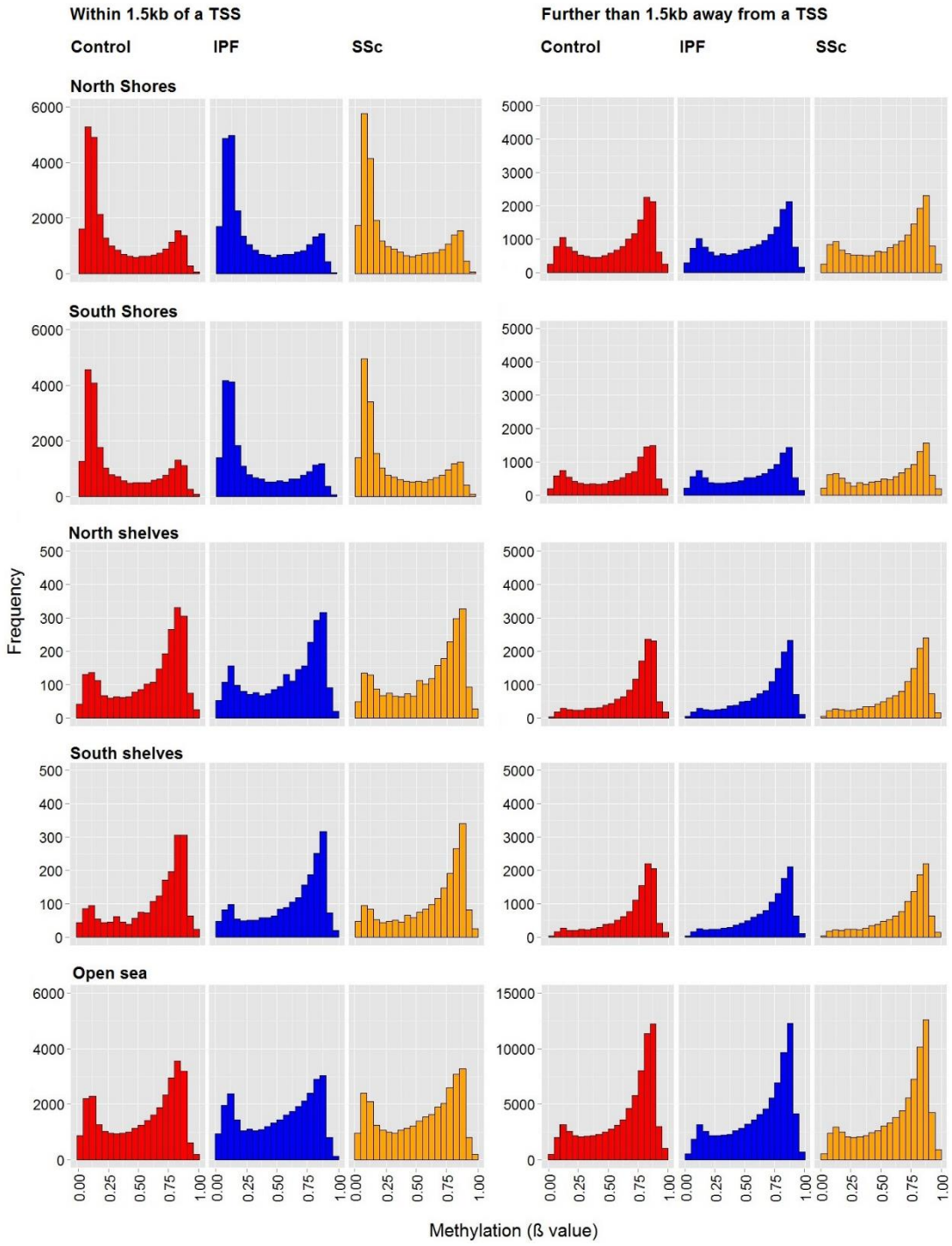


Figure 3.2.3. Distribution of CpG methylation across the genome in shore, shelf and open sea regions in relation to the distance from the corresponding gene's TSS. Each bar represents the number of CpGs within each boundary. The β value represents the level of methylation (0= unmethylated, 1= methylated). Red: Control (n=6), blue: IPF (n=5) and yellow: SSc (n=7).

3.3. Microarray analysis of differentially methylated genes in IPF and SSc compared to control primary human lung fibroblasts

Multiple CpGs were identified as having significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF and SSc compared to control lung fibroblasts. The distribution of these CpGs showed a partially methylated pattern of methylation in control and IPF lung fibroblasts with the highest frequency of CpGs having 45-65% methylation in IPF compared to 60-80% in control lung fibroblasts (**Figure 3.3.1**). The distribution of CpG methylation showed SSc also had a partially methylated pattern of methylation (**Figure 3.3.2**) with the highest frequency of CpGs having 35-65% methylation in SSc compared to 20-55% in control lung fibroblasts.

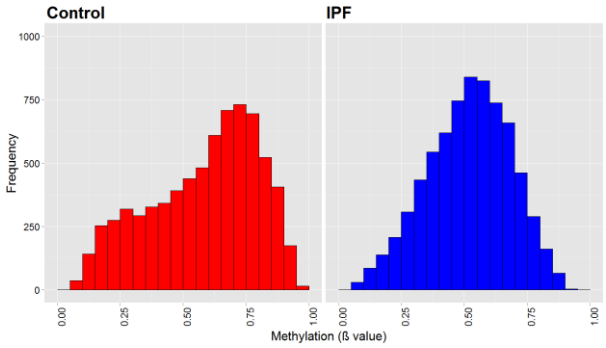


Figure 3.3.1. Distribution of CpGs which had altered methylation in IPF compared to control lung fibroblasts. Each bar represents the number of CpGs within each boundary which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) compared to control (n=6) lung fibroblasts. The β value represents the level of methylation (0= unmethylated, 1= methylated). Red: Control (n=6), blue: IPF (n=5).

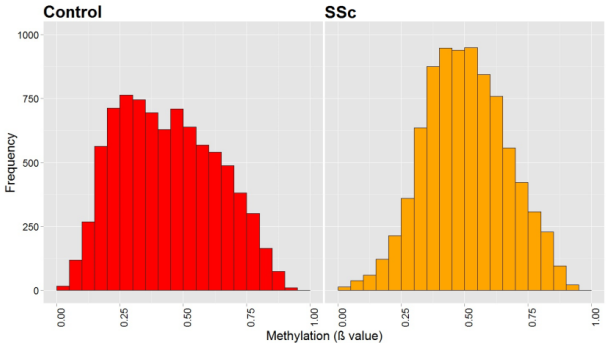


Figure 3.3.2. Distribution of CpGs which had significantly altered methylation in SSc compared to control lung fibroblasts. Each bar represents the number of CpGs within each boundary which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in SSc (n=7) compared to control (n=6) lung fibroblasts. The β value represents the level of methylation (0= unmethylated, 1= methylated). Red: Control (n=6), yellow: SSc (n=7).

7153 CpGs corresponding to 4563 genes in IPF and 8392 CpGs corresponding to 5294 genes in SSc had significantly altered methylation (Δβ ≥ 0.136; P<0.05) compared to control lung fibroblasts. The average methylation of each CpG with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to control and SSc compared to control lung fibroblasts is shown in (**Figure 3.3.3**). 4908 CpGs (69%) of the 7153 significantly differentially methylated (Δβ ≥ 0.136; P<0.05) CpGs in IPF compared to control lung fibroblasts, had significantly decreased methylation (Δβ ≥ 0.136; P<0.05). This represented a predominance of CpGs with decreased methylation in IPF lung fibroblasts (**Table 3.3.1**). Conversely in SSc lung fibroblasts, 5609 (67%) of the 8392 significantly differentially methylated (Δβ ≥ 0.136; P<0.05) CpGs in SSc compared to control lung fibroblasts, had significantly increased methylation (Δβ ≥ 0.136; P<0.05). This represented a predominance of CpGs with increased methylation in SSc lung fibroblasts (**Table 3.3.1**).

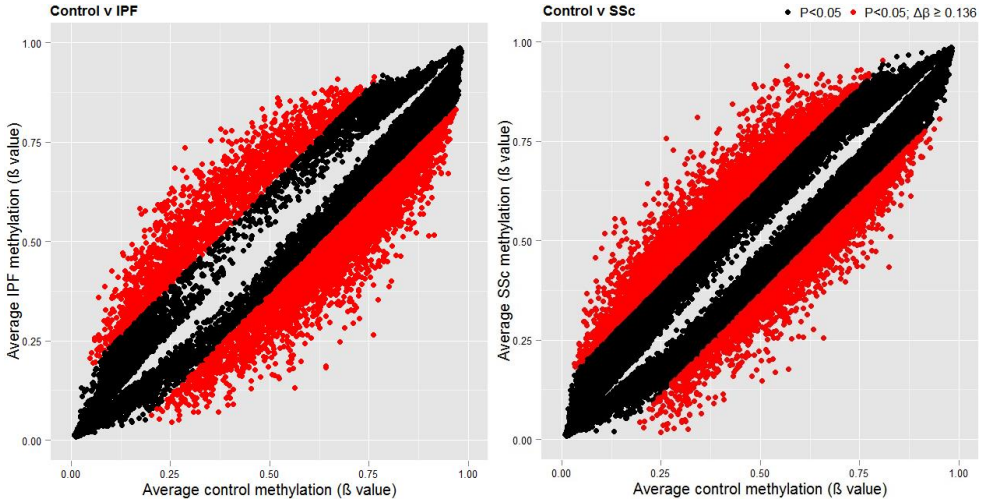


Figure 3.3.3. Scatter plots of CpGs with significantly altered methylation in IPF and SSc compared to control lung fibroblasts. Scatter plots show the average methylation in IPF (n=5) compared to control (n=6) and SSc (n=7) compared to control (n=6) lung fibroblasts. Each dot represents a CpG with either a P<0.05; Δβ< 0.136 (<13.6% change in methylation) (●) or a Δβ ≥ 0.136; P<0.05 (≥ 13.6% change in methylation) (●).

Fibroblast source	CpG sites with increased methylation	CpG sites with decreased methylation	Total number of CpG sites
IPF	2245	4908	7153
SSc	5609	2783	8392

Table 3.3.1. Number of CpG with altered methylation in IPF/SSc compared to control lung fibroblasts. The number of CpG sites which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) and SSc (n=7) compared to control (n=6) lung fibroblasts.

Hierarchical cluster analysis on the 7153 CpGs which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to control lung fibroblasts identified IPF cell lines cluster together (**Figure 3.3.4**). Hierarchical cluster analysis on the 8392 CpGs which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in SSc compared to control lung fibroblasts identified all SSc cell lines clustered together, however, control cell line 6 had a similar methylation profile to SSc cell lines (**Figure 3.3.4**). The top ten CpGs with the greatest increased and decreased methylation in IPF and SSc compared to control lung fibroblasts are shown in **Table 3.3.2** and **Table 3.3.3**.

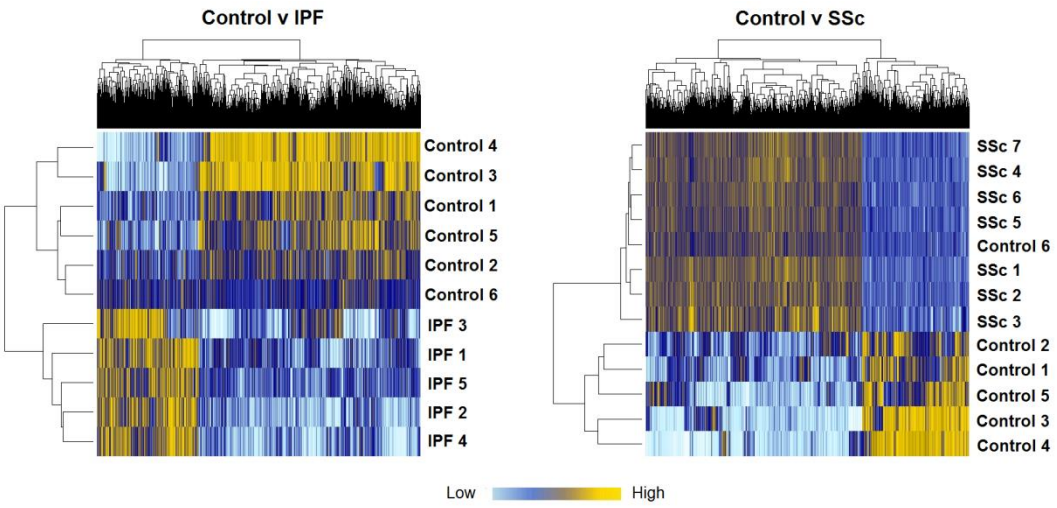


Figure 3.3.4. Hierarchical clustering based on CpGs with altered methylation in IPF and SSc compared to control lung fibroblasts. Heat-maps show CpGs with significantly altered (Δβ ≥ 0.136; P<0.05) methylation (n=7153) in IPF (n=5) and (n=8392) in SSc (n=7) compared to control (n=6) lung fibroblasts. Light blue represents low methylation, yellow represents high methylation with respect to each CpG across all cell lines.

Top 10 CpGs with increased methylation in IPF compared to control lung fibroblasts						
TSS	**Gene name**	**Symbol**	**Chr**	**Location**	**P value**	**Δβ**
2585	Forkhead box P1	FOXP1	3	N_Shelf	0.001472	0.44837
1188	Wingless-type MMTV integration site family, member 5A	WNT5A	3	N_Shore	0.002809	0.42303
2221	Proteasome assembly chaperone 3	PSMG3	7	N_Shelf	0.01675	0.40959
3187	Zinc finger protein, multitype 2	ZFPM2	8	S_Shelf	0.007444	0.40484
17980	Forkhead box P1	FOXP1	3	Open_sea	0.002096	0.4022
46045	Long intergenic non-protein coding RNA 284	LINC0284	13	Open_sea	0.000222	0.39743
3958	Synaptopodin 2	SYNPO2	4	Open_sea	0.042721	0.39036
40713	Collagen, type IV, alpha 2	COL4A2	13	Open_sea	0.003676	0.3892
56446	AX748239	AX748239	8	Island	0.021265	0.38691
70596	Homolog of rat pragma of Rnd2	SGK223	8	Open_sea	0.005544	0.38075
Top 10 CpGs with decreased methylation in IPF compared to control lung fibroblasts						
-291	Bone morphogenetic protein 4	BMP4	14	S_Shore	0.002545	-0.497547
285961	Chromosome 22 open reading frame 34	C22orf34	22	Island	0.005924	-0.459054
13	Chromosome 19 open reading frame 59	C19orf59	19	N_Shelf	0.000128	-0.455819
354	Sidekick cell adhesion molecule 1	SDK1	7	Open_sea	0.002093	-0.430344
-52	Ret finger protein-like 2	RFPL2	22	Open_sea	0.010819	-0.430104
-24273	AL832737	AL832737	6	S_Shore	0.008077	-0.420629
1160	Ankyrin repeat and SOCS box containing 2	ASB2	14	Island	0.026261	-0.419376
-99424	AK126852	AK126852	16	Island	0.007678	-0.414982
-2411	Frizzled class receptor 7	FZD7	2	N_Shore	0.030301	-0.412765
-470	Estrogen receptor 1	ESR1	6	N_Shelf	9.69E-05	-0.40573

Table 3.3.2. Top 10 most differentially methylated CpGs with increased and decreased methylation in IPF compared to control lung fibroblasts. The distance in base pairs (bp) the CpG is from its corresponding genes transcription start site (TSS), the gene which the CpG corresponds to, which chromosome the gene is on, the location of where the CpG is in relation to CpG islands, and the average difference in methylation (Δβ) between control (n=6) and IPF (n=5) is shown.

Top 10 CpGs with increased methylation in SSc compared to control lung fibroblasts						
TSS	**Gene name**	**Symbol**	**Chr**	**Location**	**P value**	**Δβ**
-388	Solute carrier family 37 member 2	SLC37A2	11	N_Shore	0.000514	0.49385
34647	Long intergenic non-protein coding RNA 523	LINC00523	14	Open_sea	0.001112	0.46801
-125	Solute carrier family 25 member 2	SLC25A2	5	Island	0.000472	0.44709
-8091	Dihydrouridine synthase 3-like	DUS3L	19	Island	0.000307	0.42738
-99840	PR domain zinc finger protein 1	PRDM1	6	Island	0.000194	0.42273
9846	BC061632	BC061632	3	Open_sea	0.005037	0.41355
3098	LIM homeobox 4	LHX4	1	Island	0.000138	0.41105
20098	Testis specific protein, Y-linked 1	TSPY1	Y	Island	0.002806	0.40858
8078	Collagen, type IV, alpha 1	COL4A1	13	Open_sea	7.46E-05	0.4011
-2049	Neuroligin 4, Y-linked	NLGN4Y	Y	N_Shelf	0.002027	0.40061
Top 10 CpGs with decreased methylation in SSc compared to control lung fibroblasts						
201986	LOC286083	LOC286083	8	Island	0.000301	-0.403498
-10189	Chromosome 10 open reading frame 11	C10orf11	10	Open_sea	0.002153	-0.394424
64108	Rabphilin 3A-like	RPH3AL	17	N_Shore	0.000114	-0.391314
-1325	Peptidylprolyl isomerase A (cyclophilin A)-like 4A	PPIAL4A	1	Open_sea	0.000155	-0.36739
54267	BC039356	BC039356	1	Open_sea	0.001861	-0.357374
6930	Coiled-coil domain containing 102B	CCDC102B	18	Open_sea	0.006385	-0.356484
30437	Semaphorin 3E	SEMA3E	7	Open_sea	0.000939	-0.351348
190871	AK126491	AK126491	10	Open_sea	0.003145	-0.335878
8377	Cat eye syndrome chromosome region, candidate 1	CECR1	22	Open_sea	0.001681	-0.329455
-202	MX dynamin like GTPase 2	MX2	21	Open_sea	0.000812	-0.324892

Table 3.3.3. Top 10 most differentially methylated CpGs with increased and decreased methylation in SSc compared to control lung fibroblasts. The distance in base pairs (bp) the CpG is from its corresponding genes transcription start site (TSS), the gene which the CpG corresponds to, which chromosome the gene is on, the location of where the CpG is in relation to CpG islands, and the average difference in methylation (Δβ) between control (n=6) and SSc (n=7) is shown.

3.3.1. Overlapping differentially methylated CpGs and genes in IPF and SSc compared to control primary human lung fibroblasts

Multiple genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) had CpGs which overlapped between IPF and SSc lung fibroblasts. Of the 2245 CpG sites which had increased methylation in IPF compared to control lung fibroblasts, 729 (33%) were the same CpG sites which had increased methylation in SSc compared to control lung fibroblasts. These 2245 CpG sites with increased methylation in IPF compared to control lung fibroblasts corresponded to 1767 genes of which 983 (55%) were the same genes with increased methylation in SSc compared to control lung fibroblasts (**Figure 3.3.1.1**). Of the 2783 CpG sites which had decreased methylation in SSc compared to control lung fibroblasts, 735 (26%) were the same CpG sites which had decreased methylation in IPF compared to control lung fibroblasts. These 2783 CpGs with decreased methylation in SSc compared to control lung fibroblasts corresponded to 2109 genes of which 942 (45%) were the same genes with decreased

methylation in IPF compared to control lung fibroblasts (**Figure 3.3.1.1**). These data suggest there are multiple common differentially methylated CpGs and genes in IPF and SSc compared to control lung fibroblasts (**Figure 3.3.1.2**). Furthermore, more genes overlap in IPF and SSc compared to CpGs, suggesting that for some genes, different CpGs of the same genes are affected. However, many differentially methylated CpGs and genes were distinct to either IPF or SSc (**Figure 3.3.1.3**) which may in part explain the phenotypic differences of each disease.

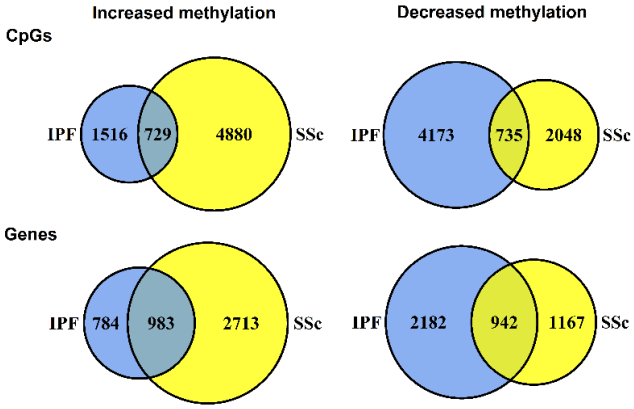


Figure 3.3.1.1. Overlapping and distinct CpGs/genes in IPF and SSc compared to controls. Number of distinct and overlapping CpGs and their corresponding genes which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) and SSc (n=7) compared to control (n=6) lung fibroblasts.

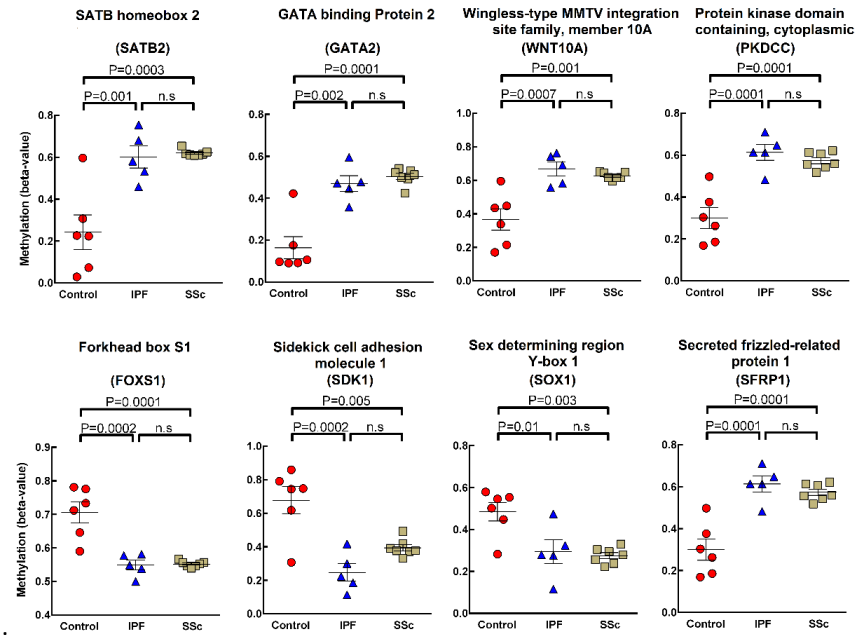


Figure 3.3.1.2. Overlapping CpGs which have altered methylation in both IPF and SSc compared to control lung fibroblasts. Examples of genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF and/or SSc compared to control lung fibroblasts. ● Control (n=6), ▲IPF (n=5), ■SSc (n=7). Each point represents a different cell line. Data presented as the mean ± the SEM. One-way ANOVA was performed with Tukey's post hoc test for statistical analysis.

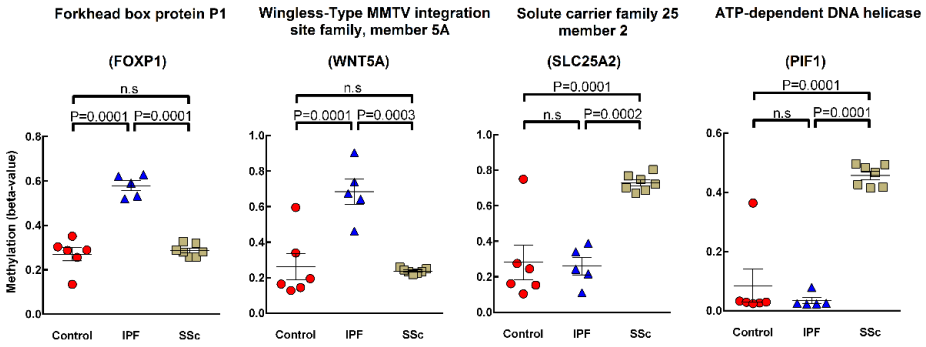


Figure 3.3.1.3. Distinct CpGs which have altered methylation in IPF or SSc compared to control lung fibroblasts. Examples of genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF or SSc compared to control lung fibroblasts. ● Control (n=6), ▲IPF (n=5), ■ SSc (n=7). Each point represents a different cell line. Data presented as the mean ± the SEM. One-way ANOVA was performed with Tukey's post hoc test for statistical analysis. Differentially methylated CpGs and genes in IPF compared to SSc primary human lung fibroblasts.

3.4. Differentially methylated CpGs in IPF compared to SSc lung fibroblasts

Multiple CpGs were identified as having significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts. The distribution of these CpGs showed that the majority had a partially methylated pattern of methylation which was skewed towards lower methylation in IPF compared to SSc lung fibroblasts (**Figure 3.4.1**). Open sea regions distal to a genes TSS had the highest frequency of CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts (**Figure 3.4.2**). There was a higher frequency of CpGs with low methylation (<30%) and a lower frequency of CpGs with high methylation (>80%) in IPF compared to SSc lung fibroblasts in all regions, which was most apparent in CpG islands within 1.5kb of a genes TSS (**Figure 3.4.2**).

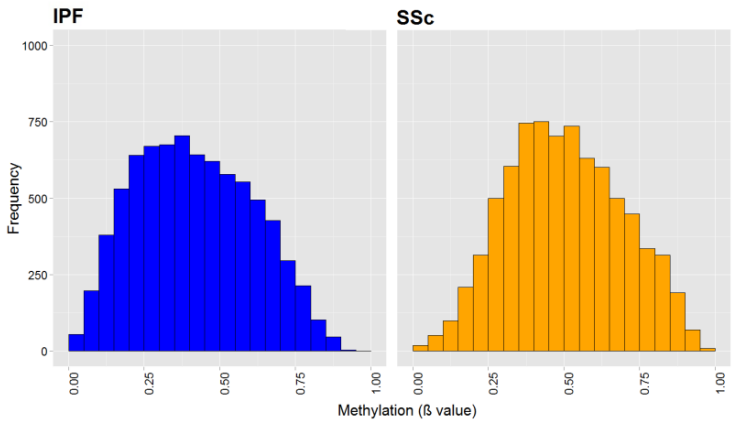


Figure 3.4.1. Distribution of CpGs which have altered methylation in IPF compared to SSc lung fibroblasts. Each bar represents the number of CpGs within each boundary with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. The β value represents the level of methylation (0= unmethylated, 1= methylated).

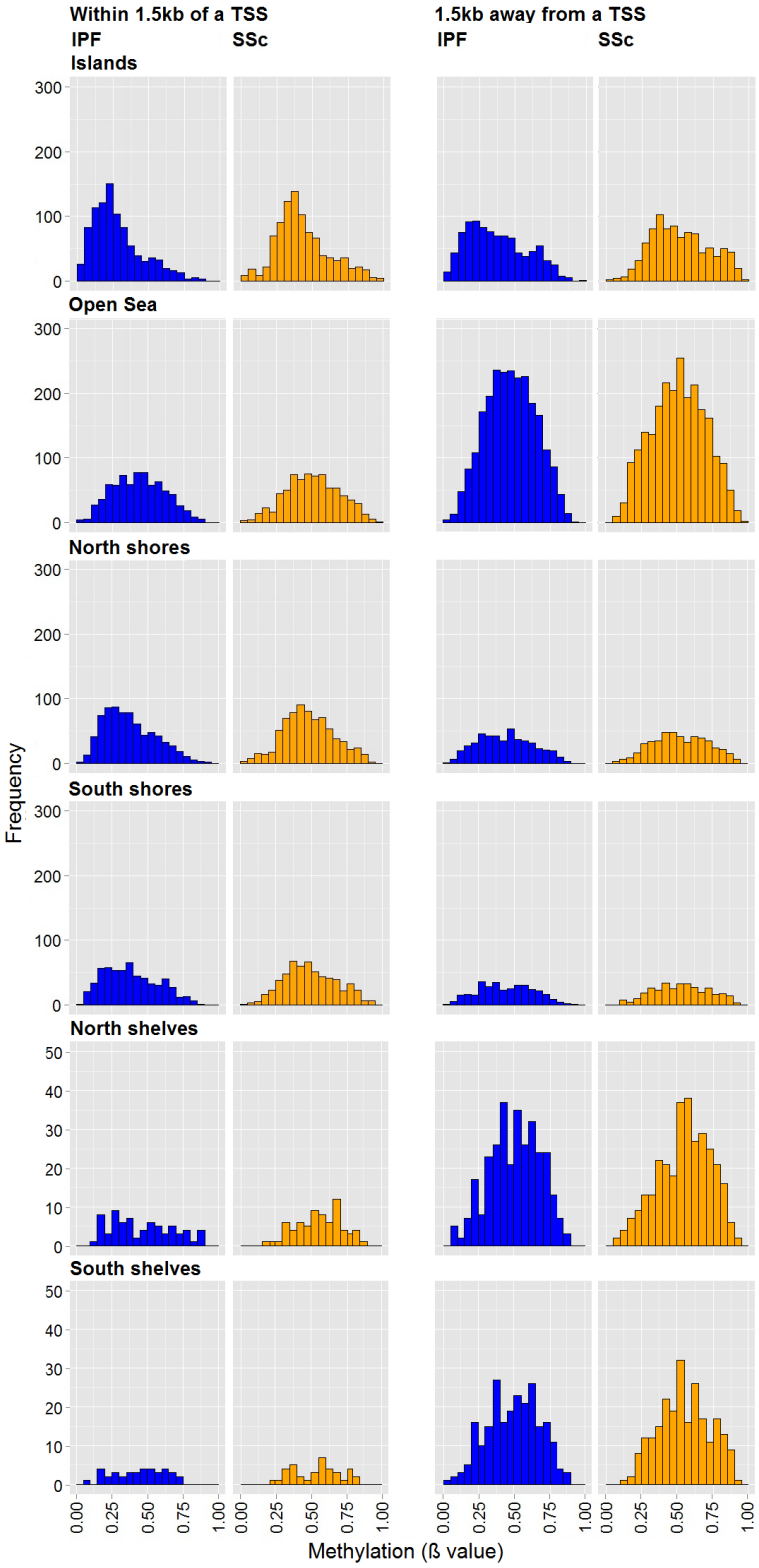


Figure 3.4.2. Distribution of CpGs which had altered methylation in IPF compared to SSc lung fibroblasts in relation to their corresponding gene's TSS. Each bar represents the number of CpGs within each boundary with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. The β value represents the level of methylation (0= unmethylated, 1= methylated).

7827 CpGs corresponding to 5082 genes in IPF had significantly altered methylation (Δβ ≥ 0.136; P<0.05) compared to SSc lung fibroblasts. 2085 CpGs corresponding to 1701 genes had significantly increased methylation (Δβ ≥ 0.136; P<0.05) and 5742 CpGs corresponding to 3835 genes had significantly decreased methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts. Cluster analysis identified all IPF lung fibroblast cell lines clustered separately from all SSc lung fibroblast cell lines (**Figure 3.4.3**). Examples of CpGs and their corresponding genes which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts are shown in **Figure 3.4.4**.

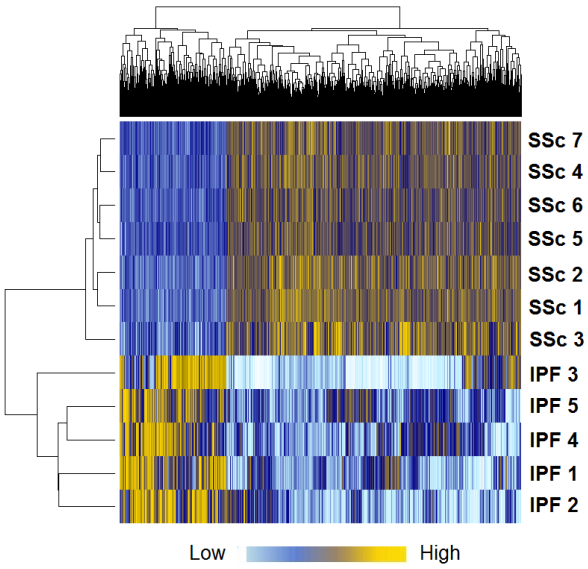


Figure 3.4.3. Hierarchical clustering based on CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts. Heat-map shows the differentially methylated CpGs (n=7827) in IPF (n=5) and compared to SSc (n=7) lung fibroblasts. Light blue represents low methylation, yellow represents high methylation.

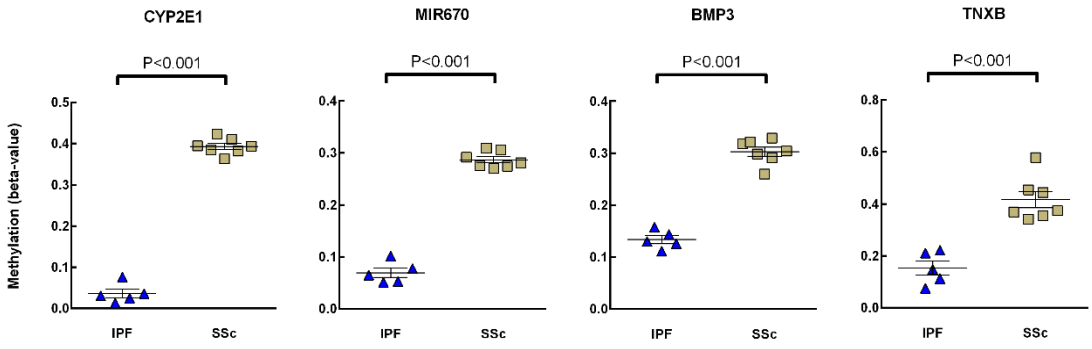


Figure 3.4.4. CpGs which have decreased methylation in IPF compared to SSc lung fibroblasts. Examples of genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts. ▲IPF (n=5), ■ SSc (n=7). Each point represents a different cell line. Data presented as the mean ± the SEM.

3.5. Location of differentially methylated CpGs in IPF and SSc compared to control primary human lung fibroblasts

As previously discussed (**see chapter 1: section 1.7.4**), the location of CpG methylation in relation to CpG islands, shores, shelves and open sea regions may be important in regulating gene expression. The average methylation and location of the 7153 CpGs with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF compared to control lung fibroblasts is shown in **Figure 3.5.1**. In order to determine whether any region was overrepresented, the observed (O) number of CpGs in each region were compared to the number of CpGs expected (E) in each region. North shore (O/E: 1.19), south shore (O/E: 1.18) and open sea (O/E: 1.46) regions had more than the expected number of CpGs with increased methylation in IPF compared to control lung fibroblasts whereas island (O/E: 0.27) regions had fewer CpG sites than expected. North shelf (O/E: 1.38), south shelf (O/E: 1.38) and open sea (O/E 1.52) regions had more than the expected number of CpGs sites with decreased methylation in IPF compared to control lung fibroblasts whereas as island (O/E: 0.41) regions had fewer CpG sites than expected (**Table 3.5.1**).

The average methylation and location of the 8392 CpGs with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in SSc compared to control lung fibroblasts is shown in **Figure 3.5.2**. North shore (O/E: 1.20), south shore (O/E: 1.31) and open sea (O/E: 1.14) regions had more than the expected number of CpGs with increased methylation in SSc compared to control lung fibroblasts whereas island (O/E: 0.74) regions had fewer CpG sites than expected. North shelf (O/E: 1.27), south shelf (O/E: 1.21) and open sea (O/E 1.78) regions had more than the expected number of CpGs sites with decreased methylation in IPF compared to control lung fibroblasts whereas island (O/E: 0.26) regions had fewer CpG sites than expected (**Table 3.5.2**).

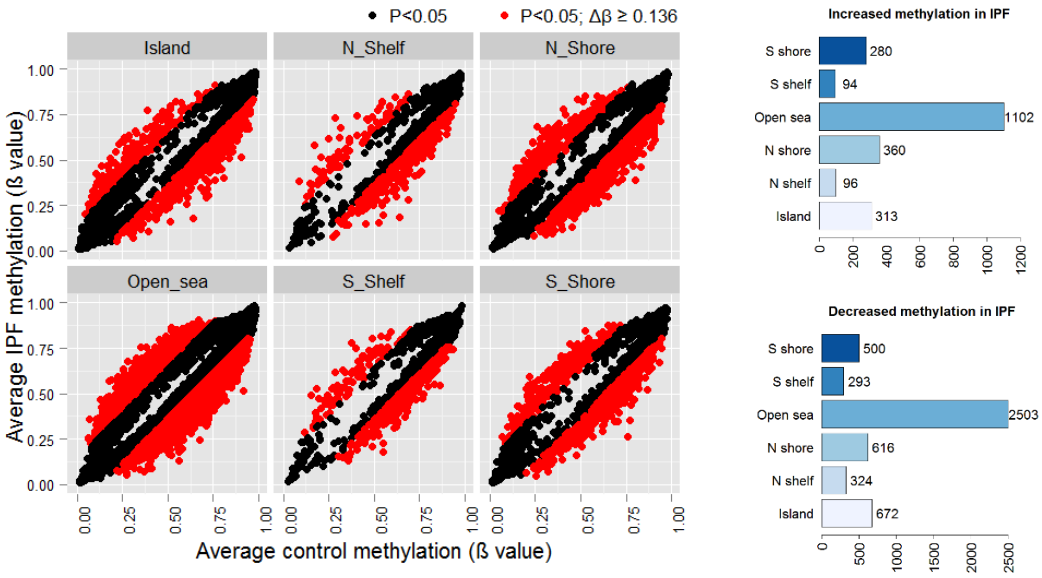


Figure 3.5.1. CpGs with altered methylation in relation to CpG islands in IPF compared to control lung fibroblasts. Left: scatter plots showing the average methylation and the location of each CpG with a significant difference (P<0.05) in methylation between IPF (n=5) and control (n=6) lung fibroblasts. Each dot represents a CpG with either a Δβ< 0.136 (<13.6% change in methylation) (●) or a Δβ ≥ 0.136 (≥ 13.6% change in methylation) between IPF and control lung fibroblasts (●). Right: bar graphs showing the number of CpGs in different locations with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to control lung fibroblasts.

Location	Total no. of CpGs	CpGs with increased methylation in IPF			CpGs with decreased methylation in IPF		
		Observed	Expected	O/E	Observed	Expected	O/E
Island	108197	313	747	0.42*	672	1634	0.41*
N_Shelf	15598	96	108	0.89	324	236	1.38*
N_Shore	43922	360	303	1.19*	616	663	0.93
Open_sea	108961	1102	753	1.46*	2503	1646	1.52*
S_Shelf	14040	94	97	0.97	293	212	1.38*
S_Shore	34255	280	237	1.18*	500	517	0.97
Total	**324973**						

Table 3.5.1. Observed to expected ratio (O/E) of the number of CpGs in each location with altered methylation in IPF compared to control lung fibroblasts. CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) compared to control (n=6) lung fibroblasts. *= Chi-square value; P<0.05.

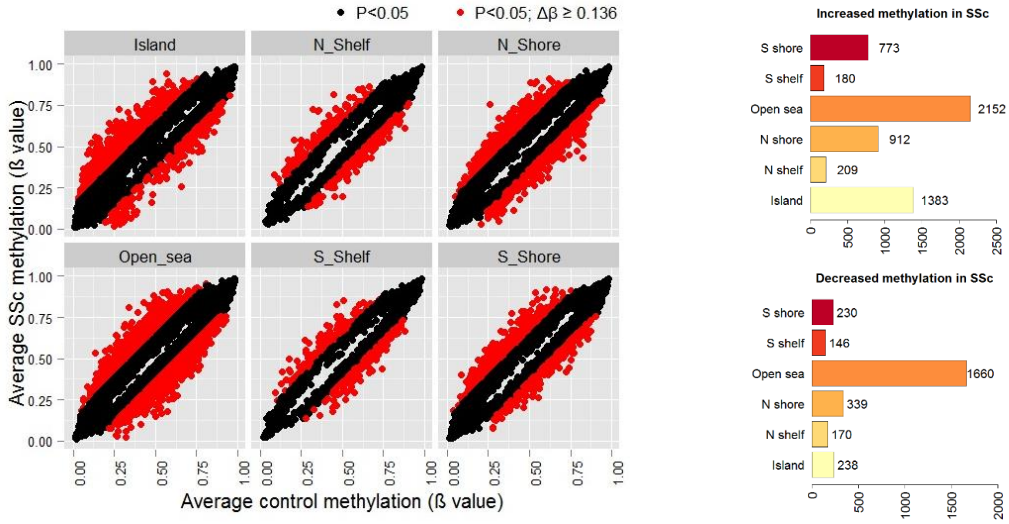


Figure 3.5.2. CpGs with altered methylation in relation to CpG islands in SSc compared to control lung fibroblasts. Left: scatter plots showing average methylation and the location of each CpG with a significant difference (P<0.05) in methylation between SSc (n=7) and control (n=6) lung fibroblasts. Each dot represents a CpG with either a Δβ< 0.136 (<13.6% change in methylation) (●) or a Δβ ≥ 0.136 (≥ 13.6% change in methylation between IPF and control lung fibroblasts (●). Right: bar graphs showing the number of CpGs in different locations with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in SSc compared to control lung fibroblasts.

Location	Total no. of CpGs	CpGs with increased methylation in SSc			CpGs with decreased methylation in SSc		
		Observed	Expected	O/E	Observed	Expected	O/E
Island	108197	1383	1867	0.74*	238	927	0.26*
N_Shelf	15598	209	269	0.78*	170	134	1.27*
N_Shore	43922	912	758	1.20*	339	376	0.90
Open_sea	108961	2152	1881	1.14*	1660	933	1.78*
S_Shelf	14040	180	242	0.74*	146	120	1.21*
S_Shore	34255	773	591	1.31*	230	293	0.78*
Total	**324973**						

Table 3.5.2. Observed to expected ratio (O/E) of the number of CpGs in each location with altered methylation in SSc compared to control lung fibroblasts. CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in SSc (n=7) compared to control (n=6) lung fibroblasts. *= Chi-square P<0.05.

3.6. Location of differentially methylated CpGs in IPF compared to SSc primary human lung fibroblasts

The average methylation and location of the 7827 CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts is shown in **Figure 3.6.1**. North shelf (O/E: 1.36) and open sea (O/E: 1.91) regions had more than the expected number of CpGs with increased

methylation in IPF compared to SSc lung fibroblasts. The number of CpGs observed in island regions with increased methylation in IPF compared to SSc lung fibroblasts were fewer than expected (O/E: 0.27). CpGs sites with decreased methylation in IPF compared to SSc lung fibroblasts were more abundant than expected in north shore (O/E: 1.22) and south shore (O/E: 1.20) regions. Island (O/E: 0.88), north shelf (O/E: 0.89) and south shelf (O/E: 0.73 had fewer than expected CpG sites with decreased methylation in IPF compared to SSc lung fibroblasts (**Table 3.6.1**).

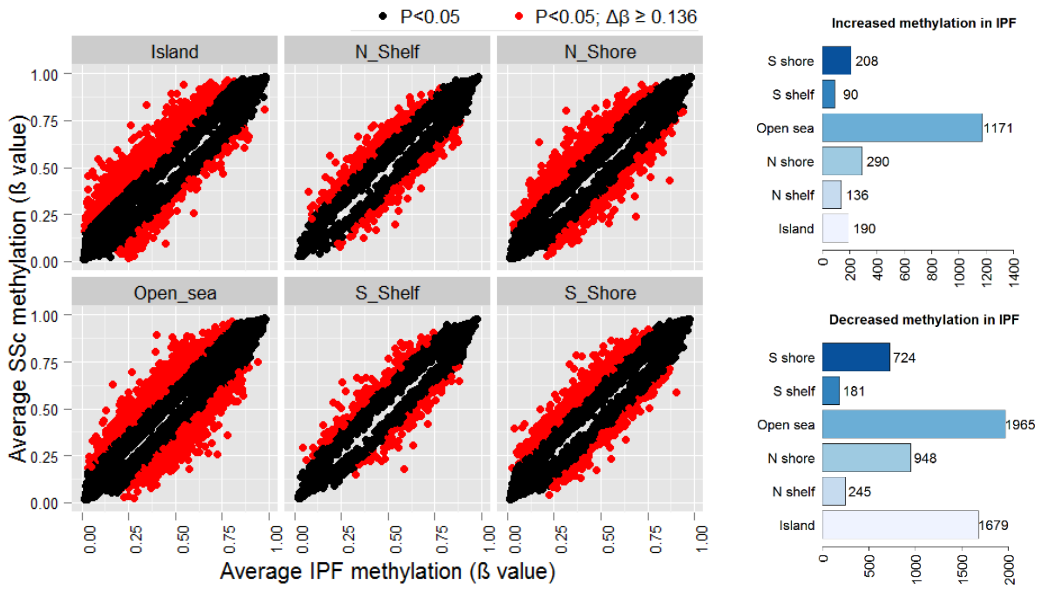


Figure 3.6.1. CpGs with altered methylation in relation to CpG islands in IPF compared to SSc lung fibroblasts. Left: scatter plots showing average methylation and the location of each CpG with a significant difference (P<0.05) in methylation between IPF (n=5) and SSc (n=7) lung fibroblasts. Each dot represents a CpG with either a Δβ< 0.136 (<13.6% change in methylation) (●) or a Δβ ≥ 0.136 (≥ 13.6% change in methylation (●).Right: bar graphs showing the number of CpGs in different locations with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts.

Location	Total no. of CpGs	CpGs with increased methylation in IPF			CpGs with decreased methylation in IPF		
		Observed	Expected	O/E	Observed	Expected	O/E
Island	108197	190	694	0.27*	1679	1912	0.88
N_Shelf	15598	136	100	1.36*	245	276	0.89
N_Shore	43922	290	282	1.03	948	776	1.22*
Open_sea	108961	1171	699	1.68*	1965	1925	1.02
S_Shelf	14040	90	90	1.00	181	248	0.73*
S_Shore	34255	208	220	0.95	724	605	1.20*
Total	**324973**						

Table 3.6.1. Observed to expected ratio (O/E) of the number of CpGs in each location with altered methylation in IPF compared to SSc lung fibroblasts. CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. *= Chi-square P<0.05.

3.7. Bisulfite sequencing validation of the Illumina Infinium Human Methylation 450 BeadChip microarray using TNXB.

The Illumina Infinium 450k methylation microarray identified TNXB as having a large number of CpGs with significantly altered methylation (P<0.05), many of which had large (≥13.6%) changes. As previously discussed, TNXB belongs to the tenascin family of ECM which consists of four family members: TNXB, TNR, TNC, and TNW. The TNXB gene contains 44 exons and encodes a large (464kDa) ECM, TNX, which is involved in collagen deposition, matrix stiffness and cell adhesion. Hypomethylation of a CGI located in exon 3 (-931 from the TNXB TSS) of the TNXB gene has previously been shown to correlate with increased tenascin-X expression in muscle tissue (Rakyan et al, 2004), suggesting that methylation could be important for tenascin-X gene activity. SP1 and SP3 binding sites in the TNXB promoter region have also been suggested to be important in regulating TNXB transcription (Wijesuriya et al, 2002). Other regions such as enhancers or other CpGs may also be important but the mechanisms of TNXB regulation remain poorly understood and no studies to date have examined the methylation or expression of TNXB in IPF or SSc lung fibroblasts.

To validate the Illumina Infinium 450k methylation microarray, bisulfite sequencing was performed on the TNXB gene. Microarray analysis identified 88 CpG sites in the TNXB gene with altered (P<0.05) methylation (14 increased, 74 decreased) in IPF compared to control lung fibroblasts (**Figure 3.7.1**). Ten of these CpGs had ≥ 13.6% increases in methylation, whilst 17 of these CpGs had ≥ 13.6% decreases in methylation. 151 CpG sites were identified as having altered (P<0.05) methylation (73 increased, 78 decreased) in SSc compared to control lung fibroblasts (**Figure 3.7.2**). 38 of these CpGs had ≥ 13.6% increases in methylation, whilst only 3 of these CpGs had ≥ 13.6% decreases in methylation. These data suggest that TNXB has decreased methylation in IPF compared to control lung fibroblasts whereas SSc lung fibroblasts have increased methylation of TNXB. Furthermore, 139 CpGs were identified as having significantly (P<0.05) altered methylation (35 increased, 104 decreased) methylation in IPF compared to SSc lung fibroblasts (**Figure 3.7.3**). CpG sites with increased methylation (≥ 13.6%) in IPF compared with control lung fibroblasts were mainly located in a north shelf region flanking the CGI in exon 10. CpGs with decreased methylation (≥ 13.6%) were predominantly found in the open sea regions and in a south shelf region flanking the CGI in exon 10 (**Figure 3.7.4**). Only 4 CpG sites had increased (≥ 13.6%) methylation in IPF compared to SSc lung fibroblasts, all of which were located in the north shore regions flanking the CGI located in intron 6 of the TNXB gene. CpG sites with increased methylation (≥ 13.6%) in SSc compared to control lung fibroblasts were mainly located in CGIs, north shelf regions and open sea regions (**Figure 3.7.4**).

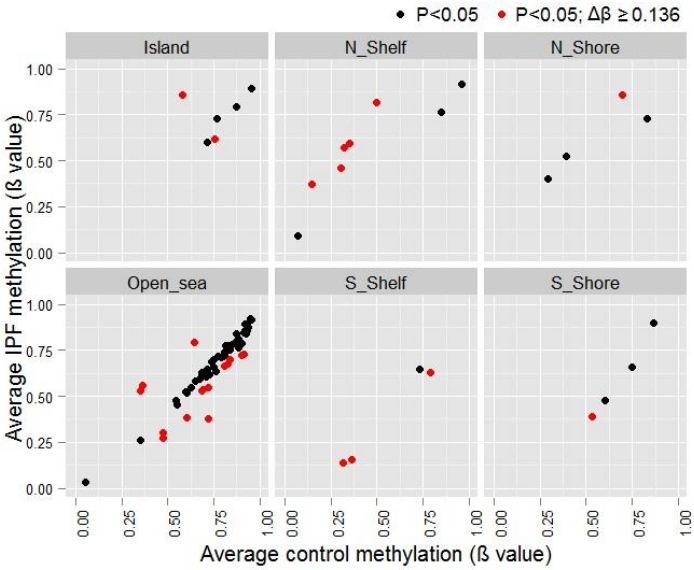


Figure 3.7.1. Location of CpGs in the TNXB gene which have altered methylation in IPF compared to control lung fibroblasts. Each dot represents a different CpG which had significantly (P<0.05) altered methylation in IPF (n=5) compared to control (n=6) lung fibroblasts. ● CpGs with a Δβ ≤ 0.136 (≤13.6% change in methylation), ● CpGs with a Δβ ≥ 0.136 (≥ 13.6% change in methylation). The β value represents the level of methylation (0= unmethylated, 1= methylated).

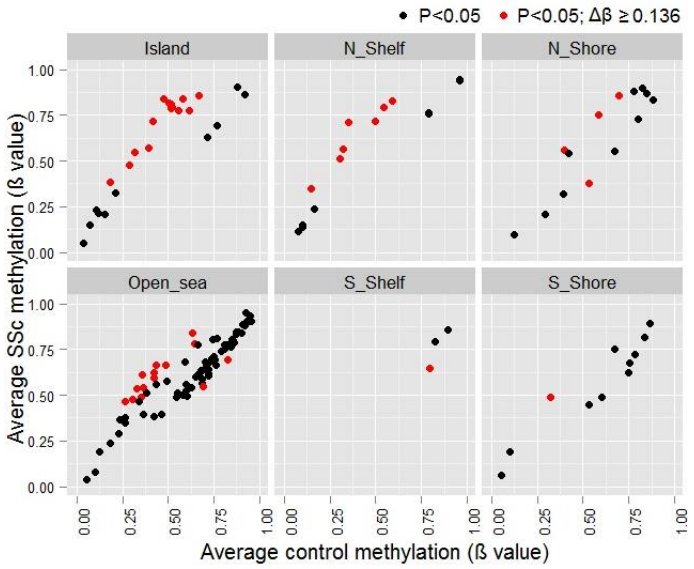


Figure 3.7.2. Location of CpGs in the TNXB gene which have altered methylation in SSc compared to control lung fibroblasts. Each dot represents a different CpG which had significantly (P<0.05) altered methylation in SSc (n=7) compared to control (n=6) lung fibroblasts. ● CpGs with a Δβ ≤ 0.136 (≤13.6% change in methylation), ● CpGs with a Δβ ≥ 0.136 (≥ 13.6% change in methylation). The β value represents the level of methylation (0= unmethylated, 1= methylated).

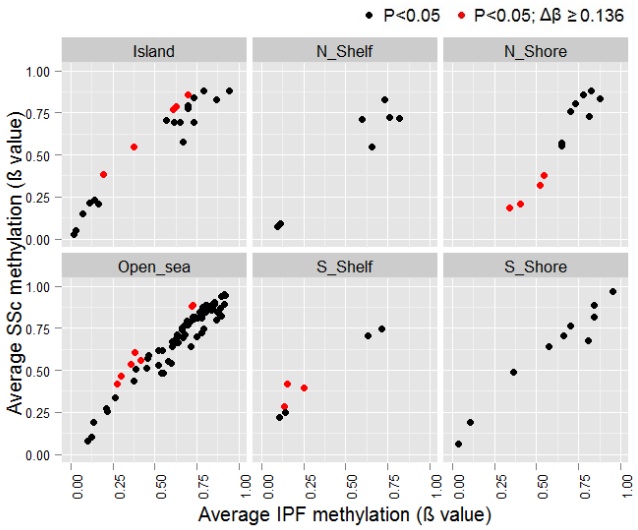


Figure 3.7.3. Location of CpGs in the TNXB gene which have altered methylation in IPF compared to SSc lung fibroblasts. Each dot represents a different CpG which had significantly (P<0.05) altered methylation in IPF (n=5) compared to SSc (n=7) lung fibroblasts. ● CpGs with a Δβ ≤ 0.136 (≤13.6% change in methylation), ● CpGs with a Δβ ≥ 0.136 (≥ 13.6% change in methylation). The β value represents the level of methylation (0= unmethylated, 1= methylated).

The microarray identified 2 CpGs within the exon 3 CGI which contained significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to control lung fibroblasts. In the exon 3 CGI, 14 CpG sites had significantly increased methylation (Δβ ≥ 0.136; P<0.05) in SSc compared to control lung fibroblasts (**Figure 3.7.4**). A 349bp region within this CGI was bisulfite sequenced to validate the microarray data. This region included 7 CpGs which were identified on the microarray. Bisulfite sequencing showed the differences in methylation between control, IPF and SSc lung fibroblasts was in agreement with microarray data (Pearsons correlation: r=0.77), although these differences in methylation were generally smaller (**Figure 3.7.5**).

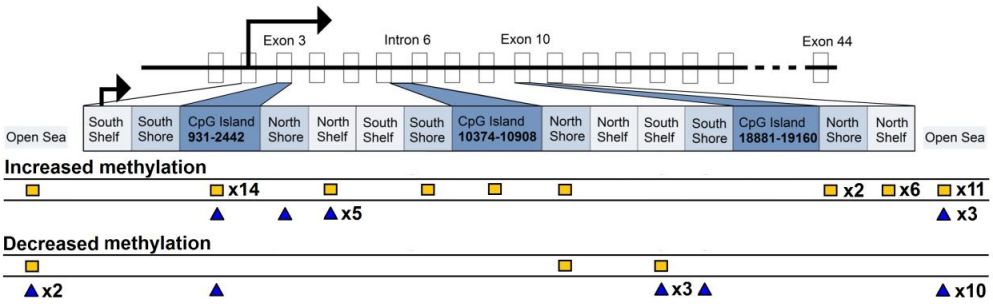


Figure 3.7.4. Diagram of the TNXB gene. The location of 27 CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) and 41 CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in SSc (n=7) compared to control (n=6) lung fibroblasts. ▲ represents IPF CpGs, ■ represents SSc CpGs. 3 CGIs located in exon 3, intron 6 and exon 10 were identified by the microarray.

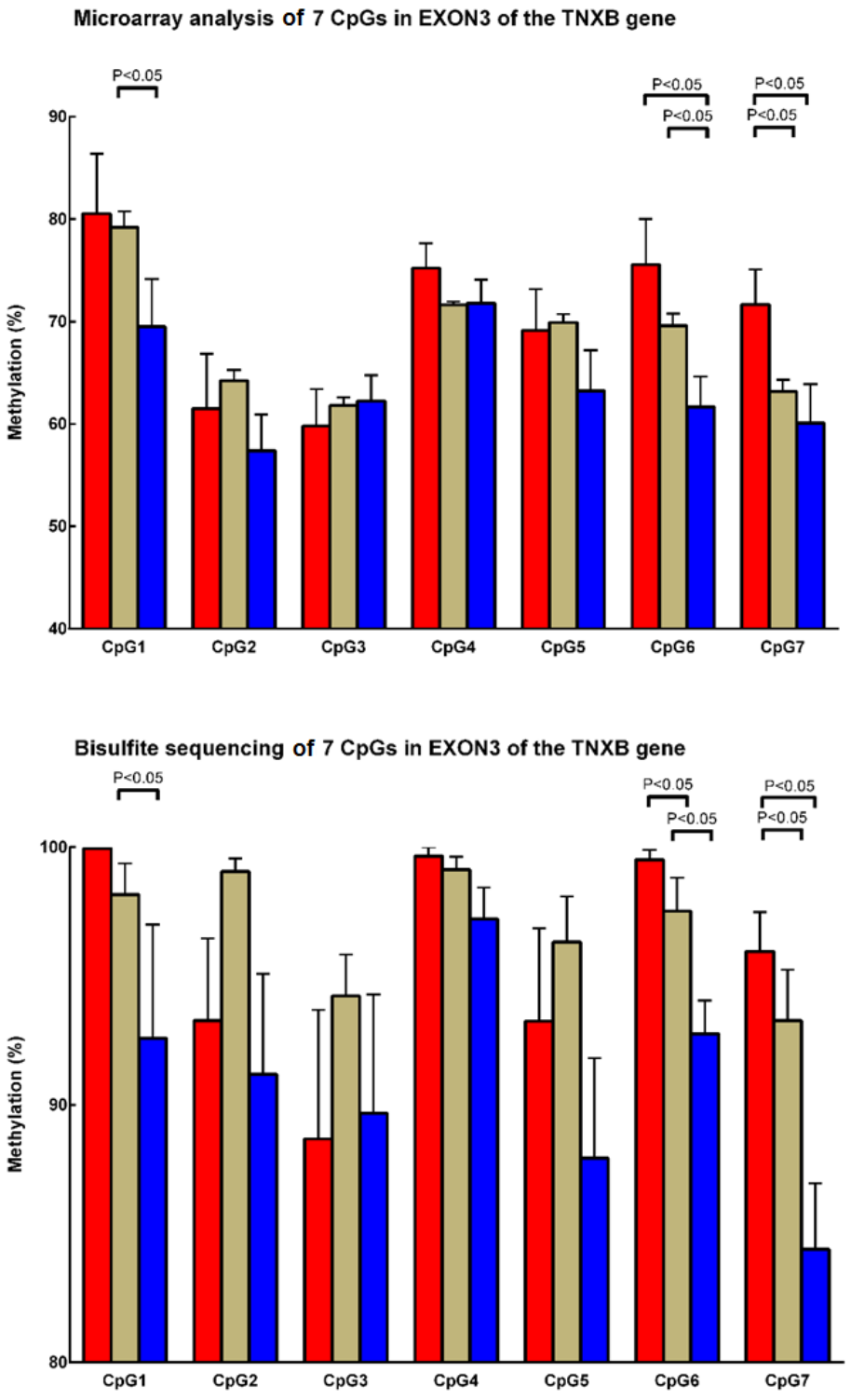


Figure 3.7.5. Validation of microarray data using bisulfite sequencing on 7 CpGs located in the exon 3 CpG island. CpGs with a significant difference in methylation in IPF (blue; n=5) and/or SSc (yellow; n=7) compared to control (red; n=6) lung fibroblasts using, top; microarray data, bottom; bisulfite sequencing. Data presented as the mean ± S.E.M.

The CpG island in exon 10 had 5 CpG sites which were covered by the array, 1 of which had a significant (P<0.05) decrease in methylation in IPF compared to control lung fibroblasts. A 440bp region covering more CpGs than the array (19 CpGs) within this CGI was bisulfite sequenced to further explore changes in TNXB methylation in IPF lung fibroblasts. Seven of these CpGs had significantly (P<0.05) decreased methylation in IPF compared to control lung fibroblasts. Five of these CpGs also had significantly (P<0.05) decreased methylation in IPF compared to SSc lung fibroblasts (**Figure 3.7.6**).

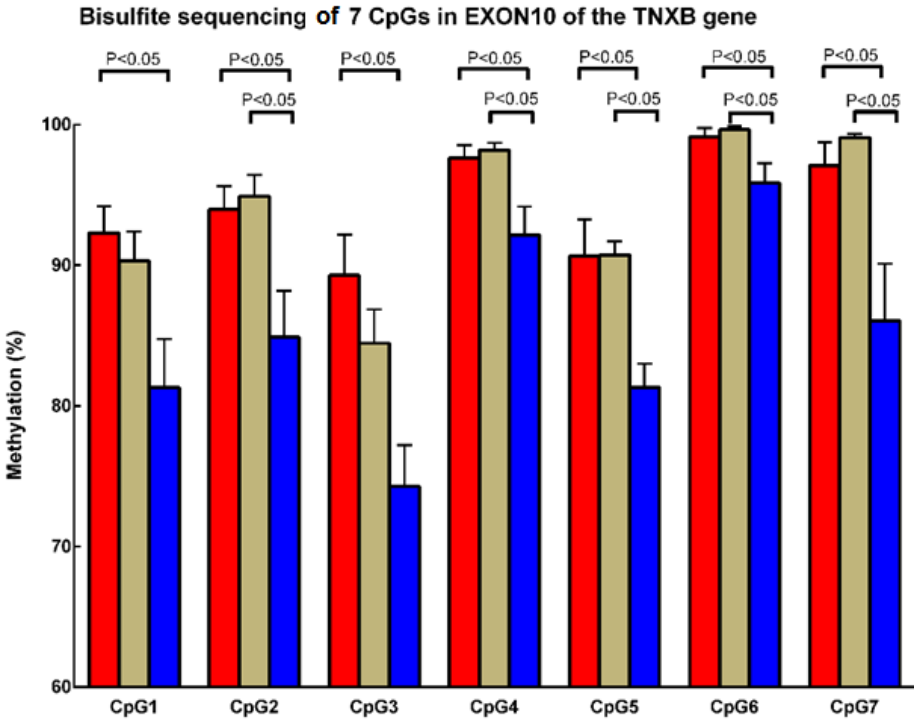


Figure 3.7.6. Bisulfite sequencing of CpGs located in the exon 10 CpG island. Seven CpGs with significantly (P<0.05) altered methylation in IPF (blue; n=5) compared to control (red; n=6) lung fibroblasts. Five of these CpGs also had significantly altered (P<0.05) methylation in IPF compared to SSc (yellow; n=7) lung fibroblasts. Data presented as the mean ± S.E.M.

3.8. Summary

- Distribution of CpG methylation was bimodal in control, IPF and SSc-PF lung fibroblasts with the highest frequency of CpGs having 0-15% methylation and 80-90% methylation on autosomes and 25-35% and 75-80% on the X-chromosome.

- CpG islands within 1.5kb of their corresponding gene's TSS had a very high frequency of CpGs with 0-15% methylation whereas CGIs further than 1.5kb had a bimodal distribution of CpG methylation with the highest frequencies of CpGs having 0-15% and 80-85% methylation.

- 7153 and 8392 CpGs had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF and SSc-PF compared to control lung fibroblasts, respectively.

- 69% (4908) of the 7153 CpGs in IPF had decreased methylation compared to control lung fibroblasts.

- 67% (5609) of the 8392 CpG in SSc-PF had increased methylation compared to control lung fibroblasts.

- Multiple CpGs and genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF and SSc-PF compared to control lung fibroblasts overlap including Wnt genes WNT10A and SFRP1 and novel genes potentially important in PF including, SDK1, SATB2, FOXS1 and GATA2, however multiple CpGs and genes are also distinct to each disease.

- Open sea regions consistently had significantly ($P<0.05$) more than the expected number of CpGs whereas CGIs had less than the expected number of CpGs with increased and decreased methylation in IPF and SSc-PF compared to control lung fibroblasts. This suggests that altered methylation in open sea regions is common in IPF and SSc-PF and could potentially be important and overlooked by studies solely focusing on methylation of CpGs within promoter CGIs.

- Bisulfite sequencing of the TNXB gene confirmed methylation microarray data and highlighted CpGs not on the array which also have significantly altered ($P<0.05$) methylation in IPF compared to control and SSc-PF lung fibroblasts.

3.9. Distribution of gene expression in lung fibroblasts and in IPF and SSc compared to control primary human lung fibroblasts

Genome-wide distribution of gene expression in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts identified the highest frequency of genes with a Log2 transformed normalised value <7.0 (**Figure 3.9.1**). Genes below this level were judged as having low expression as genes with a log2 transformed normalised value were rarely detected. The highest frequency of genes with significantly altered expression (TNoM ≤1; P<0.05) in IPF and SSc compared to control lung fibroblasts were also those which had low expression (Log2 transformed normalised value <7.0) (**Figure 3.9.2**).

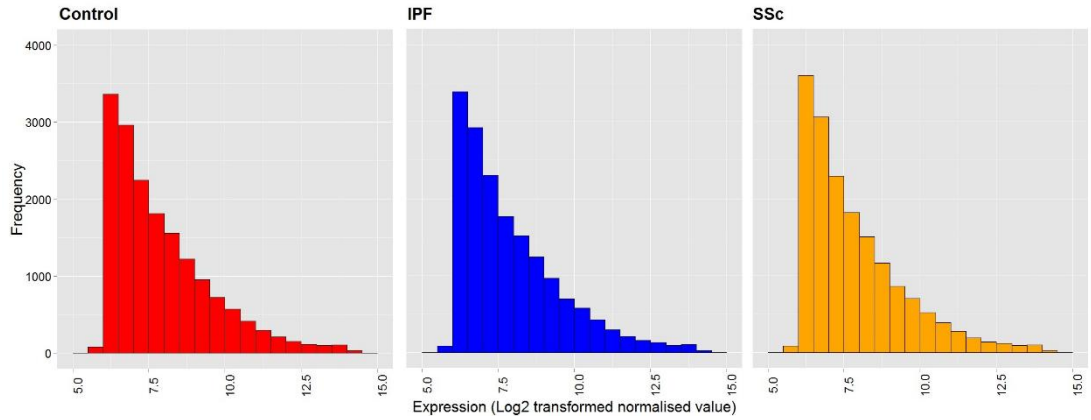


Figure 3.9.1. Genome-wide distribution of gene expression in lung fibroblasts. Each bar represents the number of genes within each boundary. Red: Control (n=6), blue: IPF (n=5) and yellow: SSc (n=7) lung fibroblasts.

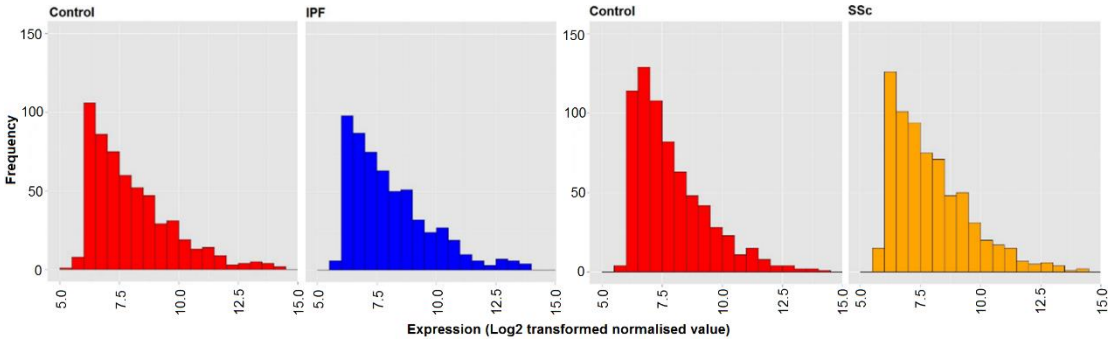


Figure 3.9.2. Distribution of genes which had altered expression in IPF and SSc compared to control lung fibroblasts. Each bar represents the number of genes within each boundary which had significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) or SSc (n=7) compared to control (n=6) lung fibroblasts. Red = control, blue = IPF and yellow = SSc lung fibroblasts.

3.10. Microarray analysis of differentially-expressed genes in IPF and SSc compared to control primary human lung fibroblasts

The Illumina Human Expression array identified 568 genes (267 decreased, 301 increased) with altered expression (TNoM ≤1; P<0.05) in IPF and 688 genes (324 decreased, 364 increased) with altered expression (TNoM ≤1; P<0.05) in SSc compared to control lung fibroblasts. Ninety-six genes including WNT2B (Bayle et al, 2008) and PPARγ (Lakatos et al, 2007, López-Isac et al, 2014) which have previously been associated with fibrosis, had significantly altered expression (47 increased, 49 decreased) in both IPF and SSc compared to control lung fibroblasts (**Figure 3.10.1**). The average expression for each gene which had significantly altered expression (TNoM ≤1; P<0.05) in IPF or SSc compared to control lung fibroblasts and which chromosomes the genes are located on is shown in **Figure 3.10.2**.

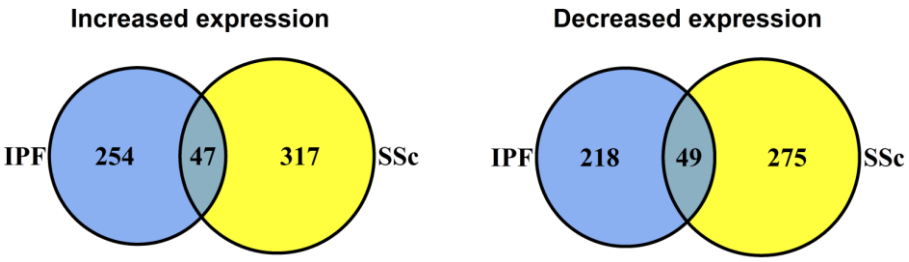


Figure 3.10.1. The number of distinct and overlapping genes with altered expression in IPF and SSc compared with control lung fibroblasts. Genes with significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) and SSc (n=7) compared to control (n=6) lung fibroblasts.

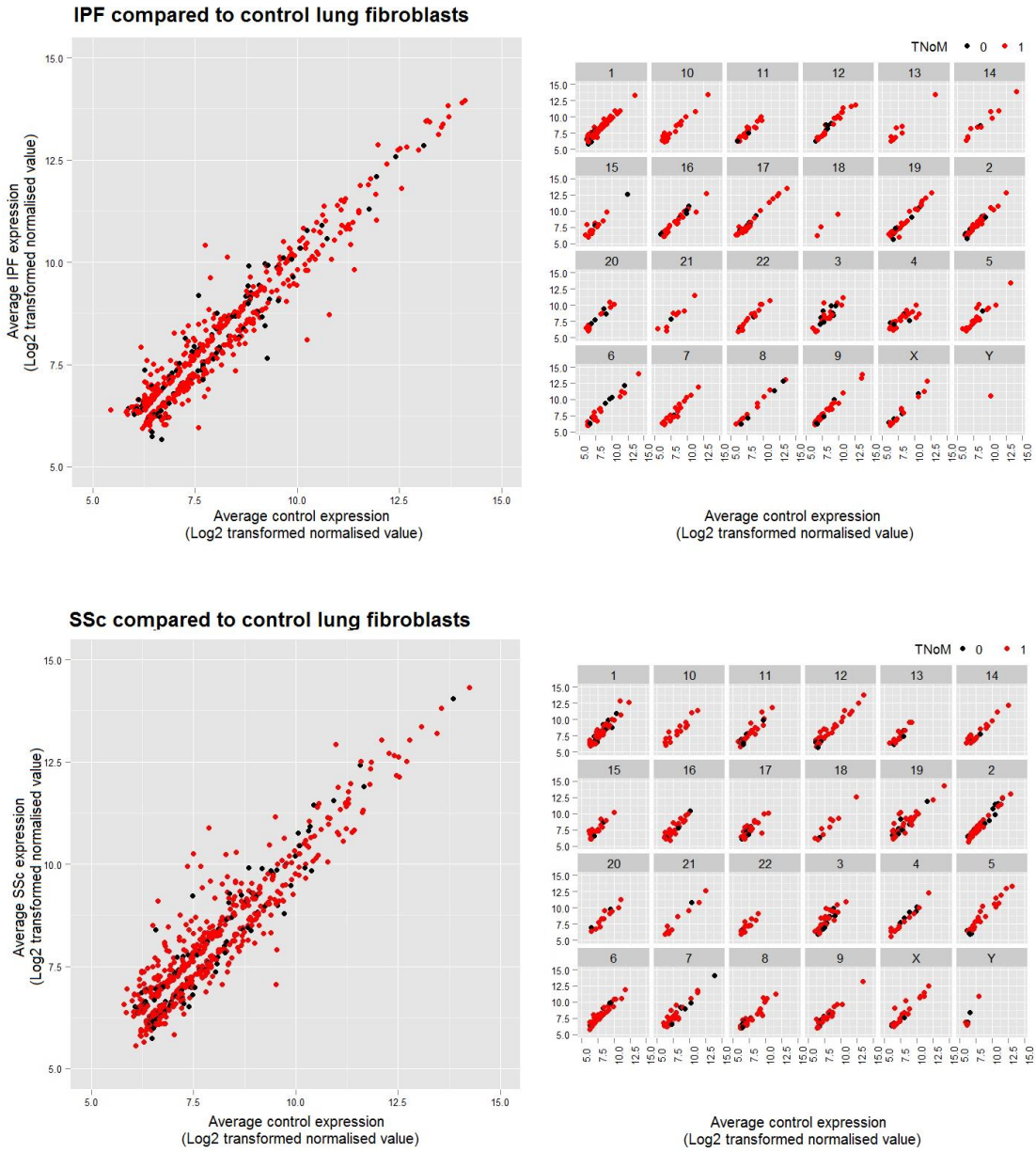


Figure 3.10.2. Gene expression in IPF and SSc compared to control lung fibroblasts. Top: shows the average expression of genes with significantly increased or decreased expression (TNoM ≤1; P<0.05) in IPF (n=5) compared with control (n=6) lung fibroblasts and which chromosome they are located on. Below: shows the average expression of genes with significantly increased or decreased expression (TNoM ≤1; P<0.05) in SSc (n=7) compared with control (n=6) lung fibroblasts and which chromosome they are located on. ● shows genes with a TNoM=0, ● shows genes with a TNoM=1.

Genes previously shown to have altered expression in IPF including IL8, WNT2B, FBLN2, PPARγ, SEPP1, CXCL1 and CXCL6 all had significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to control lung fibroblasts (**Figure 3.10.3**). Genes previously shown to have aberrant expression in SSc including S100A4 (Tomcik et al, 2014), NOTCH3 (Dees et al, 2011), IGFBP7 (Hsu et al, 2011), CCL13 (Yanaba et

al, 2010), IL7R (Grigoryev et al, 2008) and TIMP4 (Elias et al, 2008) all had significantly altered expression (TNoM ≤1; P<0.05) in SSc compared to control lung fibroblasts (**Figure 3.10.4**).

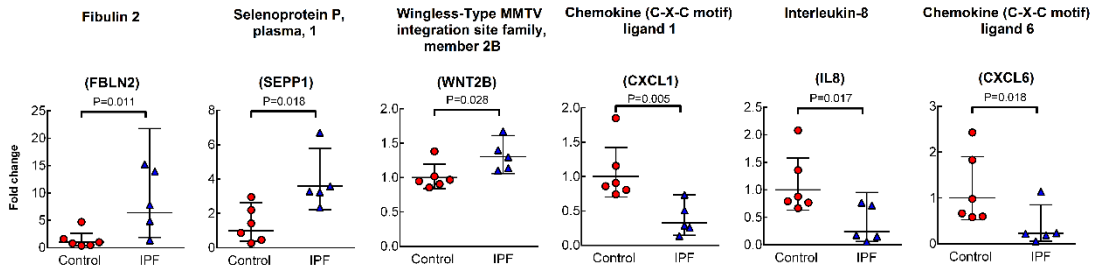


Figure 3.10.3. Genes previously implicated in IPF which had altered expression compared to control lung fibroblasts. Genes which had significantly altered expression (TNoM ≤1; P<0.05) IPF (n=5) compared to control (n=6) lung fibroblasts. Data presented as the geometric mean fold-change relative to the average control ± 95% CI. Each data point represents a different cell line.

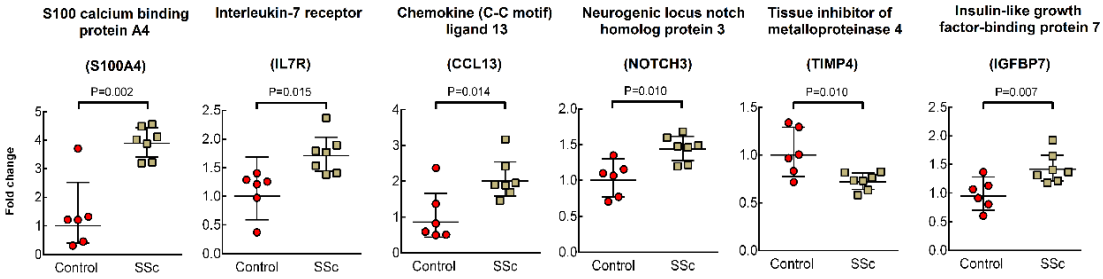


Figure 3.10.4. Genes previously implicated in SSc which had altered expression compared to control lung fibroblasts. Genes which had significantly altered expression (TNoM ≤1; P<0.05) SSc (n=7) compared to control (n=6) lung fibroblasts. Data presented as the geometric mean fold-change relative to the average control ± 95% CI. Each data point represents a different cell line.

Cluster analysis of the 568 and 688 differentially expressed genes in IPF and SSc respectively, clearly distinguished IPF and SSc from control lung fibroblasts however, there was some heterogeneity between cell lines (**Figure 3.10.5**). Although all control cell lines clustered together, control 6 had a gene expression profile that appeared intermediate between an SSc and a control. This was the same control which had a methylation profile which clustered with SSc cell lines (**see section 3.3, figure 3.3.4**).

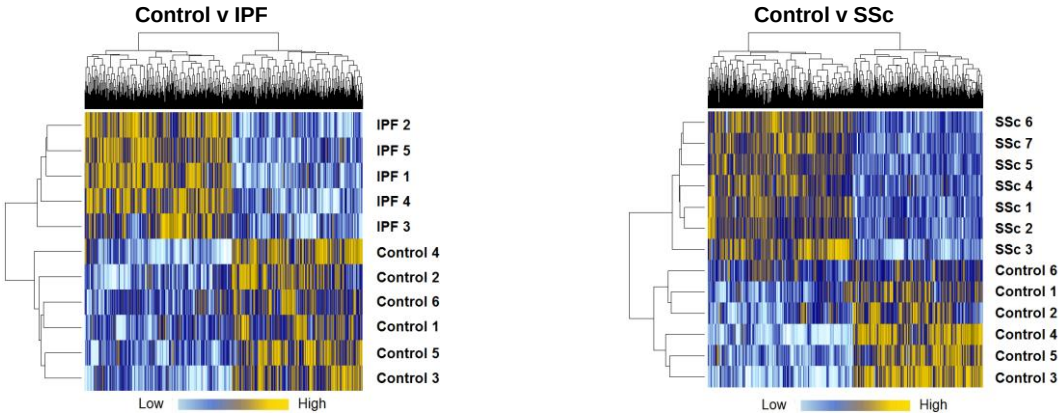


Figure 3.10.5. Hierarchical cluster analysis on differentially-expressed genes in IPF and SSc compared to control lung fibroblasts. Genes with significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) and SSc (n=7) compared to control (n=6) lung fibroblasts. Light blue represents low expression, yellow represents high expression with respect to each gene across all cell lines.

3.11. Differentially-expressed genes in IPF compared to SSc lung fibroblasts

The highest frequency of genes with significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to SSc lung fibroblasts were those which had a Log2 transformed normalised value <6.5 (**Figure 3.11.1**). The Illumina Human Expression array identified 1117 genes (499 decreased, 618 increased) with significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to SSc lung fibroblasts (**Figure 3.11.2**). IPF and SSc cell lines clustered separately from each other (**Figure 3.11.3**) suggesting they may have distinct gene expression profiles. Examples of strongly differentially-expressed genes in IPF compared to SSc lung fibroblasts, some of which have previously been linked to PF (Kohno, 1999, Königshoff et al, 2009, Ishikawa et al, 2012), are shown in **Figure 3.11.4**.

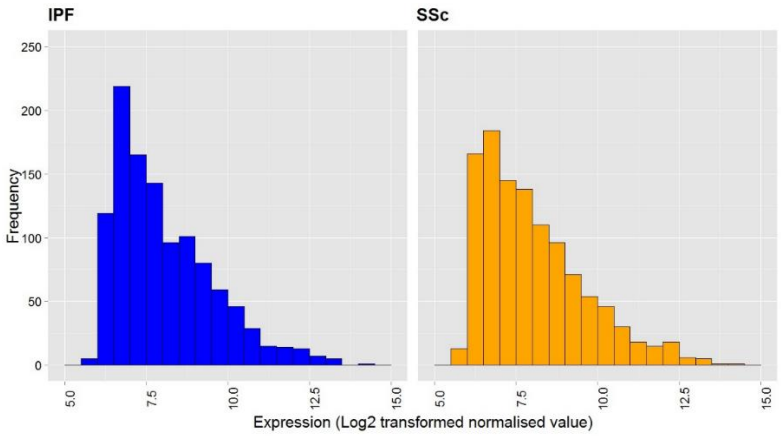


Figure 3.11.1. Distribution of genes which had altered expression in IPF compared to SSc lung fibroblasts. Each bar represents the number of genes within each boundary which had significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. Blue = IPF, orange = SSc lung fibroblasts.

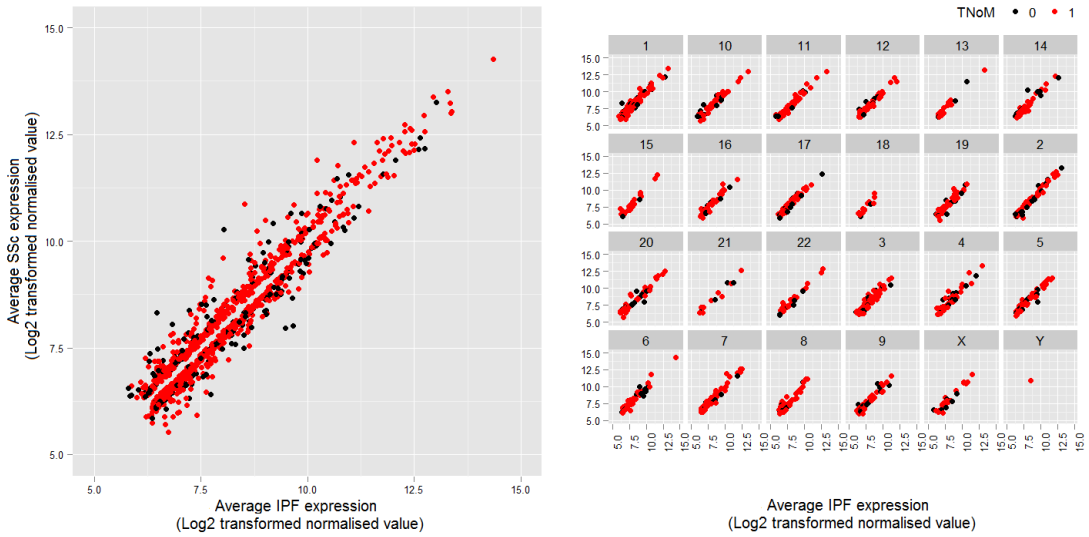


Figure 3.11.2. Gene expression in IPF compared to SSc lung fibroblasts. Scatter graphs show the average expression of genes with significantly increased or decreased expression (TNoM ≤1; P<0.05) in IPF (n=5) compared with SSc (n=7) lung fibroblasts and which chromosome they are located on. ● Show genes with a TNoM=0, ● show genes with a TNoM=1.

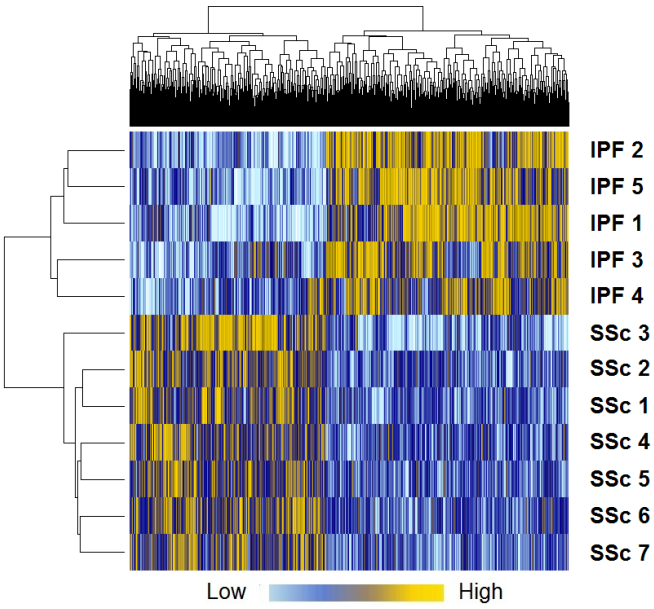


Figure 3.11.3. Hierarchical cluster analysis on differentially expressed genes in IPF compared to SSc lung fibroblasts. Genes (n=1117) with significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. Light blue represents low expression, yellow represents high expression with respect to each gene across all cell lines.

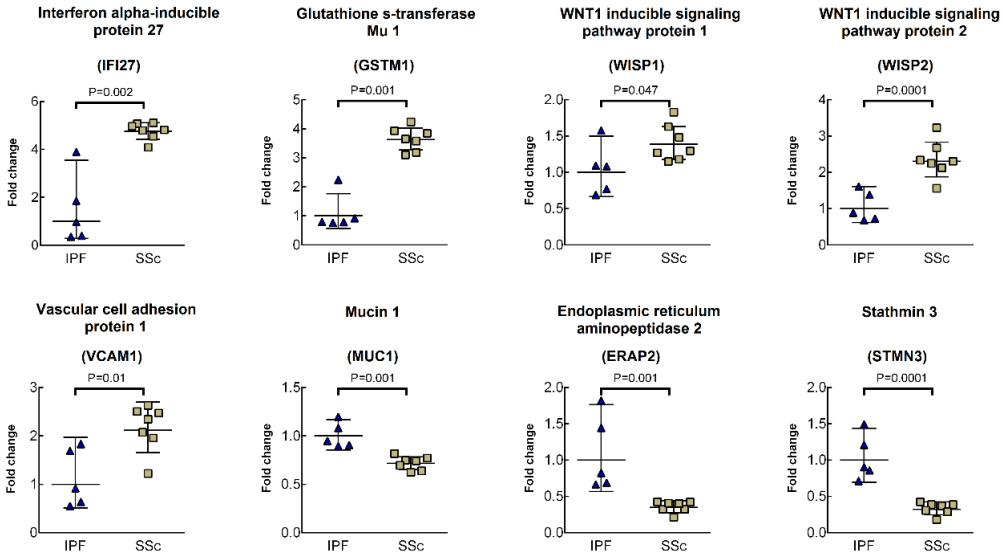


Figure 3.11.4. Differences in gene expression in IPF compared to SSc lung fibroblasts. Genes which had significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. Data presented as the geometric mean fold-change relative to the average IPF expression ± 95% CI. Each data point represents a different cell line.

3.12. qRT-PCR validation of the Illumina Infinium gene expression microarray

Genes including Cell adhesion molecule 1 (CADM1), Eukaryotic Translation Initiation Factor 1A, Y-Linked (EIF1AY) and IL8 were selected to validate the expression microarray based on all genes having significantly altered expression in IPF and/or SSc compared to control lung fibroblasts, which inversely correlated to methylation status according to microarray analysis. Other genes including metalloproteinases (MMP10, MMP12) which have been linked to fibrosis and can be activated by hypomethylation (Couillard et al, 2006) and TNXB, a gene identified which had multiple CpGs with significantly altered methylation in IPF and SSc compared to control (but was not detected on the expression array), were also analysed.

qRT-PCR confirmed the microarray data by identifying similar significant differences in gene expression between IPF, SSc and control lung fibroblasts (**Figure 3.12.1**). qRT-PCR and microarray data strongly correlated for all genes (IL8: r=0.91, CADM1: r=0.74, EIF1AY: r=0.85). In general, microarray and qRT-PCR analysis of fold-changes in gene expression and variation between cell lines in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts cell lines were consistent. TNXB, a gene which encodes the large 464kDa ECM glycoprotein, TNX, was of particular interest based on its previously described role in the ECM, disease and its interactions with other ECM proteins (**see Chapter 1: section 1.3**). Although a TNXB probe was on the Illumina expression microarray, it was not detected. qRT-PCR did detect TNXB and identified a significant increase in TNXB expression in IPF (n=5) compared to control (n=8) and SSc (n=8) lung fibroblasts. There was no significant difference (P=0.791) in TNXB gene expression between control and SSc lung fibroblasts (**Figure 3.12.2**).

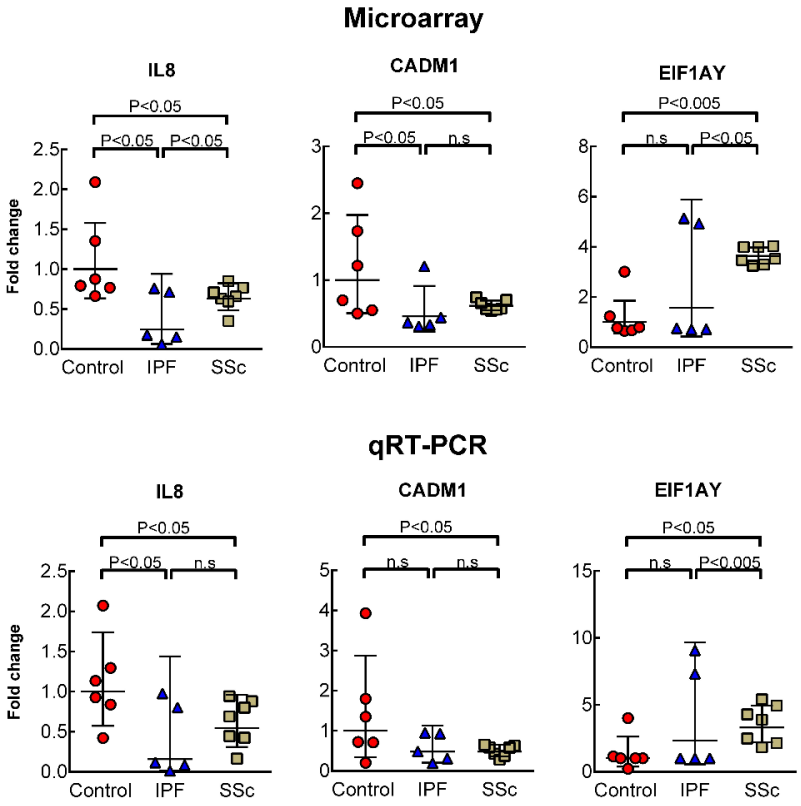


Figure 3.12.1. Validation of microarray data using qRT-PCR. Comparison of gene expression data for 3 genes; IL8, CADM1 and EIF1AY in control (n=6), IPF (n=5) and SSc (n=7) using microarray and qRT-PCR data. Data presented as the geometric mean fold-change relative to the average control expression ± 95% confidence intervals. Each data point represents a different cell line.

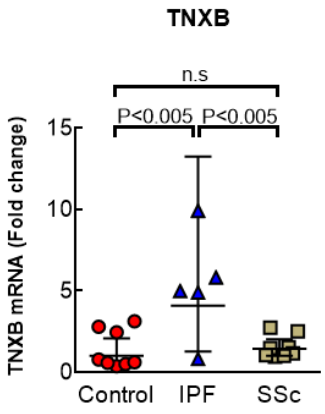


Figure 3.12.2. TNXB expression in lung fibroblasts. TNXB expression in control (n=8), IPF (n=5) and SSc (n=8) lung fibroblasts. Data presented as the geometric mean fold-change relative to the average basal control expression ± 95% confidence intervals. Each data point represents a different cell line.

3.12.1. Immunolocalisation of TNX in control, IPF and SSc lung tissue confirms qRT-PCR

Following confirmation of increased TNXB at the mRNA level in IPF compared to control and SSc lung fibroblasts, immunohistochemistry experiments were performed to determine the immunolocalisation of TNX in human control (n=7), IPF (n=6) and SSc (n=3) lung tissue. To optimise the concentration of the primary anti-TNX antibody, a dilution series was first performed on human lung tissue sections. Antibody at 5µg/ml showed strong staining of TNX. Specificity was confirmed using tissue sections incubated with serum only or with an IgG isotype control which showed no obvious staining. All IPF lung tissue sections showed strong staining for TNX in the ECM and fibroblasts, whereas control and SSc lung tissue section showed weak staining confirming that TNX is also increased in IPF lung (**Figure 3.12.1.1**).

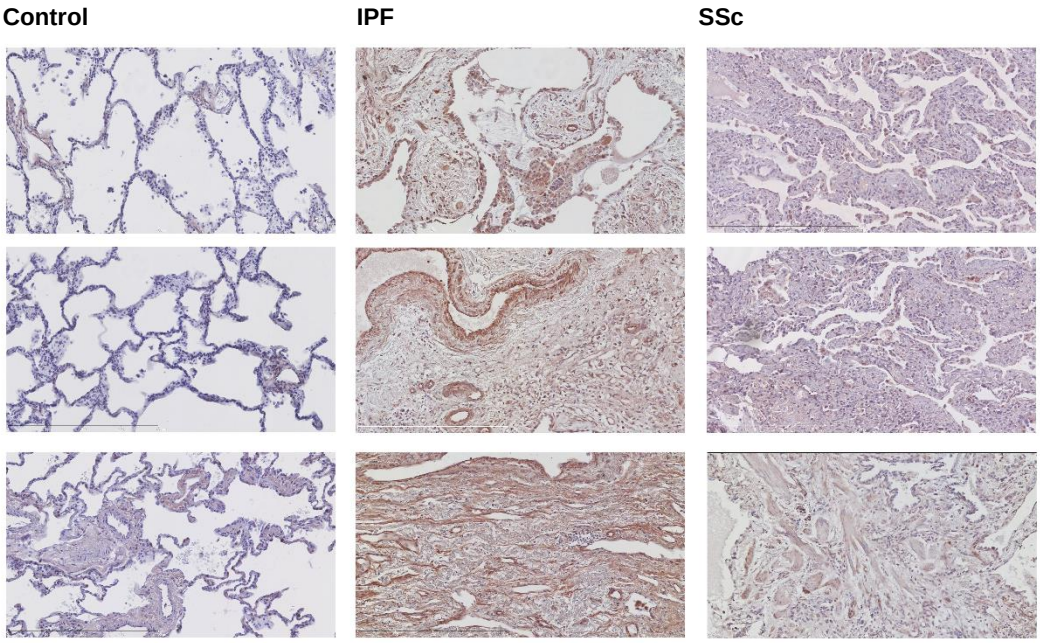


Figure 3.12.1.1. TNX immunolocalisation in lung tissue. Representative histological sections showing TNX immunolocalisation in control (n=3), IPF (n=3) and SSc (n=3) lung tissue specimens. Brown staining depicts TNX. 20x original magnification. Scale bar = 400µm.

3.13. Summary

- Multiple genes previously associated with IPF and/or SSc-PF, including PPARγ, IL8, CCL13, CXCL1, SEPP1, S100A4 and TIMP4, have significantly altered expression (TNoM ≤1; P<0.05) in IPF and SSc compared to control lung fibroblasts respectively. Multiple other genes identified with significantly altered expression (TNoM ≤1; P<0.05) in IPF and SSc-PF compared to control lung fibroblasts may be novel to fibrosis.

- 96 genes overlapped between IPF and SSc-PF suggesting multiple common genes are affected in both diseases.

- 1117 (618 increased, 499 decreased) genes had significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to SSc-PF lung fibroblasts suggesting IPF and SSc-PF have a different expression pattern.

- qRT-PCR gene expression strongly correlated with microarray expression.

- TNXB expression was identified using qRT-PCR and had increased expression in IPF compared to control and SSc-PF lung fibroblasts.

- IHC confirmed TNX expression in control, IPF and SSc-PF lung tissue.

- Data suggest TNXB is a novel gene in IPF which may be regulated by methylation.

3.14. Correlation between methylation and gene expression in primary human lung fibroblasts

3.14.1. Overview

It remains unknown what change in methylation is sufficient to have a biological effect and whether the change in methylation is the same for all genes or cell types. For example, Yang et al, showed that 8 CpGs on the Castor zinc finger 1 (CASZ1) gene, had on average a 3.5% change in methylation in IPF compared to control alveolar type II epithelial cells which corresponded with increased gene expression (Yang et al, 2014). Huang et al, showed that an average decrease of 6.9% over 28 CpGs of the gene Cyclin-dependent kinase 4 inhibitor B (CDKN2B) resulted in increased expression (Huang et al, 2014) whereas Sanders et al, showed that genes including Dimethylarginine dimethylaminohydrolase DDAH and Tumor protein p53-inducible nuclear protein 1 (TP53INP1) had large methylation differences (>20%) in IPF compared to control lung tissue which corresponded with altered expression (Sanders et al, 2012). Other studies with Illumina arrays have used a 13.6% cut as this can detect differences with 95% confidence (Bibikova et al, 2009, Lokk et al, 2012).

DNA methylation is one of many epigenetic mechanisms which control transcriptional programmes, thus it is difficult to distinguish the direct effects from the indirect effects of methylation on gene expression. Direct effects of methylation on gene expression refer to CpGs belonging to a gene which, when altered, affect the expression of that gene. Indirect effects of methylation on gene expression can refer to genes which have been activated/deactivated by changes in their methylation, which in turn regulate other genes. A good example of this is the miR-17-92 cluster which has reduced expression due to promoter hypermethylation in IPF lung tissue and fibroblasts. This cluster can regulate DNMT1 expression which in turn can regulate methylation levels of multiple targets (Dakhalallah et al, 2013).

To determine which CpGs in a gene may have a direct effect on that gene's expression, CpG methylation and gene expression were correlated together using all 18 cells lines (control; n=6, IPF: n=5 and SSc: n=7). In the following section I show correlation between basal methylation and basal expression levels. In Chapter 6 I show the effects of 5-Aza and how small changes in methylation may correlate to large changes in expression.

3.14.2. Distribution of CpG methylation which correlated with gene expression

Methylation of multiple CpGs strongly correlated (R^2= ≥0.5; P<0.05) with expression of their corresponding gene in lung fibroblasts. These CpGs were mainly distributed in island regions within 1.5kb of their corresponding genes TSS or in open sea regions further than 1.5kb from their corresponding genes TSS (**Figure 3.14.2.1**). The highest frequency of genes had low expression (**Figure 3.14.2.2**).

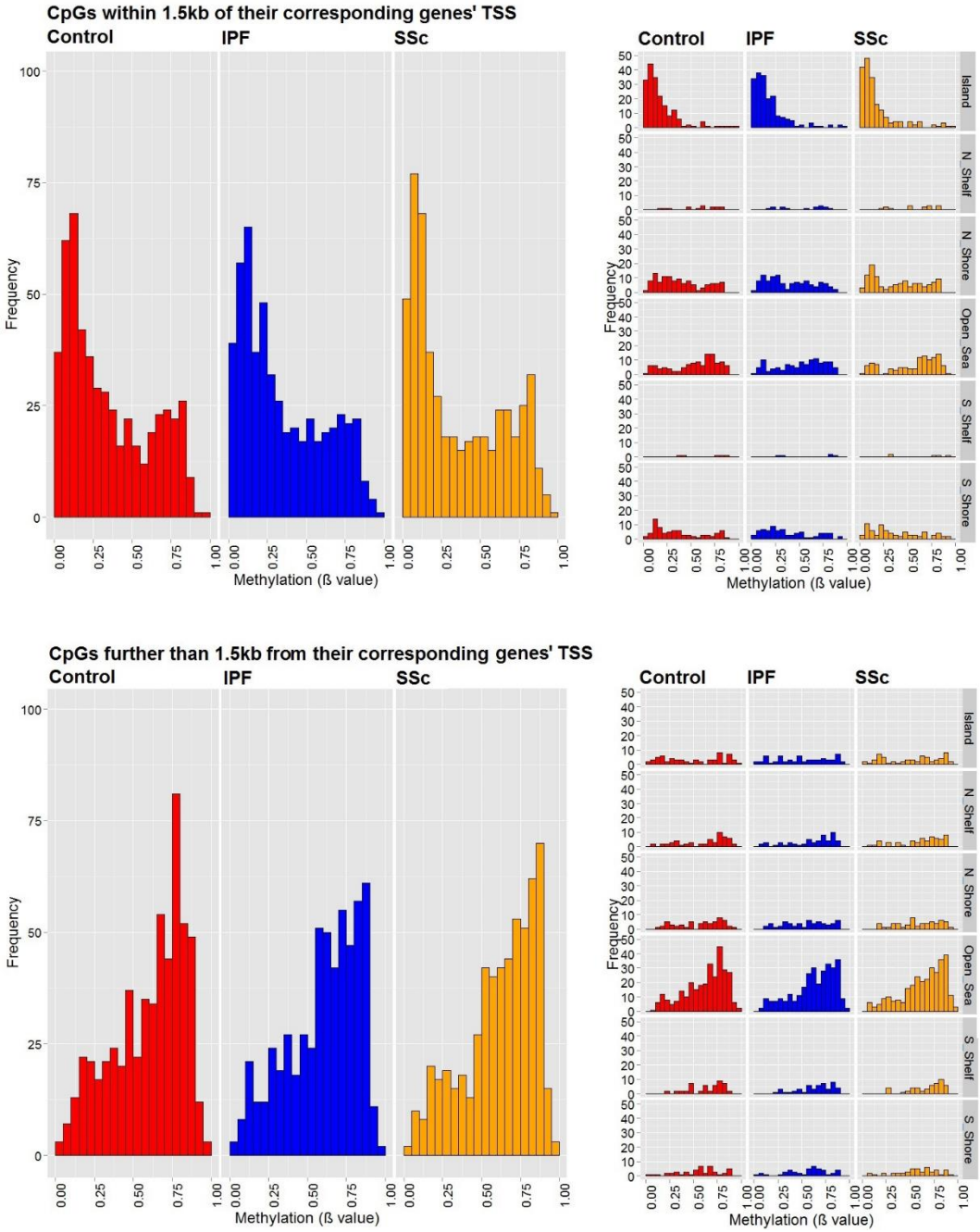


Figure 3.14.2.1. Distribution of CpG methylation of CpGs which correlated with expression of the genes. Top: distribution and location of CpGs within 1.5kb and bottom: further than 1.5kb of their

corresponding genes TSS, which had significant correlation between CpG methylation and expression in control (red; n=6), IPF (blue; n=5) and SSc (orange; n=7) lung fibroblasts.

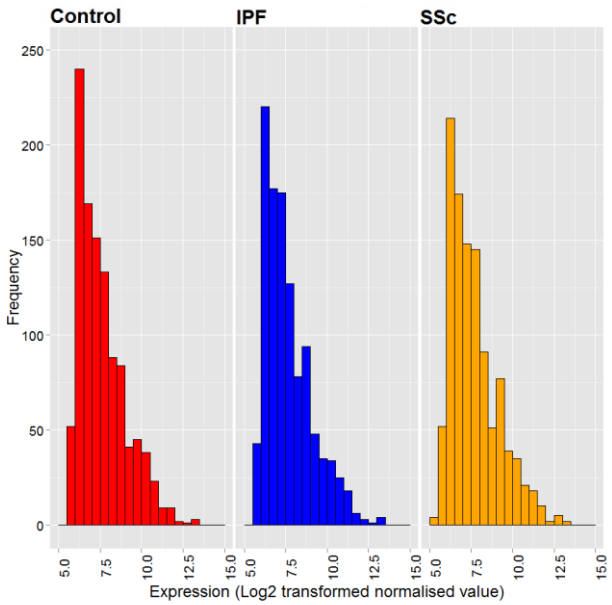


Figure 3.14.2.2. Distribution of genes which had expression levels which correlated with CpG methylation. Each bar represents the number of genes within each boundary. Red: Control (n=6), blue: IPF (n=5) and yellow: SSc (n=7) lung fibroblasts.

3.14.3. Microarray analysis of CpG methylation which correlated to expression

Methylation of 1088 CpGs (724 genes) correlated (R^2= ≥0.5; P<0.05) with expression of their respective gene across all lung fibroblast cell lines (n=18). Of these CpGs, 585 (412 genes) negatively correlated and 503 (358 genes) positively correlated (R^2= ≥0.5; P<0.05) with expression of their respective gene (**Table 3.14.3.1**). Methylation of 46 CpGs had both positive and negative correlation with expression, suggesting multiple locations of CpG methylation in the same gene may be important in regulating expression (**Table 3.14.3.2**). Examples of genes which had a significant correlation (R^2= ≥0.5; P<0.05) between CpG methylation and gene expression are shown in **Figure 3.14.3.1**.

	Positive correlation	Negative correlation	Overlap	Total
CpGs	503	585	0	1088
Genes	358	412	46	724

Table 3.14.3.1. The number of CpGs/genes with methylation which correlated with gene expression. CpG methylation which significantly correlated (R^2= ≥0.5; P<0.05) with expression in all 18 cell lines (control: n=6, IPF: n=5 and SSc: n=7).

Genes with both positive and negative correlation between methylation and expression					
ACVRL1	*CDH13*	*FST*	*MYOM2*	*PLAGL1*	*STX18*
ADAM15	*CHST15*	*GPER*	*NETO2*	*PPP3CA*	*TANC1*
ADAMTS2	*CLEC14A*	*GSTT1*	*NLGN4Y*	*PRKY*	*TNFAIP8L3*
ALDH3A1	*CPNE8*	*ICMT*	*NPTX1*	*RAMP1*	*TRIM56*
ANO1	*CRIPAK*	*IL16*	*PAX8*	*RPS4Y1*	*ZFHX4*
C13orf15	*DLL1*	*MACF1*	*PGM3*	*SAMD14*	*ZFY*
C1orf159	*EIF1AY*	*MAPRE1*	*PLA2G5*	*SASH1*	
CA12	*FAM13A*	*MGMT*	*PLAG1*	*SPON2*	

Table 3.14.3.2. Genes with both positive and negative correlation. CpG methylation which significantly correlated (R^2= ≥0.5; P<0.05) with expression in all 18 cell lines (control: n=6, IPF: n=5 and SSc: n=7). Full gene names can be found on appendice A.

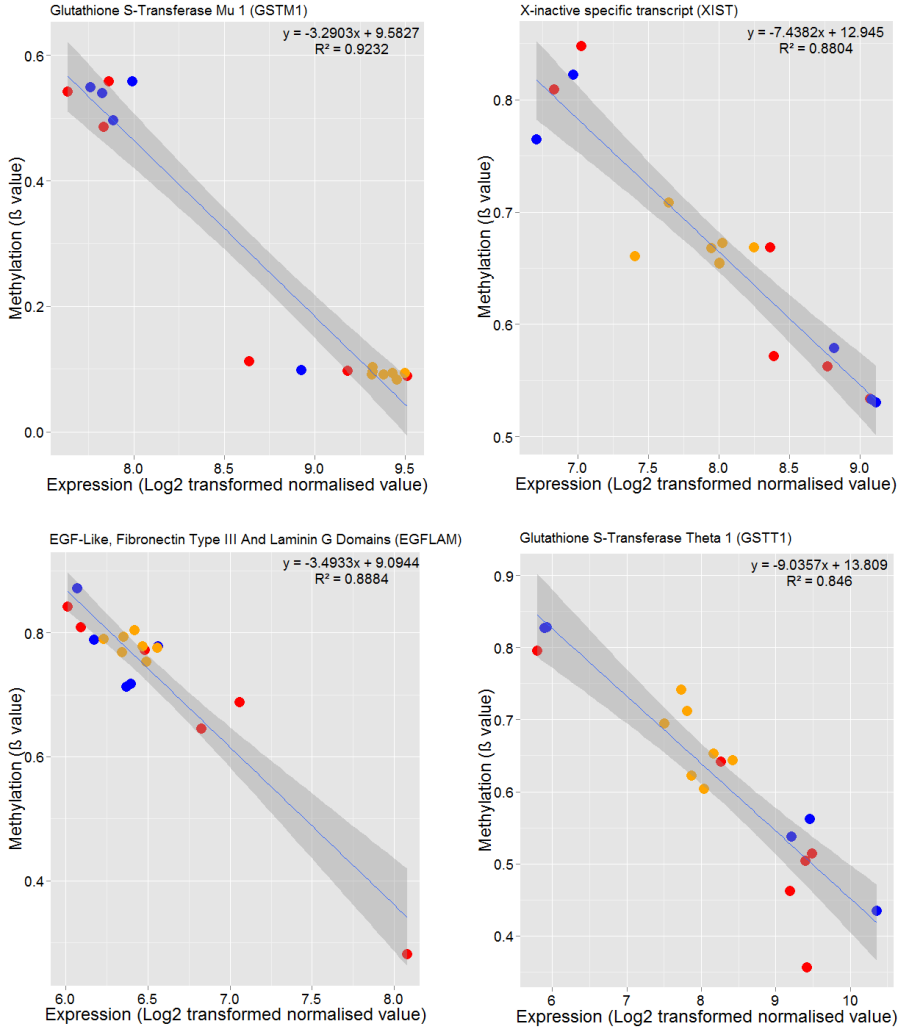


Figure 3.14.3.1. Scatter plots showing methylation of 4 CpGs which correlated with their respective gene expression level. Example of CpGs which significantly correlated (R^2≥0.5, P<0.05) across all cell lines (n=18). The methylation level is shown by the beta-value (0= 0% methylated, 1= 100% methylated). The expression level is the log2 transformed normalised value. Shaded areas indicate 95% confidence regions. ● Control (n=6), ● IPF (n=5), ● SSc (n=7).

Whilst 724 genes had a significant correlation (R^2= ≥0.5; P<0.05) between CpG methylation and gene expression, there was no overall correlation between methylation and gene expression (R=-0.067). This was evident when plotting the beta-methylation values against the log2 transformed normalised expression values (**Figure 3.14.3.2**) and suggests that different levels of methylation at specific CpGs may be important in determining the level of expression rather than a 'one rule that fits all'.

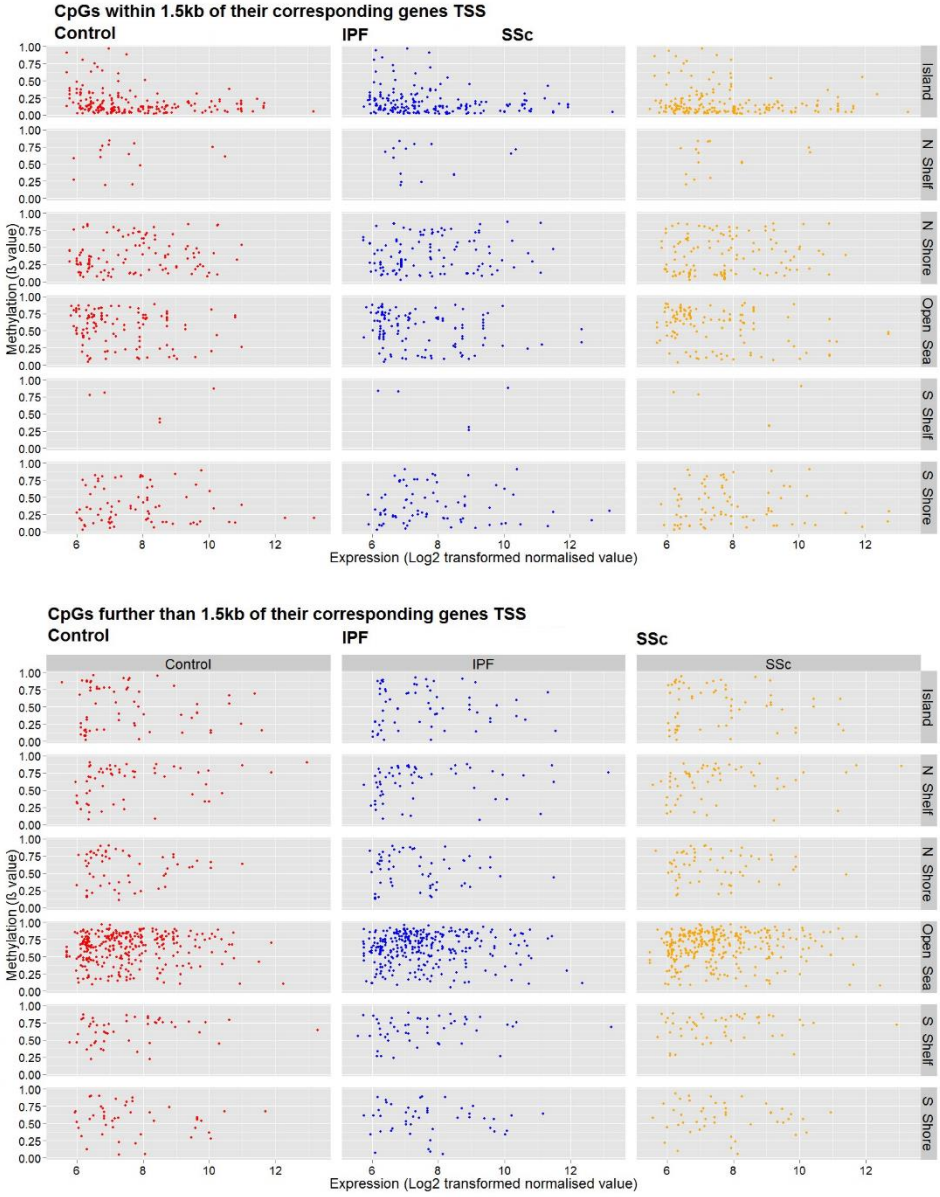


Figure 3.14.3.2. The location of CpGs which correlated with their corresponding genes expression in lung fibroblasts. Top: CpGs within 1.5kb of their corresponding genes TSS, bottom: CpGs further than 1.5kb from their corresponding genes TSS. Each dot represents a different CpG which significantly correlated (R2≥0.5; P<0.05) with gene expression in control (red; n=6), IPF (blue; n=6) and SSc (orange; n=7).

3.15. Differentially-methylated and expressed genes in IPF and SSc compared to control fibroblasts in which CpG methylation correlated with gene expression

From 1088 CpGs (724) genes which had methylation which correlated with gene expression, 100 CpGs corresponding to 85 genes had significantly (P<0.05) altered methylation and 41 genes had significantly (P<0.05) altered expression in IPF compared to control lung fibroblasts. Of these genes, 20 had both significantly altered methylation and expression in IPF compared to control lung fibroblasts (**Table 3.15.1, Table 3.15.2 and Figure 3.15.1**).

CvIPF	Expression (P<0.05)	Methylation (P<0.05)	Expression and methylation (P<0.05)
CpGs	N/A	75	25
Genes	21	65	20

Table 3.15.1. Genes/CpGs which correlated and had altered expression/methylation in IPF compared to control lung fibroblasts. Expression which significantly correlated with basal expression in all cell lines (n=18) where significant difference in expression and/or methylation were present between IPF (n=5) compared to control (n=6) lung fibroblasts.

Gene symbol	Gene name
ACSS2	Acyl-coa synthetase short-chain family member 2
ALDH4A1	Aldehyde dehydrogenase 4 family, member A1
ATP5EP2	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit pseudogene 2
C1QTNF9B	C1q and tumor necrosis factor related protein 9B
DGAT1	Diacylglycerol O-acyltransferase 1
FASTK	Fas-activated serine/threonine kinase
FBLN2	Fibulin 2
FOXP1	Forkhead box P1
GPC1	Glypican 1
IGF2BP3	Insulin-like growth factor 2 mRNA binding protein 3
ISLR	Immunoglobulin superfamily containing leucine-rich repeat
KLHL21	Kelch-like 21
LPCAT2	Lysophosphatidylcholine acyltransferase 2
MAPK8	Mitogen-activated protein kinase 8
PDE9A	Phosphodiesterase 9A
PLA2G5	Phospholipase A2, group V
PMP22	Peripheral myelin protein 22
PRKD1	Protein kinase D1
RAPGEF4	Rap guanine nucleotide exchange factor (GEF) 4
SESN3	Sestrin 3

Table 3.15.2. List of gene names which correlated and had altered expression/methylation in IPF compared to control lung fibroblasts. Genes which had expression which significantly correlated with methylation in all cell lines (n=18), which also had significant differences in expression and/or methylation in IPF (n=5) compared to control (n=6) lung fibroblasts.

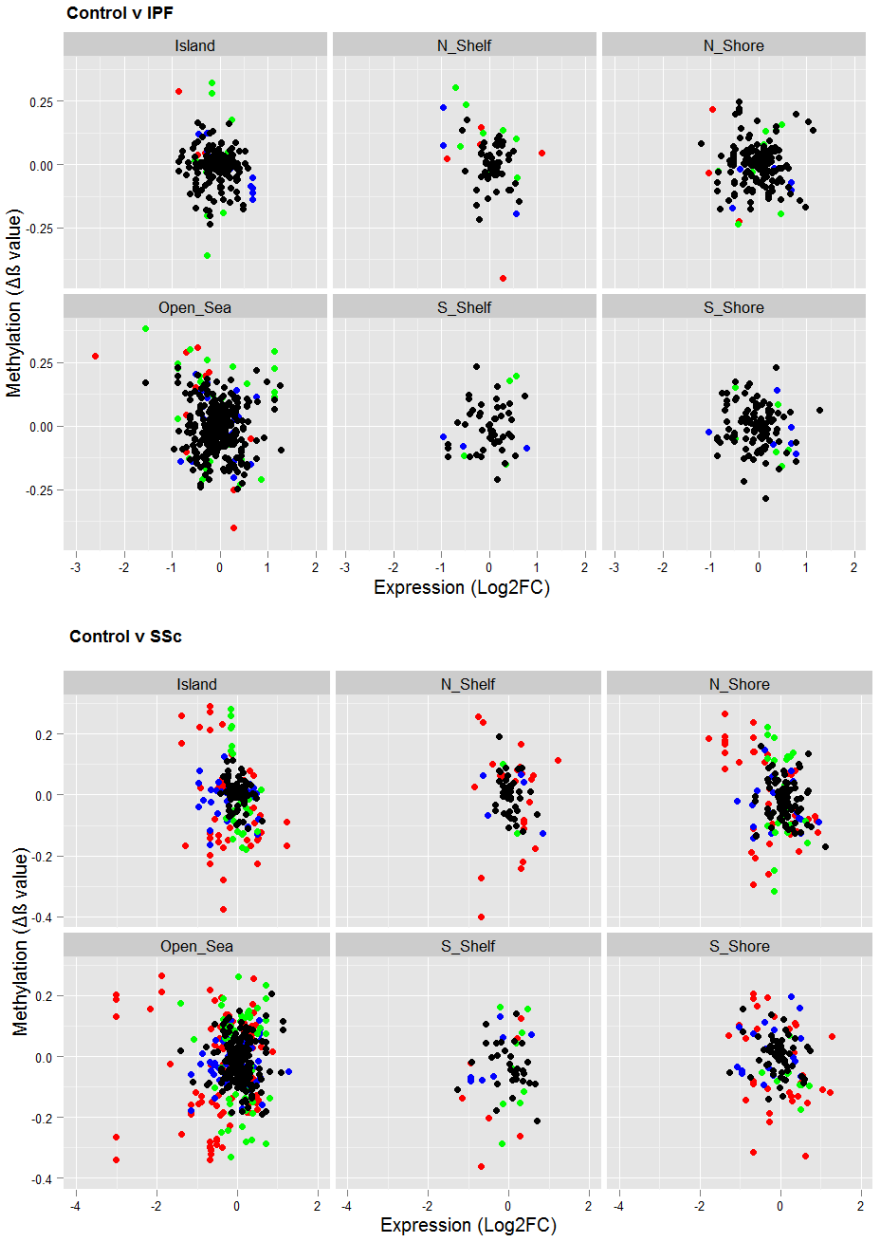


Figure 3.15.1. The number and location of the CpGs which had correlation between CpG methylation and expression which also had differences in methylation and/or expression in IPF/SSc compared to control lung fibroblasts. Top: IPF (n=5) compared to control (n=6) lung fibroblasts. Bottom: SSc (n=7) compared to control (n=6) lung fibroblasts. Delta-beta value: difference in methylation between IPF and control or SSc and control lung fibroblasts. Log2FC: the log2 fold-change in expression between IPF/SSc compared to control lung fibroblasts. Each dot represents a different CpG and its corresponding gene. Different coloured dots represent the following in IPF (top) or SSc (bottom) compared to controls: ● CpGs corresponding to genes with significantly altered methylation and expression ● CpGs corresponding to genes with significantly altered expression only ● CpGs corresponding to genes with significantly altered methylation only ● CpGs corresponding to genes which had no significant change in methylation or expression.

More genes in SSc reached the statistical threshold (P<0.05) likely due to male/female differences in IPF samples (see **Chapter 5**). Multiple (397) CpGs corresponding to 263 genes had significantly (P<0.05) altered methylation and 210 genes had significantly (P<0.05) altered expression in SSc compared to control lung fibroblasts. Of these genes, 154 had both significantly altered methylation and expression in SSc compared to control lung fibroblasts (**Figure 3.15.1, Table 3.15.3 and Table 3.15.4**).

CvSSc	Expression (P<0.05)	Methylation (P<0.05)	Expression and methylation (P<0.05)
CpGs	-	137	260
Genes	56	109	154

Table 3.15.3. Genes/CpGs which correlated and had altered expression/methylation in SSc compared to control lung fibroblasts. Basal expression which significantly correlated with basal expression in all cell lines (n=18) where significant differences in expression and/or methylation were present between SSc (n=7) compared to control (n=6) lung fibroblasts.

Gene symbol	Gene name
ACACB	Acetyl-coa carboxylase beta
ACSS2	Acyl-coa synthetase short-chain family member 2
ACVRL1	Activin A receptor type II-like 1
ADAMTS2	ADAM metallopeptidase with thrombospondin type 1 motif, 2
ADAMTSL1	ADAMTS-like 1
AFF3	AF4/FMR2 family, member 3
AMPH	Amphiphysin
ANO1	Anoctamin 1, calcium activated chloride channel
ANTXR2	Anthrax toxin receptor 2
ATP1B1	ATPase, Na+/K+ transporting, beta 1 polypeptide
ATP5EP2	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit pseudogene 2
AUTS2	Autism susceptibility candidate 2
C13orf15	Chromosome 13 open reading frame 15
C13orf16	Chromosome 13 open reading frame 16
CACNA1A	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit
CCL11	Chemokine (C-C motif) ligand 11
CD248	CD248 molecule, endosialin
CD47	CD47 molecule
CD9	CD9 molecule
CHD1L	Chromodomain helicase DNA binding protein 1-like
CHERP	Calcium homeostasis endoplasmic reticulum protein
CHRNA1	Cholinergic receptor, nicotinic, alpha 1
CHST15	Carbohydrate (N-acetylgalactosamine 4-sulfate 6-O) sulfotransferase 15
CLEC14A	C-type lectin domain family 14, member A
COL9A2	Collagen, type IX, alpha 2
CPNE8	Copine VIII
CPXM2	Carboxypeptidase X (M14 family), member 2
CRLF1	Cytokine receptor-like factor 1
CTNNA1	Catenin (cadherin-associated protein), alpha 1
CTNNB1	Catenin (cadherin-associated protein), beta 1
CYFIP2	Cytoplasmic FMR1 interacting protein 2
CYTH3	Cytohesin 3
DAB2	Disabled homolog 2, mitogen-responsive phosphoprotein
DCTD	Dcmp deaminase
DGAT1	Diacylglycerol O-acyltransferase 1
DOCK2	Dedicator of cytokinesis 2
DRG2	Developmentally regulated GTP binding protein 2
EFCAB4A	EF-hand calcium binding domain 4A

EIF1AY	Eukaryotic translation initiation factor 1A, Y-linked
EPHB2	EPH receptor B2
ERH	Enhancer of rudimentary homolog
F3	Coagulation factor III (thromboplastin, tissue factor)
FAM105A	Family with sequence similarity 105, member A
FBLN2	Fibulin 2
FNBP1L	Formin binding protein 1-like
FOXP1	Forkhead box P1
FST	Follistatin
GFRA1	GDNF family receptor alpha 1
GLDN	Gliomedin
GPC1	Glypican 1
GPER	G protein-coupled estrogen receptor 1
GPNMB	Glycoprotein (transmembrane) nmb
GPR56	G protein-coupled receptor 56
GSTM1	Glutathione S-transferase mu 1
GTPBP4	GTP binding protein 4
H19	H19, imprinted maternally expressed transcript (non-protein coding)
HIATL1	Hippocampus abundant transcript-like 1
HLA-DMA	Major histocompatibility complex, class II, DM alpha
HOXB6	Homeobox B6
HOXC4	Homeobox C4
ICAM2	Intercellular adhesion molecule 2
ISLR	Immunoglobulin superfamily containing leucine-rich repeat
JUP	Junction plakoglobin
KCNMA1	Potassium large conductance calcium-activated channel, subfamily M, alpha 1
KLHL21	Kelch-like 21
LPCAT2	Lysophosphatidylcholine acyltransferase 2
LRIG1	Leucine-rich repeats and immunoglobulin-like domains 1
LRRN4CL	LRRN4 C-terminal like
MAGEC2	Melanoma antigen family C, 2
MAPK8	Mitogen-activated protein kinase 8
MAPKAP1	Mitogen-activated protein kinase associated protein 1
MGMT	O-6-methylguanine-DNA methyltransferase
MRI1	Methylthioribose-1-phosphate isomerase homolog
MRPS6	Mitochondrial ribosomal protein S6
MSX1	Msh homeobox 1
MYOM2	Myomesin (M-protein) 2
NAPRT1	Nicotinate phosphoribosyltransferase domain containing 1
NCAM2	Neural cell adhesion molecule 2
NEDD9	Neural precursor cell expressed, developmentally down-regulated 9
NETO2	Neuropilin (NRP) and tolloid (TLL)-like 2
NFIB	Nuclear factor I/B
NLGN4Y	Neuroligin 4, Y-linked
NPTX1	Neuronal pentraxin I
NTN1	Netrin 1
NUMA1	Nuclear mitotic apparatus protein 1
OLFM1	Olfactomedin 1
OLFML1	Olfactomedin-like 1
OPLAH	5-oxoprolinase (atp-hydrolysing
OSBPL10	Oxysterol binding protein-like 10
PAGE2B	P antigen family, member 2B
PAGE5	P antigen family, member 5
PCBP3	Poly(rc) binding protein 3
PCF11	PCF11, cleavage and polyadenylation factor subunit, homolog
PGF	Placental growth factor
PGM3	Phosphoglucomutase 3
PLA2G5	Phospholipase A2, group V
PLAC9	Placenta-specific 9
PLAG1	Pleiomorphic adenoma gene 1
PLAGL1	Pleiomorphic adenoma gene-like 1
PLXNB1	Plexin B1
PMP22	Peripheral myelin protein 22

PPP2CB	Protein phosphatase 2, catalytic subunit, beta isozyme
PPP2R3A	Protein phosphatase 2, regulatory subunit B, alpha
PPP3CA	Protein phosphatase 3, catalytic subunit, alpha isozyme
PRDM8	PR domain containing 8
PRKCZ	Protein kinase C, zeta
PRKY	Protein kinase, y-linked, pseudogene
PTPRH	Protein tyrosine phosphatase, receptor type, H
PUM1	Pumilio homolog 1
RAPGEF4	Rap guanine nucleotide exchange factor (GEF) 4
REEP3	Receptor accessory protein 3
RPL29	Ribosomal protein L29
RPL34	Ribosomal protein L34
RPS4Y1	Ribosomal protein S4, Y-linked 1
RPS4Y2	Ribosomal protein S4, Y-linked 2
S100A4	S100 calcium binding protein A4
SARS	Seryl-trna synthetase
SASH1	SAM and SH3 domain containing 1
SATB2	SATB homeobox 2
SCARA3	Scavenger receptor class A, member 3
SESN2	Sestrin 2
SESN3	Sestrin 3
SGCE	Sarcoglycan, epsilon
SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1
SHANK2	SH3 and multiple ankyrin repeat domains 2
SHANK3	SH3 and multiple ankyrin repeat domains 3
SHISA3	Shisa homolog 3
SP2	Sp2 transcription factor
SPATA18	Spermatogenesis associated 18 homolog
ST3GAL2	ST3 beta-galactoside alpha-2,3-sialyltransferase 2
ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1
STK32B	Serine/threonine kinase 32B
SYTL2	Synaptotagmin-like 2 [Source:HGNC Symbol;Acc:15585]
TECR	Trans-2,3-enoyl-coa reductase
TGFB2	Transforming growth factor, beta 2
THBS1	Thrombospondin 1
THSD4	Thrombospondin, type I, domain containing 4
TIPARP	TCDD-inducible poly(ADP-ribose) polymerase
TMED1	Transmembrane emp24 protein transport domain containing 1
TMEM26	Transmembrane protein 26
TMTC1	Transmembrane and tetratricopeptide repeat containing 1
TPM1	Tropomyosin 1 (alpha)
TRAF3IP2	TRAF3 interacting protein 2
TRIM56	Tripartite motif containing 56
TSC22D1	TSC22 domain family, member 1
TSC22D2	TSC22 domain family, member 2
UBIAD1	Ubia prenyltransferase domain containing 1
VGLL4	Vestigial like 4
VPS11	Vacuolar protein sorting 11 homolog
WDR25	WD repeat domain 25
XRCC6BP1	XRCC6 binding protein 1
YPEL1	Yippee-like 1
ZFHX4	Zinc finger homeobox 4
ZSCAN16	Zinc finger and SCAN domain containing 16

Table 3.15.4. List of gene names which correlated and had altered expression/ methylation in SSc compared to control lung fibroblasts. Gene names of genes which had basal expression which significantly correlated with basal expression in all cell lines (n=18) where significant differences in expression and/or methylation were present between SSc (n=7) compared to control (n=6) lung fibroblasts.

Cluster analysis of the 20 genes which had significantly altered methylation which correlated with gene expression in IPF compared to control lung fibroblasts, identified heterogeneity between IPF and control samples. Methylation of control 6 clustered with IPF samples. (**Figure 3.15.2**). In contrast, cluster analysis of the 154 genes which had CpG methylation that correlated with gene expression and had significantly altered methylation and expression in SSc compared to control lung fibroblasts identified less heterogeneity between SSc samples, although control 6 also clustered with SSc lung fibroblasts (**Figure 3.15.3**).

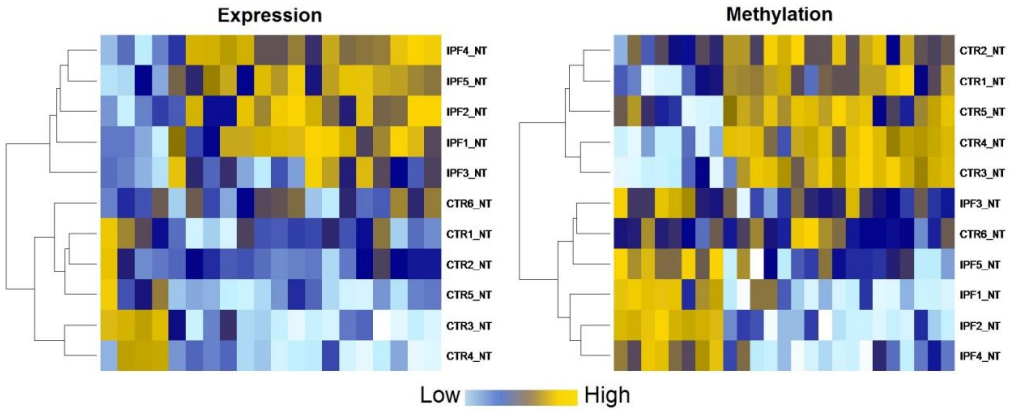


Figure 3.15.2. Hierarchical cluster analysis of genes with altered methylation and expression in IPF compared to control lung fibroblasts. Heat maps show CpGs which had significantly altered (P<0.05) methylation in IPF (n=5) compared to control (n=6) lung fibroblasts and the genes which they corresponded to which also had significantly altered expression in IPF (n=5) compared to control (n=6) lung fibroblasts. Light blue represents low expression/methylation, yellow represents high expression/methylation with respect to each gene across all cell lines.

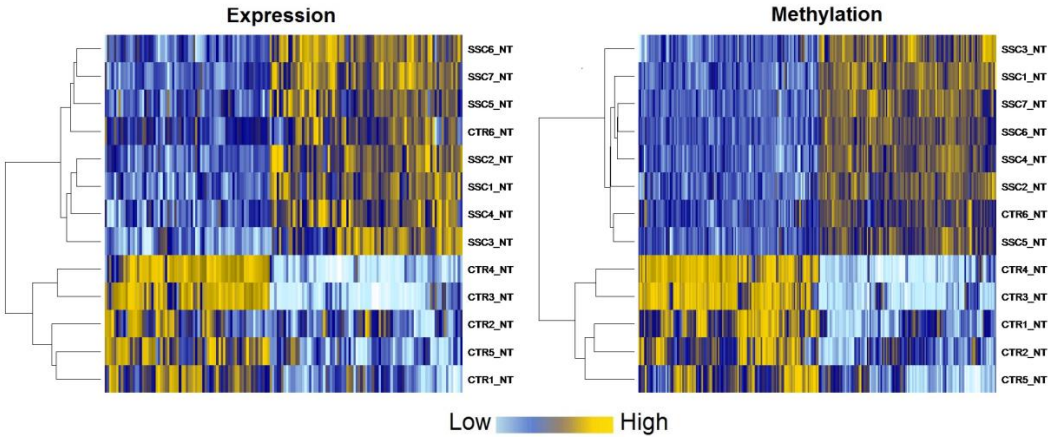


Figure 3.15.3. Hierarchical cluster analysis of genes with altered methylation and expression in SSc compared to control lung fibroblasts. Heat maps show the methylation of CpGs which had significantly altered (P<0.05) methylation in IPF (n=5) compared to control (n=6) lung fibroblasts and the genes which they corresponded to which also had significantly altered expression in IPF (n=5) compared to control (n=6) lung fibroblasts. Light blue represents low expression/methylation, yellow represents high expression/methylation with respect to each gene across all cell lines.

3.16. Summary

- Methylation of multiple CpGs strongly correlated ($R^2 \geq 0.5$; $P < 0.05$) with their corresponding genes expression in human lung fibroblasts, identifying multiple novel genes potentially regulated directly by methylation.

- No overall positive or inverse relationship between methylation and expression was found, suggesting that different levels of methylation affect genes to different extents.

- Multiple genes had both negative and positive correlation of CpG methylation with gene expression, suggesting that some genes may have multiple CpGs which can directly regulate expression.

- 20 genes in IPF and 154 genes in SSc-PF which had significant ($R^2 \geq 0.5$; $P < 0.05$) correlation between methylation and expression in all cell lines, had significantly altered methylation ($P < 0.05$) and expression ($P < 0.05$) compared to control lung fibroblasts, suggesting multiple novel genes involved in PF are potentially regulated directly by methylation in lung fibroblasts.

- The number of genes identified as having CpG methylation correlating with their expression is likely to be largely underestimated due to probes which were not detected by the expression array as a potential result of high/low methylation levels completely silencing gene expression, other epigenetic mechanisms masking DNA methylation and CpGs not interrogated by the array. Further studies using bisulfite sequening/arrays which interrogate more CpGs are required to more accurately determine methylation and expression correlation.

Chapter 4. Network and functional analysis of genes with altered methylation and expression in IPF and SSc compared to control lung fibroblasts.

4.1. Overview

To determine which biological processes may be important and relevant in IPF and/or SSc, two main types of enrichment analyses were conducted; protein family (PFAM) domain-centric and gene-GO term enrichment analysis. PFAMs describe proteins grouped into families which are represented by multiple sequence alignments and hidden Markov models (HMMs), which are used to measure probability distributions over multiple observations (Finn et al, 2014). Protein domains are functional regions and genes which share common domains can be easily identified and mapped to their respective PFAM using the Ensembl genome database (**see Chapter 2: section 2.7.2**). PFAM domain-centric analysis uses pre-defined protein-level GO annotations to determine if a specific biological process is enriched in a given list of PFAMs. It does this by determining whether there is a significantly greater number of PFAMs observed for a specific biological process compared to what would be expected by chance. Thus, the genes mapped to PFAMs which are associated with a specific biological process may not themselves be enriched in the biological process but the domains they share are. This allows the potential to identify genes which share domains but have not yet been associated with the specific biological process.

Gene-GO term enrichment analysis compares genes from a given list to pre-defined gene lists associated with a specific biological process. Thus, gene-GO term enrichment analysis identifies enriched biological processes by determining, from a given list of genes, whether there is a significantly greater number of genes observed for a specific biological process compared to what would be expected by chance. Combining data from these two types of enrichment analyses makes it possible to identify genes associated with a biological process enriched in PFAMs and a biological processes enriched in specific genes with altered methylation/expression in IPF and SSc compared to control lung fibroblasts.

PFAM enrichment was conducted using the dcGO: database of gene ontologies. Genes belonging to different PFAM groups which had altered methylation were analysed using Cytoscape 3.2.1 (**see Chapter 2 section 2.7.2**). The networks of genes linked to their respective PFAM domains generated by Cytoscape were too large to be viewed as a document, however, specific PFAMs and enriched biological processes could be visualised. The following chapter goes into detailed analysis of biological processes enriched in PFAMs containing genes with altered methylation and expression and biological processes enriched in genes with altered methylation and expression.

4.2. Biological processes enriched in PFAMs containing genes with altered methylation in IPF and SSc compared to control lung fibroblasts.

Genes which had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF and SSc compared to control lung fibroblasts mapped to 1328 and 1461 different PFAM domains, respectively. Multiple (633 and 611) biological processes were significantly (FDR $P<0.01$) enriched in these PFAMs in IPF and SSc compared to control lung fibroblasts, respectively. Of the 611 enriched biological processes in SSc, 523 (86%) overlapped with IPF. Many of these biological processes have previously been associated with the pathobiology of pulmonary fibrosis including EMT, ECM organisation, apoptosis and Wnt signalling (**Table 4.2.1**).

Control v IPF

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
ECM organization	2	4.36E-07	45	63
Wnt receptor signalling pathway	2	1.93E-06	30	38
Regulation of FGFR signalling pathway	3	1.27E-04	11	11
Integrin-mediated signalling pathway	3	9.61E-04	14	17
Tissue remodelling	3	1.36E-03	12	14
Response to hypoxia	3	2.57E-03	28	45
Induction of apoptosis by extracellular signals	3	3.56E-03	9	10
Collagen fibril organization	3	5.28E-03	10	12
Bile acid metabolic process	4	8.30E-03	6	6
Blood vessel remodelling	4	8.30E-03	6	6
EMT	3	8.78E-03	13	18
Regulation of JAK-STAT cascade	3	8.78E-03	13	18

Control v SSc

Extracellular matrix organization	2	2.41e-10	52	63
Immune response	2	1.41e-09	95	140
Wnt receptor signalling pathway	2	8.73E-09	34	38
Integrin-mediated signalling pathway	3	2.55e-06	17	17
Response to hypoxia	3	8.43e-05	33	45
Inflammatory response	3	1.42e-04	22	27
Regulation of FGFR signalling pathway	3	3.32e-04	11	11
Tissue remodelling	3	4.66e-04	13	14
JAK-STAT cascade	3	9.26E-04	12	13
Collagen fibril organization	3	1.84e-03	11	12
Induction of apoptosis by extracellular signals	3	6.86e-03	9	10

Table 4.2.1. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered methylation in IPF and SSc compared to control lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate ($P<0.01$). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

Multiple PFAMs overlapped in biological processes such as ECM organisation and collagen fibril organisation. ECM, classed as a general biological process, was associated with 63 different PFAMs, of which 45 and 52 PFAMs contained genes which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF and SSc compared to control lung fibroblasts, respectively. Collagen fibril organisation, classed as a specific biological process, was associated with 12 different PFAMs, all of which were associated with the broader gene ontology of ECM organisation. PFAMs (10 and 11) associated with collagen fibril organisation contained genes which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (**Figure 4.2.1**) and SSc (**Figure 4.2.2**) compared to control lung fibroblasts, respectively.

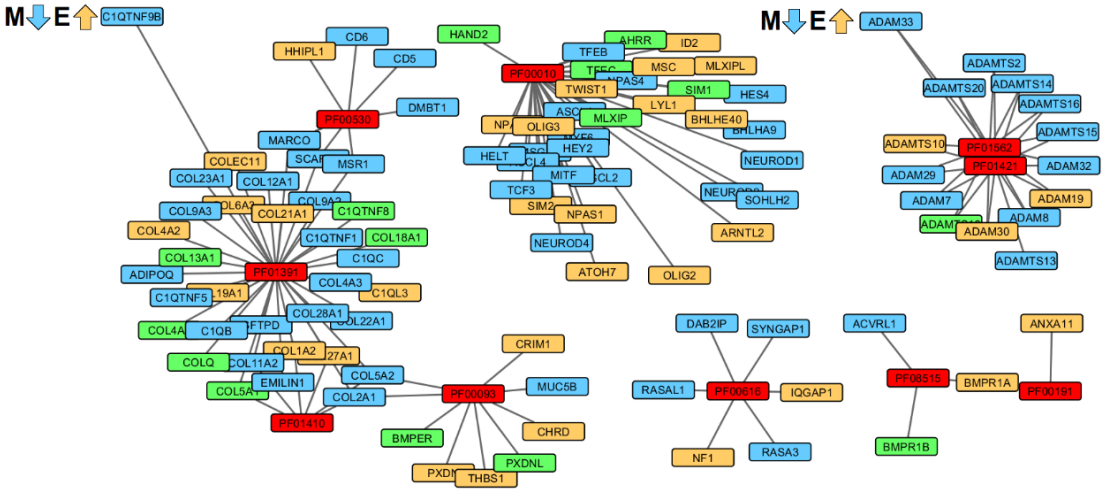


Figure 4.2.1. Genes belonging to PFAMs associated with collagen fibril organisation which have altered methylation in IPF compared to control lung fibroblasts. Genes belonging to PFAMs (red boxes) which had significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF (n=5) compared to control (n=6) lung fibroblasts. Coloured boxes represent direction of methylation, orange; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in IPF compared to control lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to control lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in IPF (n=5) relative to the average control lung fibroblast methylation/expression value.

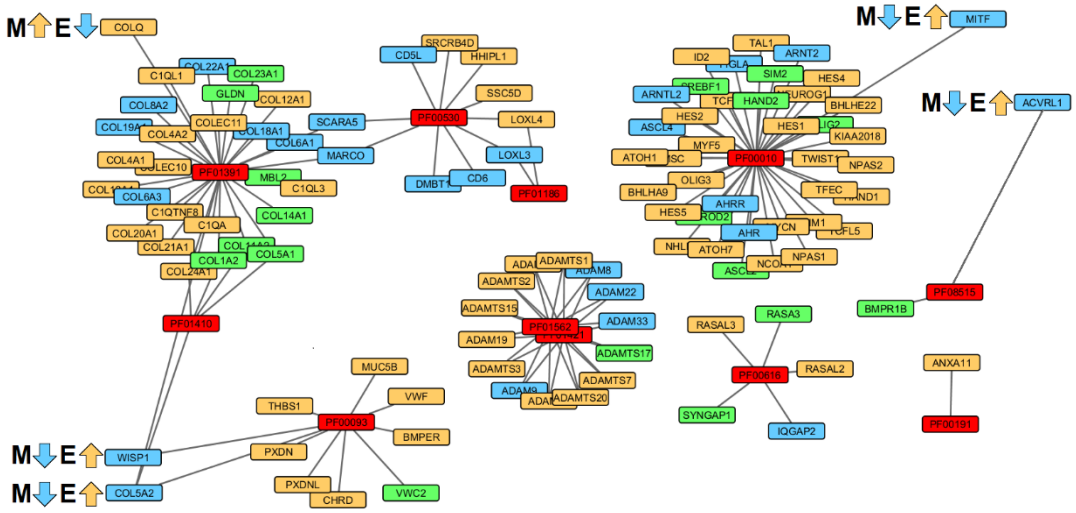


Figure 4.2.2. Genes belonging to PFAMs associated with collagen fibril organisation which have altered methylation in SSc compared to control lung fibroblasts. Genes belonging to PFAMs (red boxes) which had significantly altered methylation (Δβ ≥0.136; P<0.05) in SSc (n=7) compared to control (n=6) lung fibroblasts. Coloured boxes represent direction of methylation, orange; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in SSc compared to control lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in SSc compared to control lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in SSc (n=7) relative to the average control lung fibroblast methylation/expression value.

Gastroesophageal reflux has previously been associated with IPF and SSc (**Chapter 1 section 1.6.3**). 6 PFAMs were associated with the bile acid metabolic process, (labelled as a highly specific process), all of which contained genes which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to control lung fibroblasts. PF00067 describes proteins belonging to the cytochrome P450 family (CYPs) which are major enzymes involved in drug metabolism and important for steroid (Nebert and Russell, 2002), arachidonic acid (Rifkind et al, 1995) and bile acid metabolism (Norlin and Wikvall, 2007, Chiang, 2009). PF00104 describes proteins with a ligand-binding domain of nuclear hormone receptors. Nuclear receptors can directly bind DNA and regulate gene expression (Aranda and Pascual, 2001), some of which can also regulate cytochrome P450 enzymes and function as metabolic sensors. Dysregulation of genes belonging to these PFAMs may therefore play an important role in IPF (**Figure 4.2.3**). Although bile acid metabolic process did not reach the statistical significant cut off of P<0.01 FDR in SSc compared to control lung fibroblasts, all 6 PFAMs had genes belonging to them which had altered methylation in SSc compared to control lung fibroblasts (**Figure 4.2.4**).

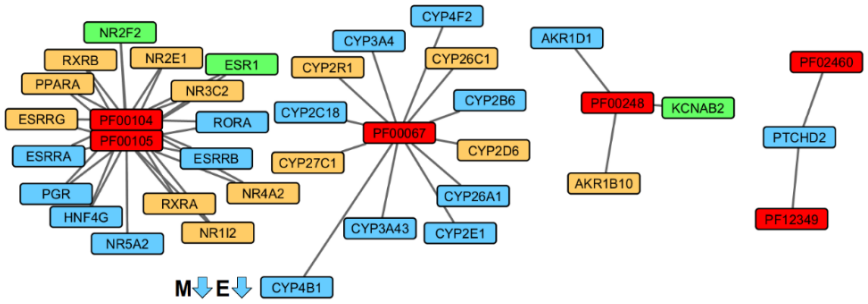


Figure 4.2.3. Genes belonging to PFAMs associated with bile acid metabolic process which have altered methylation in IPF compared to control lung fibroblasts. Genes belonging to PFAMs (red boxes) which had significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF (n=5) compared to control (n=6) lung fibroblasts. Coloured boxes represent direction of methylation, orange; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in IPF compared to control lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to control lung fibroblasts. M= methylation, E= expression. Blue = decreased relative to the average control lung fibroblast methylation/expression value.

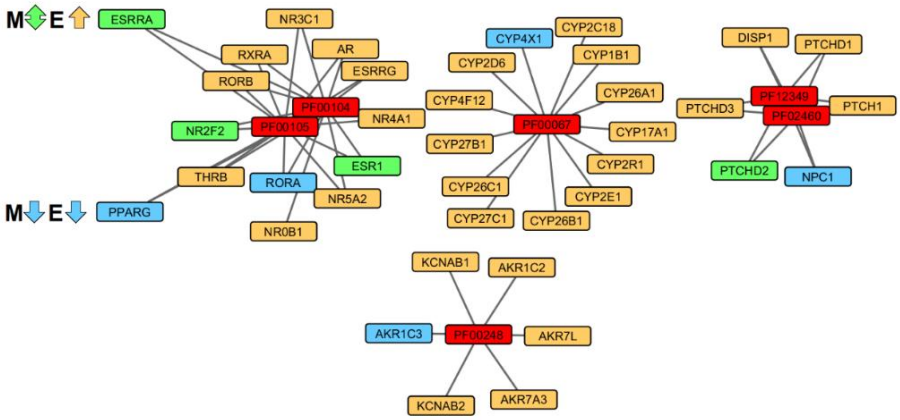


Figure 4.2.4. Genes belonging to PFAMs associated with bile acid metabolic process which have altered methylation in SSc compared to control lung fibroblasts. Genes belonging to PFAMs (red boxes) which had significantly altered methylation (Δβ ≥0.136; P<0.05) in SSc (n=7) compared to control (n=6) lung fibroblasts. Coloured boxes represent direction of methylation, orange; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in SSc compared to control lung fibroblasts. Boxes annotated with bar graphs indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in SSc compared to control lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in SSc compared to control lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in SSc (n=7) relative to the average control lung fibroblast methylation/expression value.

Multiple PFAMs such as PF00096 were identified containing a large number of genes with significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF and SSc compared to control lung fibroblasts (**Figure 4.2.5**) PF00096 describes proteins which contain a C2H2 zinc-finger domain. Zinc-finger proteins play important roles in multiple biological processes including EMT (Savagner et al, 1997, Nieto, 2002, Lamouille et al, 2014). Multiple (91 and 88) significantly (P<0.01 FDR) enriched biological processes contained PF00096 in IPF and SSc compared to control lung fibroblasts, respectively. Many of these biological processes were broadly associated with cell changes such as cell morphogenesis, cell fate, mesenchyme development and pathways such as notch signalling and serine/threonine signalling which play important roles in regulating EMT (Larue and Bellacosa, 2005, Wang and Zhou, 2011).

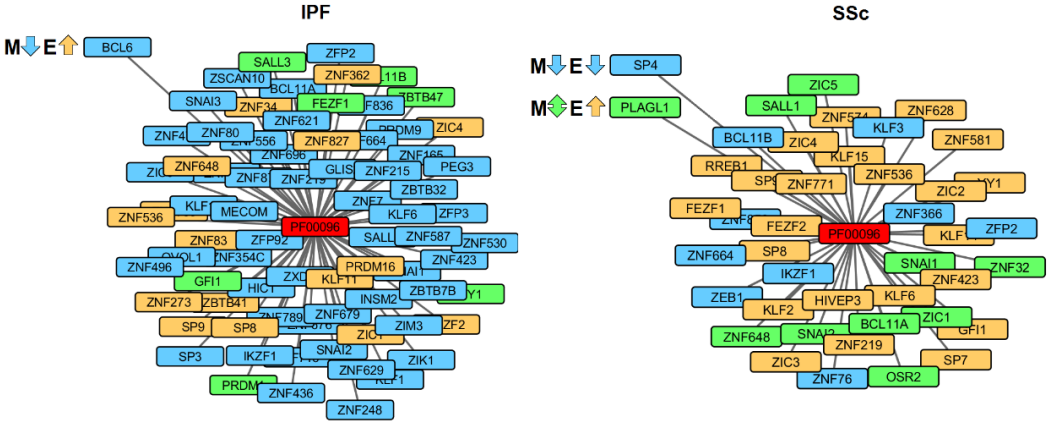


Figure 4.2.5. Zinc-finger C2H2 family (PF00096) containing multiple genes with altered methylation in IPF and SSc compared to control lung fibroblasts. Genes belonging to PFAMs (red boxes) which had significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF (n=5) and/or SSc (n=7) compared to control (n=6) lung fibroblasts. Coloured boxes represent direction of methylation, orange; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in IPF/SSc compared to control lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in IPF/SSc compared to control lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in IPF (n=5) or SSc (n=7) relative to the average control lung fibroblast methylation/expression value.

4.3. Biological processes enriched in PFAMs containing genes with altered expression in IPF and SSc compared to control lung fibroblasts.

Genes which had significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to control and SSc compared to control lung fibroblasts mapped to 419 and 527 different PFAM domains respectively. Multiple (676 and 767) biological processes were significantly (FDR P<0.01) enriched in these PFAMs in IPF and SSc compared to control lung fibroblast respectively. Of the 676 enriched biological processes in IPF, 534 (79%) overlapped with SSc. These included many biological processes previously implicated in IPF and/or SSc such as JAK-STAT signalling, ECM organisation, apoptosis,

interleukin-4 signalling, lung alveolus development and Wnt signalling (**Table 4.3.1**). The biological processes enriched in PFAMs which were shared between IPF and SSc often involved different genes belonging to the same PFAMs. This may indicate that similar processes are affected by different genes with altered methylation and/or expression in IPF and SSc. This is exemplified in the lung alveolus development process enriched in PFAMs where only 5 genes overlapped between IPF and SSc which had significantly altered expression compared to control lung fibroblasts (**Figure 4.3.1**). This observation could potentially have wide-ranging benefits in IPF and SSc such as acting as a method for disease phenotyping and a method for the identification of shared and distinct biomarkers.

Control v IPF

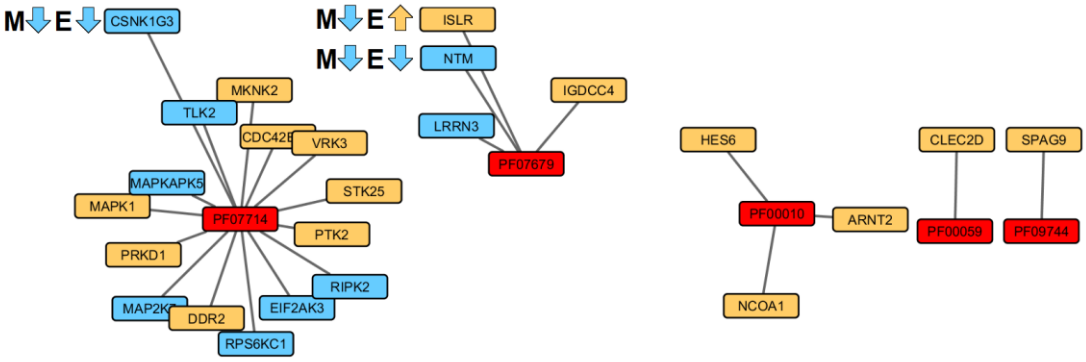
Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
Wnt receptor signalling pathway	2	2.33E-04	14	38
Regulation of tissue remodelling	3	2.31e-03	4	5
Positive regulation of tyrosine phosphorylation of Stat3 protein	4	2.31E-03	4	5
Positive regulation of epithelial cell proliferation involved in lung morphogenesis	4	3.73e-03	3	3
Interleukin-4 production	4	3.73e-03	3	3
Lung alveolus development	3	4.26e-03	5	9
ECM organization	2	4.64e-03	16	63
Signal transduction involved in regulation of gene expression	4	5.04e-03	4	6
Bile acid metabolic process	4	5.04e-03	4	6
Inflammatory response	3	5.84e-03	9	27
Response to FGFR stimulus	3	6.88e-03	5	10
Induction of apoptosis by extracellular signals	3	6.88e-03	5	10
Chromatin silencing	3	9.28e-03	8	24
Histone phosphorylation	3	9.55e-03	4	7
RNA methylation	3	9.75E-03	6	15

Control v SSc

ECM organization	2	5.37e-12	34	63
Wnt receptor signalling pathway	2	1.78E-04	16	38
ECM disassembly	3	2.73e-04	5	5
Induction of apoptosis by extracellular signals	3	4.55e-04	7	10
Histone phosphorylation	3	2.88e-03	5	7
JAK-STAT cascade	3	2.91E-03	7	13
Positive regulation of anti-apoptosis	3	3.09e-03	6	10
Chromatin silencing	3	3.19e-03	10	24
Positive regulation of interleukin-4 production	4	4.75e-03	4	5
Positive regulation of collagen biosynthetic process	4	6.33e-03	3	3
Positive regulation of epithelial cell proliferation involved in lung morphogenesis	4	6.33e-03	3	3
TGFß receptor complex assembly	4	6.33e-03	3	3
NK T cell differentiation	4	6.33E-03	3	3
Interleukin-4 production	4	6.33e-03	3	3
Lung alveolus development	3	9.72e-03	5	9

Table 4.3.1. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered expression in IPF and SSc compared to control lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate (P<0.01). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

PFAMs containing genes with significantly altered expression (TNoM ≤1; P<0.05) in IPF compared with control lung fibroblasts



PFAMs containing genes with significantly altered expression (TNoM ≤1; P<0.05) in SSc compared with control lung fibroblasts

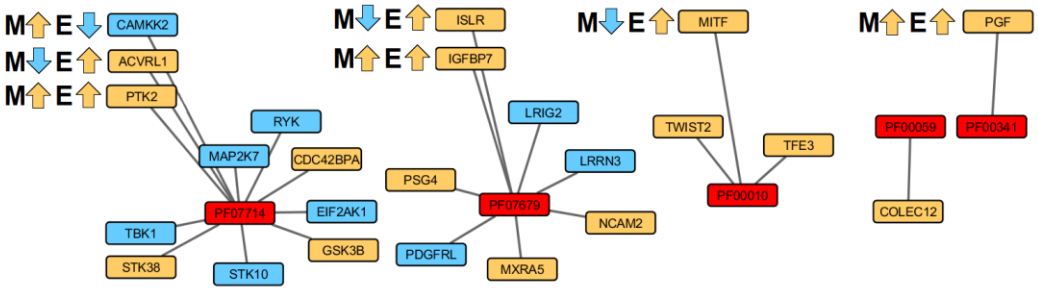


Figure 4.3.1. Genes belonging to PFAMs associated with lung alveolar development which have altered expression in IPF/SSc compared to control lung fibroblasts. Genes which had significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) and/or SSc (n=7) compared to control (n=6) lung fibroblasts. Colour boxes represent expression direction, orange; genes with increased expression, blue; genes with decreased expression in IPF/SSc compared to control lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in IPF/SSc compared to control lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in IPF (n=5) or SSc (n=7) relative to the average control lung fibroblast methylation/expression value.

Multiple biological processes significantly (FDR P<0.05) enriched in PFAMs containing genes with significantly altered expression (TNoM ≤1; P<0.05) in IPF and SSc were the same biological processes enriched in PFAMs containing genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF and SSc compared to control lung fibroblasts. This adds further evidence that aberrantly methylated/expressed genes in IPF/SSc are associated with common and specific PFAMs and suggests a strong link between methylation and expression. Of the 676 biological processes enriched in PFAMs containing genes with altered expression, 445 (66%) were the same PFAMs which contained genes with altered methylation in IPF compared to control lung fibroblasts. Similarly, of the 767 biological

processes enriched in PFAMs containing genes with altered expression, 481 (63%) were the same biological processes enriched in PFAMs containing genes with altered methylation in SSc compared to control lung fibroblasts. Furthermore, PFAMs, such as PF00096, which contained multiple genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF and SSc compared to control lung fibroblasts (**Section 4.2**, **Figure 4.2.5**) also had multiple genes with significantly altered expression (**Figure 4.3.2**), showing that multiple PFAMs contain genes with both altered methylation and expression.

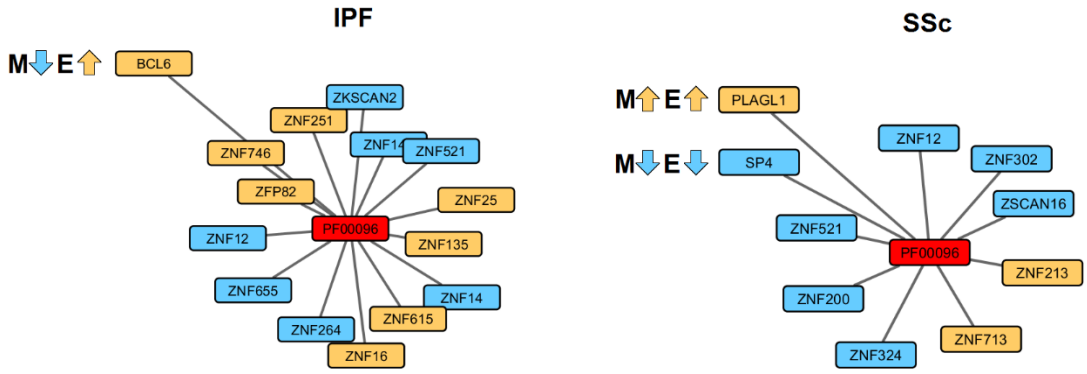


Figure 4.3.2. Genes associated with the C2H2 zinc finger domain (PF00096) which have altered expression in IPF and SSc compared to control lung fibroblasts. Genes which had significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) and/or SSc (n=7) compared to control (n=6) lung fibroblasts. Colour boxes represent expression direction, orange; genes with increased expression, Blue; genes with decreased expression in IPF/SSc compared to control lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation (Δβ ≥0.136; P<0.05) and significantly altered expression (TNoM ≤1; P<0.05) in IPF/SSc compared to control lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in IPF (n=5) or SSc (n=7) relative to the average control lung fibroblast methylation/expression value.

4.4. Biological processes enriched in PFAMs containing genes with altered methylation in IPF compared to SSc lung fibroblasts.

Genes which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts mapped to 1468 different PFAM domains. Multiple (601) biological processes were enriched in these PFAMs, including Wnt signalling, ECM organisation, epithelial cell differentiation and induction of apoptosis by extracellular signals (**Table 4.4.1**). Of these, 488 (81%) biological processes were the same as those which were enriched in PFAMs containing genes with altered methylation in IPF and SSc compared to control lung fibroblasts (**Figure 4.4.1**). The identification of these processes adds further evidence that multiple pathways overlap between IPF and SSc which may involve both common and distinct genes in each disease. Furthermore, specific PFAMs associated with multiple

biological processes which may be relevant to pulmonary fibrosis but were not significantly enriched, including collagen fibril organisation and bile acid metabolic process, also contained multiple genes which had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF compared to SSc lung fibroblasts (**Figure 4.4.2**).

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
Epithelial cell differentiation	2	6.05e-12	58	69
ECM organization	2	2.02e-09	51	63
Wnt receptor signalling pathway	2	6.81E-07	32	38
Response to hypoxia	3	1.27e-06	36	45
Regulation of histone modification	3	1.57e-06	23	25
Chromatin remodelling	3	3.01e-04	24	31
Trachea development	3	3.53e-04	11	11
Tissue remodelling	3	5.03e-04	13	14
Inflammatory response	3	7.17e-04	21	27
Patterning of blood vessels	3	1.64e-03	9	9
Cell aging	3	3.82e-03	15	19
Regulation of FGFR signalling pathway	3	3.82e-03	10	11
Response to virus	2	5.56e-03	29	45
Regulation of respiratory gaseous exchange	3	6.90e-03	7	7
Induction of apoptosis by extracellular signals	3	7.25e-03	9	10

Table 4.4.1. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered methylation in IPF compared to SSc lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate ($P<0.01$). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

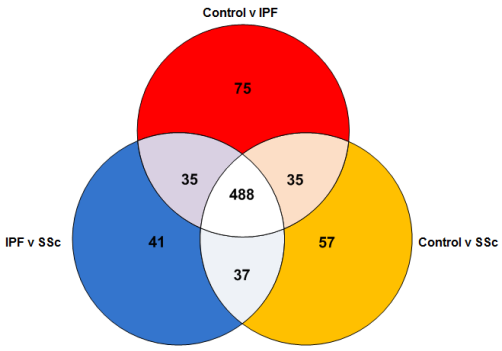


Figure 4.4.1. Number of distinct and overlapping biological processes enriched in PFAMs containing genes with altered methylation in IPF and SSc compared to each other and compared to control lung fibroblasts. Biological processes (488) were identified as being enriched in PFAMs containing genes with altered methylation in IPF and SSc compared to controls and IPF compared to SSc lung fibroblasts.

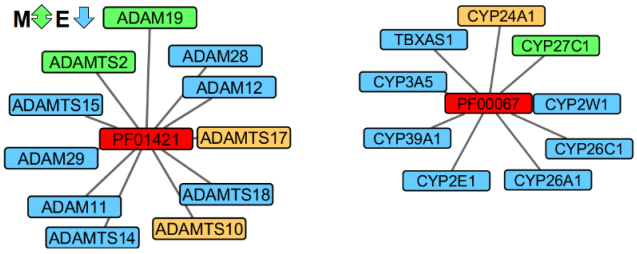


Figure 4.4.2. PFAMs associated with biological processes potentially relevant to pulmonary fibrosis. PF01421 ADAM and ADAMTS domain family and PF00067 cytochrome P450 domain family which, among other biological processes, are associated with collagen fibril organisation and bile metabolic process respectively. Collagen fibril organisation and bile acid metabolic process were not significantly enriched in PFAMs with altered methylation in IPF compared to SSc lung fibroblasts but multiple genes belonging to PFAMs associated with these biological processes did have significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$). Coloured boxes represent direction of methylation, orange; genes with increased methylation, Blue; genes with decreased methylation, green; genes with both increased and decreased methylation in IPF compared to SSc lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and significantly altered expression (TNoM ≤1; $P<0.05$) in IPF compared to SSc lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in IPF (n=5) relative to the average SSc lung fibroblast methylation/expression value.

4.5. Biological processes enriched in PFAMs containing genes with altered expression in IPF compared to SSc lung fibroblasts.

Genes which had significantly altered expression in IPF compared to SSc lung fibroblasts mapped to 765 PFAMs. Biological processes (636) including Wnt signalling, ECM organisation and bile acid metabolic process were enriched in these PFAMs (**Table 4.5.1**). Of these 636 enriched biological process, 447 (70%) overlapped with those enriched in PFAMs containing genes which had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF compared to SSc lung fibroblasts. These included response to hypoxia, Wnt signalling, ECM organisation, chromatin remodelling and induction of apoptosis by extracellular signals.

As previously shown, multiple biological processes enriched in PFAMs containing genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and expression overlapped (66% and 63% respectively) between IPF and SSc compared to control lung fibroblasts and 70% overlapped between IPF and SSc lung fibroblasts. This shows that the multiple biological processes are commonly enriched in PFAMs containing genes with both significantly altered methylation and expression. Many of these genes have both significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and expression suggesting a direct effect of methylation on expression. For example the biological process 'induction of apoptosis

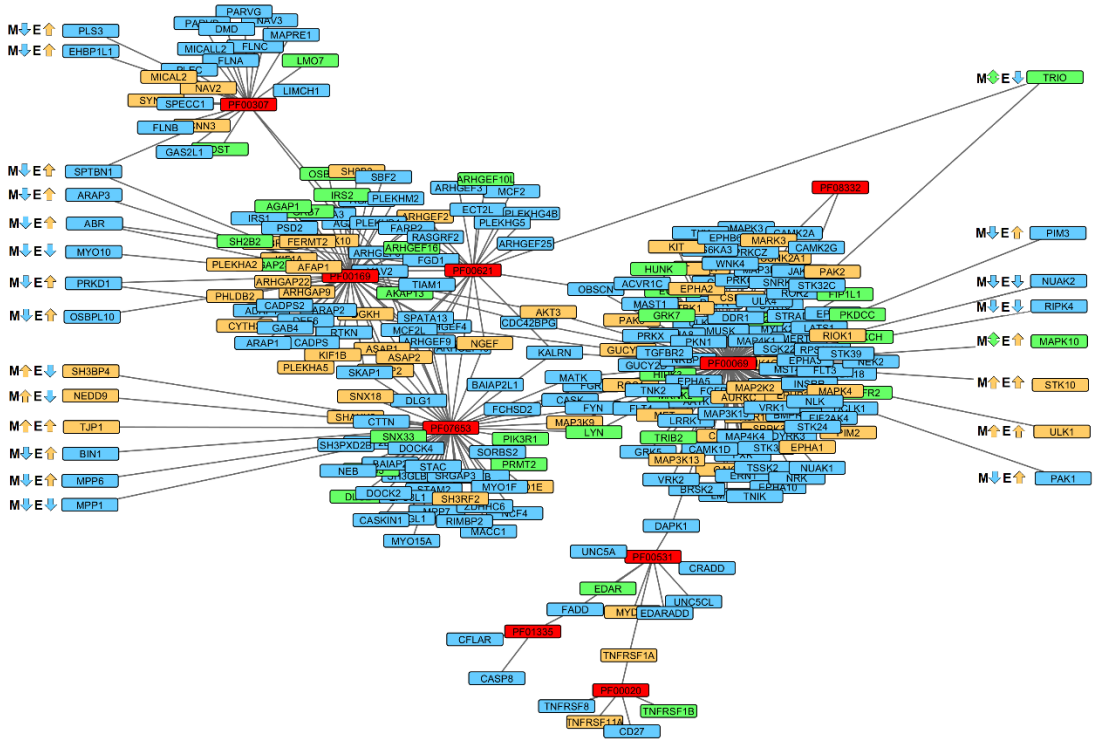
by extracellular signals' was associated with 10 PFAMs; 9 of these PFAMs contained 271 genes with altered methylation in IPF compared to SSc and 7 of these PFAMs contained 57 genes with altered expression in IPF compared to SSc lung fibroblasts (**Figure 4.5.1**). Twenty-two genes overlapped which had altered methylation and expression in IPF compared with SSc lung fibroblasts, suggesting multiple genes encoding proteins with functional domains associated with apoptosis could be directly regulated by methylation.

Furthermore, biological processes such as Wnt signalling were enriched in PFAMs containing genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and expression in both IPF and SSc compared to control lung fibroblasts and in PFAMs containing genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and expression in IPF compared to SSc lung fibroblasts. This suggests that Wnt signalling may be different in both IPF and SSc compared to control but also different in IPF compared to SSc lung fibroblasts.

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
ECM organization	2	3.71e-13	42	63
Response to hypoxia	3	4.82e-06	25	45
Regulation of PI3K activity	3	4.80e-05	12	16
Regulation of IL-2 production	3	5.11e-05	9	10
Regulation of epithelial cell proliferation involved in lung morphogenesis	3	4.70e-04	6	6
Bile acid metabolic process	4	4.70e-04	6	6
JAK-STAT cascade	3	1.20E-03	9	13
Wnt receptor signalling pathway	2	1.38E-03	18	38
Lung epithelium development	3	1.81e-03	7	9
Collagen fibril organization	3	3.22e-03	8	12
Induction of apoptosis by extracellular signals	3	4.18e-03	7	10
Methylation	2	4.79e-03	30	83
Collagen catabolic process	4	5.75e-03	4	4
Chromatin remodelling	3	7.32e-03	14	31
Regulation of viral reproduction	3	9.64e-03	8	14

Table 4.5.1. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered expression in IPF compared to SSc lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate ($P<0.01$). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

PFAMs containing genes with altered methylation in IPF compared with SSc lung fibroblasts



PFAMs containing genes with altered expression in IPF compared with SSc lung fibroblasts

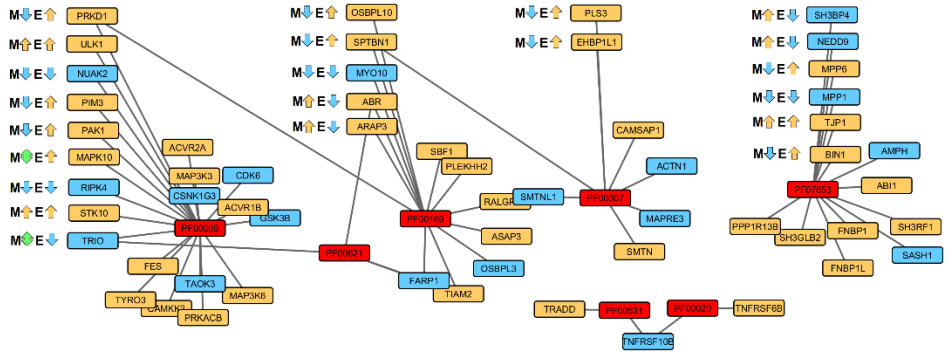


Figure 4.5.1. Genes belonging to PFAMs associated with induction of apoptosis by extracellular signals which have altered methylation and expression in IPF compared to SSc lung fibroblasts. Coloured boxes represent direction of methylation/expression, yellow; genes with increased methylation/expression, blue; genes with decreased methylation/expression, green; genes with both increased and decreased methylation in IPF compared to SSc lung fibroblasts. Annotated boxes indicate genes with both significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and significantly altered expression (TNoM ≤1; $P<0.05$) in IPF compared to SSc lung fibroblasts. M= methylation, E= expression. Blue = decreased, orange = increased in IPF (n=5) relative to the average SSc lung fibroblast methylation/expression value.

4.6. Functional analysis of differentially methylated and expressed genes in IPF and SSc compared to control lung fibroblasts

As previously described, Genecodis 3 web analysis tool (Tabas-Madrid et al, 2012) was used to determine enriched functional groups and understand the biological meaning of the differentially methylated genes in IPF and SSc compared to control lung fibroblasts. Multiple biological processes (550 and 309) were significantly (FDR P<0.05) enriched in genes with significantly altered methylation ($\Delta\beta \geq 0.136$; P<0.05) in IPF and SSc compared to control lung fibroblasts, respectively. Of the 309 enriched biological processes in SSc, 203 (66%) overlapped with IPF. These included apoptosis, Wnt signalling, blood coagulation, regulation of transcription, small GTPase-mediated signal transduction and ECM organisation (**Table 4.6.1**). Many of these processes have previously been implicated in the pathogenesis of pulmonary fibrosis (Morrisey, 2003, Chilosi et al, 2003, Königshoff et al, 2008, Chambers, 2008, Chambers and Scotton, 2012, Mercer et al, 2013).

Multiple biological processes (31 and 17) were identified as being significantly (FDR P<0.05) enriched in genes that had significantly altered expression (TNoM ≤1; P<0.05) in IPF and SSc compared to control lung fibroblasts, respectively. These included apoptosis, regulation of transcription, gene expression and small GTPase-mediated signal transduction (**Table 4.6.2**). Only 4 biological processes overlapped between IPF and SSc. Wnt signalling, ECM organisation and lung development were not enriched in genes that were differentially expressed in IPF and SSc compared to control lung fibroblasts, however, there were specific genes which did have altered expression belonging to these processes. Furthermore, although these biological processes were not significantly enriched in genes with significantly altered expression (TNoM ≤1; P<0.05) in IPF and SSc, they were significantly (P<0.01) enriched in PFAMs containing genes with significantly altered expression (TNoM ≤1; P<0.05) (**see section 4.3, Table 4.3.1**).

Control v IPF

Biological process (BP)	Number of genes	Total number of genes associated with the BP	Hypergeometric P value
Cell adhesion	170	556	4.53E-35
Regulation of transcription	290	1609	4.00E-17
Regulation of small GTPase mediated signal transduction	54	172	2.99E-11
Apoptotic process	122	594	1.86E-10
Blood coagulation	96	457	1.02E-08
ECM organization	25	73	5.68E-06
Regulation of cell proliferation	30	101	1.01E-05
Wnt receptor signalling pathway	31	110	2.03E-05
Immune response	74	382	2.08E-05
Negative regulation of BMP signalling pathway	13	26	3.48E-05
Wound healing	22	68	5.90E-05
Response to glucocorticoid stimulus	25	84	6.53E-05
Inflammatory response	53	259	0.000113
EMT	11	25	0.000667
Lung development	21	75	0.000772

Control v SSc

Cell adhesion	199	556	9.13e-43
Regulation of transcription	373	1609	3.20E-31
Apoptotic process	159	594	7.65e-19
Regulation of small GTPase mediated signal transduction	70	172	3.60E-18
Blood coagulation	125	457	1.87e-15
Wnt receptor signalling pathway	42	110	9.31E-10
Chromatin modification	65	224	4.14e-09
Integrin-mediated signalling pathway	31	70	4.95e-09
ECM organization	29	73	3.47e-07
Lung development	28	75	2.43e-06
Chemotaxis	35	126	0.000102
FGFR signalling pathway	25	78	0.00015
Inflammatory response	58	259	0.000182
Immune response	76	382	0.000604
Pregnancy	19	72	0.013026

Table 4.6.1. Functional analysis of differentially methylated genes in SSc compared to control lung fibroblasts. Multiple biological processes and pathways were identified which included genes with significantly increased or decreased ($\Delta\beta \geq 0.136$; $P<0.05$) methylation in IPF (n=5) and SSc (n=7) compared to control (n=6) lung fibroblasts. The number of genes associated with each biological process and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

Control v IPF

Biological process	Number of genes	Total number of genes associated with the BP	Hypergeometric P value
Apoptotic process	28	594	2.57e-05
Signal transduction	39	1176	0.000296
Viral reproduction	15	329	0.008767
Ubiquitin-dependent protein catabolic process	10	150	0.009025
Viral transcription	7	82	0.014664
Gene expression	16	408	0.019243
Small GTPase-mediated signal transduction	13	312	0.028138
Platelet activating factor biosynthetic process	2	4	0.043839
Negative regulation of cell proliferation	13	341	0.047919
Regulation of transcription, DNA-dependent	38	1609	0.048648

Control v SSc

Pregnancy	10	72	0.000691
Transport	29	604	0.000802
Apoptotic process	12	156	0.009583
Negative regulation of cell proliferation	17	341	0.020969
Signal transduction	39	1176	0.025865
Translation	14	241	0.027217
Copper ion transport	4	14	0.028483
Regulation of transcription, DNA-dependent	48	1609	0.031969
Negative regulation of cell volume	2	2	0.03632
Microtubule cytoskeleton organization	6	55	0.036543

Table 4.6.2. Functional analysis of differentially expressed genes in IPF/SSc compared to control lung fibroblasts. Genes which had significantly altered expression (TNoM ≤1; P<0.05) in IPF (n=5) or SSc (n=7) compared to control (n=6) lung fibroblasts. The number of genes associated with each biological process and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

4.7. Functional analysis of differentially methylated and expressed genes in IPF compared to SSc lung fibroblasts

Multiple (639) biological processes, many of which have previously been implicated in pulmonary fibrosis including apoptosis, Wnt signalling, blood coagulation, regulation of transcription, lung development and ECM organisation, were identified as being significantly (FDR P<0.05) enriched in genes that had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts (**Table 4.7.1**).

Biological process	Number of genes	Total number of genes associated with the BP	Hypergeometric P value
Multicellular organism development	339	945	8.36e-78
Regulation of transcription, DNA-dependent	367	1609	2.70E-31
Cell adhesion	174	556	2.41e-30
Blood coagulation	121	457	1.48e-14
Regulation of small GTPase mediated signal transduction	61	172	5.76E-13
Apoptotic process	135	594	6.97e-11
Platelet activation	60	234	1.03e-06
Response to hypoxia	49	175	1.04e-06
Lung development	27	75	5.75e-06
Wnt receptor signalling pathway	33	110	2.66E-05
ECM organization	25	73	4.02e-05
Negative regulation of BMP signalling pathway	13	26	0.000115
Cellular response to TGFß stimulus	11	24	0.001203
Female pregnancy	20	72	0.004635
DNA methylation	7	19	0.044627

Table 4.7.1. Functional analysis of differentially methylated genes in IPF compared to SSc lung fibroblasts. Multiple biological processes and pathways identified which included genes with significantly increased or decreased (Δβ ≥ 0.136; P<0.05) methylation in IPF (n=5) compared to SSc (n=7) lung fibroblasts. The number of genes associated with each biological process and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

Multiple (45) biological processes were identified as being significantly (FDR P<0.05) enriched in genes which had significantly altered expression (TNoM ≤1; P<0.05) in IPF compared to SSc lung fibroblasts. O these biological processes, 19, including apoptosis, regulation of transcription, blood coagulation, TGFß signalling and small GTPase-mediated signal transduction, overlapped with those which were significantly (FDR P<0.05) enriched in genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts (**Table 4.7.2**).

Biological process	**Number of genes (Methylation)**	**Number of genes (expression)**	**Total number of genes associated with the BP**
Placenta development	10	5	28
Regulation of transcription, DNA-dependent	367	78	1609
Proteolysis	101	35	543
Ubiquitin-dependent protein catabolic process	31	15	150
Transport	110	45	604
Endocytosis	29	14	110
Apoptotic process	135	41	594
Mitosis	41	14	187
Signal transduction	308	72	1176
Small GTPase-mediated signal transduction	80	25	312
Nervous system development	137	29	410
Axon guidance	115	20	307
Pregnancy	20	10	72
Blood coagulation	121	29	457
Negative regulation of cell proliferation	101	27	341
Cellular nitrogen compound metabolic process	48	18	200
Intracellular signal transduction	83	26	303
Negative regulation of apoptotic process	79	20	272
Cellular lipid metabolic process	30	13	128

Table 4.7.2. Functional analysis of biological processes which contained differentially expressed genes which overlapped with biological processes containing differentially methylated genes in IPF compared to SSc lung fibroblasts. Nineteen overlapping biological processes which contained genes which had significantly altered expression (TNoM ≤1; P<0.05) and significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. The number of genes associated with each biological process and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

4.8. Summary

- PFAM and gene-term enrichment analysis identified multiple biological processes enriched in genes with significantly altered methylation in IPF and/or SSc compared to control lung fibroblasts, many of which have previously been implicated in pulmonary fibrosis and many of which are novel which may be relevant to fibrosis.

- Multiple biological processes enriched in genes with significantly altered methylation were the same as those enriched in genes with significantly altered expression in IPF and SSc compared to control lung fibroblasts. This suggests that methylation is important in regulating gene expression in pulmonary fibrosis.

- Many of these biological processes were also significantly enriched in genes which had significantly altered methylation and/or expression in IPF compared to SSc lung fibroblasts.

- Taken together this data suggests multiple common biological processes are enriched with genes which have altered methylation and expression in IPF and SSc compared to control lung fibroblasts, however, multiple genes are distinct to each disease.

4.9. Pathway analysis of differentially methylated and expressed genes in IPF and SSc compared to control lung fibroblasts

GO term enrichment analysis identified multiple biological processes associated with the lung and/or fibrosis, including Wnt signalling, ECM organisation and lung development which were enriched in genes that were differentially methylated in both IPF and SSc compared to control lung fibroblasts. There were also multiple genes in commonly enriched processes that were distinct to each disease which was concurrent with data from the PFAM enrichment analysis. Biological processes including Wnt signalling, ECM organisation and lung development were not enriched in genes that were differentially expressed in IPF and SSc compared to control lung fibroblasts, potentially because the cut-off for inclusion was too stringent. However, there were specific genes which did have altered expression belonging to these processes. Furthermore, although these biological processes were not significantly enriched in genes with significantly altered expression in IPF and SSc, PFAMs containing genes with significantly altered expression were significantly enriched in these processes.

Aberrant Wnt signalling has previously been implicated in a multitude of fibroses including IPF and SSc (Chilosi et al, 2003, Selman et al, 2008, Königshoff and Eickelberg, 2010, Lam and Gottardi, 2011). B-catenin is a downstream component of the Wnt signalling pathway and has been implicated in fibroses including SSc-PF where its expression is increased (Lam and Gottardi, 2011). Active β-catenin acts as a transcription coactivator for the transcription factors Lymphoid enhancer-binding factor 1 (LEF1)/ T-cell factor 1 (TCF1) which mediate nuclear responses to Wnt signals (Eastman and Grosschedl, 1999). Glycogen synthase kinase 3β (GSK3β) can degrade β-catenin (Nakamura et al, 1998), however, the binding of Wnt ligands to low-density lipoprotein receptor-related protein (LRP)/Frizzled (FZD) receptors inhibits GSK3β which subsequently inhibits the phosphorylation and degradation of β-catenin (Liu et al, 2002). Expression of Wnt-related genes including WNT5A, frizzled class receptor 2 (FZD2) and frizzled class receptor 3 (FZD3) are increased in IPF (Königshoff et al, 2008, Vuga et al, 2009). Expression of FZD2 and the Wnt target LEF1 are increased in SSc skin fibroblasts whereas Wnt antagonists DKK1, Dickkopf-related protein 2 (DKK2), SFRP1 and WNT inhibitory factor 1 (WIF1) all have decreased expression which has been associated with promoter methylation (Wei et al, 2012, Dees et al, 2013), however, the full extent and role of methylation in Wnt signalling in pulmonary fibrosis is unknown.

GO term enrichment analysis identified the WNT receptor signalling pathway as being significantly enriched in 31 genes that were differentially methylated in IPF (P=2.03E-05) and 42 genes that were differentially methylated in SSc (P=9.31E-10) compared to control lung fibroblasts. Eighteen genes were the same in both IPF and SSc, 13 were distinct to IPF and 24 were distinct to SSc (**Table 4.9.1**). Whilst GO enrichment is a useful tool for finding enriched biological processes and potentially in the use of generating user-defined pathways, GO terms do not correspond directly with well-defined pathways such as KEGG pathways (Mao et al, 2005). For already well-defined pathways such as the

KEGG Wnt signalling pathway, KEGG enrichment was also used. KEGG enrichment identified 45 genes that were differentially methylated in IPF (P=1.835E-09) and 59 genes that were differentially methylated in SSc (P=1.56E-14) compared to control lung fibroblasts (**Table 4.9.1**).

GO-term enrichment

Differentially methylated genes in fibrotic lung fibroblasts involved in WNT signalling							
CALCOCO1	***FBXW11***	***SFRP5***	*CSNK1D*	*TNIK*	*C1orf187*	*GSK3A*	*TNKS*
CCDC88C	*GRK6*	*TLE3*	*CSNK1G3*	***WNT4***	*CSNK1G1*	*LRRFIP2*	***WIF1***
CCNY	***LEF1***	***WNT11***	*FAM123B*	***WNT8A***	***CTNNB1***	*MCC*	*WISP1*
CD44	***LRP5***	***WNT5A***	*FBXW4*	***APC***	*CYLD*	*NDRG2*	***WNT10B***
CDK14	*MITF*	*BRD7*	*KREMEN2*	*ARL6*	*DRD2*	*RSPO2*	***WNT16***
CPZ	***NKD2***	*CELSR2*	*RNF146*	*BCL9*	***FZD6***	***TCF7L1***	***WNT7B***
DACT1	*NXN*	***CSNK1A1L***	*SOSTDC1*	***BTRC***	*GRK5*	*TLE2*	

KEGG pathway enrichment

Differentially methylated genes in fibrotic lung fibroblasts involved in WNT signalling							
AXIN2	*FZD7*	***SFRP5***	***CSNK1A1L***	*SMAD2*	***CTNNB1***	*PPP3CA*	***WIF1***
CAMK2A	***LEF1***	*TBL1XR1*	*CUL1*	***WNT4***	*CTNNBIP1*	*PRICKLE1*	***WNT10B***
CAMK2B	***LRP5***	*TCF7L2*	*CXXC4*	*WNT7A*	*DAAM2*	*PRKCG*	***WNT16***
CAMK2G	*MAPK10*	*VANGL2*	*FZD9*	***WNT8A***	*FOSL1*	*SIAH1*	*WNT6*
CSNK2A1	*NFATC1*	*WNT10A*	*PLCB1*	*WNT8B*	*FZD1*	*SKP1*	***WNT7B***
CTBP2	***NKD2***	***WNT11***	*PLCB2*	***APC***	***FZD6***	*SMAD3*	
DKK4	*PRKCA*	*WNT3*	*PPP2R5A*	*BAMBI*	*GPC4*	*SOST*	
FBXW11	*PRKCB*	***WNT5A***	*PPP2R5C*	***BTRC***	*JUN*	*SOX17*	
FZD10	*SFRP1*	*WNT9A*	*PRICKLE2*	*CACYBP*	*NFATC2*	*TBL1Y*	
FZD5	*SFRP2*	*CCND3*	*SENP2*	*CREBBP*	*PLCB3*	***TCF7L1***	

Table 4.9.1. Differentially methylated genes in fibrotic compared to control lung fibroblasts involved in the Wnt signalling pathway as determined by GO-term and KEGG enrichment. Genes highlighted in red are differentially methylated in both IPF (n=5) and SSc (n=7) compared to control (n=6) lung fibroblasts. Genes specifically differentially methylated in IPF (n=5) or SSc (n=7) compared to control (n=6) lung fibroblasts are highlighted in blue and yellow respectively. Genes in bold text overlap between both enrichment analyses. Full gene names can be found on appendice B.

In IPF, multiple Wnts had significantly altered methylation compared to control lung fibroblasts, although WNT2B was the only Wnt which had significantly increased expression in IPF compared to control lung fibroblasts. Wnt ligand receptors LRP5 had significantly increased methylation and FZD4 had significantly increased expression in IPF compared to control lung fibroblasts, however, no change in methylation or expression of GSK3β or CNNTB1 was observed. The Wnt target gene, LEF1, also had significantly increased methylation in IPF compared to control lung fibroblasts but no change in LEF1 expression was observed. Secreted frizzled-related protein 5 (SFRP5), a Wnt signalling inhibitor (Stuckenholz et al, 2013), had significantly decreased methylation in IPF compared to control lung fibroblasts. In breast cancer cell lines/tissue, promoter methylation correlates with decreased expression (Veeck et al, 2008) suggesting SFRP5 expression can be regulated by methylation, however, no change in SFRP5 expression was observed in IPF compared to control lung fibroblasts.

In SSc multiple Wnts also had significantly altered methylation compared to control lung fibroblasts, and like IPF, WNT2B was the only Wnt with significantly increased expression. In contrast with IPF, SSc lung fibroblasts had increased methylation of Wnt signalling inhibitors including SFRP5 and WIF1, compared to control lung fibroblasts. GSK3β expression was significantly increased in SSc compared to control lung fibroblasts. Downstream targets CNNTB1, LEF1 and Transcription factor 7 like 1 (TCF7L1) all had significantly increased methylation in SSc compared to control lung fibroblasts but had no change in expression, suggesting that methylation of these genes does not directly regulate their expression.

Using genes identified by GO term enrichment and KEGG enrichment analyses, known functional interactions between genes associated with Wnt signalling which had significantly altered methylation and/or expression in IPF and SSc compared to control lung fibroblasts were identified by STRING 10 analysis and are shown in **Figure 4.9.1 and Figure 4.9.3**. These genes were mapped to the Wnt signalling KEGG pathway using cytoscape (**see Chapter 2: section 2.7.2**) to identify where they fit and how they potentially interact with other genes in the pathway (**Figure 4.9.2 and Figure 4.9.4**).

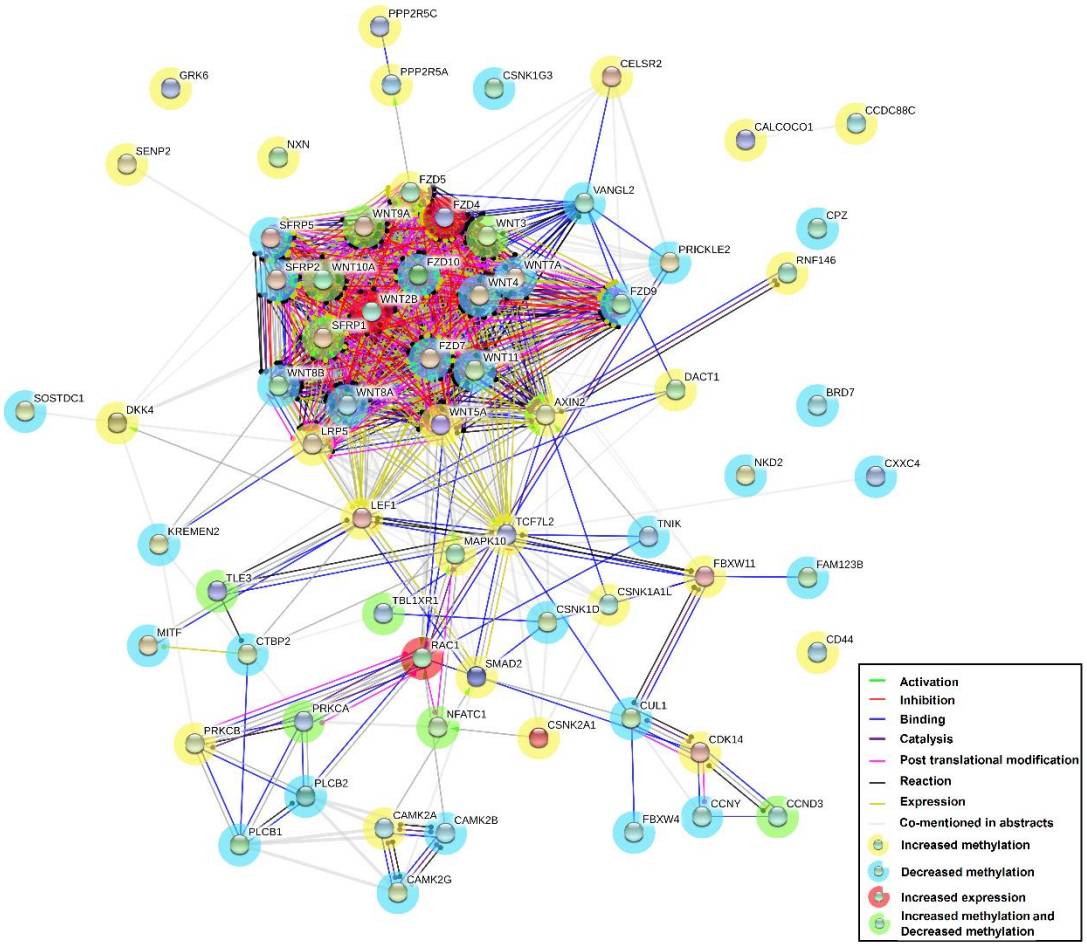


Figure 4.9.1. Protein-protein interactions of genes involved in Wnt signalling which have altered methylation and/or expression in IPF compared to control lung fibroblasts. Network analysis showing genes which interact with each other which have significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF (n=5) compared to control (n=6) lung fibroblasts. Predicted functional links are indicated by the colour of adjoining lines: each colour represents a different type of evidence. Blue; binding, green; activation, red; inhibition, yellow; expression, black; reaction, purple; catalyst, pink; post-translational modification, grey; co-mentioned in abstracts. Coloured nodes represent genes directly linked to the input list. Coloured circles around the nodes indicate direction of methylation and expression: yellow = genes with CpGs which had increased methylation, blue = genes with CpGs which had decreased methylation, green = genes with CpGs which had both increased and decreased methylation in IPF compared to control lung fibroblasts. Red = genes with significantly increased (TNoM ≤ 1; P<0.05) expression in IPF compared to control lung fibroblasts.

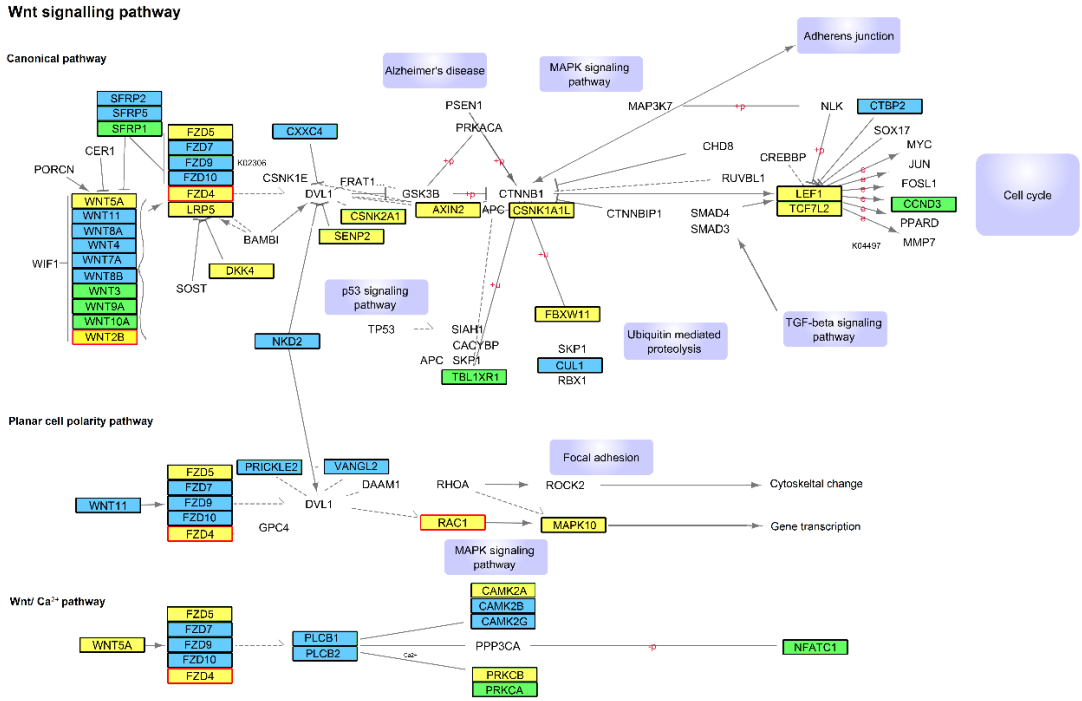


Figure 4.9.2. Genes associated with the KEGG pathway of Wnt signalling which have altered methylation and/or expression in IPF compared to control lung fibroblasts. Coloured nodes outlined in black represent genes with significantly altered methylation (Δβ ≥0.136; P<0.05) and in red represent genes with significantly altered expression (TNoM ≤ 1; P<0.05) in IPF compared to control lung fibroblasts. For methylation: yellow = genes with CpGs which had increased methylation, blue = genes with CpGs which had decreased methylation, green = genes with CpGs which had both increased and decreased methylation in IPF compared to control lung fibroblasts. For expression: yellow = genes which had increased expression in IPF compared to control lung fibroblasts. +P = phosphorylation, e = gene expression relationship, dotted lines = indirect effect on gene, purple boxes = joining pathways.

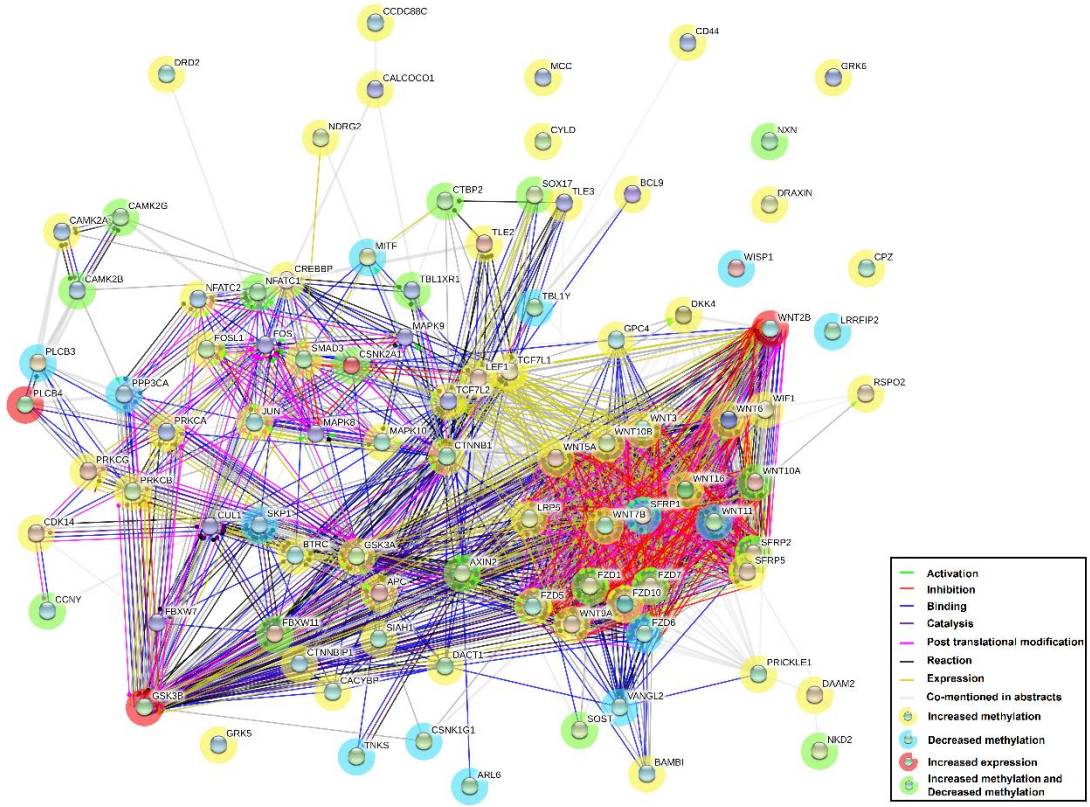


Figure 4.9.3. Protein-protein interactions of genes involved in Wnt signalling which have altered methylation and/or expression in SSc compared to control lung fibroblasts. Network analysis showing genes which interact with each other which have significantly altered methylation (Δβ ≥0.136; P<0.05) in SSc (n=7) compared to control (n=6) lung fibroblasts. Predicted functional links are indicated by the colour of adjoining lines: each colour represents a different type of evidence. Blue; binding, green; activation, red; inhibition, yellow; expression, black; reaction, purple; catalyst, pink; post-translational modification, grey; co-mentioned in abstracts. Coloured nodes represent genes directly linked to the input list. Coloured circles around the nodes indicate direction of methylation and expression: yellow = genes with CpGs which had increased methylation, blue = genes with CpGs which had decreased methylation, green = genes with CpGs which had both increased and decreased methylation in SSc compared to control lung fibroblasts. Red = genes with significantly (TNoM ≤ 1; P<0.05) increased expression in SSc compared to control lung fibroblasts.

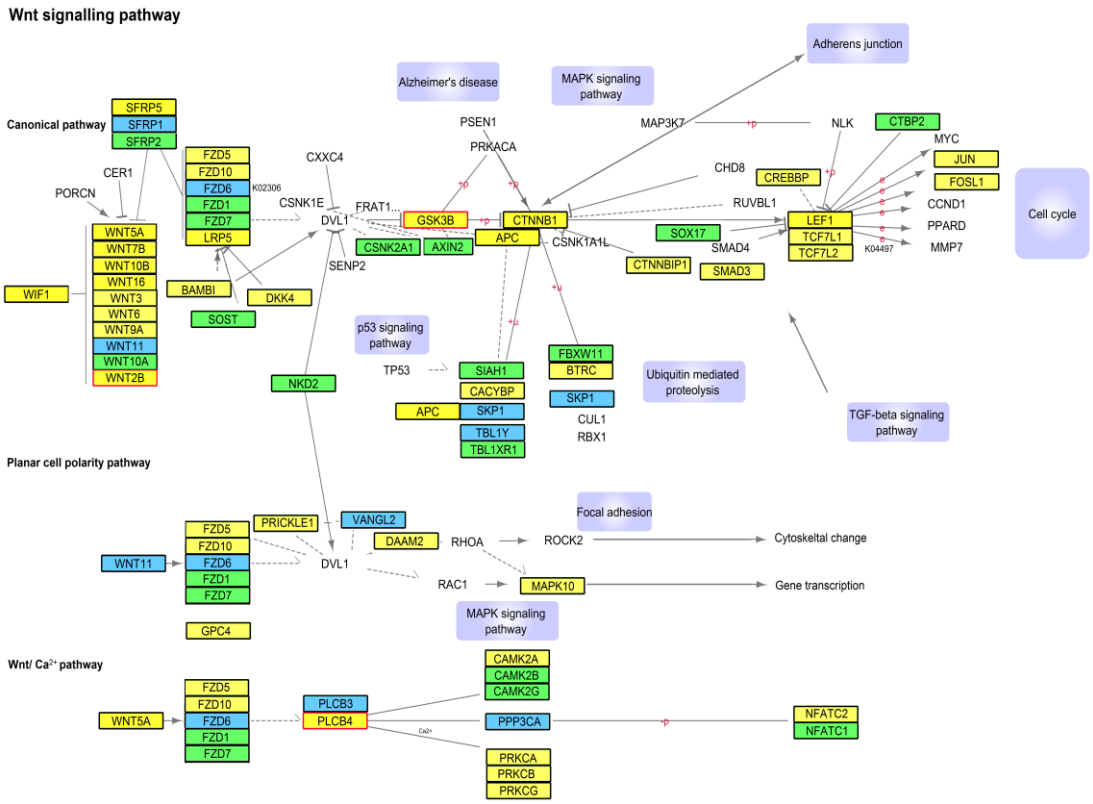


Figure 4.9.4. Genes associated with the KEGG pathway of Wnt signalling which have altered methylation and/or expression in SSc compared to control lung fibroblasts. Coloured nodes outlined in black represent genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and in red represent genes with significantly altered expression (TNoM ≤ 1; $P<0.05$) in SSc compared to control lung fibroblasts. For methylation: yellow = genes with CpGs which had increased methylation, blue = genes with CpGs which had decreased methylation, green = genes with CpGs which had both increased and decreased methylation in SSc compared to control lung fibroblasts. For expression: yellow = genes which had increased expression in SSc compared to control lung fibroblasts. +P = phosphorylation, e = gene expression relationship, dotted lines = indirect effect on gene, purple boxes = joining pathways.

4.10. Pathway analysis of differentially methylated and expressed genes in IPF compared to SSc lung fibroblasts

The WNT receptor signalling pathway was significantly enriched in 33 genes that were differentially methylated in IPF (P=2.66E-05) compared to SSc lung fibroblasts (**Table 4.10.1**).

Differentially-methylated genes in IPF compared to SSc lung fibroblasts associated with Wnt signalling

Symbol	Name	Symbol	Name
BCL9	B-cell CLL/lymphoma 9	**NXN**	Nucleoredoxin
CCDC88C	Coiled-coil domain containing 88C	**PORCN**	Porcupine homolog
CCNY	Cyclin Y	**PYGO2**	Pygopus homolog 2
CELSR2	Cadherin, EGF LAG seven-pass G-type receptor 2	**SOSTDC1**	Sclerostin domain containing 1
CSNK1G1	Casein kinase 1, gamma 1	**TCF7**	Transcription factor 7 (T-cell specific, HMG-box)
CTNNB1	Catenin (cadherin-associated protein), beta 1	**TCF7L1**	Transcription factor 7-like 1 (T-cell specific, HMG-box)
CYLD	Cylindromatosis	**TLE3**	Transducin-like enhancer of split 3 (E(sp1) homolog, Drosophila)
DRD2	Dopamine receptor D2	**TLE4**	Transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)
DVL3	Dishevelled, dsh homolog 3	**TNIK**	TRAF2 and NCK interacting kinase
ETV2	Ets variant 2	**WIF1**	WNT inhibitory factor 1
FAM123B	Family with sequence similarity 123B	**WNT1**	Wingless-type MMTV integration site family, member 1
FBXW11	F-box and WD repeat domain containing 11	**WNT2**	Wingless-type MMTV integration site family member 2
FZD6	Frizzled homolog 6	**WNT5A**	Wingless-type MMTV integration site family, member 5A
GRK5	G protein-coupled receptor kinase 5	**WNT5B**	Wingless-type MMTV integration site family, member 5B
LRP5	Low density lipoprotein receptor-related protein 5	**WNT7B**	Wingless-type MMTV integration site family, member 7B
MCC	Mutated in colorectal cancers	**WNT8A**	Wingless-type MMTV integration site family, member 8A
NKD2	Naked cuticle homolog 2		

Table 4.10.1. Differentially-methylated genes in IPF compared to SSc lung fibroblasts associated with Wnt signalling. Genes associated with WNT signalling with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts are shown.

Genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts associated with Wnt signalling known to interact with each other are shown in **Figure 4.10.1.** Compared to SSc; IPF lung fibroblasts had multiple Wnts including WNT1, WNT2, WNT5B, WNT7B and WNT8A which had significantly decreased methylation. Conversely, WNT5A had significantly increased methylation and significantly decreased expression. Although no Wnt genes had significantly increased expression, frizzled receptors FZD1 and FZD4 had significantly increased expression which could potentially facilitate greater Wnt binding. GSK3β expression was significantly decreased whereas CTNNB1 expression was significantly increased in IPF compared to SSc lung

fibroblasts (**Figure 4.10.2**). This is in agreement with previous evidence in Wnt literature showing the binding of Wnt ligands to FZD receptors leads to inhibition of GSK3β resulting in increased CTNNB1 (Liu et al, 2002). Interestingly, CTNNB1 also had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts suggesting a potential direct link between methylation and expression.

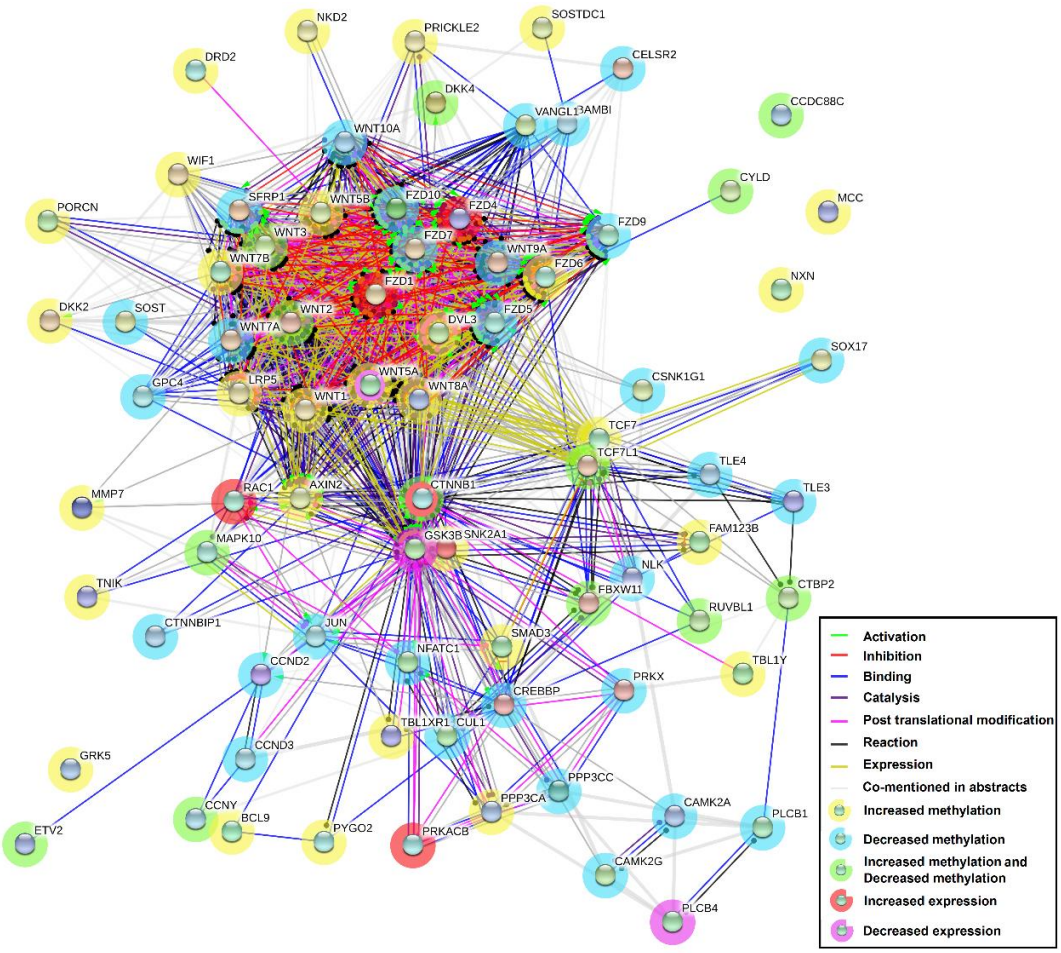


Figure 4.10.1. Protein-protein interactions of genes involved in Wnt signalling which have altered methylation and/or expression in IPF compared SSc lung fibroblasts. Network analysis showing genes which interact with each other which have significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. Predicted functional links are indicated by the colour of adjoining lines: each colour represents a different type of evidence. Blue; binding, green; activation, red; inhibition, yellow; expression, black; reaction, purple; catalyst, pink; post-translational modification, grey; co-mentioned in abstracts. Coloured nodes represent genes directly linked to the input list. Coloured circles around the nodes indicate direction of methylation and expression: yellow = genes with CpGs which had increased methylation, blue = genes with CpGs which had decreased methylation, green = genes with CpGs which had both increased and decreased methylation in IPF compared to SSc lung fibroblasts. Red = genes with increased (TNoM ≤ 1; P<0.05) expression, purple = genes with decreased (TNoM ≤ 1; P<0.05) expression in IPF compared to SSc lung fibroblasts.

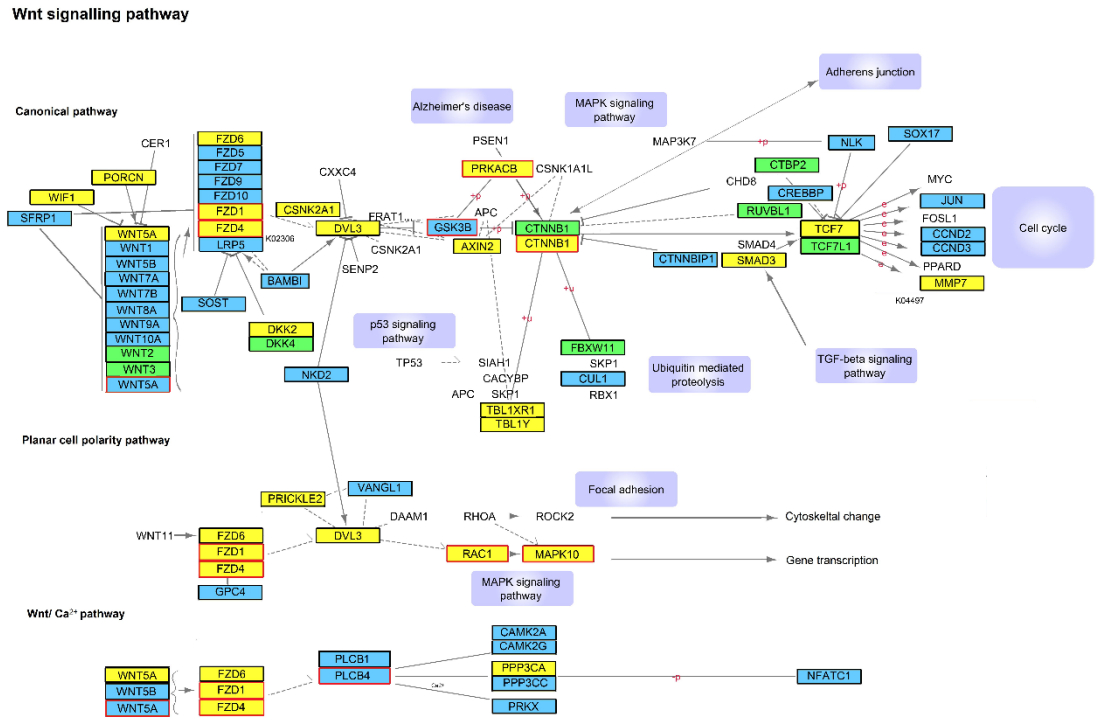


Figure 4.10.2. Genes associated with the KEGG pathway of Wnt signalling which have altered methylation and/or expression in IPF compared to SSc lung fibroblasts. Coloured nodes outlined in black represent genes with significantly altered methylation (Δβ ≥0.136; P<0.05) and in red represent genes with significantly altered expression (TNoM ≤ 1; P<0.05) in IPF (n=5) compared to SSc (n=7) lung fibroblasts. For methylation: yellow = genes with CpGs which had increased methylation, blue = genes with CpGs which had decreased methylation, green = genes with CpGs which had both increased and decreased methylation in IPF compared to SSc lung fibroblasts. For expression: yellow = genes which had increased expression, blue = genes which had decreased expression in IPF compared to SSc lung fibroblasts. +P = phosphorylation, e = gene expression relationship, dotted lines = indirect effect on gene, purple boxes = joining pathways.

4.11. Validation of functional analyses using TNXB

As previously shown TNXB had multiple CpGs with significantly altered methylation in IPF and SSc compared to control lung fibroblasts as confirmed by microarray and bisulfite sequencing (**see section 3.7, Figure 3.7.5**). TNXB expression was also significantly increased in IPF compared to SSc and control lung fibroblasts as confirmed by qRT-PCR and IHC (see s**ection 3.12.1, Figure 3.12.2**). Remodelling of the ECM involves multiple processes which include organisation and rearrangement of ECM components which can lead to matrix stiffening. TNX is known to play an important role in collagen deposition and in regulating collagen fibril density (Mao et al, 2002) and was associated with multiple biological processes including cell adhesion, signal transduction and actin cytoskeletal organisation. The ECM-receptor interaction and focal adhesion KEGG pathways were also enriched in TNXB, among other genes which had significantly altered methylation in IPF compared to control lung fibroblasts (**Figure 4.11.1**).

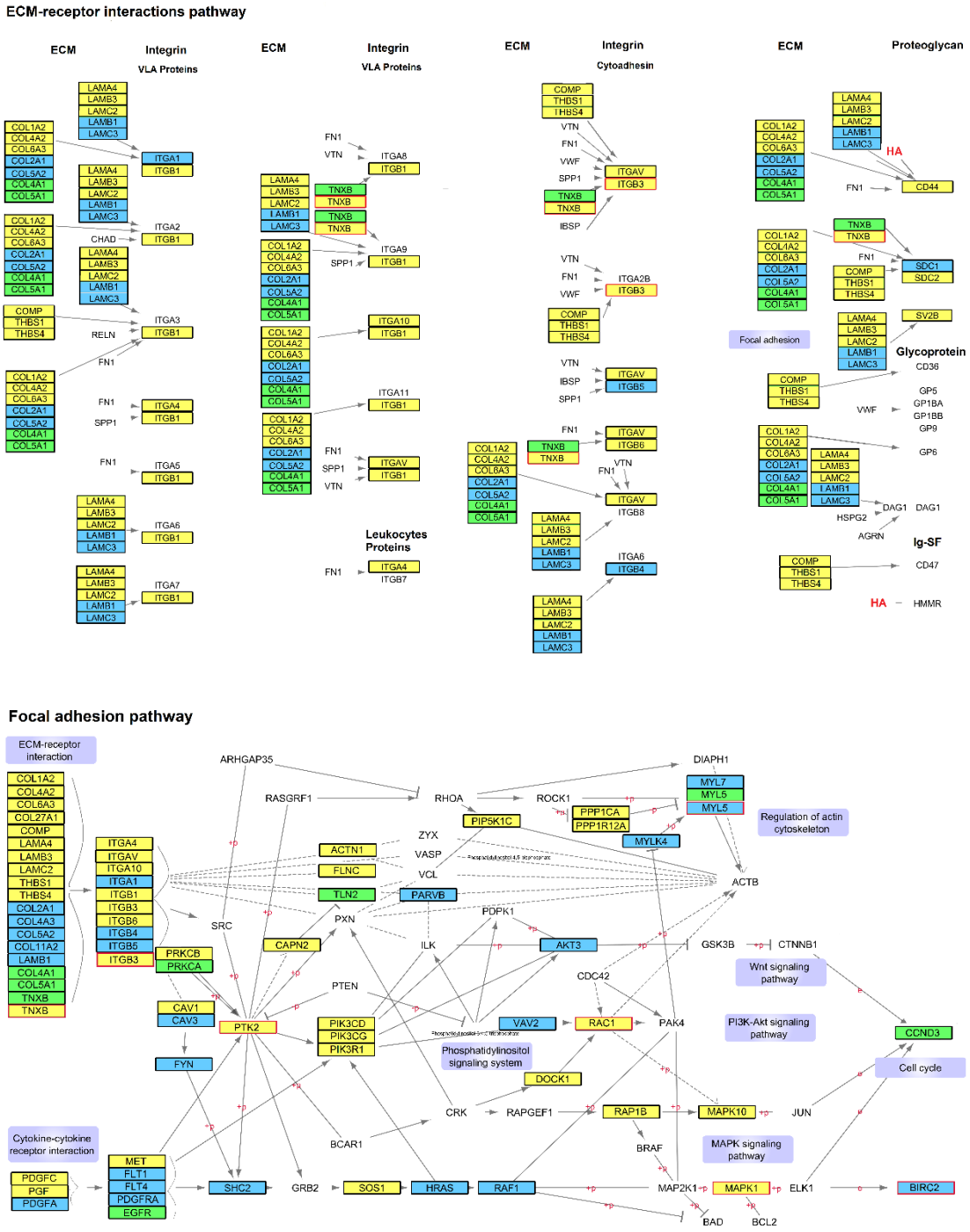


Figure 4.11.1. TNXB is associated with ECM-receptor interactions and focal adhesions KEGG pathways. Coloured nodes outlined in black represent genes with significantly altered methylation (Δβ ≥0.136; P<0.05) and in red represent genes with significantly altered expression in IPF (n=5) compared to control (n=6) lung fibroblasts. For methylation: yellow = genes with CpGs which had increased methylation, blue = genes with CpGs which had decreased methylation, green = genes with CpGs which had both increased and decreased methylation in IPF compared to control lung fibroblasts. For expression: yellow = genes which had increased expression. +P = phosphorylation, e = gene expression relationship, HA = Hyaluronic acid, dotted lines = indirect effect on gene, purple boxes = joining pathways.

Actin cytoskeletal organisation, which is strongly linked to cell contractility and focal adhesions, form mechanical links between cells and ECM and thus provide feedback loops to the ECM. To validate function analysis data, the role of TNXB in tissue contraction was determined using collagen gel contraction assays. Lung fibroblast-mediated collagen gel contraction was significantly ($P<0.05$) increased in IPF compared to control lung fibroblasts at 24, 48 and 72 hours (**Figure 4.11.2**) and TNXB knockdown significantly ($P<0.05$) decreased collagen gel contraction in all IPF lung fibroblast cell lines suggesting an important role for TNXB in tissue contractility (**Figure 4.11.3**)

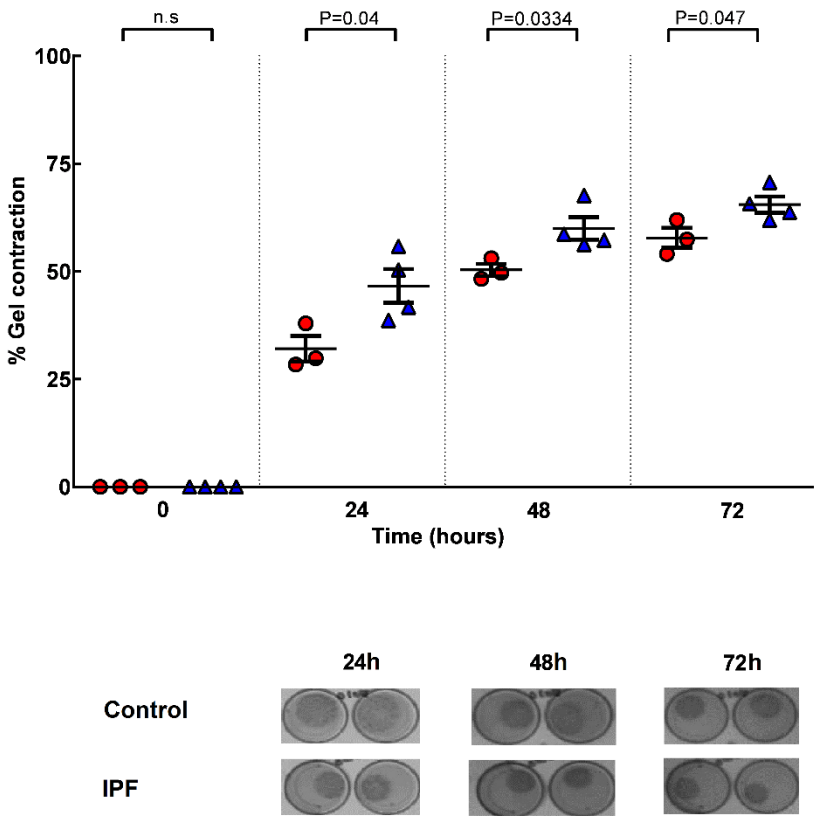


Figure 4.11.2. Lung fibroblast-mediated collagen gel contraction in control and IPF lung fibroblasts. Control ● (n=3) and IPF ▲(n=4) lung fibroblast-mediated contraction of collagen gels was measured between 0-72 hours. Each point represents data from two replicate experiments for each individual primary lung fibroblast cell line. Data presented as the mean ± SEM. Photographs show 1 control and 1 IPF cell lines contracting collagen gels at 24, 48 and 72 hours.

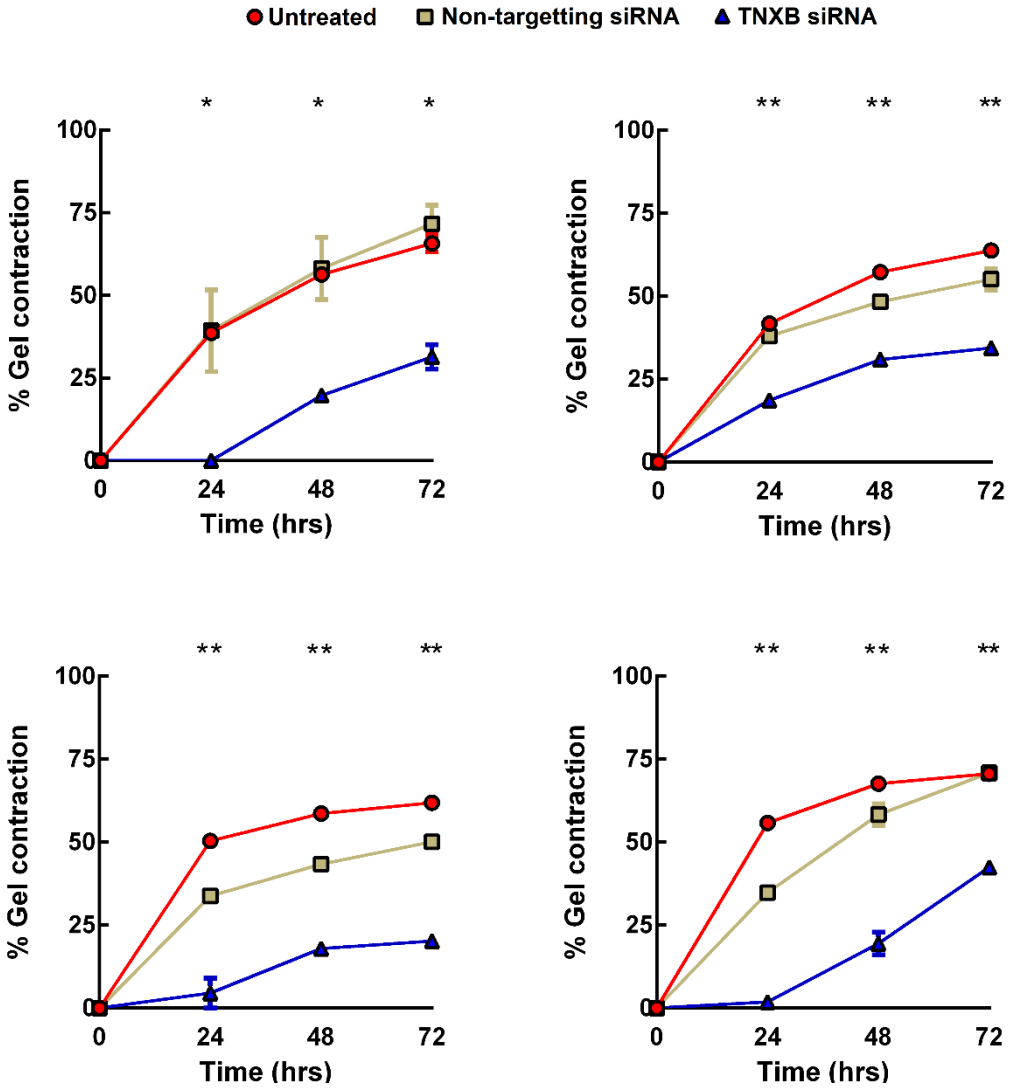


Figure 4.11.3. Lung fibroblast-mediated collagen gel contraction in IPF lung fibroblasts after TNXB knockdown. TNXB was knocked down in IPF (n=4) lung fibroblast cell lines using TNXB siRNA. Lung fibroblasts-mediated contraction of collagen gels was measured between 0-72 hours. Each point represents data from two replicate experiments for each individual primary lung fibroblast cell line. ● = untreated, ■ = non-targeting siRNA and ▲ = TNXB siRNA. Data presented as the mean ± SEM. One-way ANOVA with Tukey's post hoc test was performed for statistical analysis; *P<0.05, **P<0.01. Error bars not shown fall within the data point.

TNXB can interact with multiple ECM components including collagens, integrins and proteoglycans and deficiency has previously been associated with increased PPARγ expression and decreased TGFß expression (Jing et al, 2011). To determine whether TNXB could regulate PPARγ and/or TGFß, IPF lung fibroblasts were treated with 10nM of a non-targeting control siRNA and a TNXB siRNA for 48 hours. TNXB was significantly (P=0.002) knocked down in IPF lung fibroblasts using TNXB siRNA compared

with non-targeting control siRNA, however, there was no significant change in PPARγ (P=0.854) or TGFß (P=0.255) expression (**Figure 4.11.4**). Expression of ECM-associated genes including collagens (COL6A3, COL12A1) and decorin (DCN) were also unchanged after TNXB knockdown (**Figure 4.11.5**).

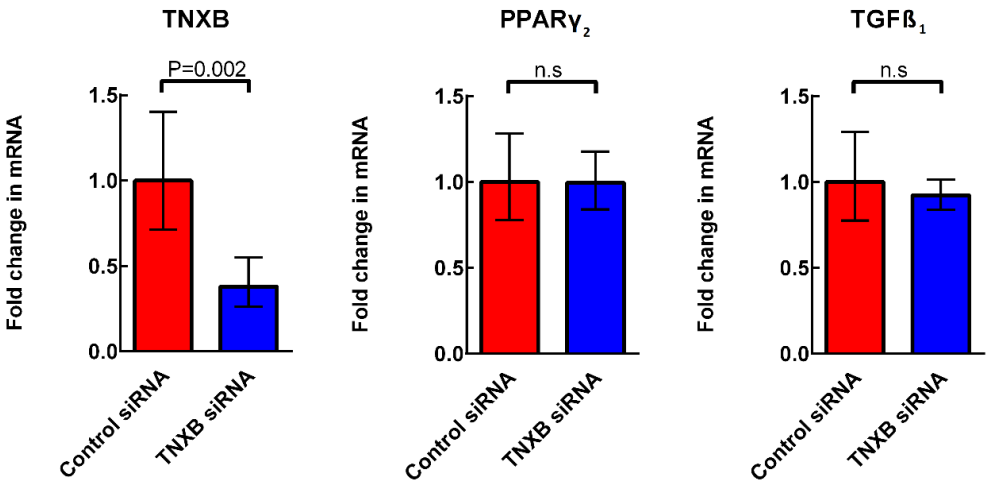


Figure 4.11.4. TNXB, PPARγ$_2$ and TGFß$_1$ expression after knockdown of TNXB. Quantitative PCR analysis of TNXB, PPARγ$_2$ and TGFß$_1$ after 48 hours using 10nM of TNXB siRNA and 10nM of a control non-targeting siRNA in IPF (n=5) fibroblasts. Data shown as the geometric mean of triplicate experiments with error bars showing ± 95% CIs.

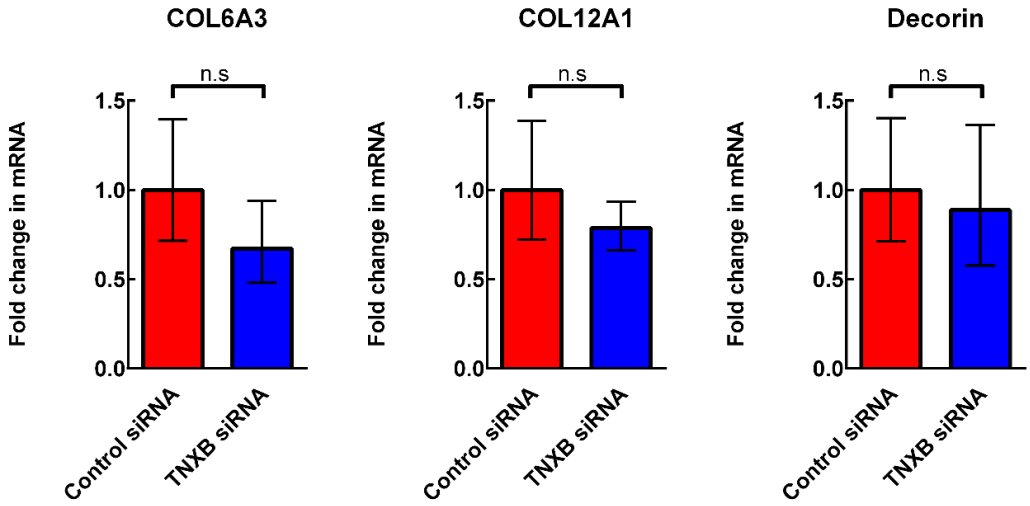


Figure 4.11.5. ECM-associated gene expression after knockdown of TNXB. Quantitative PCR analysis of COL6A3, COL12A1 and Decorin after 48 hours using 10nM of TNXB siRNA relative to a non-targeting siRNA in IPF (n=5) fibroblasts. Data shown as the geometric mean of triplicate experiments with error bars showing ± 95% CIs.

4.12. Summary

- Combining multiple bioinformatics tools including STRING 10.0, Cytoscape 3.2.1 and R packages, helped identify how genes with significantly altered methylation and expression could potentially interact with each other.

- Mapping genes with significantly altered methylation and/or expression onto KEGG pathways identified genes previously associated with fibrosis but not with altered methylation including Wnts, collagens, as well as novel genes not previously associated with IPF or SSc but associated with pathways which have been associated with fibrosis. These included, Wnt signalling including the genes, Small ubiquitin-related modifier 1 (SENP2), Nucleoside diphosphate kinase (NDK2) and C-terminal-binding protein 2 (CTBP2) and ECM-interactions including the genes, TNXB and Thrombospondin 4 (THBS4).

- PFAM and gene-term enrichment identified TNXB as being important in a number of processes including focal adhesion and actin cytoskeletal organisation both of which are important in regulating cell contractility. A model of cell-mediated extracellular matrix contraction using collagen gels showed IPF lung fibroblasts contracted collagen to a greater extent than control lung fibroblasts and knockdown of TNXB in IPF lung fibroblasts significantly reduced collagen gel contraction.

- TNXB knockdown did not have a significant effect on PPARγ_2 and TGF$ß_1$ expression, suggesting TNXB does not directly regulate these genes in lung fibroblasts.

Chapter 5. Differentially methylated genes in male compared to female lung fibroblasts

5.1. Overview

DNA methylation influences a number of cellular processes such as genomic imprinting (Nicholls et al, 1989) and X-chromosome inactivation (Mohandas et al, 1981, Payer and Lee, 2008). Although sex-differences across the epigenome remain poorly understood, during the course of my PhD it became apparent from the emerging literature that gender may influence methylation in multiple diseases. For example, Pinto et al, 2013 showed that in familial breast cancer, males had a different methylation and miRNA expression pattern compared to females. Furthermore, sex-specific differences in methylation can alter cell phenotypes (Hall et al, 2014) and sex-specific differences in methylation during brain development can affect multiple genes (Spiers et al, 2015). This has led to speculation that a number of neurological diseases with a sex-bias may be linked to altered methylation.

There is unquestionable evidence that IPF predominates in males and SSc predominates in females, however, no study has reported global methylation differences between male and female lung fibroblasts or within the context of IPF lung fibroblasts. Therefore methylation differences between males and females may be important in determining disease outcome or could be used to target different therapeutic treatments in these sex-bias diseases.

Using male (n=4) and female lung fibroblasts (n=7) derived from control (n=6, 2 male/4 female) and IPF (n=5, 2 male/3 female) lung fibroblasts, I show that multiple genes on autosomes and the X-chromosome have CpGs with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male compared to female lung fibroblasts. Much of the heterogeneity observed in control and IPF lung fibroblasts was accounted for by differences in methylation and expression between male and females. Furthermore, multiple biological processes enriched in genes with altered methylation in male compared to female lung fibroblasts were the same biological processes which were enriched in genes with altered methylation in IPF and SSc compared to control lung fibroblasts. This data suggests that multiple biological processes potentially involved in the pathogenesis of IPF are enriched in CpGs/genes with altered methylation in male compared to female lung fibroblasts.

5.2. Genome wide distribution of methylation in male and female lung fibroblasts

Male (n=4) and female (n=7) lung fibroblasts had a similar genome-wide bimodal distribution of methylation patterns on all autosomes, with the highest frequencies of CpGs having 0-15% and 85-90% methylation (**Figure 5.2.1**). This was also true for the X-chromosome in male lung fibroblasts, however, in female lung fibroblasts the CpG distribution on the X-chromosome showed a partially methylated methylation pattern with the highest frequency of CpGs having 40-50% methylation (**Figure 5.2.1**) which is consistent with X-chromosome inactivation. This partially methylated pattern of CpG methylation which was apparent in island, shore, shelf and open sea regions, within and beyond 1.5kb of their corresponding genes TSS (**Figure 5.2.2**). In male lung fibroblasts, island regions within 1.5kb of their corresponding genes TSS had a unimodal distribution of CpG methylation with the highest frequency of CpGs having low (0-20%) methylation. In contrast, there was a bimodal pattern of methylation for CpGs located in island regions further than 1.5kb away from their gene's corresponding genes TSS in males (**Figure 5.2.2**)

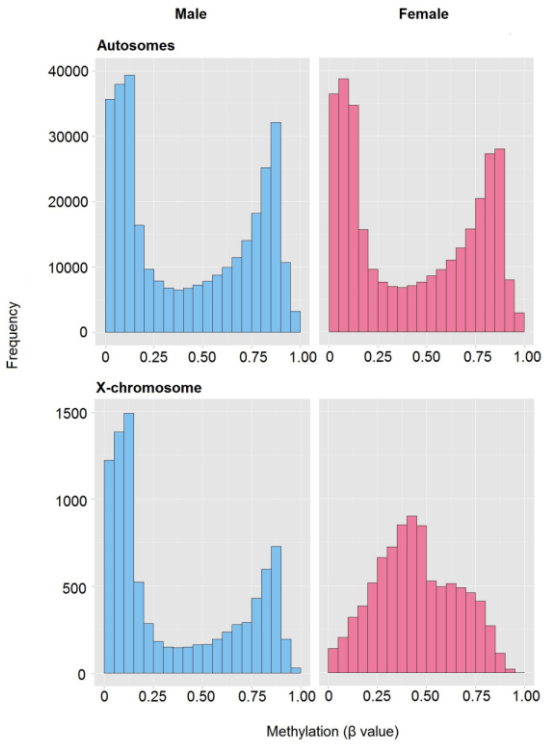


Figure 5.2.1. Distribution of CpG methylation in autosomes and the X-chromosome in male and female lung fibroblasts. Each bar represents the number of CpGs within each boundary. The β value represents the level of methylation (0= unmethylated, 1= methylated). Blue: male (n=4) and pink: female (n=7) lung fibroblasts.

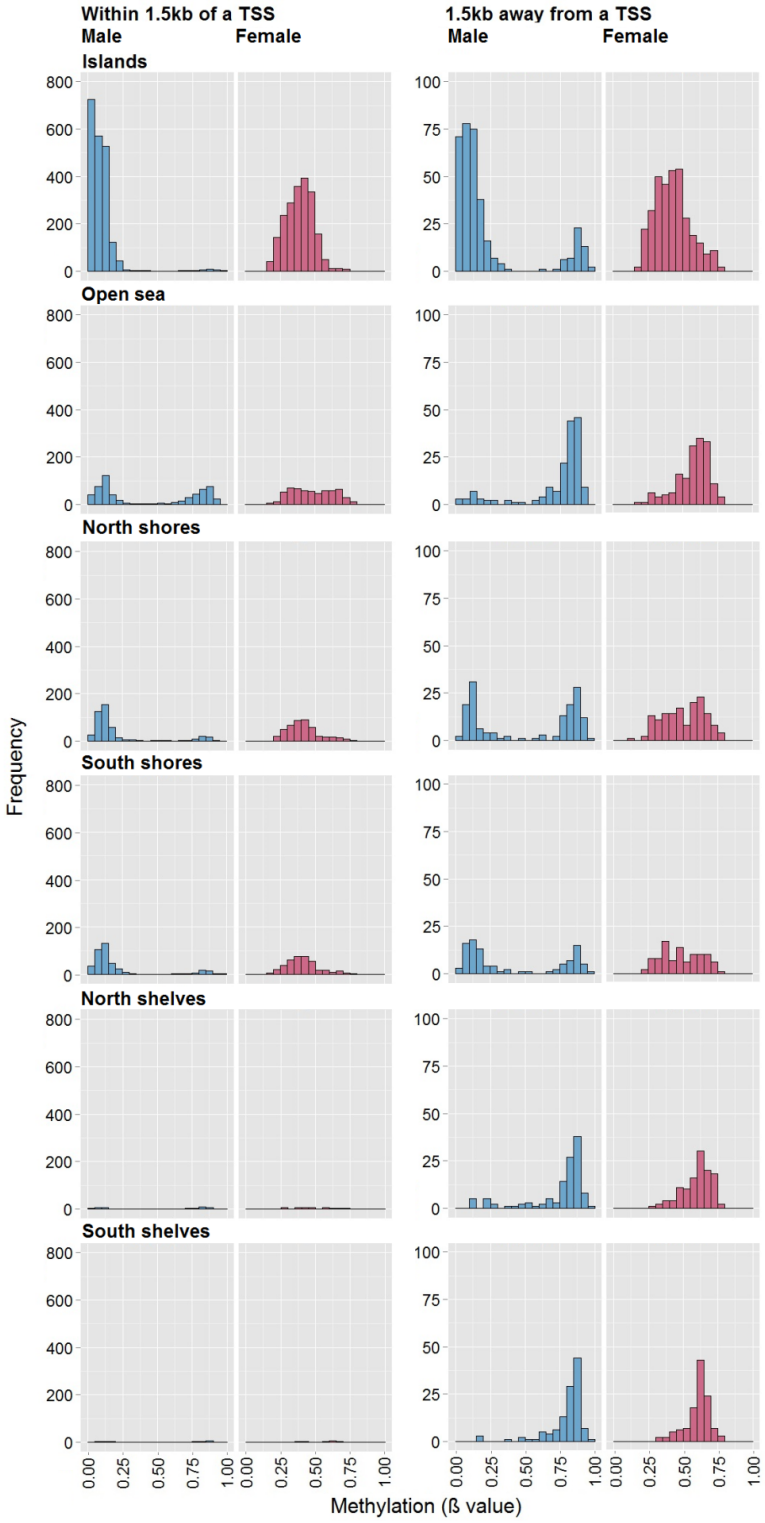


Figure 5.2.2. CpG distribution of CpGs in relation to their corresponding genes TSS. CpGs which had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male (n=4) compared to female (n=7) lung fibroblasts. The β value represents the level of methylation (0= unmethylated, 1= methylated).

5.3. Microarray analysis of differentially methylated genes in male compared to female lung fibroblasts

The Illumina Infinium Human Methylation 450k BeadChip microarray identified multiple CpG sites with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in male (n=4) compared to female (n=7) lung fibroblasts. In male lung fibroblasts, 8979 CpGs corresponding to 4131 genes had significantly altered methylation (Δβ ≥ 0.136; P<0.05) compared to female lung fibroblasts. 50% were located on the X-chromosome, 49% on autosomes and 1% on the Y-chromosome. Of these 8979 CpGs, 2404 CpGs corresponding to 1730 genes had significantly increased methylation (Δβ ≥ 0.136; P<0.05) and 6575 CpGs corresponding to 2928 genes had significantly decreased methylation (Δβ ≥ 0.136; P<0.05) in male compared to female lung fibroblasts (**Figure 5.3.1**). Examples of genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in male compared to females lung fibroblasts are shown in **Figure 5.3.2**.

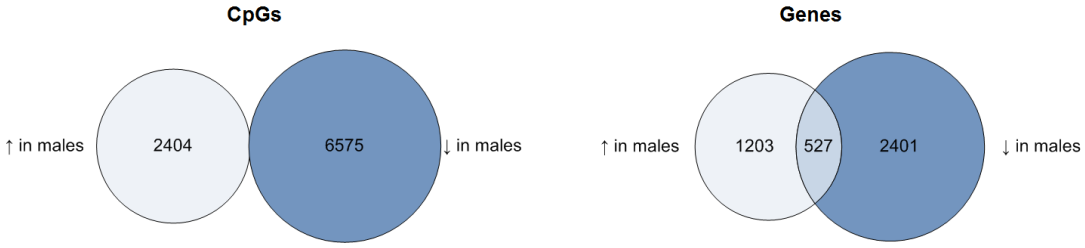


Figure 5.3.1. Number of CpGs/genes with altered methylation in male compared to female lung fibroblasts. CpGs/genes with significantly (Δβ ≥ 0.136; P<0.05) increased or decreased methylation in male (n=4) compared with female (n=7) lung fibroblasts.

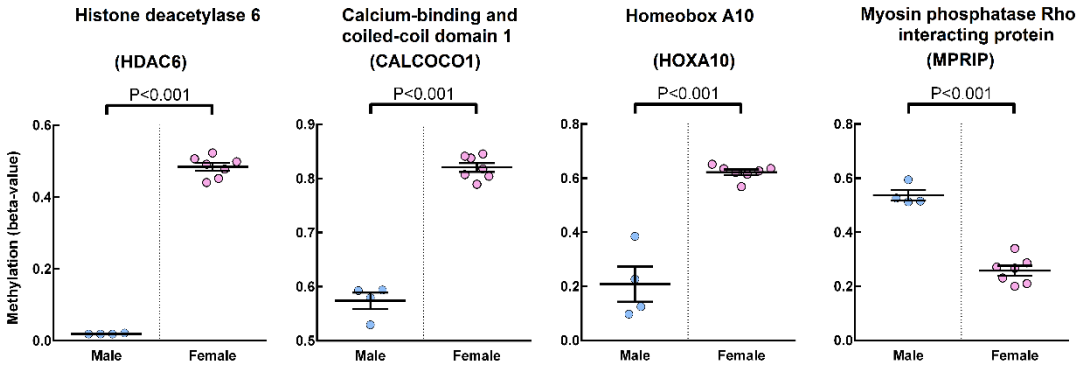


Figure 5.3.2. CpGs with altered methylation in male compared to female lung fibroblasts. Example of 4 CpGs/genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in male (n=4) compared to female (n=7) lung fibroblasts. Beta-value (β) indicates methylation. 0 = 0% methylated, 1= 100% methylated. Error bars represent the mean +-S.E.M.

The number of differentially methylated CpGs between male and female lung fibroblasts and where they are located in relation to CpG islands is shown in **Figure 5.3.3**. North shelf (O/E: 1.90) and south shelf (O/E: 1.91) regions had almost double the expected number of CpGs with increased methylation in male compared to female lung fibroblasts. North shore (O/E: 1.25) and south shore (O/E: 1.09) and open sea (O/E: 1.26) regions also had more than the expected number of CpGs with increased methylation in male compared to female lung fibroblasts. The number of CpGs observed in island regions with increased methylation in male compared to female lung fibroblasts were fewer than expected (O/E: 0.36). In contrast, CpGs sites with decreased methylation in males were more abundant than expected in island regions (O/E: 1.35) and fewer than expected in north shelf (O/E: 0.61) and south shelf (O/E: 0.51) and open sea (O/E: 0.67) regions (**Table 5.3.1**). Cluster analysis clearly identified male lung fibroblast cell lines cluster separately from female lung fibroblast cell lines, independent of whether they were derived from IPF or control patients (**Figure 5.3.4**).

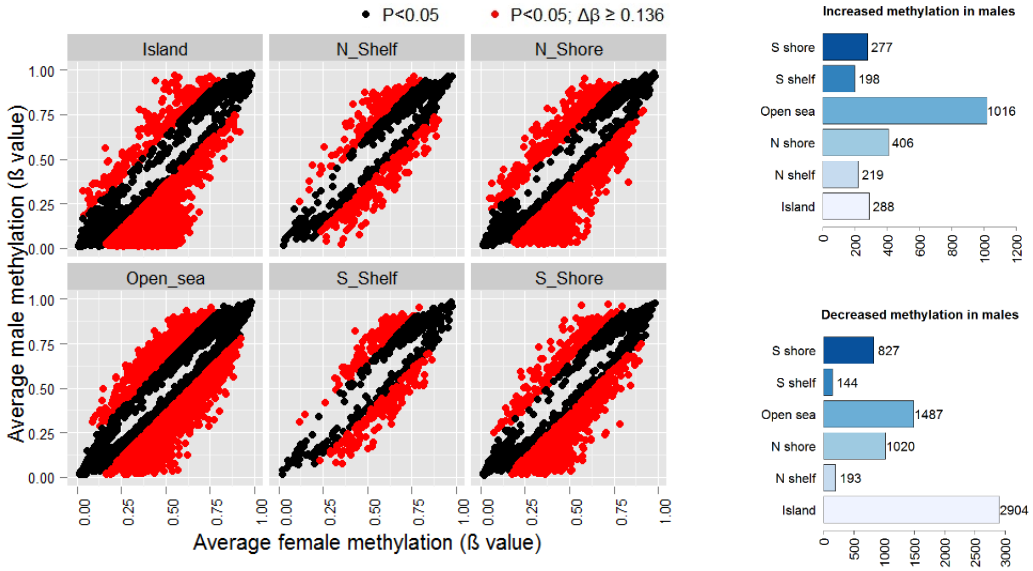


Figure 5.3.3. CpGs with altered methylation in relation to CpG islands in male compared to female lung fibroblasts. Left: scatter plots showing the average methylation and the location of each CpG with a significant difference (P<0.05) in methylation between male (n=4) and female (n=7) lung fibroblasts. Each dot represents a CpG with either a Δβ< 0.136 (<13.6% change in methylation) (●) or a Δβ ≥ 0.136 (≥ 13.6% change in methylation) between male and female lung fibroblasts (●). Right: bar graphs showing the number of CpGs in different locations with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in male compared to female lung fibroblasts.

Location	Total no. of CpGs	CpGs with increased methylation in males			CpGs with decreased methylation in males		
		Observed	Expected	O/E	Observed	Expected	O/E
Island	108197	288	800	0.36*	2904	2189	1.33*
N_Shelf	15598	219	115	1.90*	193	316	0.61*
N_Shore	43922	406	325	1.25*	1020	889	1.15*
Open_sea	108961	1016	806	1.26*	1487	2205	0.67*
S_Shelf	14040	198	104	1.91*	144	284	0.51*
S_Shore	34255	277	253	1.09	827	693	1.19*
Total	**324973**						

Table 5.3.1. Observed to expected ratio (O/E) of the number of CpGs in each location with altered methylation in male compared to female lung fibroblasts. CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in male (n=4) compared to female (n=7) lung fibroblasts. *= Chi-square value; P<0.05.

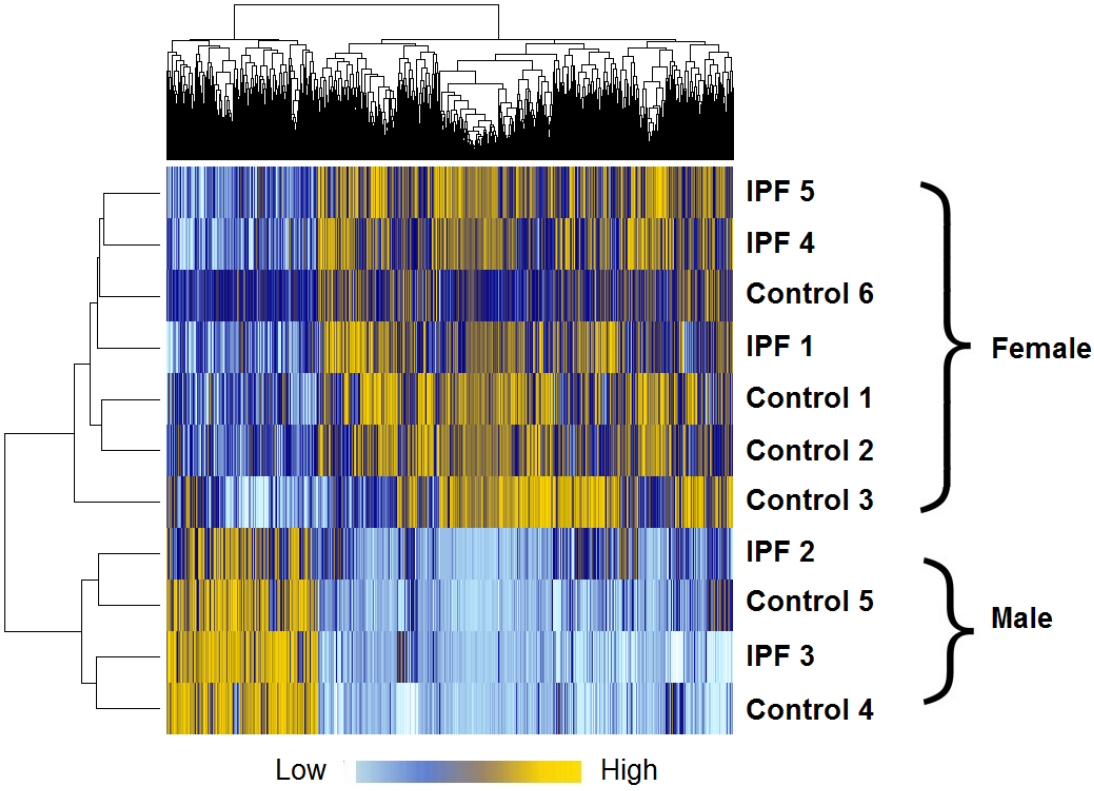


Figure 5.3.4. Hierarchical clustering based on CpGs with altered methylation in male compared to female lung fibroblasts. Heat-map showing the CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in male (n=4) compared to female (n=7) lung fibroblasts. Light blue represents low methylation, yellow represents high methylation.

5.4. Genome-wide distribution of gene expression in male and female lung fibroblasts

The distribution of gene expression in male and female lung fibroblasts was similar with the highest frequency of genes having low expression (**Figure 5.4.1**). The highest frequency of genes with significantly altered expression (TNoM ≤1; P<0.05) in male compared to female lung fibroblasts had low expression (**Figure 5.4.2**).

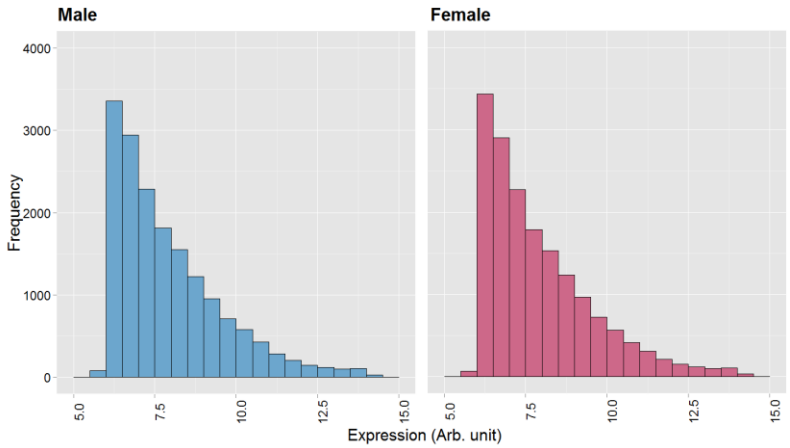


Figure 5.4.1. Genome-wide distribution of gene expression in male and female lung fibroblasts. Each bar represents the number of genes within each boundary. Blue = male; pink = female lung fibroblasts.

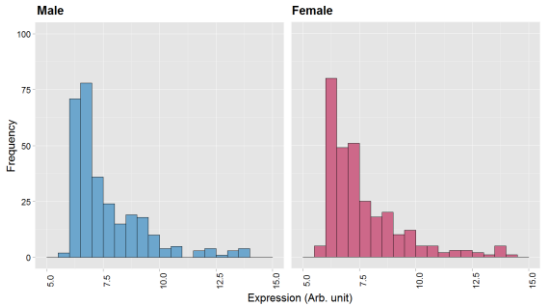


Figure 5.4.2. Distribution of genes which had altered expression in male compared to female lung fibroblasts. Each bar represents the number of genes within each boundary which had significantly altered expression (TNoM ≤1; P<0.05) in male (n=4) compared to female (n=7) lung fibroblasts. The raw value for expression is shown on the x-axis. Blue = male; pink = female lung fibroblasts.

5.5. Microarray analysis of differentially expressed genes in males compared to female primary human lung fibroblasts

The Illumina Human Expression array identified 297 genes (137 increased, 160 decreased) with significantly altered expression (TNoM ≤1; P<0.05) in male compared to female lung fibroblasts. The average expression for each gene differentially expressed in male and female lung fibroblasts and which chromosomes the genes mapped to is shown in **Figure 5.5.1**. Examples of genes with large significant changes in expression (TNoM ≤1; P<0.05) in male compared to female lung fibroblasts are shown in **Figure 5.5.2**.

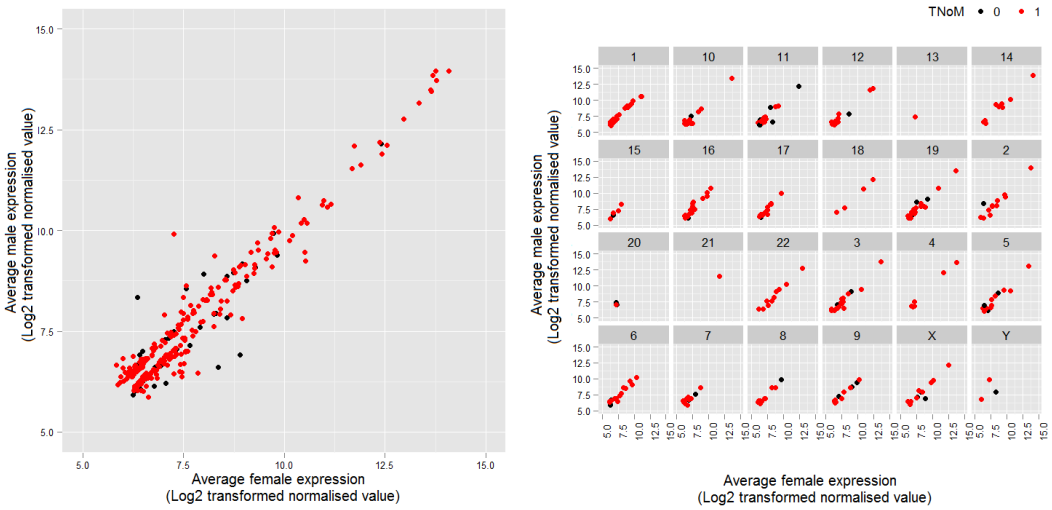


Figure 5.5.1. Gene expression in male compared with female lung fibroblasts. Scatter graphs show the average expression of genes with significantly increased or decreased expression (TNoM ≤1; P<0.05) in male (n=4) compared with female (n=7) lung fibroblasts. Right: shows which chromosome each differentially expressed gene is on. ● Show genes with a TNoM=0, ● show genes with a TNoM=1.

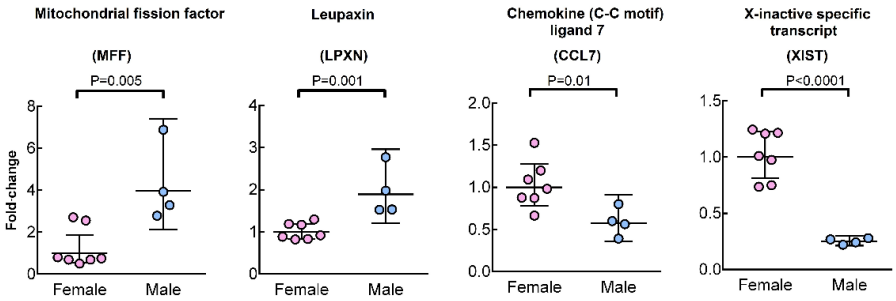


Figure 5.5.2. Genes with altered expression in male compared to female lung fibroblasts. Four examples of individual genes including MFF, LPXN, CCL7 and XIST which have significantly altered expression (TNoM ≤1; P<0.05) in male (n=4) compared to female (n=7) lung fibroblasts. Fold-changes are relative to the geometric mean of female gene expression. Error bars represent 95% confidence intervals.

Cluster analysis of the 297 differentially expressed genes clearly distinguished male from female lung fibroblasts irrespective of whether they were dervived from control or IPF lung fibroblasts. (**Figure 5.5.3**).

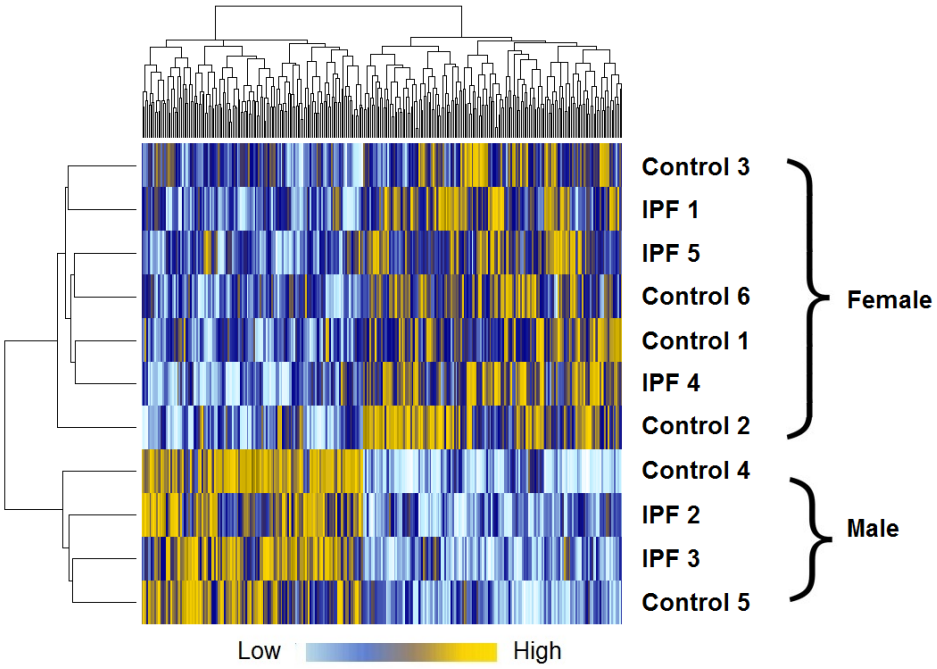


Figure 5.5.3. Cluster analysis showing the differentially expressed genes in male compared to female lung fibroblasts. Genes (n=297) with significantly altered expression (TNoM ≤1; P<0.05) in male (n=4) compared to female (n=7) lung fibroblasts. Light blue represents low expression, yellow represents high expression with respect to each gene across all cell lines.

5.6. Differentially methylated CpGs in IPF males compared to control males and IPF females compared to control females

As previously shown, IPF and SSc have multiple CpG sites which have altered methylation compared to control lung fibroblasts and multiple CpG sites have altered methylation in male compared to female lung fibroblasts. Differences in methylation of male and female lung fibroblasts could underlie sex differences in vulnerability of developing sex-biased diseases such as IPF or SSc and potentially underlie sex differences in vulnerabilities to drugs in order to treat such diseases. To explore whether these methylation difference exist between male and female IPF lung fibroblasts, male IPF lung fibroblasts were compared to male control lung fibroblasts and female IPF lung fibroblasts were compared to female control lung fibroblasts.

Genome-wide distribution of CpG methylation in IPF males compared to control males was similar on autosomes and on the X-chromosome with both control and IPF lung fibroblasts displaying a bimodal distribution. However, the distribution of CpG methylation in male IPF compared with male control

lung fibroblasts was different on the Y-chromosome (**Figure 5.6.1**). Genome-wide distribution of CpG methylation in IPF female and control female was similar on autosomes and the X-chromosome, with the highest frequency of CpGs having low (0-15%) and (high 80-90%) methylation across autosomes and 40-50% methylation on the X-chromosome (**Figure 5.6.2**).

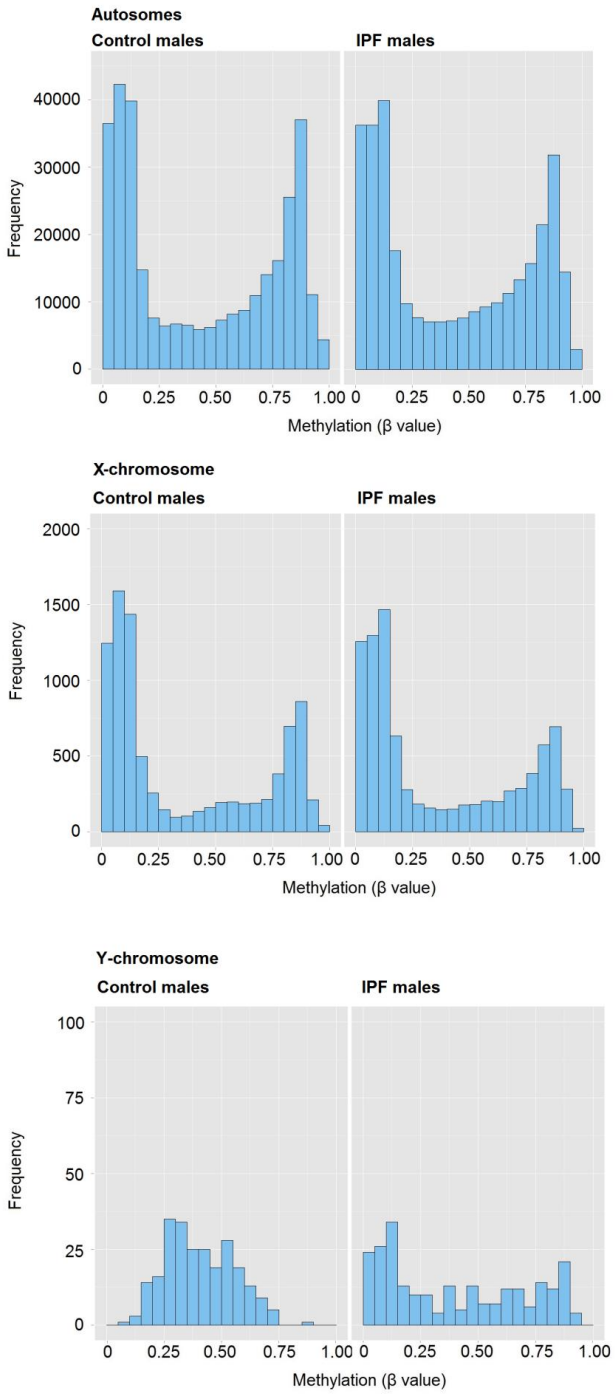


Figure 5.6.1. Genome-wide distribution of CpGs on autosomes, the X-chromosome and the Y-chromosome in IPF male and control male lung fibroblasts. Each bar represents the number of CpGs within each boundary. Average methylation values (β value) are shown for IPF male (n=2) and control male (n=2) lung fibroblasts.

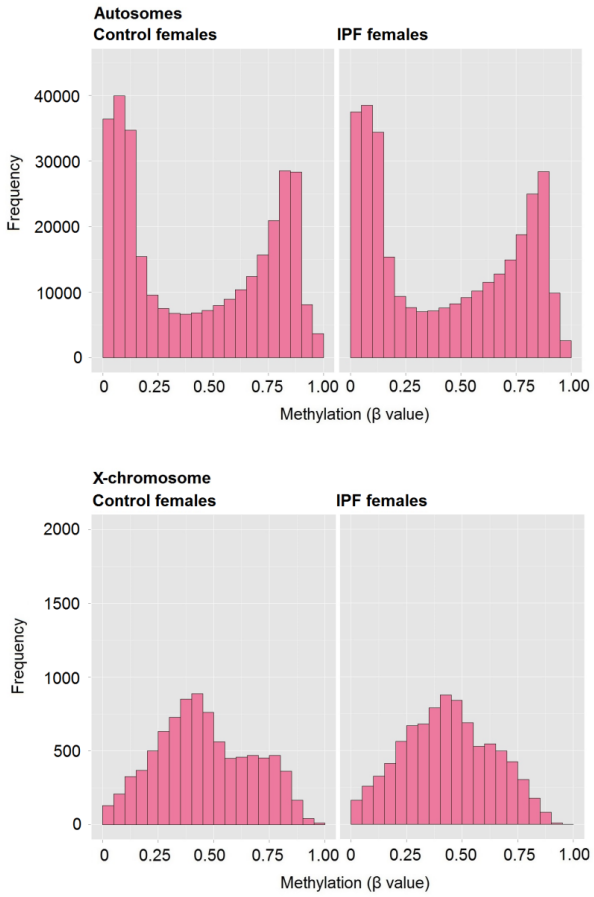


Figure 5.6.2. Genome-wide distribution of CpGs on autosomes and the X-chromosome and in IPF female and control female lung fibroblasts. Each bar represents the number of CpGs within each boundary. Average methylation values (β value) are shown for IPF female (n=3) and control female (n=4) lung fibroblasts.

5.7. Microarray analysis of differentially methylated genes in male IPF compared to male control and female IPF compared to female control primary human lung fibroblasts

CpGs (4467) corresponding to 3493 genes had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male IPF compared to male control lung fibroblasts. Of these, 1823 CpGs corresponding to 1588 genes had significantly increased methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and 2644 CpGs corresponding to 2147 genes had significantly decreased methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male IPF compared to male control lung fibroblasts (**Figure 5.7.1**). CpGs (5437) corresponding to 3853 genes had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in female IPF compared to female control lung fibroblasts. Of these, 2302 CpGs corresponding to 1816 genes had significantly increased methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and 3135 CpGs corresponding to 2331 genes had significantly decreased methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in female IPF compared to female control lung fibroblasts (**Figure 5.7.1**).

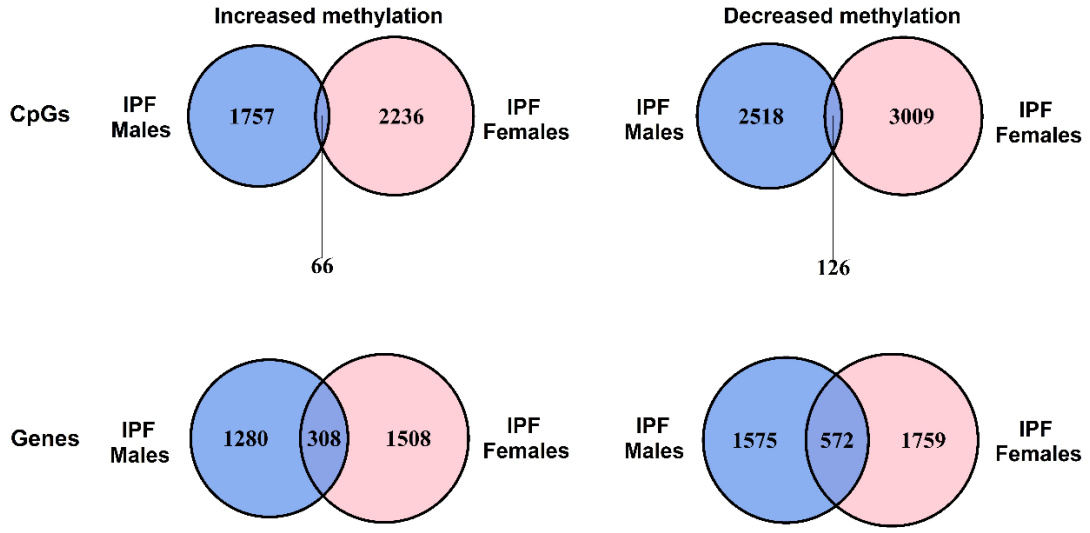


Figure 5.7.1. The number of distinct and overlapping CpGs/genes which have altered methylation in IPF male compared to control male and IPF female compared control female lung fibroblasts. CpGs/genes which have significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF male (n=2) compared to control male (n=2) and IPF female (n=3) compared to control female (n=4) lung fibroblasts.

Of the 1823 CpG sites which had increased methylation in male IPF compared to male control lung fibroblasts, only 66 were the same CpG sites which had increased methylation in female IPF compared to female control lung fibroblasts. The 1823 CpG sites with increased methylation in male IPF compared to male control lung fibroblasts corresponded to 1588 genes of which 308 were the same genes with increased methylation in female IPF compared to female control lung fibroblasts. Of the 2644 CpG sites which had decreased methylation in male IPF compared to male control lung fibroblasts, 126 were the same CpG sites which had decreased methylation in female IPF compared to female control lung fibroblasts. The 2644 CpG sites with decreased methylation in male IPF compared to male control lung fibroblasts corresponded to 2147 genes of which 572 were the same genes with decreased methylation in female IPF compared to female control lung fibroblasts. Very few CpGs and only 25% of genes overlapped suggesting that differences in CpG methylation between male IPF and male controls are not the same CpGs with differences between female IPF and female controls including multiple ECM associated gene (**Figure 5.7.4**). Representative genes are shown in **Figure 5.7.2**, **Figure 5.7.3 and Figure 5.7.4**. The number of differentially methylated CpGs between male IPF compared to male control and female IPF compared to female control lung fibroblasts and their location in relation to CpG islands is shown in **Figure 5.7.5**.

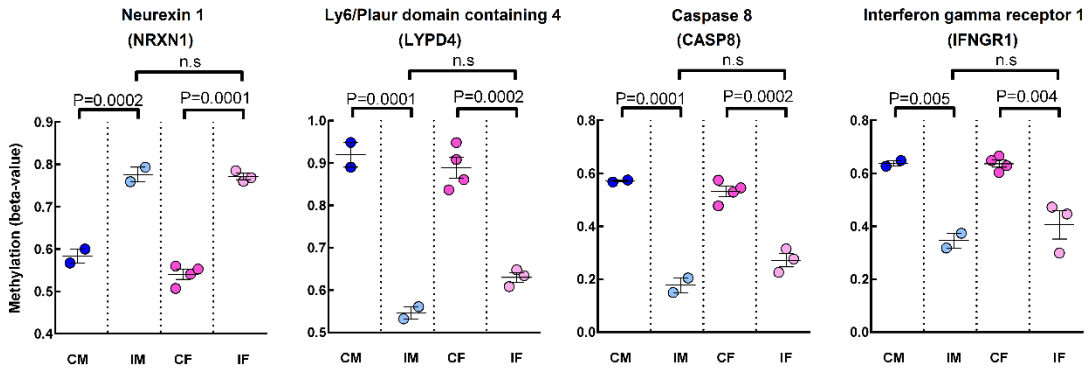


Figure 5.7.2. CpGs and their corresponding gene which had altered methylation in male IPF male compared to control male and in IPF female compared to control female lung fibroblasts. Statistical analysis performed using 1-way ANOVA analysis with Tukey's multiple comparison test identified some CpGs which had significantly altered (P<0.05) methylation in IPF males (IM) (n=2) compared to control males (CM) (n=2) which also had significantly altered (P<0.05) methylation in IPF females (IF) (n=3) compared with control females (CF) (n=4). Error bars represent the mean ± S.E.M.

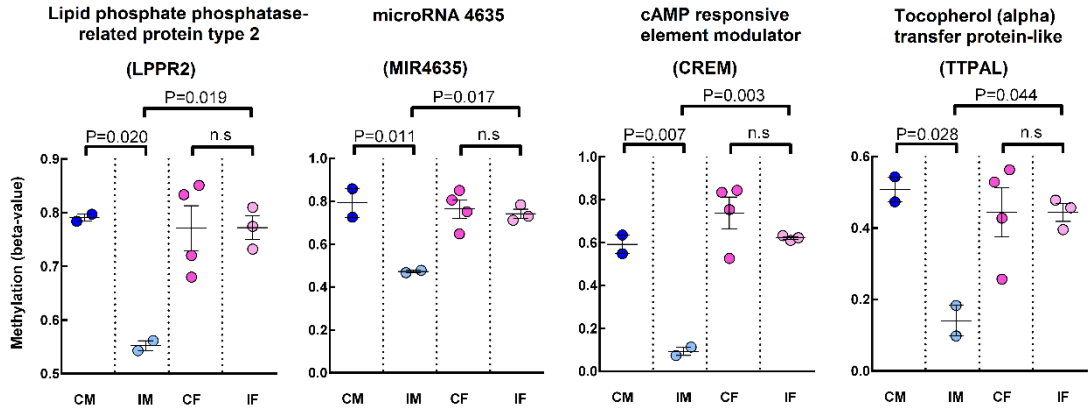


Figure 5.7.3. CpGs and their corresponding genes which had altered methylation in IPF male compared to control male but not in IPF female compared to control female lung fibroblasts. Statistical analysis performed using 1-way ANOVA analysis with Tukey's multiple comparison test identified multiple CpGs which had significantly altered (P<0.05) methylation in IPF males (IM) (n=2) compared to control males (CM) (n=2) and significantly altered (P<0.05) methylation in IPF males (IM) (n=2) compared with IPF females (IF) (n=3). Error bars represent the mean ± S.E.M.

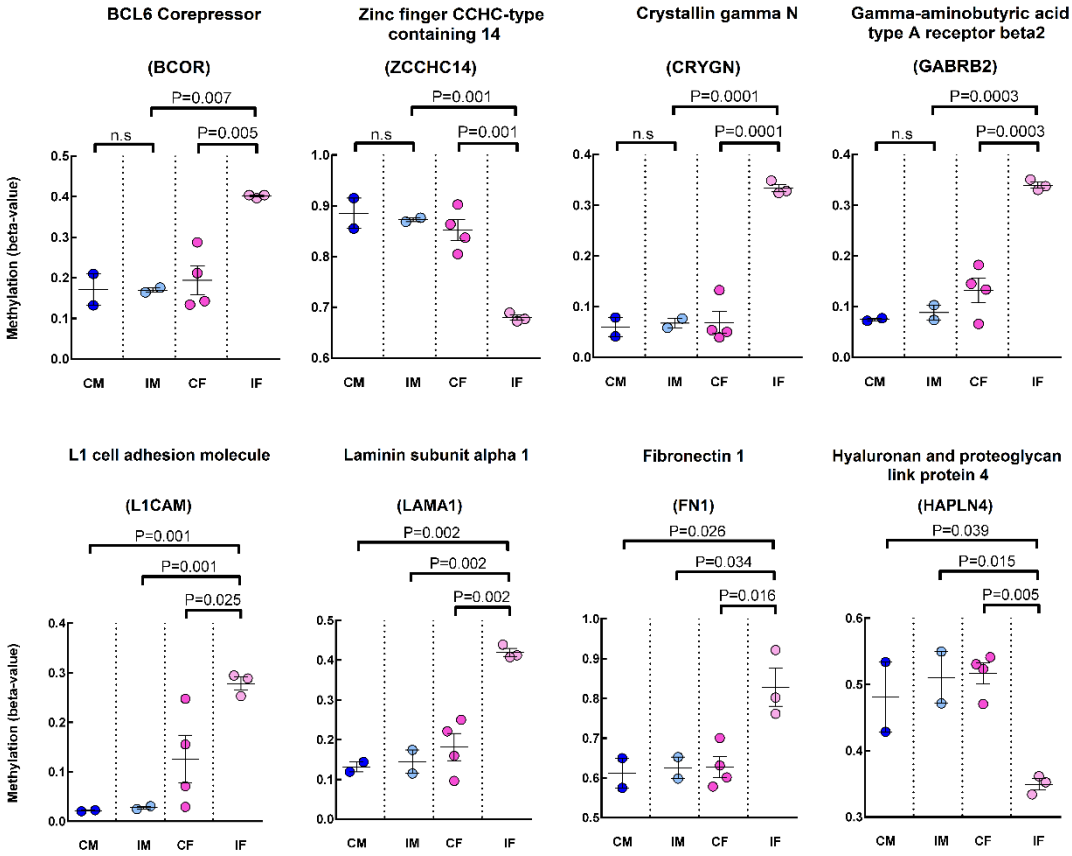


Figure 5.7.4. CpGs and their corresponding genes which had altered methylation in IPF female compared to control female but not in IPF male compared to control male lung fibroblasts. Statistical analysis performed using 1-way ANOVA analysis with Tukey's multiple comparison test identified multiple CpGs which had significantly altered (P<0.05) methylation in IPF females (IF) (n=3) compared to control females (CF) (n=4) and significantly altered (P<0.05) methylation in IPF females (IF) (n=3) compared with IPF males (IF) (n=2). Error bars represent the mean ± S.E.M.

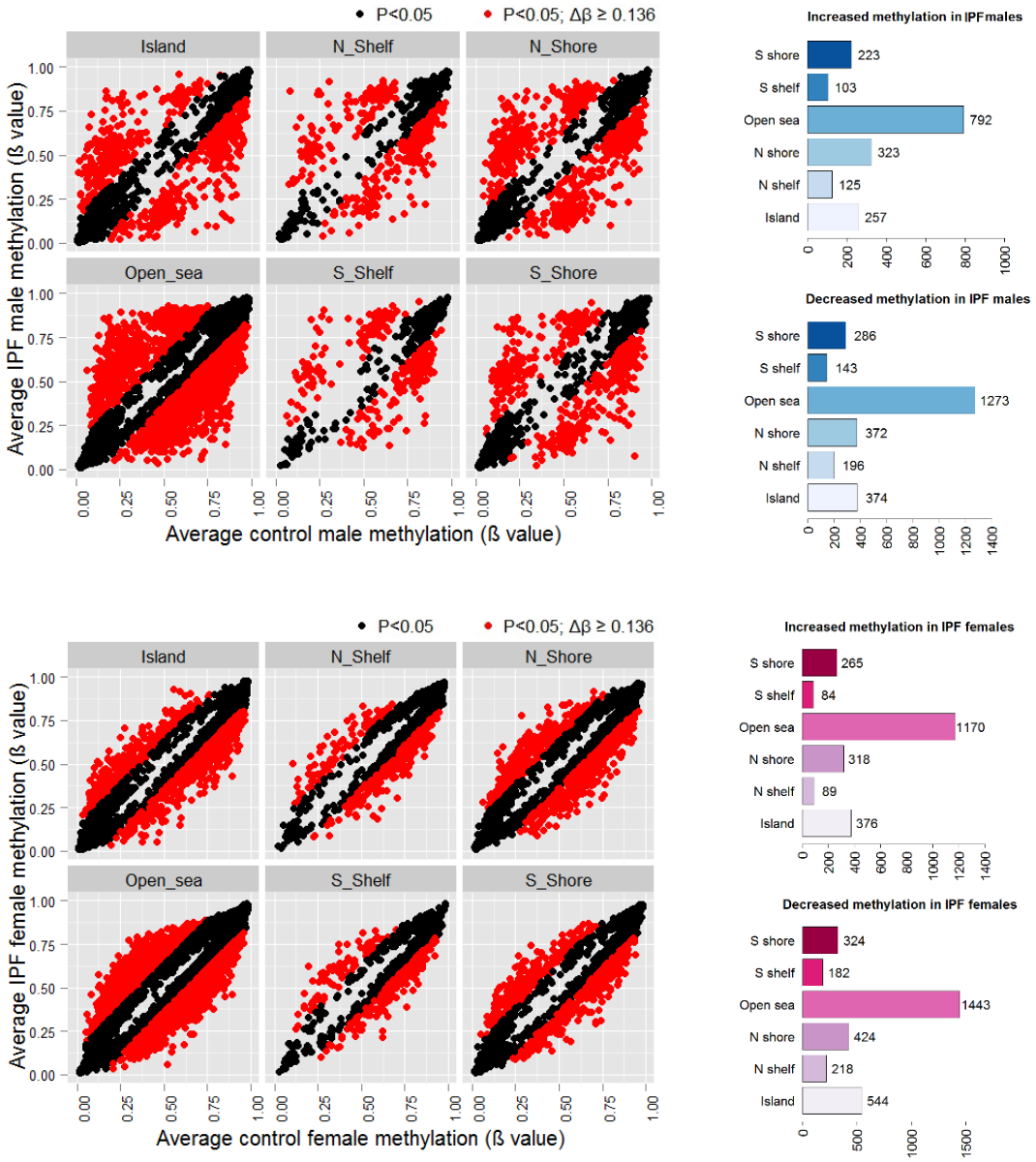


Figure 5.7.5. CpGs with altered methylation in relation to CpG islands in IPF male compared to control male and IPF female compared to control female lung fibroblasts. Scatter plots show average methylation and the location of each CpG with a significant difference (P<0.05) in methylation between IPF male (n=2) compared to control male (n=2) and IPF female (n=3) compared to control female (n=4) lung fibroblasts. Each dot represents a CpG with either a Δβ< 0.136 (<13.6% change in methylation) (●) or a Δβ ≥ 0.136 (≥ 13.6% change in methylation (●). Bar graphs show the number of CpGs and their location in respect to CpG islands.

Shelf, open sea (more than expected) and islands (less than expected) had significantly different O/E ratios in the number of CpGs with decreased methylation in IPF male compared to control male (**Table 5.7.1**) and IPF female compared control female (**Table 5.7.2**) lung fibroblasts. North shore, shelves, open sea (more than expected) and islands (less than expected) had significantly different O/E ratios (Chi^2 = <0.05) in the number of CpGs with increased methylation in IPF male compared to control male whereas in IPF female compared control female lung fibroblasts, only island regions (less than expected) and open sea regions (more than expected) had significantly different O/E ratios (Chi^2 <0.05).

Location	Total no. of CpGs	CpGs with increased methylation in IPF males			CpGs with decreased methylation in IPF males		
		Observed	Expected	O/E	Observed	Expected	O/E
Island	108197	257	607	0.42*	374	880	0.42*
N_Shelf	15598	125	88	1.43*	196	127	1.54*
N_Shore	43922	323	246	1.31*	372	357	1.04
Open_sea	108961	792	611	1.30*	1273	887	1.44*
S_Shelf	14040	103	79	1.31*	143	114	1.25*
S_Shore	34255	223	192	1.16	286	279	1.03
Total	**324973**						

Table 5.7.1. Observed to expected ratio (O/E) of the number of CpGs in each location with altered methylation in IPF male compared to control male lung fibroblasts. CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF male (n=2) compared to control male (n=2) lung fibroblasts. *= Chi-square P<0.05.

Location	Total no. of CpGs	CpGs with increased methylation in IPF females			CpGs with decreased methylation in IPF females		
		Observed	Expected	O/E	Observed	Expected	O/E
Island	108197	376	766	0.49*	544	1044	0.52*
N_Shelf	15598	89	110	0.81	218	150	1.45*
N_Shore	43922	318	311	1.02	424	424	1.00
Open_sea	108961	1170	772	1.52*	1443	1051	1.37*
S_Shelf	14040	84	99	0.84	182	135	1.34*
S_Shore	34255	265	243	1.09	324	330	0.98
Total	**324973**						

Table 5.7.2. Observed to expected ratio (O/E) of the number of CpGs in each location with altered methylation in IPF female compared to control female lung fibroblasts. CpGs with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF female (n=3) compared to control female (n=4) lung fibroblasts. *= Chi-square P<0.05.

5.8. Comparing expression profiles of IPF males with control males

As IPF affects males more than females and clear differences in gene expression were observed when comparing male with female lung fibroblasts, IPF males (n=2) were compared with control males (n=2) and IPF females (n=3) were compared with control females (n=4) in order to determine gene expression differences between IPF males and IPF females. Because of low N numbers in the male group, a TNoM of ≤1 and a large fold-change (≥2FC) instead of a P value <0.05 was used to find differentially expressed genes. 407 genes (204 decreased, 203 increased) had altered expression (TNoM ≤1; ≥2FC) in IPF male lung fibroblasts compared to control male lung fibroblasts. Seventy-five genes (28 decreased, 47 increased) had altered expression (TNoM ≤1; ≥2FC) in IPF female lung fibroblasts compared to control female lung fibroblasts. Interestingly, 5 genes located on the Y-chromosome (Ribosomal protein S4, Y isoform 1 (RPS4Y1), Eukaryotic translation initiation factor 1A, Y-linked (EIF1AY), Taxilin gamma pseudogene, Y-linked, (CYorf15A), Neuroligin 4, Y-linked (NLGN4Y) and Jumonji/ARID domain-containing protein 1D (JARID1D) (a histone demethylase) all had large significant increases (≥2FC) in IPF male compared to control male lung fibroblasts and compared to IPF and control female lung fibroblasts (**Figure 5.8.1**).

Multiple genes involved in interferon signalling were downregulated in male IPF compared to male control lung fibroblasts although these differences were not significant. Gene expression differences between fibroblasts of the same sex and origin appeared highly heterogeneous, unlike CpG methylation where fibroblasts of the same sex and origin were homogenous. Nonetheless, these data identified some genes which had large changes in expression and suggest that sex-differences in gene expression in patients with IPF may exist. Further studies using more cell lines should delineate the full extent of sex-specific differences in gene expression in IPF lung fibroblasts.

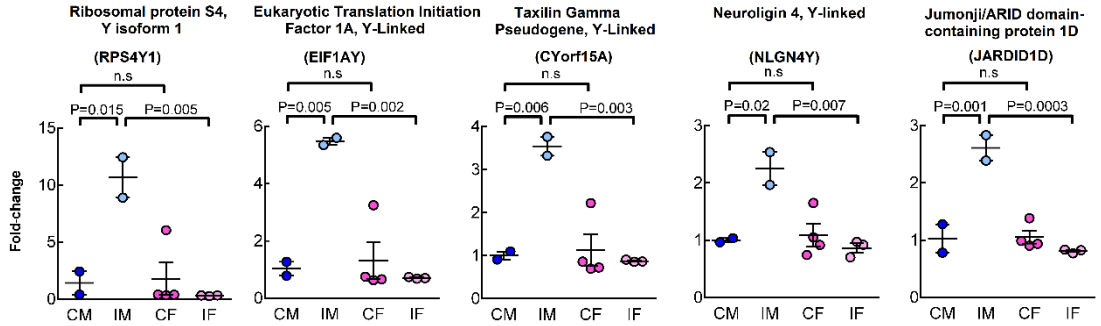


Figure 5.8.1. Genes with altered expression in IPF male compared to control male lung fibroblasts located on the Y-chromosome. Five genes including RPS4Y1, EIF1AY, CYorf15A, NLGN4Y and JARID1D had significantly altered expression (TNoM ≤1; P<0.05) in IPF male (IM) (n=2) compared control to male (CM) (n=2), control female (CF) (n=4) and IPF female (IF) (n=3) lung fibroblasts. Data presented as the geometric mean fold changes relative to the average control male gene expression ± 95% confidence intervals.

5.9. Summary

- Multiple genes have significantly altered methylation and/or expression in male compared to female lung fibroblasts.

- 49% of genes with significantly altered methylation in male compared to female lung fibroblasts were located on autosomes.

- Only 192 CpGs (4.3%) with significantly altered methylation in IPF males compared to control males were the same CpG sites which had significantly altered methylation in IPF female compared to control female lung fibroblasts, suggesting changes to methylation in IPF occur at different CpGs in male compared to female lung fibroblasts.

- 880 genes (25.2%) with significantly altered methylation in IPF males overlapped with females, suggesting multiple genes are common between both sexes, however multiple genes were distinct to each sex, including multiple ECM associated genes and may underlie sex-specific susceptibilities towards developing pulmonary fibrosis.

- Genes on the Y-chromosome including, RPS4Y1, EIF1AY, CYorf15A, NLGN4Y and JARID1D had large increases in expression in IPF male compared with control male lung fibroblasts, suggesting they may play an important role in male pulmonary fibrosis.

5.10. Biological processes enriched in PFAMs containing genes with altered methylation in male compared to female lung fibroblasts.

As previously shown in **section 5.3**, multiple genes had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male (n=4) compared to female (n=7) lung fibroblasts. These genes mapped to 1663 different PFAM domains. Biological processes (629) were significantly (FDR $P<0.01$) enriched in these PFAMs including EMT, ECM organisation, collagen fibril organisation, apoptosis, integrin signalling and Wnt signalling. Many of these biological processes have previously been associated with the pathobiology of pulmonary fibrosis and were also significantly enriched in PFAMs containing genes with significantly altered methylation in IPF/SSc compared to control lung fibroblasts (**Table 5.10.1**).

For example, the biological process 'collagen fibril organisation' was significantly enriched in 11 out of 12 PFAMs which contained 89 genes with significantly altered methylation in male compared to female lung fibroblasts and 10 out of 12 PFAMs which contained 134 genes with significantly altered methylation in IPF compared control lung fibroblasts. Of these genes, 55 overlapped (**Figure 5.10.1**). One of these genes, MUC5B, had decreased methylation in IPF compared to controls and in male compared to female lung fibroblasts. A MUC5B polymorphism has been linked to IPF (Seibold et al, 2011) and aberrant methylation surrounding the polymorphism may play a role in increased MUC5B expression in IPF (Helling et al, 2015). These data could suggest that males and females have altered methylation of genes which are associated with multiple biological processes relevant to pulmonary fibrosis, which could in part, explain the sex-bias in IPF and SSc. Furthermore, biological processes including sex determination and pregnancy were identified as being enriched in genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male compared to female lung fibroblasts, as one might expect, suggesting PFAM analysis is a suitable tool for determining enriched processes.

Genes with significantly altered expression in male compared to female lung fibroblasts were associated with 286 PFAMs. Biological process (687) were significantly (FDR<0.01) enriched in these PFAMs including the toll receptor pathway, ECM organisation, apoptosis, Wnt signalling and methylation (**Table 5.10.2**). Of the biological processes enriched in PFAMs containing genes with significantly altered methylation, 441 (70%) overlapped with those enriched in PFAMs containing genes with significantly altered expression in male compared to female lung fibroblasts. These data suggest multiple biological processes relevant to pulmonary fibrosis are enriched in PFAMs containing genes which have significantly altered methylation in male compared with female lung fibroblasts and adds further evidence that differentially methylated/expressed genes are associated with common and specific PFAMs suggesting a strong link between methylation and expression.

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
ECM organization	2	8.73e-09	47	63
Wnt receptor signalling pathway	2	5.49E-06	29	38
Integrin-mediated signalling pathway	3	9.00e-06	16	17
Collagen fibril organization	3	6.09e-04	11	12
Response to virus	2	6.34e-04	29	45
Tissue remodelling	3	1.04e-03	12	14
Regulation of FGFR signalling pathway	3	1.32e-03	10	11
EMT	3	1.76e-03	14	18
JAK-STAT cascade	3	2.07E-03	11	13
Induction of apoptosis by extracellular signals	3	2.87e-03	9	10
Sex determination	3	4.27e-03	15	21
Blood vessel remodelling	4	7.17e-03	6	6
Regulation of gene expression	4	7.17e-03	6	6
Bile acid metabolic process	4	7.17e-03	6	6
Cellular response to mechanical stimulus	3	8.49e-03	9	11

Table 5.10.1. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered methylation in male compared to female lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate (P<0.01). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
Regulation of JAK-STAT cascade	3	5.00E-07	10	18
Respiratory system development	2	8.28e-07	21	82
Epithelial cell differentiation	2	1.74e-05	17	69
Integrin-mediated signalling pathway	3	2.64e-04	7	17
Toll signalling pathway	3	3.89E-04	6	13
Female pregnancy	3	4.49e-04	8	24
Wnt receptor signalling pathway	2	6.05E-04	10	38
Histone H3-K4 methylation	3	1.30E-03	5	11
ECM organization	2	2.66e-03	12	63
Methylation-dependent chromatin silencing	4	5.37e-03	3	5
Response to FGFR stimulus	3	6.10e-03	4	10
Induction of apoptosis by extracellular signals	3	6.10e-03	4	10
Response to virus	2	6.12e-03	9	45
Tissue homeostasis	3	8.69e-03	6	24

Table 5.10.2. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered expression in male compared to female lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate (P<0.01). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

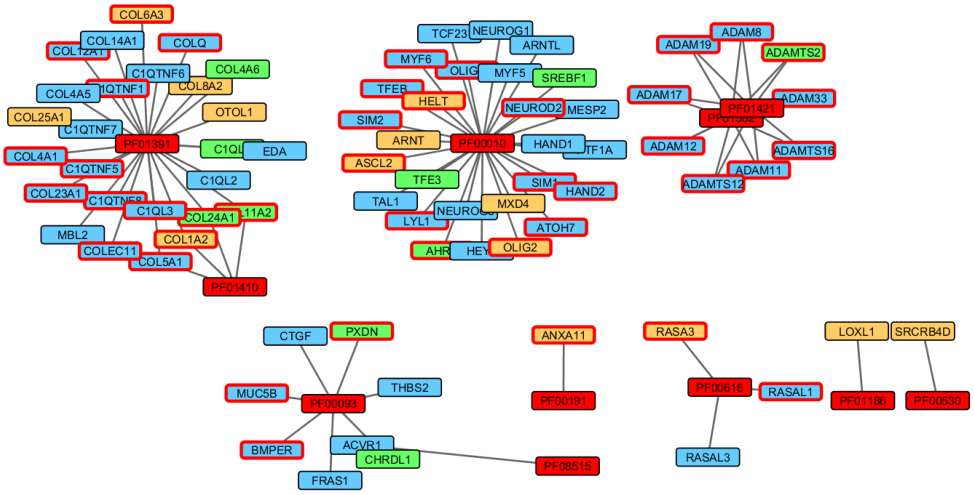


Figure 5.10.1. Collagen fibril organisation enriched in PFAMs containing genes with altered methylation in male compared to female lung fibroblasts. Coloured boxes represent direction of methylation, yellow; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in male (n=4) compared to female (n=7) lung fibroblasts. Boxes with a red border indicate genes which also have significantly altered methylation (Δβ ≥0.136; P<0.05) in IPF (n=5) compared to control (n=6) lung fibroblasts.

5.11. Biological processes enriched in PFAMs containing genes with altered methylation in IPF male compared to control male and IPF female compared to control female lung fibroblasts.

In **section 5.7**, I showed male IPF lung fibroblasts had multiple CpGs/genes with significantly altered methylation and expression compared to male control lung fibroblasts many of which were not the same as the CpGs/genes which had significantly altered methylation and expression in female IPF lung fibroblasts compared to female lung fibroblasts. This potentially suggested that male and female IPF lung fibroblasts may have different methylation and expression patterns. As multiple biological processes enriched in PFAMs containing genes with significantly altered methylation in IPF compared to control and male compared to female lung fibroblasts overlapped, IPF lung fibroblasts were compared to control lung fibroblasts based on gender. This analysis was done to elucidate whether biological processes enriched with PFAMs/genes which had significant differences in methylation/expression in male IPF compared to male control lung fibroblasts were similar to enriched biological processes in female IPF compared to female control lung fibroblasts.

As previously shown in **section 5.7**, 3493 and 3853 genes had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in male IPF (n=2) compared to male control (n=2) and female IPF (n=3) compared to female control (n=4) lung fibroblasts, respectively. These genes mapped to 1529 and 1559 different PFAM domains respectively. In males, 662 biological processes were significantly (FDR P<0.01)

enriched in PFAMs containing genes with significantly altered methylation in male IPF compared to male control lung fibroblasts including EMT, ECM organisation, apoptosis, integrin signalling and Wnt signalling (**Table 5.11.1**). Similarly in females, 673 biological processes were significantly (FDR P<0.01) enriched in PFAMs containing genes with significantly altered methylation in female IPF compared to female control lung fibroblasts including EMT, ECM organisation, collagen fibril organisation, apoptosis, integrin signalling and Wnt signalling (**Table 5.11.2**).

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
ECM organization	2	1.72e-07	43	63
Sex determination	3	7.05e-05	17	21
Wnt receptor signalling pathway	2	8.14E-04	24	38
JAK-STAT cascade	3	9.80E-04	11	13
Induction of apoptosis by extracellular signals	3	1.53e-03	9	10
Integrin-mediated signalling pathway	3	1.53e-03	13	17
Detection of virus	3	4.63e-03	6	6
Bile acid metabolic process	4	4.63e-03	6	6
Methylation	2	6.15e-03	41	83
EMT	3	7.71e-03	7	18
Lung morphogenesis	3	9.82e-03	8	10
Tissue homeostasis	3	9.82e-03	15	24

Table 5.11.1. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered expression in male IPF compared to male control lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate (P<0.01). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
Sex differentiation	2	1.08E-12	58	75
ECM organization	2	5.94E-07	43	63
Integrin-mediated signalling pathway	3	4.33E-06	16	17
Wnt receptor signalling pathway	2	9.32E-06	28	38
Collagen fibril organization	3	3.91E-04	11	12
Induction of apoptosis by extracellular signals	3	2.03E-03	9	10
EMT	3	4.62E-03	13	18
Regulation of JAK-STAT cascade	3	4.62E-03	13	18
Pregnancy	3	4.95E-03	16	24
Bile acid metabolic process	4	5.59E-03	6	6
Methylation	2	7.13E-03	42	83

Table 5.11.2. Domain-centric enrichment of biological processes enriched in PFAMs associated with genes with altered expression in female IPF compared to female control lung fibroblasts. Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate (P<0.01). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

Of biological processes enriched in PFAMs containing genes with significantly altered methylation in male IPF compared to male control fibroblasts, 548 (83%) were the same biological processes enriched in PFAMs containing genes with significantly altered methylation in female IPF compared to female control lung fibroblasts. Interestingly, biological processes distinct to males, enriched in PFAMs containing genes with significantly altered methylation in male IPF compared to male control fibroblasts included some relating to viruses/bacteria, whereas enriched biological processes distinct to females included response to hexose, response to corticosteroid stimulus and ER stress (**Table 5.11.3**).

Distinct to males

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
Response to virus	2	2.70e-03	26	45
Detection of virus	3	4.63e-03	6	6
Regulation of IL2 production	3	1.07e-04	10	10
Response to TNF	3	1.08e-04	12	13
Positive regulation of ERK1 and ERK2 cascade	3	6.55E-04	10	11
Actinobacterium-type cell wall biogenesis	3	1.53E-03	9	10
Regulation of circadian rhythm	3	3.07e-03	12	16
Proteoglycan metabolic process	3	3.27e-03	13	18
Superoxide metabolic process	3	3.27e-03	13	18
Fatty acid biosynthetic process	3	5.61e-03	22	38
Lipopolysaccharide-mediated signaling pathway	3	5.90e-03	11	15
Hormone biosynthetic process	3	5.90e-03	11	15

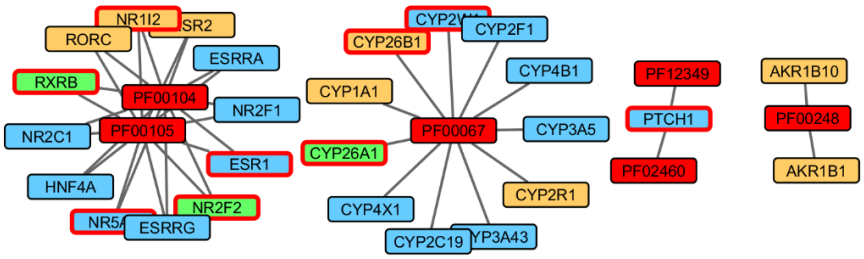
Distinct to females

Biological process	Specificity	FDR	Overlap PFAMs	Total PFAMs
Insulin receptor signaling pathway	3	2.34e-05	12	12
Response to hexose stimulus	3	8.67e-05	16	19
Phosphatidylinositol-3-phosphate biosynthetic process	3	8.95e-04	10	11
Mammary gland morphogenesis	3	2.31e-03	7	7
Cell aging	3	2.40e-03	14	19
Rho protein signal transduction	3	3.73E-03	11	14
Female genitalia development	3	4.54e-03	8	9
PDGF receptor signaling pathway	3	4.54e-03	8	9
Regulation of ph	3	4.85E-03	14	20
Response to corticosteroid stimulus	3	6.01e-03	9	11
Chromatin remodeling	3	7.25e-03	19	31
Response to ER stress	3	8.07e-03	17	27

Table 5.11.3.Domain-centric enrichment of biological processes distinct to males or females. Domain-centric enrichment of biological processes distinct to males (top) or females (bottom) enriched in PFAMs associated with genes with altered expression compared to respective control lung fibroblasts and Specificity = levels of granularity of a specific biological process (1=highly general, 2; general, 3; specific, 4; highly specific). FDR = false discovery rate (P<0.01). Overlap PFAMs = number of overlapping PFAMs associated with the biological process out of the total number of PFAMs in the input list. Total PFAMs = total number of PFAMs associated with the biological process.

Furthermore, as shown in **section 5.7** multiple genes which had different methylation in male IPF compared to male control were not the same genes which had had different methylation in female IPF compared to female control lung fibroblasts. This was also apparent from the PFAM enrichment analysis. Bile acid metabolic process was significantly enriched in PFAMs containing 26 genes with significantly altered methylation in male IPF compared to male control lung fibroblasts and 30 genes with significantly altered methylation in female IPF compared to female control lung fibroblasts. Only 9 of these genes overlapped (**Figure 5.11.1**).

IPF males compared to control males



IPF females compared to control females

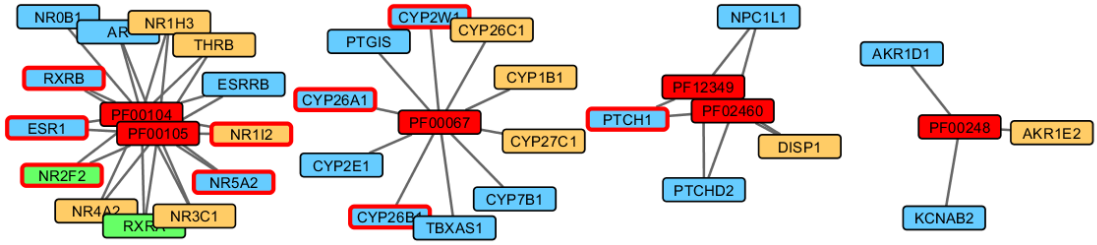
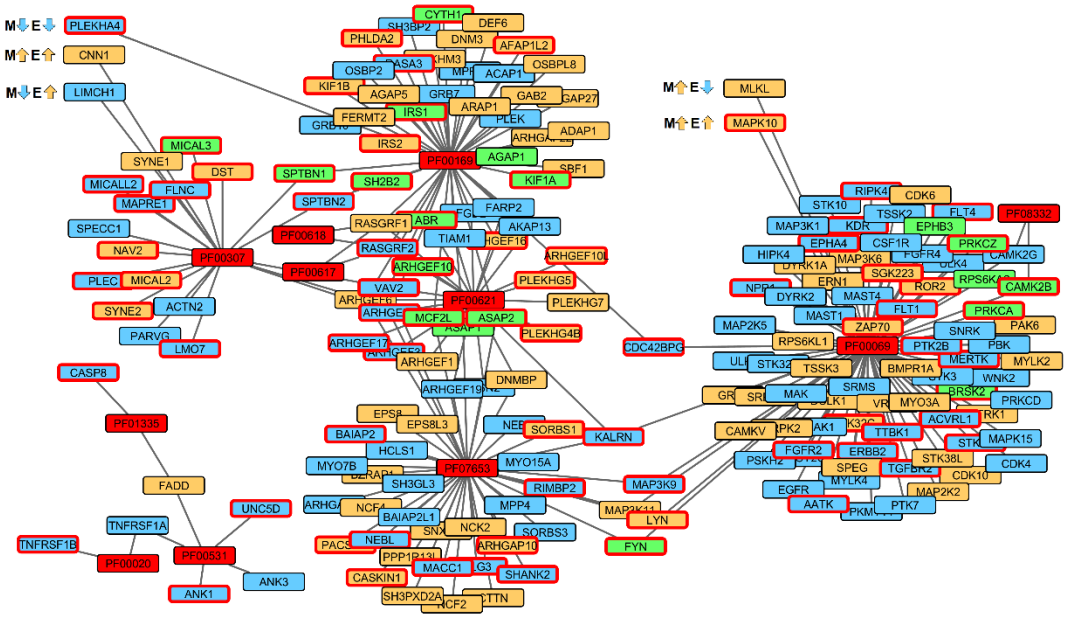


Figure 5.11.1. Bile acid metabolic process enriched in PFAMs containing genes with altered methylation in IPF male compared to control male and IPF female compared to control female lung fibroblasts. Coloured boxes represent direction of methylation, yellow; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in male IPF (n=2) compared to male control (n=2) or female IPF (n=3) compared to female control (n=4) lung fibroblasts. Boxes with a red border represent overlapping genes which also had significantly altered expression (TNoM ≤1; ≥2FC).

In another example, the biological process 'induction of apoptosis by extracellular signals' was significantly enriched in PFAMs containing 193 genes with significantly altered methylation in male IPF compared to male control lung fibroblasts and 215 genes with significantly altered methylation in female IPF compared to female control lung fibroblasts. 78 genes overlapped whereas 115 and 137 were distinct in male and females respectively (**Figure 5.11.2**). Taken together this suggests that biological processes enriched in PFAMs/genes with altered methylation in IPF compared to control lung fibroblasts may involve multiple overlapping as well as different genes between males and females.

IPF males compared to control males



IPF females compared to control females

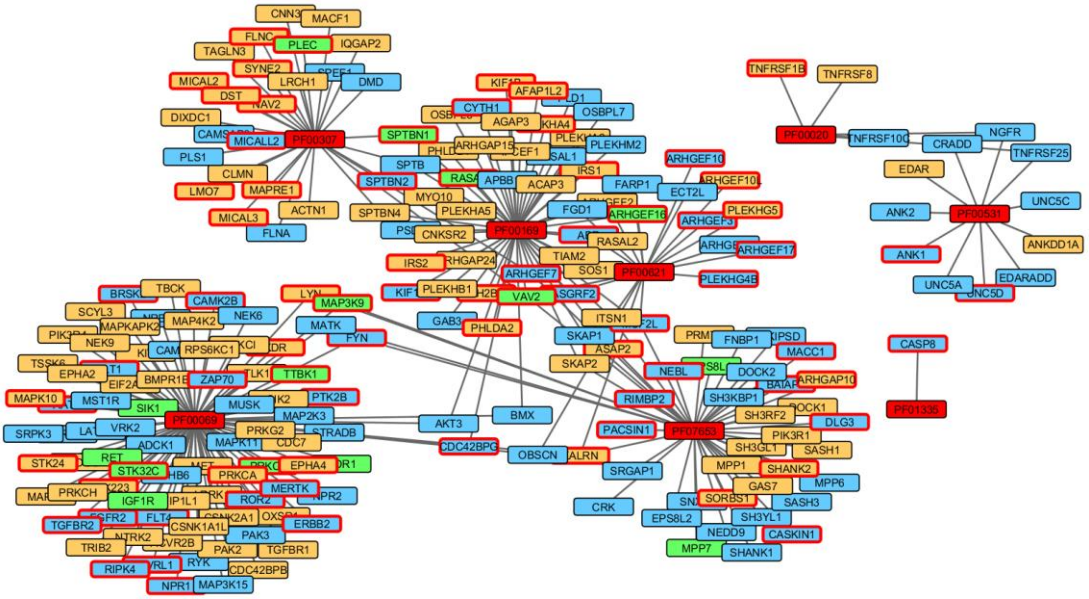


Figure 5.11.2. Induction of apoptosis by extracellular signals enriched in PFAMs containing genes with significantly altered methylation in IPF male compared to control male and IPF female compared to control female lung fibroblasts. Coloured boxes represent direction of methylation, yellow; genes with increased methylation, blue; genes with decreased methylation, green; genes with both increased and decreased methylation in male IPF (n=2) compared to male control (n=2) or female IPF (n=3) compared to female control (n=4) lung fibroblasts. Boxes with a red border represent overlapping genes. Genes which also had significantly altered expression (TNoM ≤1; ≥2FC) are annotated with the direction of change shown. M=methylation, E= expression.

5.12. Summary

- Multiple biological processes relevant to fibrosis are enriched in genes with significantly altered methylation and/or expression in male compared to female lung fibroblasts. This could, in part explain the sex-bias in IPF.

- Some biological processes enriched in PFAMs containing genes with significantly altered methylation in IPF compared to control lung fibroblasts were distinct to each sex. These included those relating to response to virus, Actinobacterium and circadian rhythm in males and ER stress, response to hexose and response to corticosteroids in females.

- 83% of biological processes enriched in PFAMs containing genes with significantly altered methylation in male IPF compared to male controls were the same biological processes enriched in PFAMs containing genes with significantly altered methylation in female IPF compared to female control lung fibroblasts. This suggests the majority of biological processes overlap.

- Whilst multiple genes with altered methylation in male IPF compared to male controls are the same genes with altered methylation in female IPF compared to female control lung fibroblasts, suggesting multiple genes overlap between both sexes, multiple genes were distinct to both sexes. This suggests that within overlapping enriched biological processes, multiple genes have sex-specific differences in methylation.

Chapter 6. Effect of DNMT inhibition on methylation in primary human lung fibroblasts

6.1. Overview

No study to date has examined the effects of DNMT inhibition on global DNA methylation and gene expression in IPF or SSc lung fibroblasts. 5-Aza-2'-deoxycytidine (5-Aza), a chemical analogue of cytosine, can incorporate itself into DNA and binds to DNMT enzymes which inhibits their activity (Christman, 2002). Previous studies suggest 5-Aza treatment could have potential benefits in treating a number diseases by reactivating genes silenced by methylation. Indeed, 5-Aza is already used to treat a number of diseases including myelodysplastic syndromes (Wijermans et al, 2000). As previously shown in **Chapter 3**, multiple genes have CpGs with altered methylation in IPF and SSc compared with control lung fibroblasts. Therefore, using 5-Aza to treat such diseases may be beneficial and warrants further investigation. However, the opposing argument is that 5-Aza is non-specific and thus, could activate multiple genes which could have adverse effects. Furthermore, for the vast majority of genes, it is unknown whether DNA methylation can directly regulate expression, thus demethylation of one gene could activate or inhibit multiple genes regulated by it and in turn these genes could do the same. Differences in methylation between genders, methylation of genes involved in epigenetic regulation (MiRs, HDACs) and other epigenetic mechanisms create a complex network of gene regulation on multiple levels. Furthermore, what change in methylation is required to have a biological effect and whether this change in methylation is the same for all genes or tissues or across different cell lines is unknown. To further try to understand the role of DNA methylation in regulating gene expression in human lung fibroblasts, control, IPF and SSc lung fibroblast cell lines were treated with 5-Aza. Basal methylation was compared to methylation levels after 5-Aza treatment for each cell line. The following section will focus on analysing the effects of 5-Aza on CpG methylation and gene expression. Furthermore, correlation between changes in methylation and gene expression after 5-Aza treatment will also be discussed.

I show that DNMT inhibition with 5-Aza alters the methylation of multiple CpGs and the expression of multiple genes in control, IPF and SSc lung fibroblasts. Analysis of individual lung fibroblast cell lines show they respond differently in relation to which CpGs/genes have altered methylation/expression and to what extent methylation and expression levels change. Furthermore, by correlating methylation changes with changes to gene expression I identify multiple genes in lung fibroblasts that are potentially directly regulated by methylation, many of which are differentially methylated or expressed in IPF and/or SSc compared with control lung fibroblasts. I also show that small changes in methylation can have an effect on gene expression

6.2. Effect of DNMT inhibition on methylation in primary human lung fibroblasts

Following 5-Aza treatment, the Illumina Infinium Human Methylation 450k BeadChip microarray identified multiple CpG sites with significantly (P<0.05) altered methylation in control (n=6) and IPF (n=5) and SSc (n=7) lung fibroblasts. 19421 CpGs corresponding to 10250 genes in control (n=6), 17412 CpGs corresponding to 9493 genes in IPF (n=5) and 16948 CpGs corresponding to 9926 genes in SSc (n=7) lung fibroblasts had significantly (P<0.05) altered methylation after 5-Aza treatment. 937 biological processes were enriched in genes with significantly altered methylation (P<0.05) after 5-Aza treatment in control lung fibroblasts. 77% of these biological processes overlapped with IPF and 67% with SSc lung fibroblasts. These included apoptosis, Wnt signalling, cell adhesion and lung development, which were previously shown to be enriched in genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF and/or SSc compared to control lung fibroblasts (**Table 6.2.1**). This suggests that multiple genes associated with biological processes relevant to PF are potentially regulated by methylation.

Biological process	Control	IPF	SSc	Ref list
Apoptotic process	276	306	283	594
Blood coagulation	200	215	202	457
Cell adhesion	243	268	267	556
Chromatin modification	119	131	115	224
Cytoskeleton organization	53	53	58	106
Gene expression	188	216	171	408
Lung development	43	49	46	75
Regulation of transcription, DNA-dependent	775	838	763	1609
TGFß signalling pathway	40	39	30	65
Viral reproduction	137	163	133	329
Wnt receptor signalling pathway	65	65	59	110

Table 6.2.1. Examples of biological processes which are enriched in genes which have altered methylation after 5-Aza treatment in lung fibroblasts. Numbers represent the number of genes with significantly altered methylation (P<0.05) after 5-Aza treatment in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts. The reference list shows the number of genes associated with each biological process.

Multiple CpG sites in control (n=6) and IPF (n=5) but not SSc (n=7) lung fibroblasts had significant (P<0.05) changes in methylation after 5-Aza treatment which were ≥13.6% (Δβ ≥0.136). 284 CpGs (283 decreased, 1 increased) in control (n=6), 670 CpGs (666 decreased, 5 increased) in IPF (n=5) and 1 CpG in SSc (n=7) lung fibroblasts had significantly altered (P<0.05; Δβ ≥0.136) methylation after 5-Aza treatment (**Figure 6.2.1**).

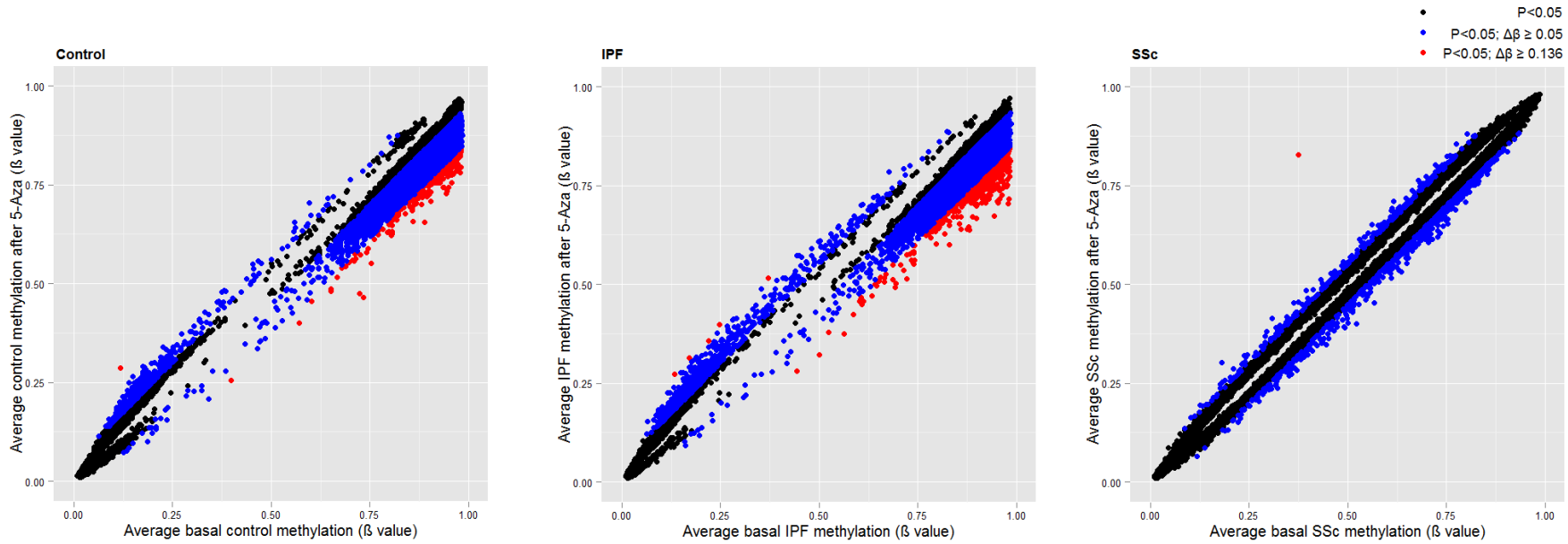


Figure 6.2.1. CpGs which had altered methylation after 5-Aza treatment in lung fibroblasts. Scatter plots show all CpGs which had significantly altered (P<0.05) methylation in control (n=6), IPF (n=5) and SSc (n=7) after treatment with 5-Aza. β value = methylation (0 = 0%, 1 = 100% methylated). ● P<0.05; Δβ<0.05, ● P<0.05; Δβ ≥0.05<0.136 ● P<0.05; Δβ ≥0.136.

This data suggested the majority of changes in CpG methylation after 5-Aza treatment were small. (<13.6%). However, cluster analysis of the CpGs with a significant change in methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in control and IPF lung fibroblast cell lines after 5-Aza treatment identified cell lines which did not respond to 5-Aza treatment to the same extent as others (**Figure 6.2.2**). It is unclear why some cell lines responded to 5-Aza more than others, however, failure to efficiently incorporate 5-Aza or activation of genes which affected DNMTs or de-methylation genes, such as TETs, could offer a feasible explanation. Three control lung fibroblast cell lines (1, 3 and 4) responded to 5-Aza treatment and clustered together, separately from their respective basal methylation states. The other three cell lines (2, 5 and 6) clustered together with their basal methylation state after 5-Aza treatment, suggesting they did not respond to the same extent as the other cell lines after 5-Aza treatment at these CpGs.

Similarly, IPF cell lines 1, 2 and 5 responded to 5-Aza and clustered together, separately from their respective basal methylation states. IPF cell line 3 responded to 5-Aza, but not to the same extent as cell lines 1, 2 and 5. Nonetheless, after treatment, IPF cell line 3 clustered separately from its basal methylation state. In contrast, after 5-Aza treatment IPF cell line 4 clustered together with its basal methylation state, suggesting it did not respond to the same extent as other cell lines after 5-Aza treatment at these CpGs (**Figure 6.2.2**).

Only 1 CpG in SSc lung fibroblasts had a $\Delta\beta \geq 0.136$ change in methylation after 5-Aza treatment. Cluster analysis on CpGs which had a significant change in methylation ($P<0.05$; $\Delta\beta \geq 0.05$) showed all non-treated cell lines (basal methylation) clustered separately from all treated cell lines. Therefore, small changes in methylation occurred after 5-Aza treatment in all cell lines, however, there was heterogeneity between samples suggesting the change in methylation at these CpGs was different between cell lines (**Figure 6.2.2**).

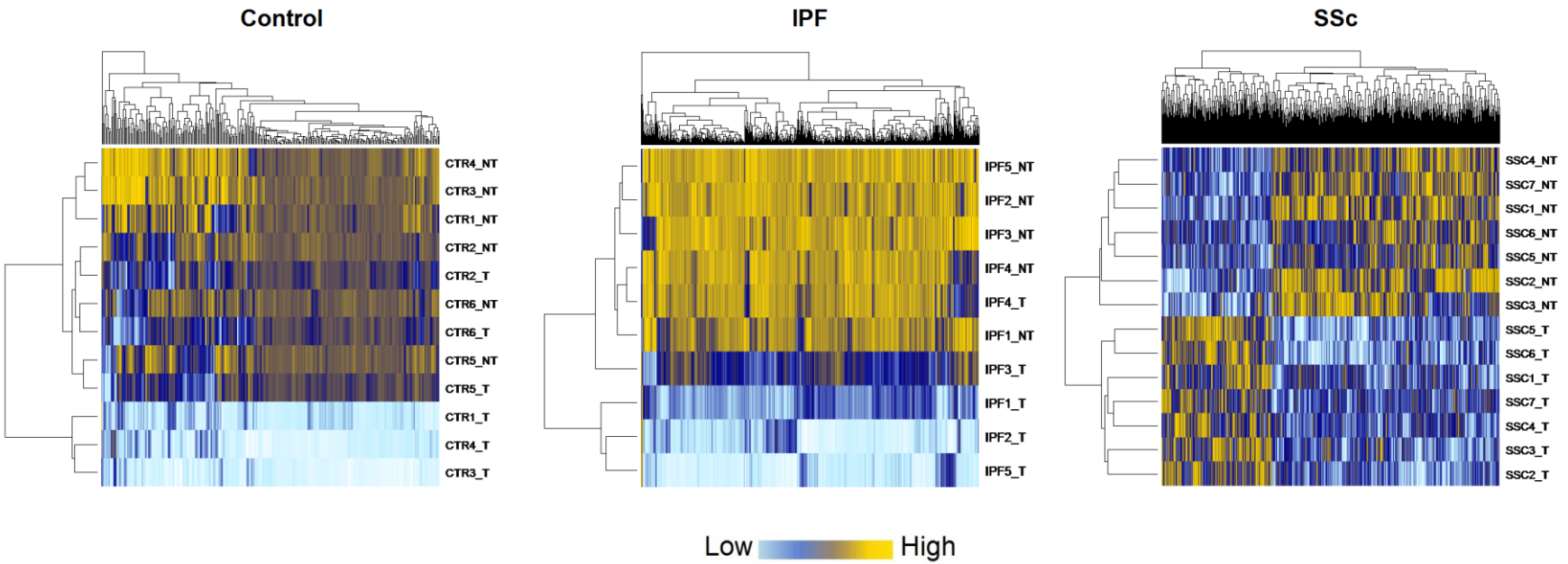


Figure 6.2.2. Hierarchical clustering based on CpGs with altered methylation in lung fibroblasts after treatment with 5-Aza. Heat-maps shows CpGs with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in control (n=6) and IPF (n=5) and ($\Delta\beta \geq 0.05$; $P<0.05$) in SSc (n=7) lung fibroblasts after 5-Aza treatment. T=treated, NT=untreated. Light blue represents low methylation, yellow represents high methylation with respect to each CpG.

These data suggest different cell lines responded differently to 5-Aza, which could account for the relatively low number of CpGs which had significant (P<0.05) changes in methylation ≥13.6% after 5-Aza treatment in control, IPF and SSc lung fibroblasts. To further investigate to what extent 5-Aza treatment had on methylation in lung fibroblasts, each cell line was analysed separately. Basal methylation of each CpG on the array (n=324973) in each control (n=6), IPF (n=5) and SSc (n=7) lung fibroblast cell line was compared to the methylation of each CpG after 5-Aza treatment.

Results confirmed that some cell lines responded to a greater extent to 5-Aza treatment than others (**Figure 6.2.3**, **Figure 6.2.4 and Figure 6.2.5**). Cell lines which had multiple CpGs with a Δβ ≥0.136 change in methylation after 5-Aza treatment were labelled as strong responding cell lines, whereas cell lines which had fewer CpGs with a Δβ ≥0.136 change in methylation after 5-Aza treatment were labelled as weaker responding cell lines. Control cell lines, 1, 3 and 4 and IPF cell lines 1, 2 and 5 had multiple CpGs which had changes (Δβ ≥0.136) in methylation after 5-Aza. As previously shown (**Figure 6.2.1**), only 1 CpG in SSc (n=7) lung fibroblasts had a significant change in methylation (P<0.05) which was ≥13.6% after 5-Aza treatment. Analysis of each individual SSc cell line identified all as having relatively few CpGs with changes (Δβ ≥0.136) in methylation after 5-Aza. However, there were CpGs which did have large changes in methylation (≥13.6%) after 5-Aza treatment. **Table 6.2.2** summarises the number of CpGs which had a Δβ ≥0.136 (13.6%) change in methylation after 5-Aza treatment in each cell line.

Cell line	Control	IPF	SSc
1	12101	8985	582
2	781	31194	2063
3	26608	1375	1235
4	14848	313	458
5	321	15942	620
6	704	-	495
7	-	-	596

Table 6.2.2. CpGs which have a change in methylation after 5-Aza treatment in each cell line. Number of CpGs in each control (n=6), IPF (n=5) and SSc (n=7) lung fibroblast cell line which have a change Δβ ≥0.136 (13.6%) in methylation after 5-Aza treatment.

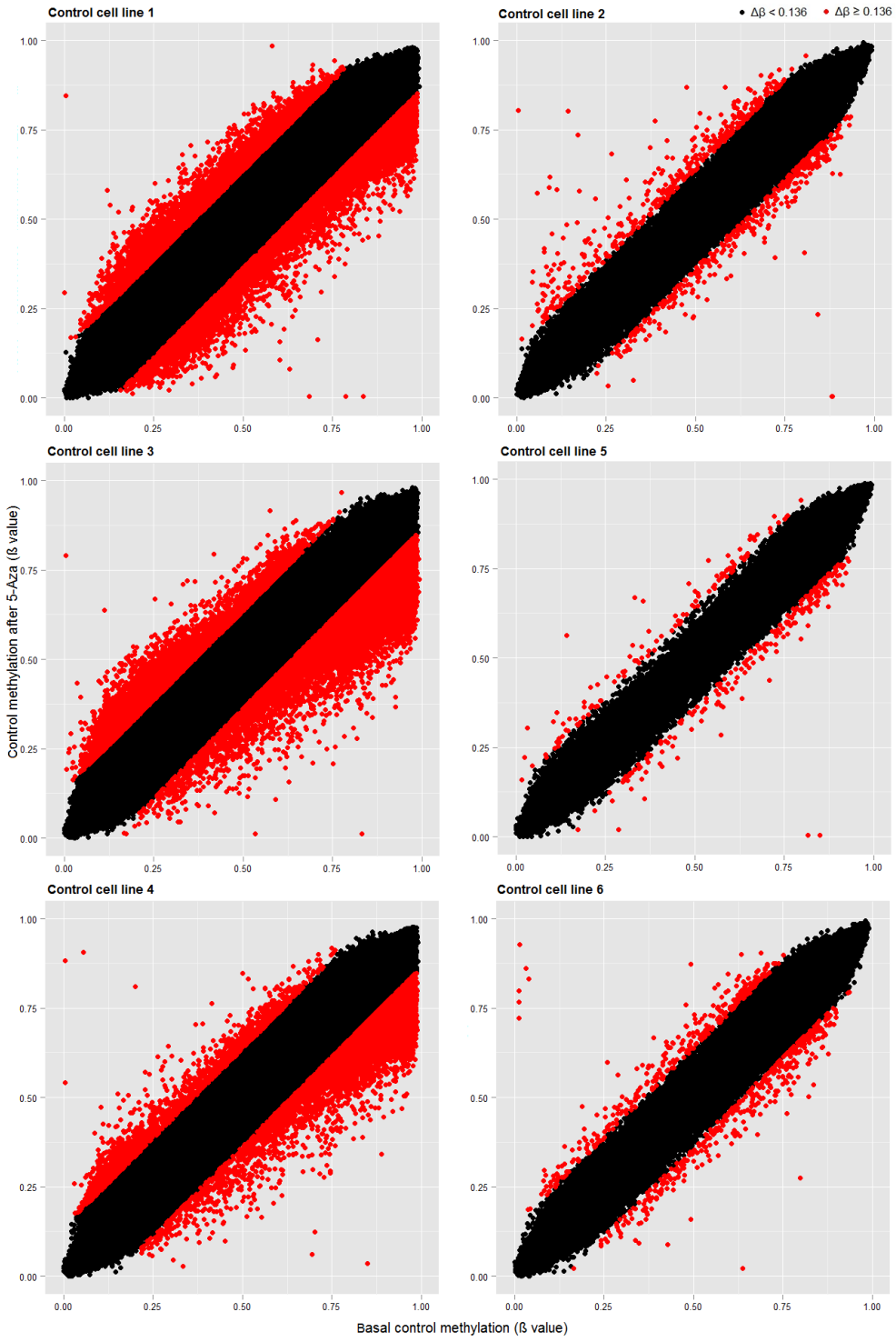


Figure 6.2.3. CpGs which have altered methylation after 5-Aza treatment in 6 different control lung fibroblast cell lines. Scatter plots show the basal methylation compared to the methylation level after treatment with 5-Aza. β value = methylation (0 = 0%, 1 = 100% methylated). Δβ value = change in methylation after 5-Aza treatment.● Δβ <0.136 (<13.6%) change in basal compared to 5-Aza treated methylation, ● Δβ ≥0.136 (≥ 13.6%) change in basal compared to 5-Aza treated methylation.

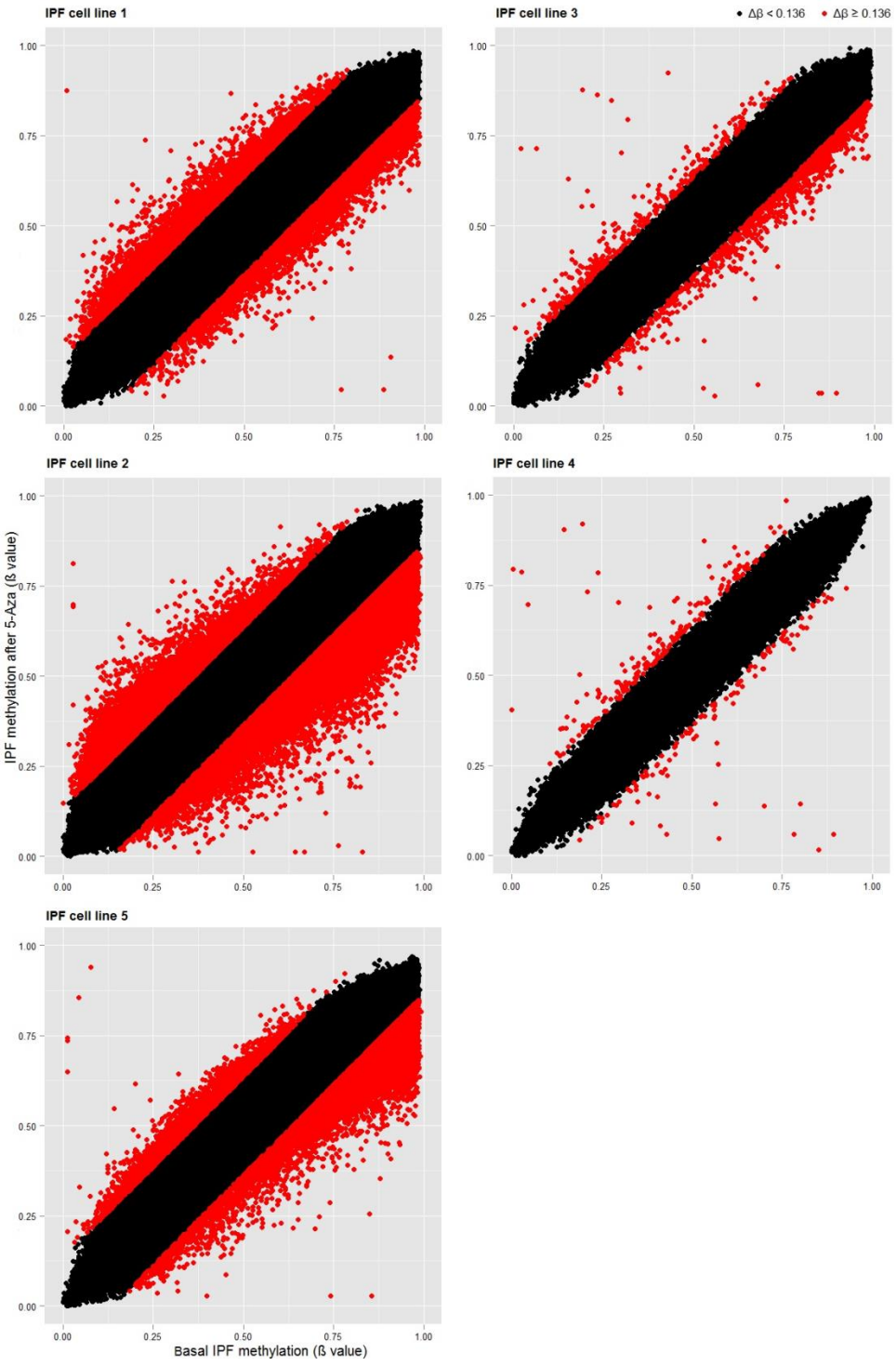


Figure 6.2.4. CpGs which had altered methylation after 5-Aza treatment in 5 different IPF lung fibroblast cell lines. Scatter plots show the basal methylation compared to the methylation level after treatment with 5-Aza. β value = methylation (0 = 0%, 1 = 100% methylated). Δβ value = change in methylation after 5-Aza treatment.● Δβ <0.136 (<13.6%) change in basal compared to 5-Aza treated methylation, ● Δβ ≥0.136 (≥ 13.6%) change in basal compared to 5-Aza treated methylation.

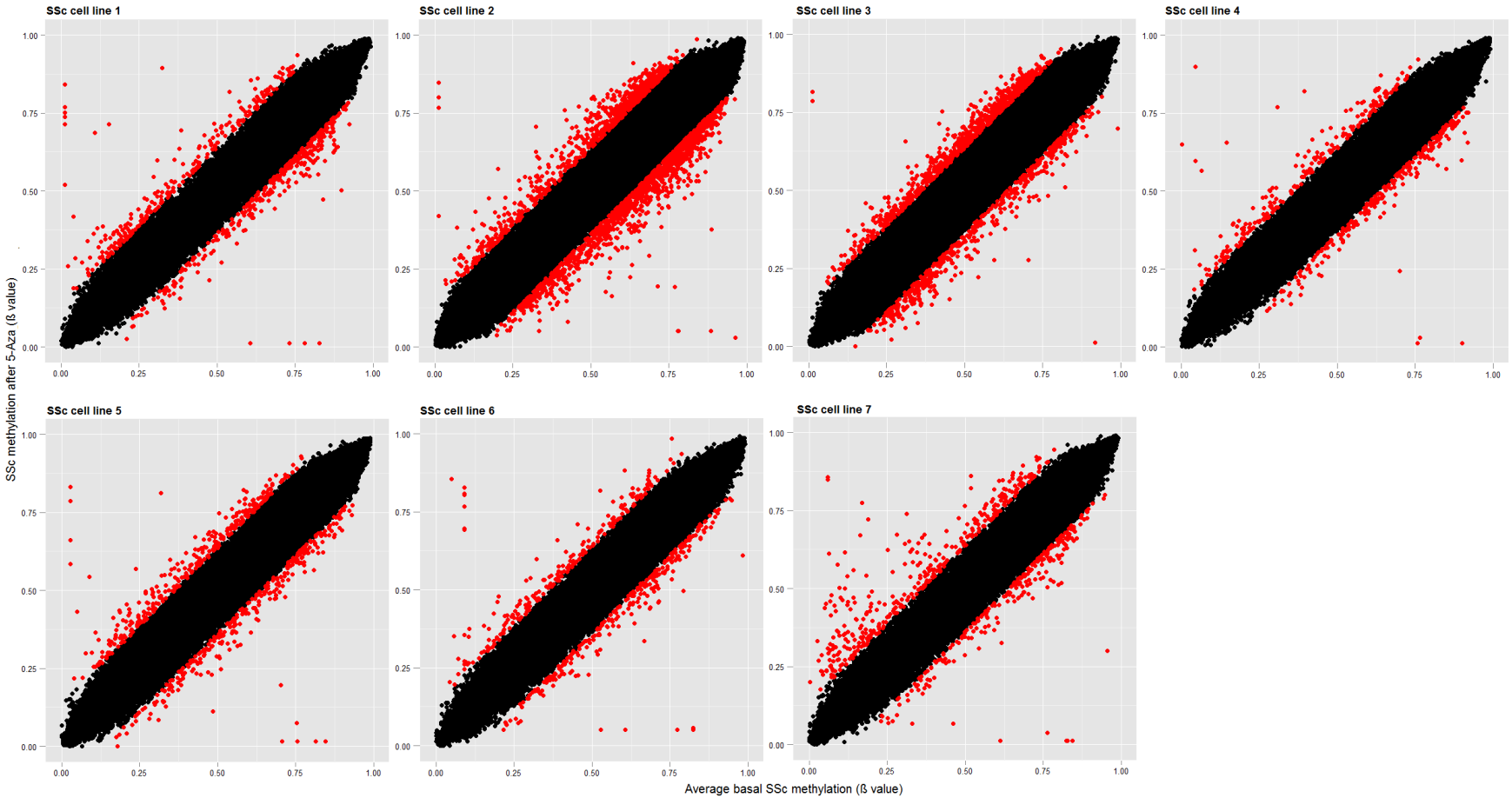


Figure 6.2.5. CpGs which had altered methylation after 5-Aza treatment in 7 different SSc lung fibroblast cell lines. Scatter plots show the basal methylation compared to the methylation level after treatment with 5-Aza. β value = methylation (0 = 0%, 1 = 100% methylated). Δβ value = change in methylation after 5-Aza treatment. ● Δβ <0.136 (<13.6%) change in basal compared to 5-Aza treated methylation, ● Δβ ≥0.136 (≥ 13.6%) change in basal compared to 5-Aza treated methylation.

6.3. Effect of DNMT inhibition on expression in primary human lung fibroblasts

Basal gene expression was compared with gene expression after 5-Aza treatment in all control (n=6), IPF (n=5) and SSc (n=7) lung fibroblast cell lines. The Illumina Infinium HT12v4 expression microarray identified 548 genes in control, 2912 genes in IPF and 940 genes in SSc with significantly ($P<0.05$; TNoM ≤1) altered expression after 5-Aza treatment compared with basal expression (**Figure 6.3.1**). Cluster analysis on genes with significantly altered expression ($P<0.05$; TNoM ≤1) after 5-Aza treatment compared with basal expression identified control cell line 5 and SSc cell line 3 as cell lines which, after 5-Aza treatment, clustered with non-treated cell lines. This suggests that multiple genes in these cell lines were not affected by 5-Aza to the same extent that other cell lines were (**Figure 6.3.1**).

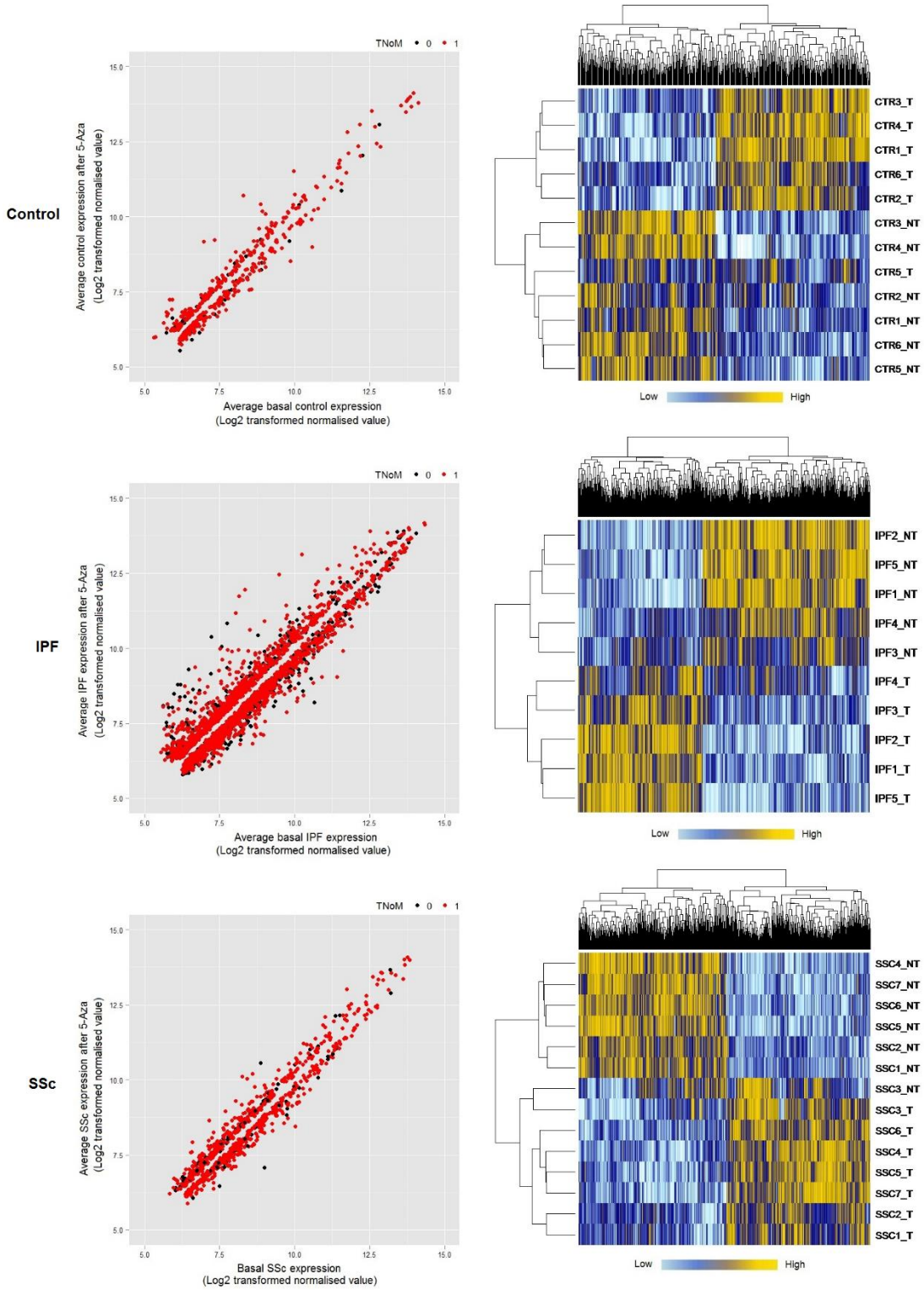


Figure 6.3.1. Effect of 5-Aza on gene expression. Scatter plots and heat-maps show the effect of DNMT inhibition on gene expression in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts. Scatter plots shows the average expression of genes which had a significant change in expression ($P<0.05$; TNoM≤1) after DNMT inhibition with 5-Aza compared to basal expression. ● TNoM=0, ● TNoM =1. Not treated (basal) = NT, treated (5-Aza) = T. Light blue represents low expression, yellow represents high expression with respect to each CpG.

To determine whether 5-Aza affected gene expression to a greater extent in different lung fibroblast cell lines, basal gene expression was compared to gene expression after 5-Aza treatment in each individual lung fibroblast cell line. Large changes in gene expression were classified as ≥ 2-fold changes (≥ 2FC) in basal gene expression compared to gene expression after 5-Aza treatment.

Concurrent with data showing multiple CpGs with large changes in methylation after 5-Aza treatment, control fibroblast cell lines 1, 3 and 4 and IPF fibroblast cell lines 1, 2 and 5 had the most genes which had a ≥2FC in gene expression after 5-Aza treatment compared with basal expression levels (**Figure 6.3.2 and Figure 6.3.3**). All SSc lung fibroblast cell lines had a relatively low number of genes with a ≥2FC in gene expression after 5-Aza treatment. Interestingly however, SSc lung fibroblast cell lines 2 and 3 (which had the most number of CpGs with altered methylation ($\Delta\beta \geq 0.136$), had the least number of genes (5 and 3 respectively) with a ≥2FC in gene expression after 5-Aza treatment (**Figure 6.3.4**). This may suggest that small changes in methylation can have a big effect on gene expression and that large changes in methylation do not necessarily result in large changes in expression. Furthermore, small changes in methylation could effect the expression of genes such as master transcription factors or other epigenetic regulators such as miRs, which subsequently could have an effect on other genes. **Table 6.3.1** summarises the number of genes in each cell line which have ≥2FC in expression after 5-Aza treatment compared with their basal expression levels.

Cell line	Control	IPF	SSc
1	250	618	7
2	9	632	5
3	496	111	3
4	656	127	102
5	0	625	117
6	115	-	53
7	-	-	151

Table 6.3.1. Genes which have a change in expression after 5-Aza treatment compared with basal expression in each cell line. Number of genes with a ≥2FC in expression in each control (n=6), IPF (n=5) and SSc (n=7) lung fibroblast cell line after 5-Aza treatment compared to basal expression.

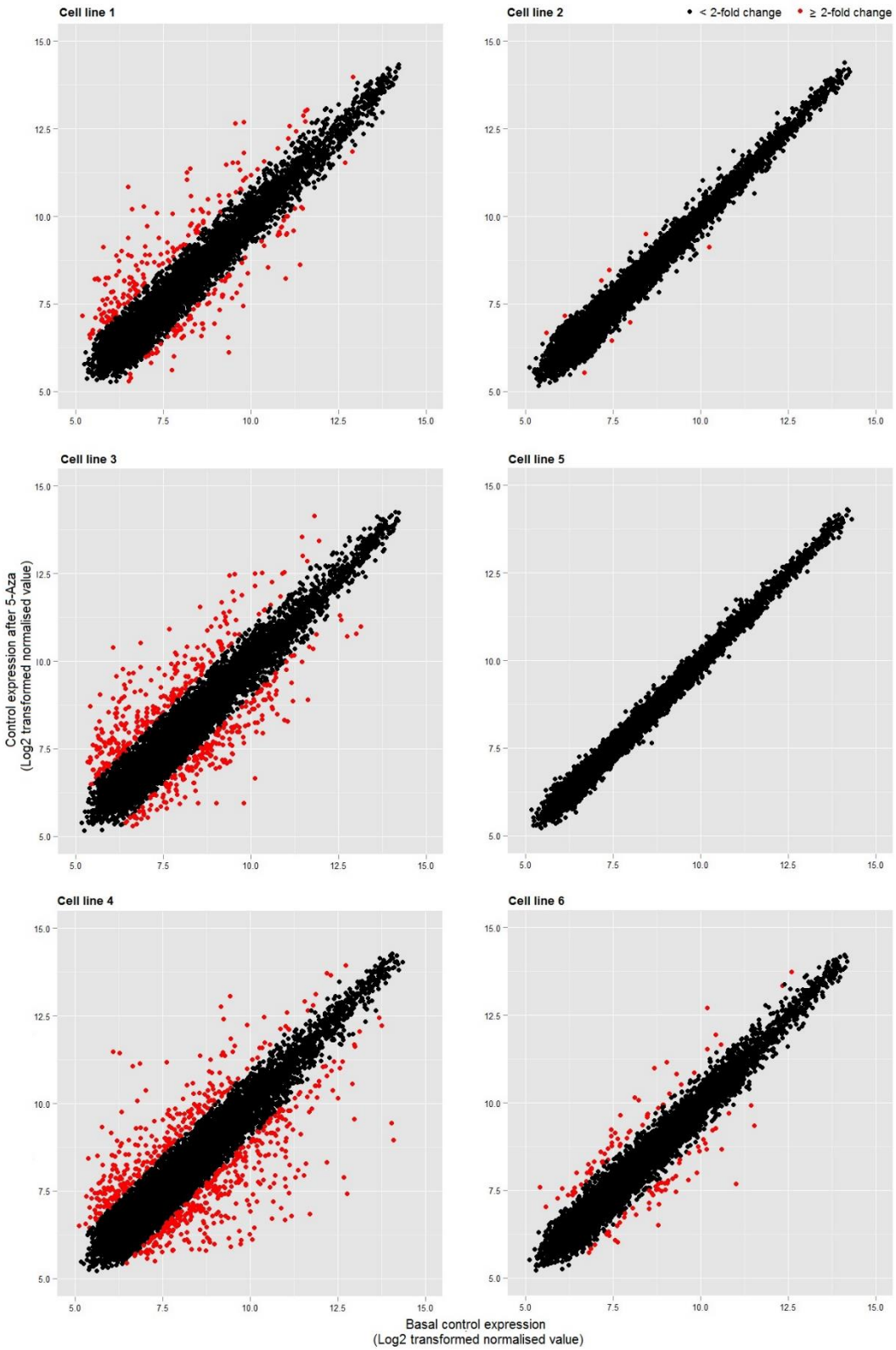


Figure 6.3.2. Genes which have altered expression after 5-Aza treatment in 6 different control lung fibroblast cell lines. ● Fold change in expression <2, ● fold change in expression ≥2.

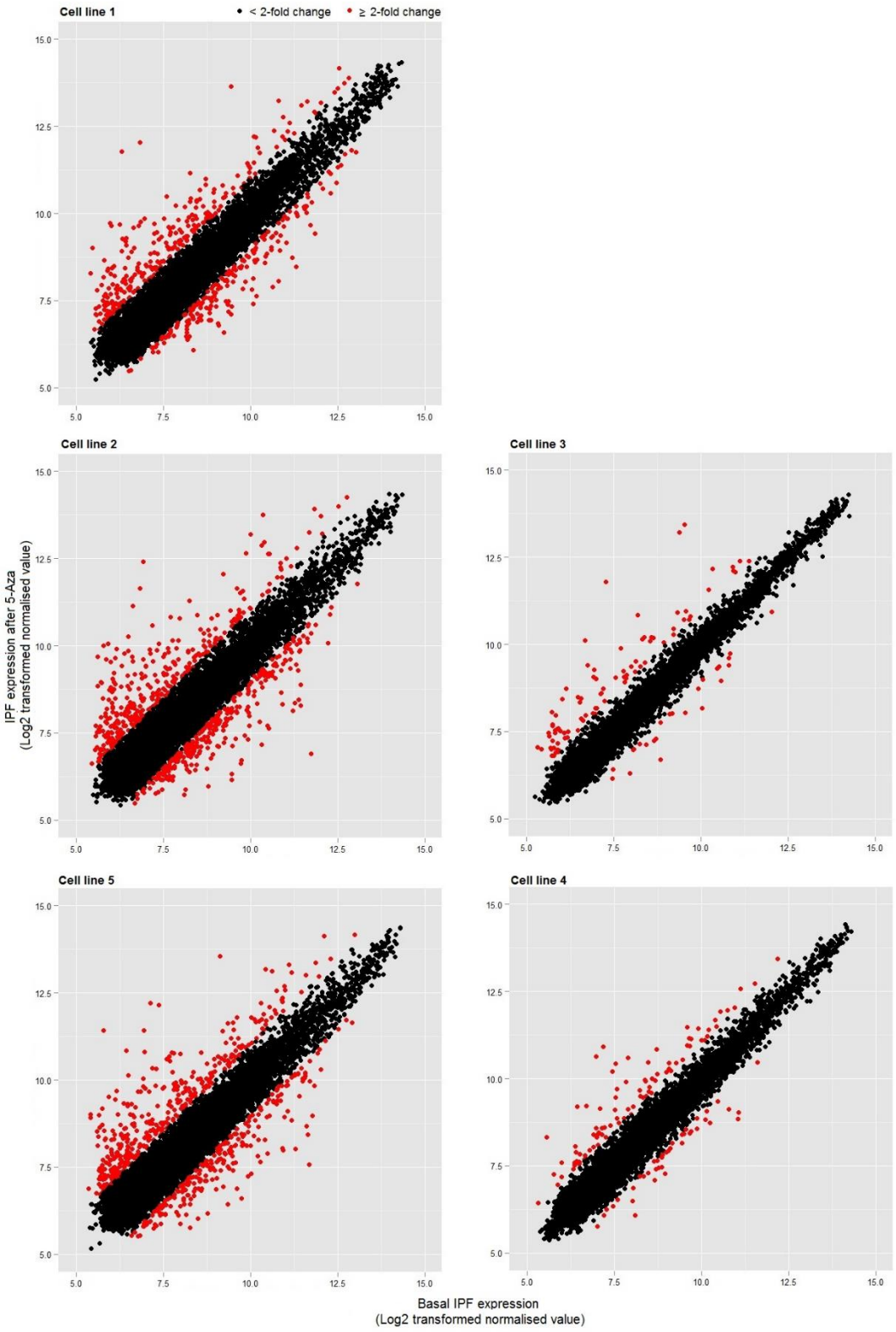


Figure 6.3.3. Genes which have altered expression after 5-Aza treatment in 5 different IPF lung fibroblast cell lines. ● Fold change in expression <2, ● fold change in expression ≥2.

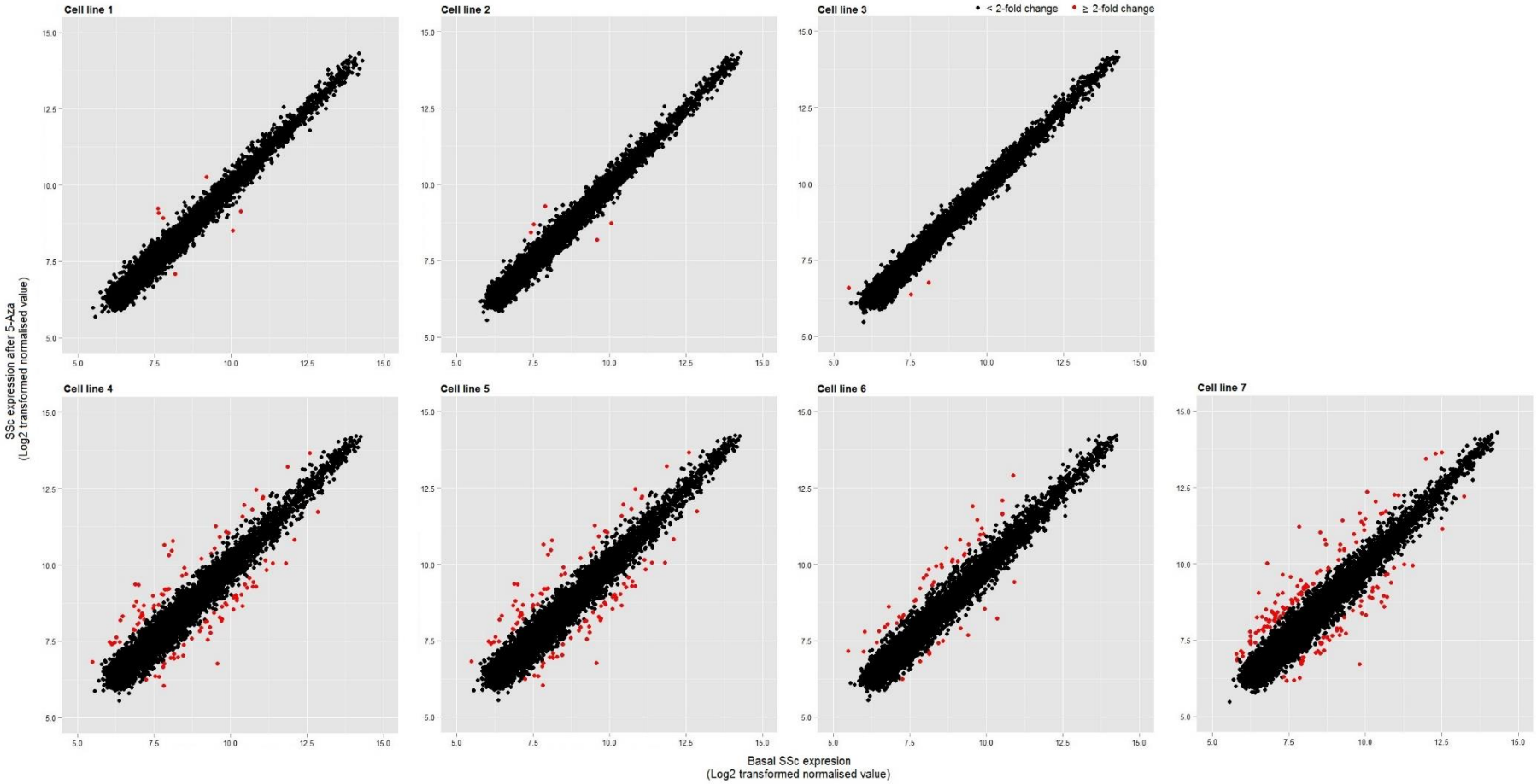


Figure 6.3.4. Genes which had altered expression after 5-Aza treatment in 7 different SSc lung fibroblast cell lines. ● Fold change in expression <2, ● fold change in expression ≥2.

6.4. Confirmation of 5-Aza expression arrays

As previously discussed, multiple genes had large changes to their expression after treatment with 5-Aza. For genes including Matrix metalloproteinase-10 (MMP10), Matrix metalloproteinase-12 (MMP12) (**Figure 6.4.1**), changes in expression after 5-Aza treatment were confirmed by qRT-PCR. Fold changes in gene expression after treatment with 5-Aza were identified as being much larger by qRT-PCR compared with microarray data, however, this was not surprising as qRT-PCR has previously been shown to be more sensitive than microarray analysis (Chen et al, 2009).

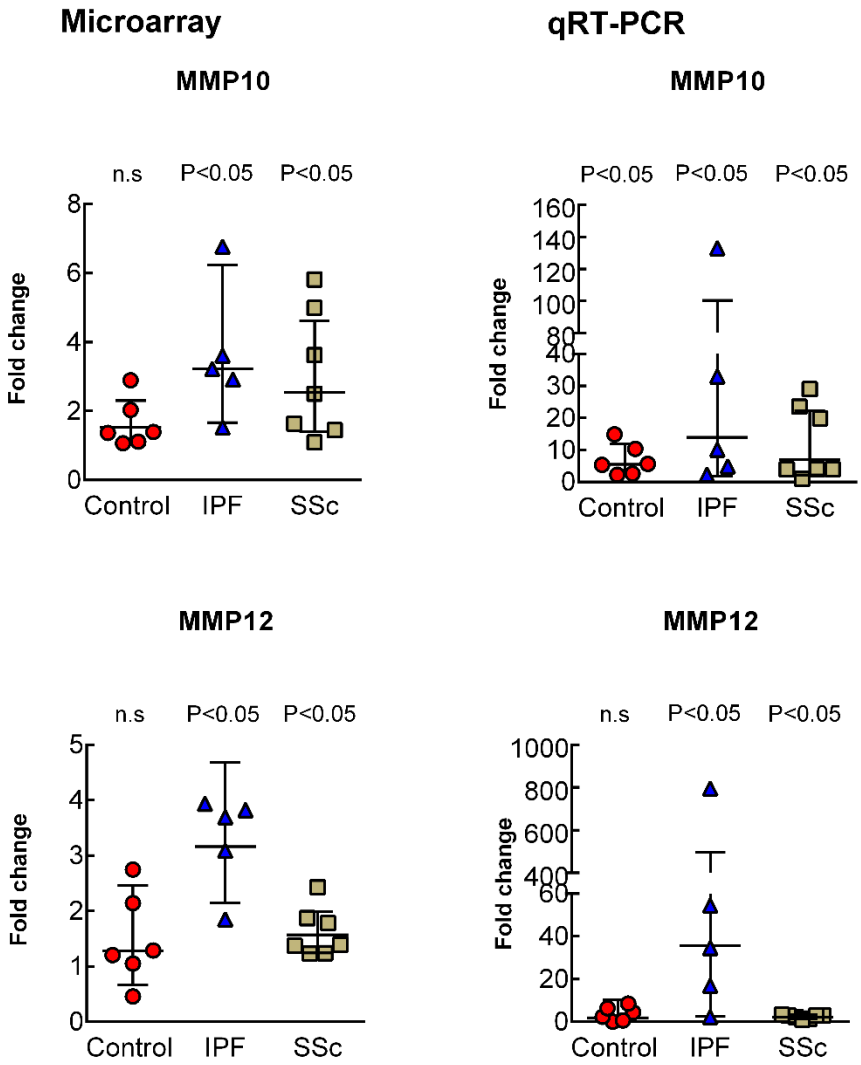


Figure 6.4.1. Validation of 5-Aza microarray data using qRT-PCR. MMP10 and MMP12 expression in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts after 5-Aza treatment. Data presented as the geometric mean fold-change relative to basal expression of each cell line ± 95% confidence intervals. Each data point represents a different cell line.

6.5. Correlation between methylation and gene expression after 5-Aza treatment

The ß value of methylation and the log2 transformed normalised expression value after 5-Aza treatment was compared with basal levels to determine what change in methylation may have an effect on gene expression. Although some cell lines responded more to 5-Aza than others, it was hypothesised that changes to methylation would still correlate with changes to expression in a linear manner. For example, small changes in methylation could result in small changes in expression and large changes in methylation with large changes in expression. However, previous studies have shown small methylation changes can induce large changes in expression (Sanders et al, 2012, Huang et al, 2014, Yang et al, 2014). Furthermore, it is unclear whether changes in methylation have a linear relationship with changes in expression. The following analysis looked at what change in methylation could have an effect on expression.

1392 CpGs corresponding to 801 genes had significant correlation between changes in methylation and changes in expression in control, IPF and SSc cell lines. The location of these CpGs, their change in methylation and their corresponding genes change in expression after 5-Aza treatment compared with basal levels in each control, IPF and SSc cell line is shown in **Figure 6.5.1**, **Figure 6.5.2** and **Figure 6.5.3**.

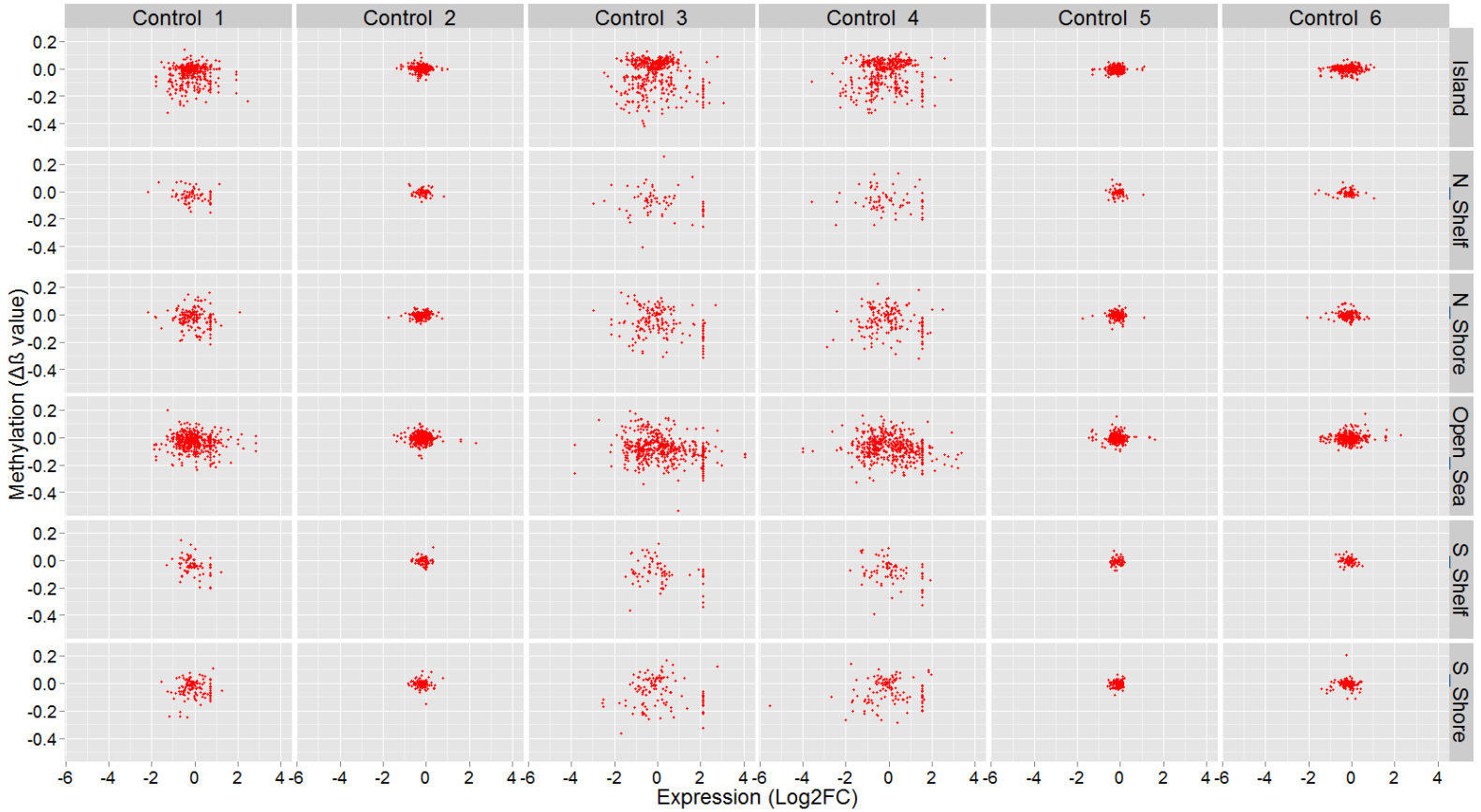


Figure 6.5.1. Control lung fibroblasts: the location of 1392 CpGs in which changes to CpG methylation correlate with changes in expression. Scatter plot shows CpGs change in methylation (Δβ) and their corresponding genes change in expression (Log2FC) after 5-Aza treatment compared with basal levels in each control lung fibroblast cell line (n=6). Each dot represents a different CpG.

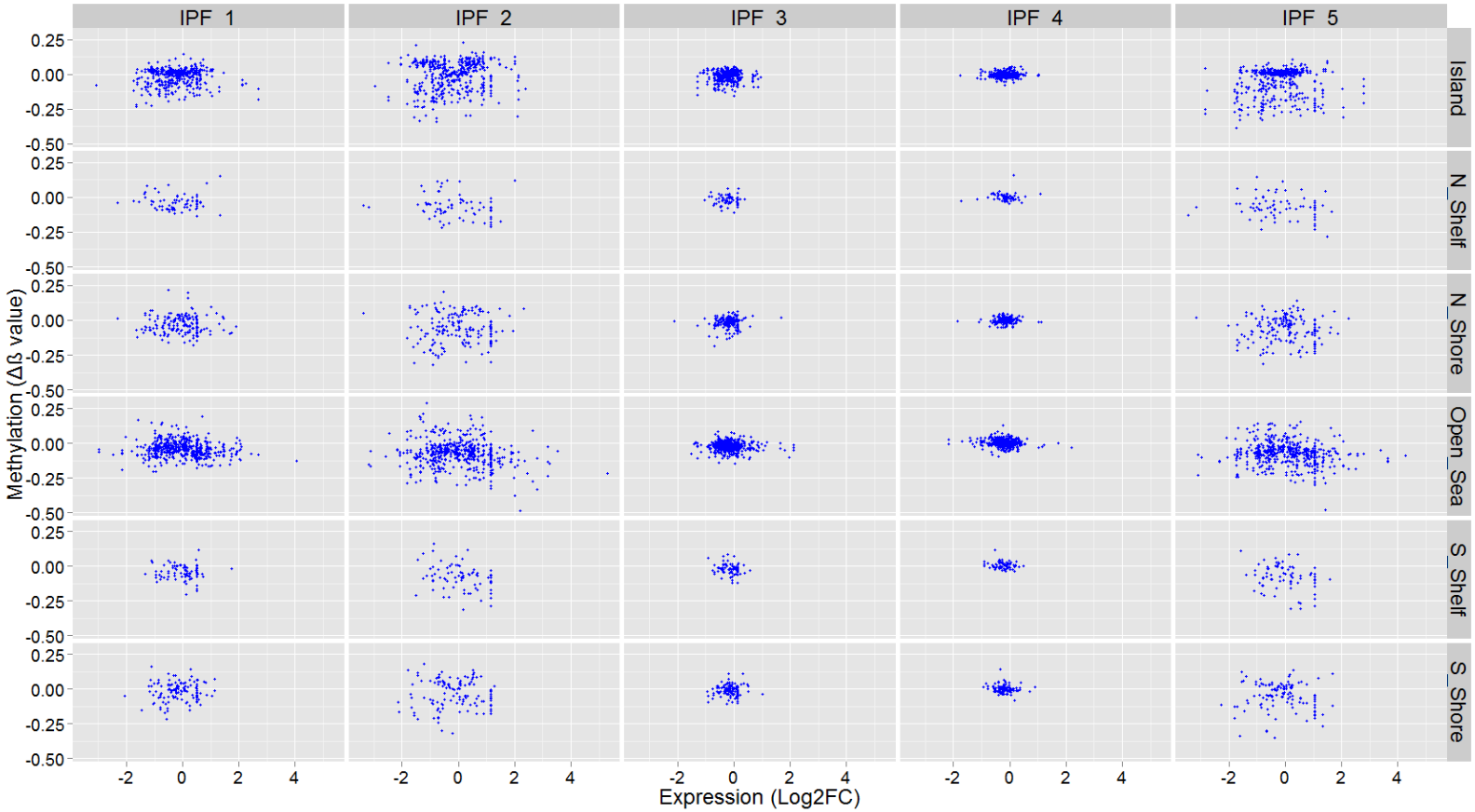


Figure 6.5.2. IPF lung fibroblasts: the location of 1392 CpGs in which changes to CpG methylation correlate with changes in expression. Scatter plot shows CpGs change in methylation (Δβ) and their corresponding genes change in expression (Log2FC) after 5-Aza treatment compared with basal levels in each IPF lung fibroblast cell line (n=5). Each dot represents a different CpG.

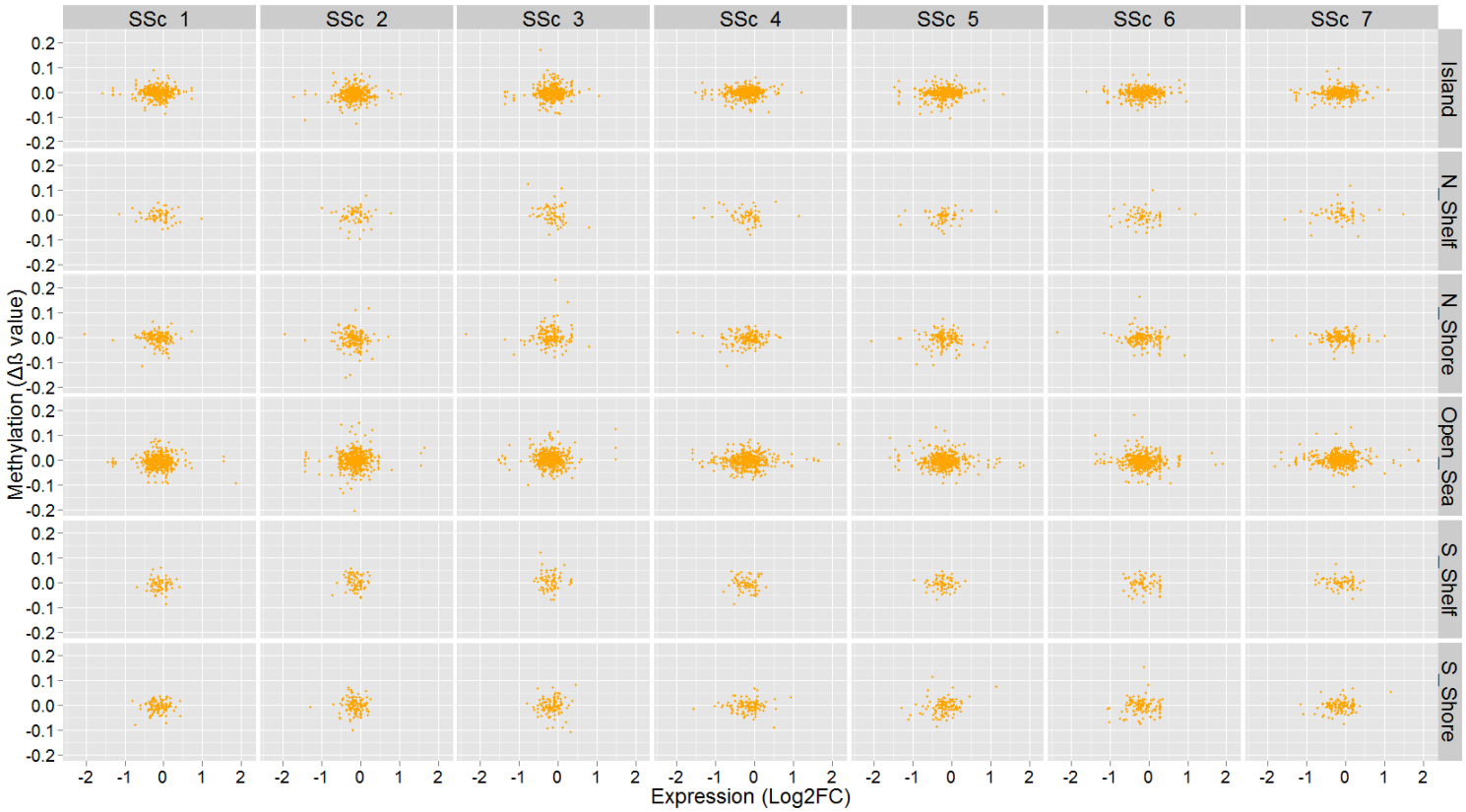


Figure 6.5.3. SSc lung fibroblasts: the location of 1392 CpGs in which changes to CpG methylation correlate with changes in expression. Scatter plot shows CpGs change in methylation (Δβ) and their corresponding genes change in expression (Log2FC) after 5-Aza treatment compared with basal levels in each SSc lung fibroblast cell line (n=7). Each dot represents a different CpG.

Multiple genes including, Mitotic spindle assembly checkpoint protein MAD1 (MAD1L1), Rho guanine nucleotide exchange factor 10 (ARHGEF10), Macrophage erythroblast attacher (MAEA) and O-6-methylguanine-DNA methyltransferase (MGMT) had >10 CpG sites to which changes in methylation correlated with changes in expression. Some examples of CpGs/genes which showed correlation between changes in methylation and changes in expression are shown in **Figure 6.5.4**.

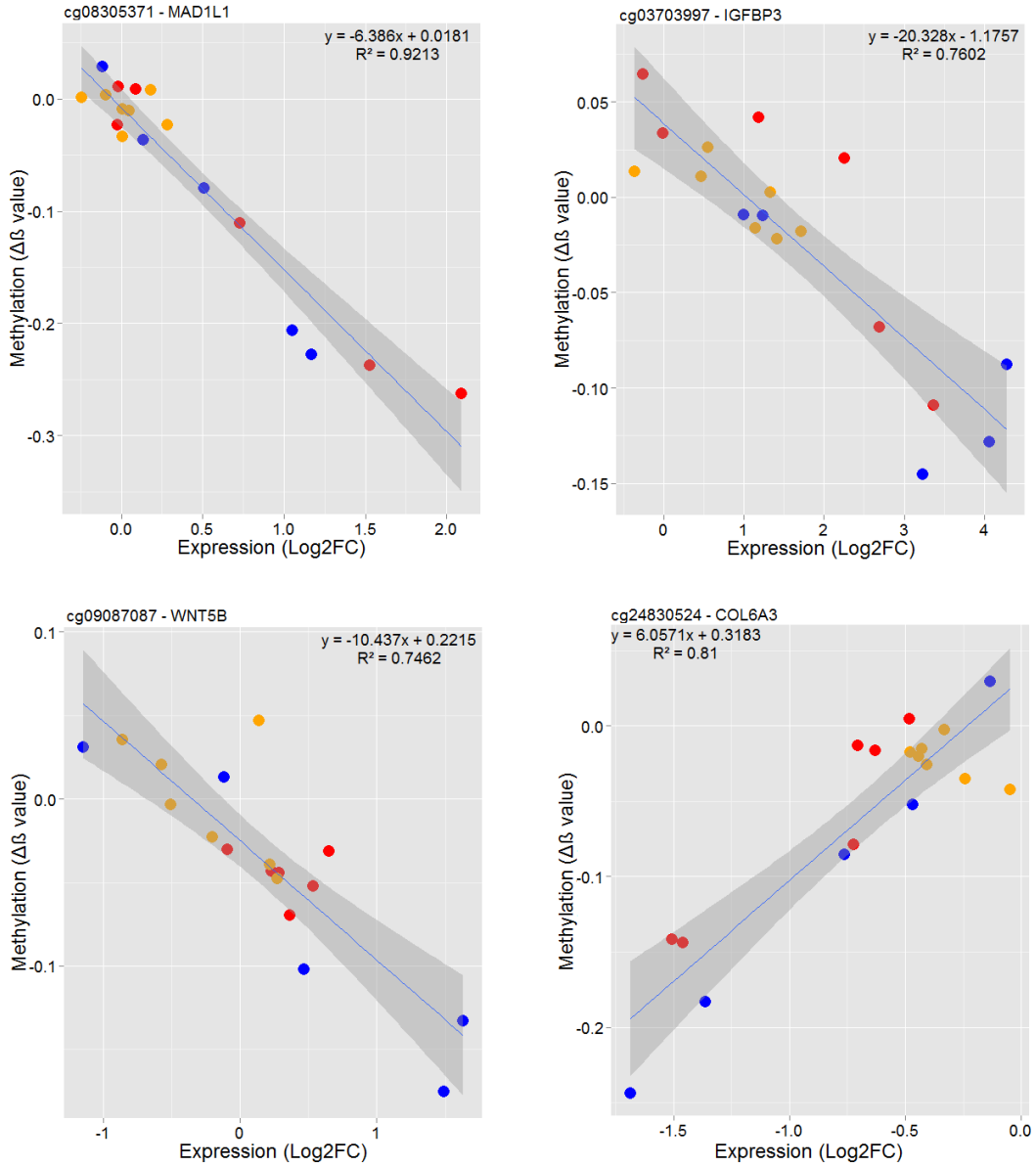


Figure 6.5.4. CpGs which showed correlation between changes in their methylation with changes to their respective genes expression level. Examples of 4 CpGs which have correlation between CpG methylation and expression across all cell lines (n=18) after 5-Aza treatment compared with basal levels. The change in methylation is shown by the Δß value. The change in expression is shown by the log2 fold-change (Log2FC). Shaded areas indicate 95% confidence regions. ● Control (n=6), ● IPF (n=5), ● SSc (n=7).

A number of CpGs appeared to have negligible changes in methylation which correlated with large changes to their respective genes expression, particularly evident in SSc cell lines. For example, changes in methylation at 9 CREB binding protein (CREBBP) CpGs correlated with changes in CREBBP expression. In 2/6 control and 6/7 SSc cell lines, the highest change in methylation across all 9 CpG sites was <5%, yet expression of CREBBP was decreased ≥2fold in all cell lines (n=18) after 5-Aza treatment compared with basal levels (**Figure 6.5.5**). Another gene, CMIP, had 2 CpGs to which changes in methylation correlated with changes in CMIP expression. All cell lines (n=18) had ≥2fold increased expression of CMIP after 5-Aza compared with basal CMIP levels, yet the highest change in methylation across both CpGs in 11/18 cell lines was <5%. Changes in methylation at other CpGs not covered on the array could be responsible for the observed changes in expression, however, these data suggest that small changes in methylation could potentially have a big impact on the expression of some genes.

Other genes appeared to require larger changes in methylation to have an effect on expression. For example, changes in methylation at 130 MAD1L1 CpGs correlated with changes in MAD1L1 expression. Methylation changes at 4 CpGs in the MAD1L1 gene which showed the strongest correlation with MAD1L1 expression are shown in **Figure 6.5.6**. 17/18 cell lines had at least 1 CpG with a ≥5% change in methylation after 5-Aza treatment compared with basal levels, yet only cell lines which had above an 18% change in methylation in at least 1 CpG had a ≥2fold change in expression. This suggests for some genes large changes in methylation are required to have an effect on gene expression.

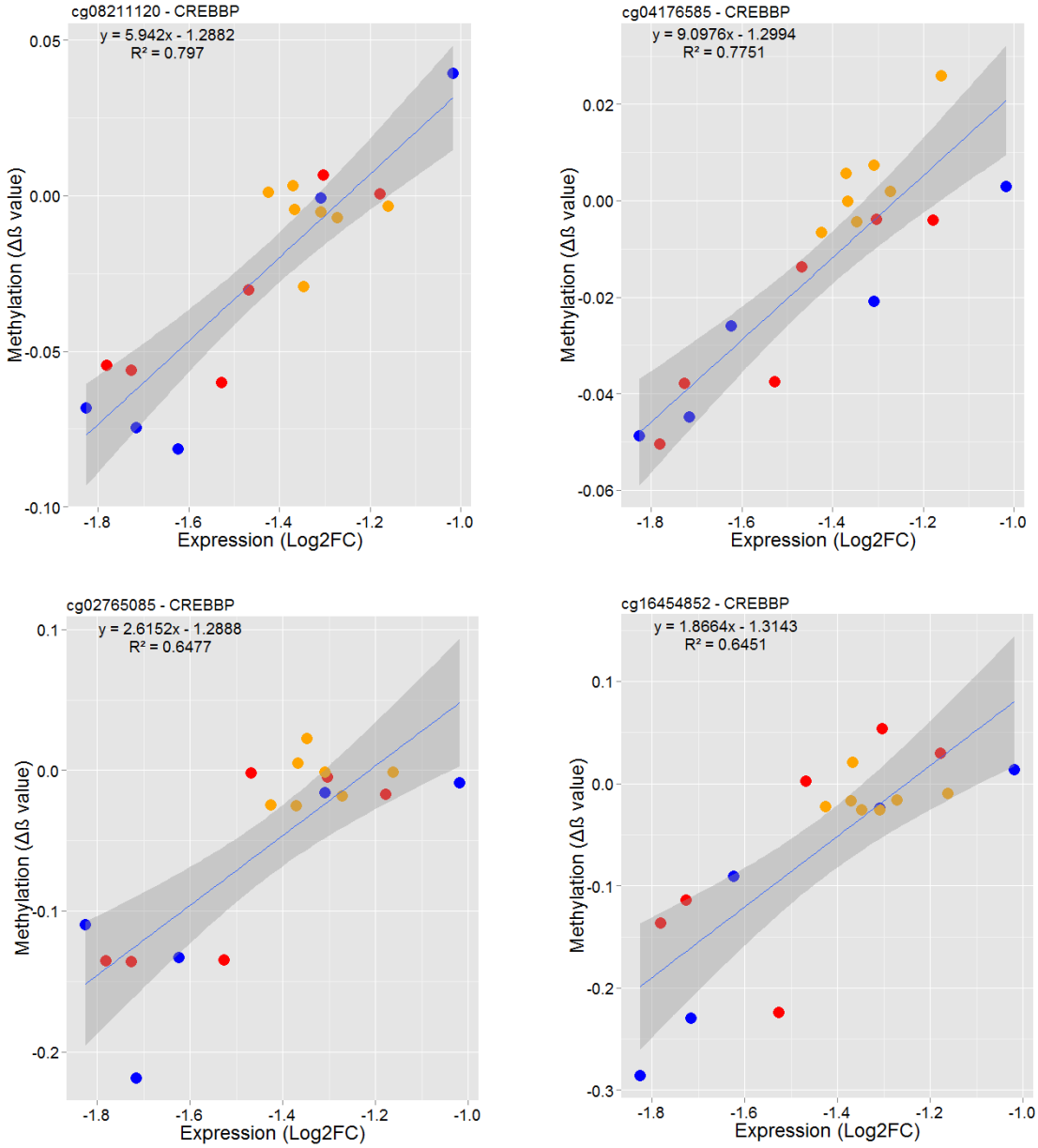


Figure 6.5.5. CREBBP CpGs which showed correlation between changes in their methylation with changes to CREBBP expression. Examples of 4 CREBBP CpGs which have significant correlation (R2=) between CpG methylation and CREBBP expression across all cell lines (n=18) after 5-Aza treatment compared with basal levels. The change in methylation is shown by the Δß value. The change in expression is shown by the log2 fold-change (Log2FC). Shaded areas indicate 95% confidence regions. ● Control (n=6), ● IPF (n=5), ● SSc (n=7).

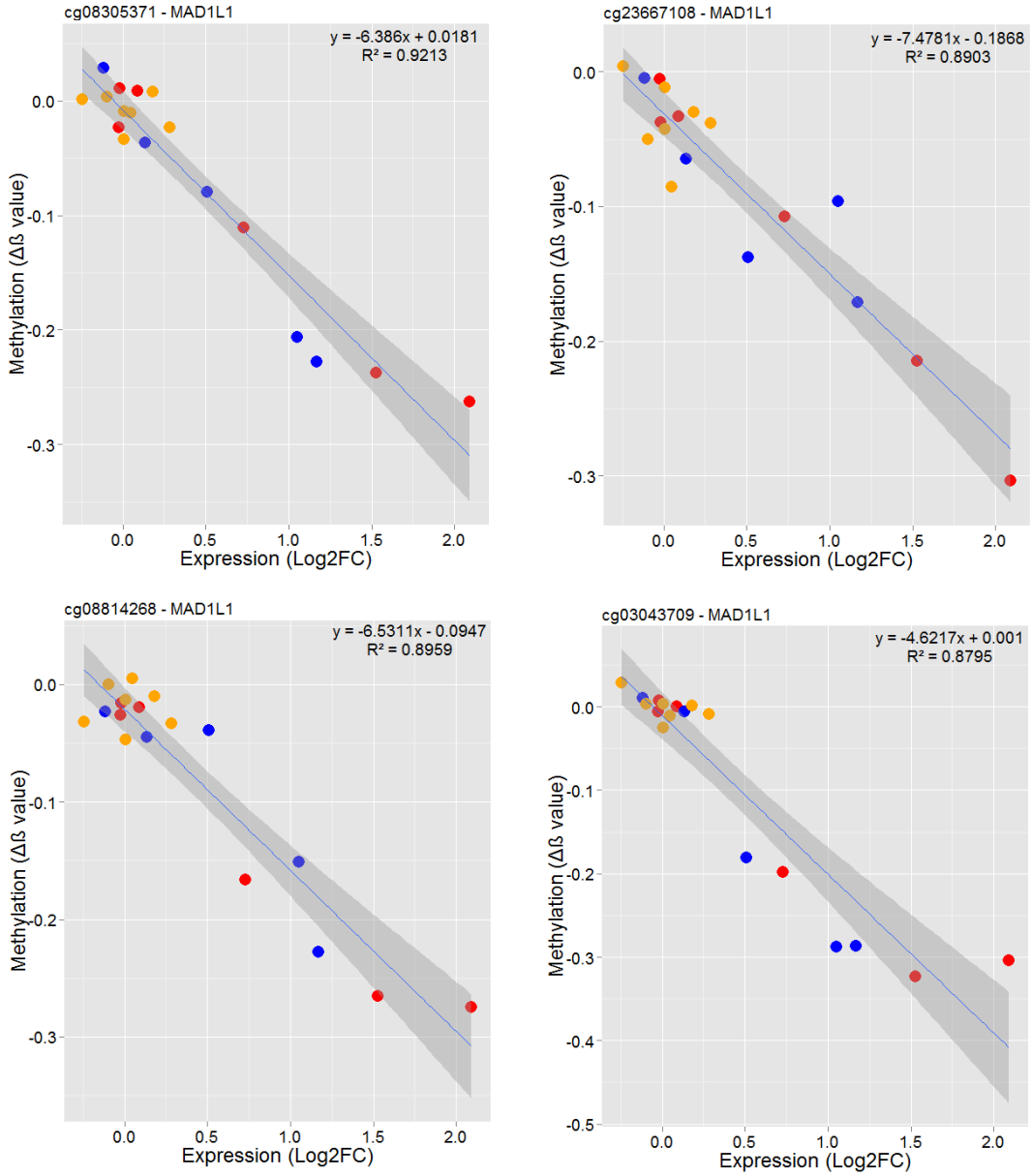


Figure 6.5.6. MAD1L1 CpGs which showed correlation between changes in their methylation with changes to MAD1L1 expression. Examples of 4 MAD1L1 CpGs which have significant correlation (R2=) between CpG methylation and MAD1L1 expression across all cell lines (n=18) after 5-Aza treatment compared with basal levels. The change in methylation is shown by the Δß value. The change in expression is shown by the log2 fold-change (Log2FC). Shaded areas indicate 95% confidence regions. ● Control (n=6), ● IPF (n=5), ● SSc (n=7).

6.6. Effect of DNMT inhibition on TNXB methylation in primary human lung fibroblasts

As previously shown in **section 6.2**, DNMT inhibition reduced methylation of multiple genes in control, IPF and SSc lung fibroblasts. 68, 34 and 36 CpG sites in the TNXB gene had decreased methylation (P<0.05) after 5-Aza treatment in control, IPF and SSc respectively (**Figure 6.6.1**). The highest frequency of CpGs with significantly (P<0.05) altered methylation after 5-Aza treatment were located in open sea regions in all lung fibroblasts (**Figure 6.6.2 and Figure 6.6.3**).

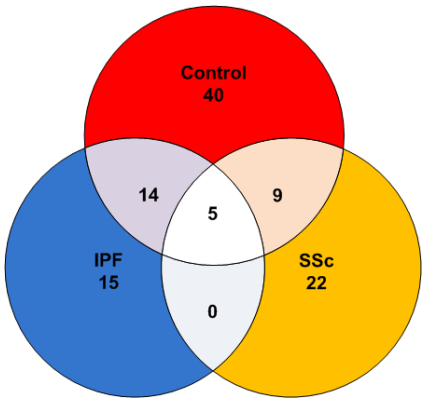


Figure 6.6.1. Distinct and overlapping CpG sites which had decreased methylation in lung fibroblasts after 5-Aza treatment. CpGs with significantly decreased (P<0.05) methylation after 5-Aza treatment in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts.

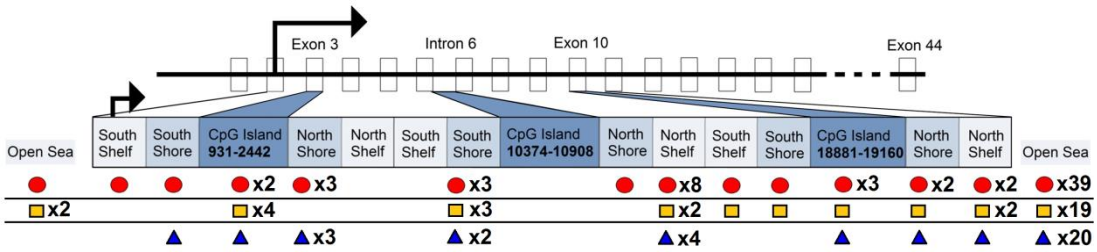


Figure 6.6.2. The location of CpG sites which had decreased methylation in lung fibroblasts after 5-Aza treatment. CpGs with significantly (P<0.05) decreased methylation after 5-Aza treatment in control ● (n=6), IPF ▲(n=5) and SSc ■ (n=7) lung fibroblasts.

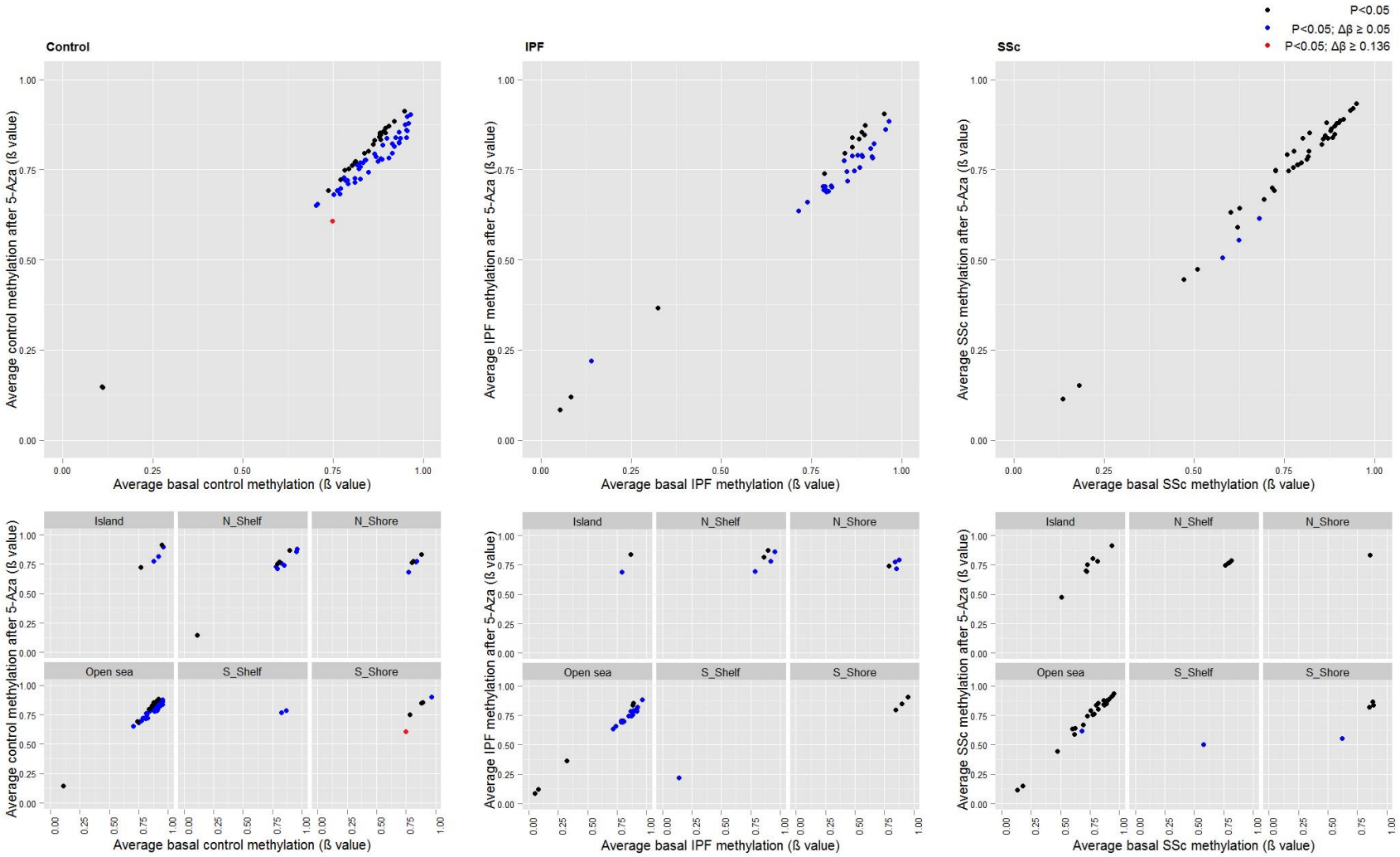


Figure 6.6.3. The location, number and degree of methylation change of CpG sites which had decreased methylation after 5-Aza treatment. CpG with significantly (P<0.05; ≥ 5%) decreased methylation after 5-Aza treatment in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts.

Three control cell lines and three IPF cell lines had large decreases in methylation in the TNXB gene after 5-Aza treatment, whereas little change in methylation was observed after 5-Aza treatment in SSc lung fibroblast cell lines (**Figure 6.6.4**). Large changes in methylation after 5-Aza treatment corresponded with large increases in TNXB mRNA expression (**Figure 6.6.4**).

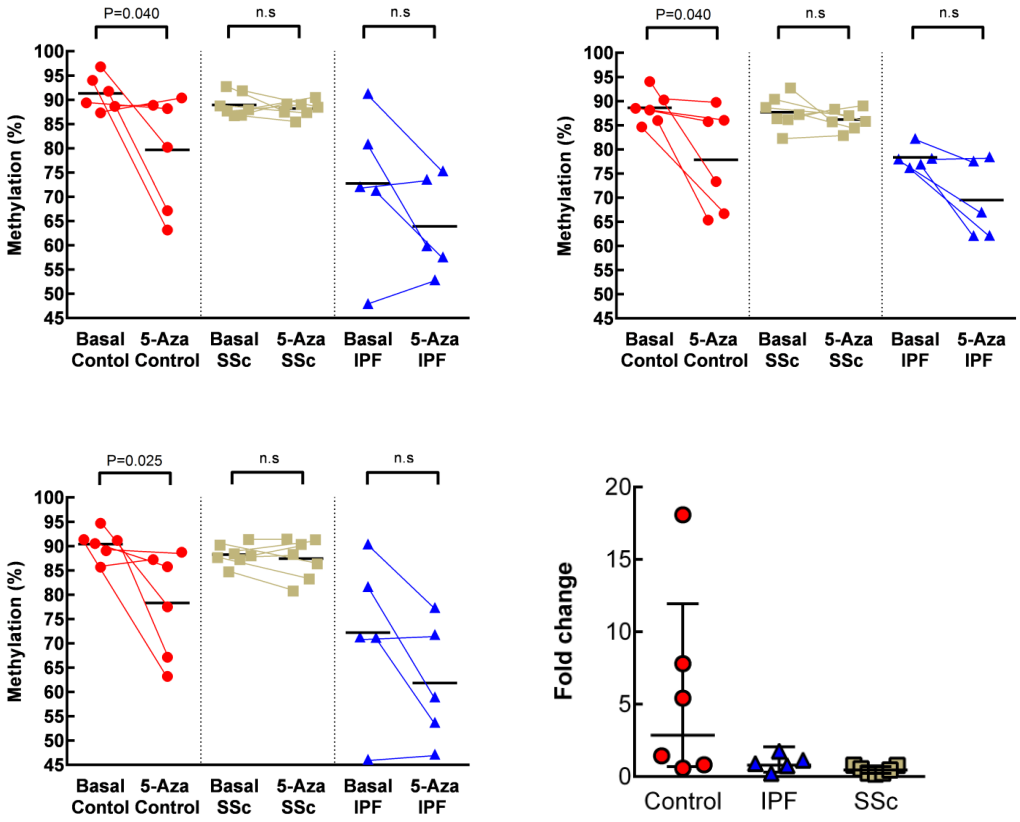


Figure 6.6.4. Three representative CpGs which had decreased in methylation after 5-Aza treatment in control lung fibroblasts. Three representative CpGs which had decreased in methylation after 5-Aza treatment in control lung fibroblasts. No significant change was observed in IPF (n=5) or SSc (n=7) lung fibroblasts. Each point represents a different lung fibroblast cell line. Bottom right: fold change in TNXB mRNA expression relative to the basal average in each fibroblast cell line after 5-Aza treatment.

6.7. Enrichment and pathway analysis of genes modulated by 5-Aza treatment

6.7.1. Biological enrichment of strong and weak responding cell lines to 5-Aza

GO-term enrichment analysis was performed on strong and weak responding cell lines to determine if similar biological processes were enriched in genes with varying levels of altered methylation. Strong responding cell lines: control 1 and IPF5 identified 867 and 835 biological processes, respectively, which were significantly enriched in genes with ≥13.6% changes in methylation after 5-Aza treatment. Weak responding cell lines: control 5, IPF4 and SSc7 identified 6, 9 and 85 biological processes, respectively, which were significantly enriched in genes with ≥13.6% changes in methylation after 5-Aza treatment. Whilst the weak responding cell lines had few enriched biological processes, the weak responding cell line SSc7, did have multiple biological processes including regulation of transcription, cell adhesion, lung development, regulation of histone deacetylation, response to TGFß and apoptosis which were significantly enriched in genes with ≥13.6% changes in methylation after 5-Aza treatment. Furthermore, the weak responding cell lines had multiple CpGs/genes with ≥5% changes in methylation after 5-Aza treatment (control 5: 19152 CpGs/10753 genes, IPF4: 21233 CpGs/11192 genes, SSc7: 24828 CpGs/12188 genes).

A strong responding IPF cell line, IPF5 (a cell line that had 15942 CpGs with ≥13.6% changes in methylation after 5-Aza treatment), and a weak responding IPF cell line, IPF4 (a cell line which only had 313 CpGs with ≥13.6% changes in methylation after 5-Aza treatment) were compared using genecodis enrichment analysis. For IPF cell line 5, genes with ≥13.6% changes in methylation after 5-Aza treatment were analysed, whereas, for IPF cell line 4, genes with ≥5% changes in methylation after 5-Aza treatment were analysed. 835 and 617 biological processes were identified as being enriched in genes with altered methylation after 5-Aza in IPF5 and IPF4, respectively, of which 407 overlapped. Furthermore, many of these biological processes were relevant to fibrosis (**Table 6.7.1.1**). These data suggest that cell lines which have a weaker response to 5-Aza still have multiple CpG/genes with altered methylation, many of which are associated with biological processes which may be relevant to pulmonary fibrosis.

IPF5 (**strong response to 5-Aza**)

Biological process (**BP**)	**Number of genes**	**Total number of genes associated with the BP**	**Hypergeometric P value**
Regulation of transcription, DNA-dependent	539	1609	2.53E-49
Apoptotic process	233	594	1.07e-31
Cell adhesion	237	556	7.72e-39
FGFR signaling pathway	37	78	2.19e-07
Chromatin modification	87	224	2.85e-11
Actin cytoskeleton organization	59	128	4.07e-11
ECM organization	34	73	1.38e-06
Collagen fibril organization	12	32	0.049419
Lung development	45	75	1.48e-13
Cellular response to TGFß stimulus	10	24	0.042908

IPF4 (**weak response to 5-Aza**)

Biological process (**BP**)	**Number of genes**	**Total number of genes associated with the BP**	**Hypergeometric P value**
Regulation of transcription, DNA-dependent	757	1609	2.36E-70
Apoptotic process	285	594	3.56e-27
Cell adhesion	310	556	3.42e-46
FGFR signaling pathway	40	78	6.89e-05
Chromatin modification	107	224	5.56e-10
Actin cytoskeleton organization	73	128	2.82e-11
ECM organization	46	73	4.47e-09
Collagen fibril organization	19	32	0.001445
Lung development	51	75	7.58e-12
Cellular response to TGFß stimulus	15	24	0.003

Table 6.7.1.1. Examples of biological processes enriched in genes with altered methylation after 5-Aza treatment in 1 strong and 1 weak responding IPF cell line. Biological processes enriched in genes with large changes in methylation (≥13.6%) in 1 strong responding cell line (IPF5) compared to biological processes enriched in genes with ≥5% changes in methylation in 1 weak responding cell line (IPF4) after 5-Aza treatment. The number of genes associated with each biological process (BP) and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

Genes (625) had large changes in expression (≥2-fold) after 5-Aza treatment in the strong responding cell line, IPF5. Enrichment analysis identified 320 biological processes enriched in these genes, many of which were relevant to fibrosis (**Table 6.7.1.2**). Genes (127) had large changes in expression (≥2-fold) after 5-Aza treatment in the weak responding cell line, IPF4. Enrichment analysis identified 311 biological processes enriched in these genes, many of which were relevant to fibrosis (**Table 6.7.1.2**) and 82 of which, overlapped with cell line IPF5. This suggests that smaller changes in methylation (<13.6) can still lead to large changes in expression (≥2-fold) and that the genes which have altered expression after 5-Aza in both strong and weak responsing cell lines, are associated with many biological processes relevant to fibrosis.

IPF cell line 5 (**strong response to 5-Aza**)

Biological process (**BP**)	**Number of genes**	**Total number of genes associated with the BP**	**Hypergeometric P value**
Prostaglandin biosynthetic process	3	15	0.016979
Anti-apoptosis	15	200	7.08e-05
Signal transduction	56	1176	7.34e-10
Cell-cell signaling	17	242	4.63e-05
Response to wounding	7	65	0.001961
Response to virus	16	144	3.70e-07
Cytokine-mediated signaling pathway	20	181	7.34e-09
Epithelial cell differentiation	4	42	0.034859
Type I interferon-mediated signaling	15	75	5.90E-10
Response to mechanical stimulus	7	51	0.000613

IPF cell line 4 (**weak response to 5-Aza**)

Biological process (**BP**)	**Number of genes**	**Total number of genes associated with the BP**	**Hypergeometric P value**
Prostaglandin biosynthetic process	2	15	0.015534
Anti-apoptosis	4	200	0.030658
Signal transduction	18	1176	2.30e-05
Cell-cell signaling	10	242	8.25e-06
Response to wounding	2	65	0.047935
Response to virus	3	144	0.045149
Cytokine-mediated signaling pathway	6	181	0.00173
Epithelial cell differentiation	2	42	0.037966
Type I interferon-mediated signaling	4	75	0.003268
Response to mechanical stimulus	2	51	0.045168

Table 6.7.1.2. Examples of biological processes enriched in genes with altered expression after 5-Aza treatment in 1 strong and 1 weak responding IPF cell line. Biological processes enriched in genes with large changes in expression (≥2-fold) in 1 strong (IPF5) and 1 weak (IPF4) responding cell line after 5-Aza treatment. The number of genes associated with each biological process and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

6.7.2. Biological enrichment and pathway analysis of genes which had correlation between changes in methylation and changes in expression after 5-Aza treatment

CpGs (1392) corresponding to 801 genes had significant correlations (R^2=≥0.5; P<0.05) between changes in methylation and changes in expression, in control, IPF and SSc lung fibroblasts. GO-term enrichment analysis identified 251 biological processes enriched in these genes, many of which are relevant to fibrosis (**Table 6.7.2.1**). Furthermore, KEGG enrichment identified 50 pathways, many of which have relevance to fibrosis, as being enriched in genes which may be directly regulated by methylation (**Table 6.7.2.2**).

Biological process (BP)	**Number of genes**	**Total number of genes associated with the BP**	**Hypergeometric P value**
ECM organization	12	73	2.87e-05
Apoptotic process	35	594	0.000103
Regulation of transcription, DNA-dependent	68	1609	0.000203
Cell adhesion	32	556	0.00029
Wnt receptor signaling pathway	12	110	0.000772
Response to virus	13	144	0.001927
Wound healing	8	68	0.006282
Actin cytoskeleton organization	11	128	0.006915
Cytoskeleton organization	9	106	0.016346
Positive regulation of fibroblast proliferation	5	39	0.026752

Table 6.7.2.1. Biological processes enriched in genes which have significant correlation between changes in methylation and changes in gene expression after 5-Aza. Biological processes enriched in genes with significant correlation (R^2=≥0.5; P<0.05) between changes in methylation and changes in expression after 5-Aza treatment, in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts. The number of genes associated with each biological process and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

KEGG pathway	Number of genes	Total number of genes associated with the BP	Hypergeometric P value
p53 signaling pathway	13	67	7.51E-07
Cell cycle	14	123	0.0001
Oocyte meiosis	12	110	0.000604
Pathways in cancer	21	324	0.000703
Focal adhesion	16	197	0.000744
Amoebiasis	11	102	0.000785
ECM-receptor interaction	10	84	0.000922
Calcium signaling pathway	14	175	0.001462
Regulation of actin cytoskeleton	15	209	0.002467
Rheumatoid arthritis	9	84	0.002669
Dilated cardiomyopathy	9	89	0.003466
Vibrio cholerae infection	7	53	0.003594
Ribosome biogenesis in eukaryotes	8	73	0.00408
MAPK signaling pathway	16	262	0.005983
TGFß signaling pathway	8	82	0.007816
Axon guidance	10	128	0.008473
Small cell lung cancer	8	84	0.008608
Malaria	6	48	0.008876
Tight junction	10	130	0.009047
Amyotrophic lateral sclerosis (ALS)	6	50	0.009395
PPAR signaling pathway	7	70	0.009956
Cytokine-cytokine receptor interaction	15	259	0.010144
Fc gamma R-mediated phagocytosis	8	92	0.010881
GAG biosynthesis - chondroitin sulfate	4	22	0.011585
GnRH signaling pathway	8	98	0.014999
Wnt signaling pathway	10	149	0.018348
Hypertrophic cardiomyopathy (HCM)	7	82	0.020279
Complement and coagulation cascades	6	65	0.023526
Notch signaling pathway	5	46	0.023884
Hepatitis C	9	133	0.024226
Gap junction	7	88	0.024609
Purine metabolism	10	158	0.025048
Ribosome	7	86	0.025537
Amino sugar and nucleotide sugar metabolism	5	48	0.027037
Bacterial invasion of epithelial cells	6	70	0.030024
Adherens junction	6	71	0.031273
Chronic myeloid leukemia	6	73	0.032186
Gastric acid secretion	6	72	0.032559
Base excision repair	4	33	0.032616
RNA transport	9	145	0.033431
Melanogenesis	7	98	0.03545
Pathogenic Escherichia coli infection	5	54	0.036099
Neurotrophin signaling pathway	8	124	0.036267
Osteoclast differentiation	8	126	0.038855
Chagas disease (American trypanosomiasis)	7	102	0.038992
Protein digestion and absorption	6	78	0.039198
Inositol phosphate metabolism	5	57	0.040367
Hematopoietic cell lineage	6	83	0.048048
Cytosolic DNA-sensing pathway	5	60	0.048733
Ubiquitin mediated proteolysis	8	135	0.049444

Table 6.7.2.2. KEGG pathways enriched in genes which have significant correlation between changes in methylation and changes in gene expression after 5-Aza. KEGG pathways enriched in genes with significant correlation between changes in methylation and changes in expression after 5-Aza treatment in control (n=6), IPF (n=5) and SSc (n=7) lung fibroblasts. The number of genes associated with each biological process (BP) and the total number of genes that belong to each process is shown. The P-value was calculated using the hypergeometric distribution and corrected for multiple testing using the Benjamini-Hochberg FDR method. A smaller adjusted P-value correlated with greater gene enrichment.

Using data combined from both GO-term and KEGG enrichment analyses, STRING analysis was used to determine how genes potentially directly regulated by methylation could interact with each other in specific pathways such as WNT signalling (**Table 6.7.2.3**) and ECM-interactions (**Table 6.7.2.4**). Furthermore, genes were highlighted which were previously identified as having significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and/or expression (TNoM ≤1; $P<0.05$) in IPF compared to control lung fibroblasts.

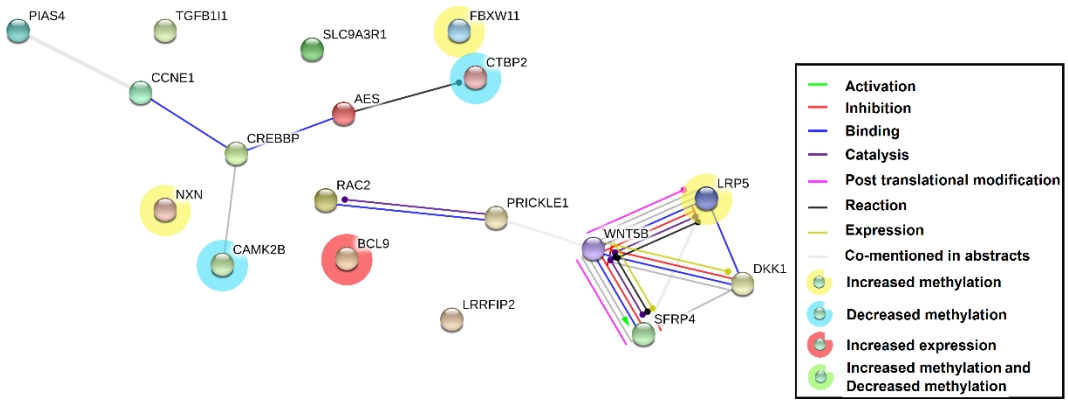


Table 6.7.2.3. The Wnt signalling pathway enriched in genes with a significant correlation between changes in methylation and changes in expression after 5-Aza treatment. Genes with significantly altered basal methylation and/or expression in IPF (n=5) compared to control (n=6) lung fibroblasts are highlighted.

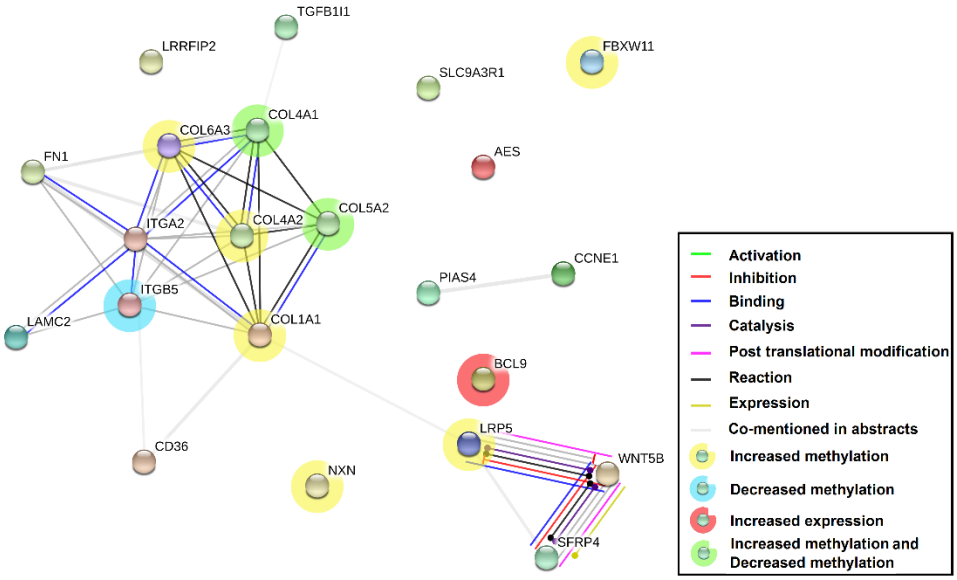


Table 6.7.2.4. The ECM-interaction pathway enriched in genes with a significant correlation between changes in methylation and changes in expression after 5-Aza . Genes with significantly altered basal methylation and/or expression in IPF (n=5) compared to control (n=6) lung fibroblasts are highlighted.

6.8. Summary

- Treatment of cell lines with 5-Aza affected methylation at multiple CpG sites and the expression of multiple genes in all cell lines.

- Multiple biological processes previously identified as having significantly altered methylation in IPF and SSc compared to control lung fibroblasts were identified as being enriched in genes with significantly altered methylation ($P<0.05$) after 5-Aza treatment.

- The extent of methylation and expression changes after 5-Aza treatment varied between cell lines.

- Correlation between changes in methylation and changes in expression, after 5-Aza treatment, identified multiple novel genes potentially regulated by methylation in lung fibroblasts and multiple novel pathways potentially important in pulmonary fibrosis.

- Multiple genes for which changes in methylation and changes in expression correlated after 5-Aza treatment, had significantly altered basal methylation and/or basal expression in IPF/SSc compared to control lung fibroblasts.

- TNXB methylation was significantly reduced in control cell lines after 5-Aza treatment and corresponded with increased TNXB expression. The three control fibroblast cell lines with large decreases in methylation at multiple CpGs within the TNXB had the largest increases in TNXB expression.

- The number of CpGs with a large (≥13.6%) change in methylation after 5-Aza treatment correlated with the number of genes with large (≥2FC) changes in expression.

- Small changes in methylation (≤5%) also correlated with large changes (≥2FC) in expression in multiple genes, suggesting small changes in methylation may have an effect on the expression of multiple genes.

Chapter 7. Discussion

7.1. Overview

Previous studies have identified multiple genes with altered methylation in lung tissue derived from IPF patients compared to non-fibrotic controls (Sanders et al, 2012, Rabinovich et al, 2012, Yang et al, 2014). Rabinovich et al, used Agilent microarrays to examine global methylation patterns in IPF lung tissue and found global methylation was decreased in IPF compared to control lung tissue. Interestingly, the pattern of methylation seen in IPF was similar to that which was seen in lung cancer tissue, where methylation patterns of 65% of CpG islands overlapped (Rabinovich et al, 2012). Sanders et al reported no global difference in methylation between IPF and control lung tissue but did find that DNMT3A and DNMT3B had increased expression (Sanders et al, 2012). Both these studies also suggested that DNA methylation may influence gene expression although neither study analysed genome-wide expression in relation to CpG methylation. Yang et al, used CHARM arrays to analyse the methylation of 4.6 million CpG sites and subsequently compared global expression using Agilent expression arrays, making this study the most comprehensive study to date looking at CpG methylation in IPF lung tissue (Yang et al, 2014). 2130 differentially methylated regions were identified in IPF compared to control tissue, of which 738 corresponded to significant changes in gene expression. Whilst these studies offer a valuable insight into how methylation differs between IPF and control lung tissue and potentially how methylation may affect gene expression, the use of lung tissue, which contains multiple cell types, makes it impossible to distinguish cell type-specific changes in methylation, as large changes in methylation in one cell type could be masked by large changes in methylation in a different cell type. Therefore, examining methylation in lung tissue could mask changes in methylation of genes which may be important in pulmonary fibrosis or result in over or under estimation of methylation. Furthermore, the use of different lung tissue samples and their degree of heterogeneity, different stringency criteria for inclusion/statistical analysis and the use of different microarray platforms could all contribute reasons as to why these previous studies had different results regarding global methylation.

To date, only one study has examined global DNA methylation in IPF lung fibroblasts (Huang et al, 2014). This study identified global changes in methylation between IPF and two control lung fibroblast groups with 58% of CpGs in IPF lung fibroblasts being hypomethylated (Huang et al, 2014). However, a number of limitations were present in this study. For example, Huang et al, utilised the Illumina Infinium HumanMethylation 27k BeadChip array which only covers 1-2 CpGs per a gene and is biased towards CpG islands. Furthermore 6-10% of probes on the array were non-specific (Chen et al, 2011) and control fibroblasts used for analysis were mainly male, whereas IPF fibroblasts were mainly female, thus making it impossible to determine if the reported differences in methylation between IPF and controls were real or due to differences between males and females. Methylation was also compared to gene expression data obtained from other studies, thus the cell lines used to determine methylation were not the same cell lines used to determine expression. This makes it difficult to

determine whether genes with altered methylation truly have changes in expression, as different culture techniques, array platforms, and heterogeneity between fibroblast cell lines all add bias to their results.

To my knowledge, this thesis is the first to examine genome-wide methylation in IPF and SSc lung fibroblasts and to examine the effects of methylation on gene expression using the same cell cultures. Furthermore, it is the first study to characterise gender-specific methylation differences in human lung fibroblasts and compare gender differences in methylation in IPF males with control males and IPF females with control females.

7.2. Illumina Infinium HumanMethylation 450k BeadChip Array

The Illumina Infinium HumanMethylation 450k BeadChip microarray is the successor to the previous Ilumina methylation array, the Infinium HumanMethylation 27k BeadChip array. The Illumina 450k methylation array was chosen for its comprehensive coverage of the human methylome. The Illumina 450k array interrogates 482421 CpGs representing ~2% of the entire methylome, with an average of 17 CpGs per a gene, covering CGIs and their flanking shore, shelf and open sea regions, which are recognised as playing an increasingly important role in disease and gene regulation (Irizarry et al, 2009, Rakyan et al 2011, Rao et al, 2013, Bockmühl et al, 2015). The 450k array represents a significant upgrade from the 27k array which is heavily biased towards CGIs and covers far fewer CpGs, although, a number of limitations have been discovered (Chen et al, 2013, Price et al, 2013). By definition, the Illumina microarrays are biased towards which CpGs they examine as the list of CpGs was compiled by a consortium of experts and aimed towards their interests. Furthermore, in 2013 two studies identified a number of probes which overlapped with SNPs or were non-specific (Chen et al, 2013, Price et al, 2013). Others have also suggested multiple probes on the 450k array are affected by a SNP (Liu et al, 2013). There is no general consensus as to whether to remove or leave probes on the array which cover a SNP, although many recent studies have removed or have suggested removing them (Price et al, 2013, Chen et al, 2013, Fortin et al, 2014). SNPs which cover a CpG site or within a probe could be compromised by sample genotype (Dedeurwaerder et al, 2011, Price et al, 2013), thus all probes covering a SNP were removed. Non-specific probes can hybridise to multiple different genomic regions and could potentially measure multiple sites of methylation. Non-specific probes represented approximately 8.6% of all probes on the array (Price et al, 2013), therefore these probes were filtered out to avoid potentially inaccurate methylation readings. Normalisation of the array was performed by Cambridge Genomic Services (CGS, UK) using the popular R package Lumi (Du et al, 2008). However, as the number of researchers using the 450k array increases so does the number of tools being developed to help with analysing array data (Dedeurwaeder et al, 2013). Therefore future studies should explore the increasing number of tools now available to analysis the 450k array.

7.2.1. Criteria for inclusion

The change in methylation required to have a significant biological effect and whether this change in methylation is the same for all genes or different cell lines remains a difficult question to answer. There is no universal way to analyse methylation data and different normalisation methods and array technologies make it hard to compare data from different studies. For example, Sanders et al, used a P<0.05 to avoid missing true positive signals and Huang et al, used a P<0.05 with a 2-fold change, whereas, Rabinovich used a FDR of P<0.05. In my studies, an FDR of P<0.05 resulted in very few CpGs/genes being detected, likely due to the heterogeneity between IPF samples and effects of sex-specific differences in methylation in control and IPF lung fibroblasts. Instead, I chose a cut-off methylation value of $\Delta\beta \geq 0.136$ (≥13.6% change in methylation) with a P value <0.05. This was based on previous studies showing a 13.6% change in methylation could be detected with 95% confidence (Bibikova et al, 2009, Lokk, 2012). The β-value, which represents the level of methylation (0= unmethylated, 1= 100% methylated), was chosen as it allows easy biological interpretation and is recommended for use when an absolute difference in methylation cut-off is applied (Dedeurwaeder et al, 2011).

Several studies have reported that altered methylation of specific genes in IPF corresponds with their expression, however, to what extent methylation must change in order to have an effect on gene expression is unknown. Sanders et al, showed that genes including DDAH and TP53INP1 had large methylation differences (>20%) in IPF compared to control lung tissue. These changes in methylation corresponded to altered gene expression (Sanders et al, 2012) however, several studies have shown that small changes (<5%) in methylation can also affect individual gene expression. For example, Yang et al, showed that 8 CpGs on the CASZ1 gene, had on average a 3.5% change in methylation in IPF compared to control type II alveolar epithelial cells which corresponded with increased gene expression (Yang et al, 2014). In 2010, Huang et al, showed <5% changes in methylation of the PTGER2 gene could affect PTGER2 expression and in 2014, showed that an average decrease of 6.9% over 28 CpGs of the gene CDKN2B resulted in increased expression (Huang et al, 2010, Huang et al, 2014). This suggests <13.6% changes in methylation could potentially be important in regulating expression. In support of this, small but significant differences in TNXB methylation identified by the 450k array were confirmed by bisulfite sequencing. Furthermore, cell lines which had a weak response to 5-Aza with relatively few CpGs reaching the 13.6% cut-off, still had multiple genes with large changes (≥2-fold) in expression (**see Chapter 6**). This suggests that small changes in methylation potentially can have an effect on gene expression and that a 13.6% cut-off may be too stringent. However, it has been reported that the 450k array may not be suitable for the detection of small differences in methylation due to technical variability in measurements (Dedeurwaerder et al, 2011). Therefore, to avoid the possibility of including a large number of false positives, only genes reaching the 13.6% cut-off were included.

For gene expression analysis, the TNoM method was applied (**see Chapter 2: section 2.6.6**) which has previously been applied to gene expression data in other IPF studies (Zuo et al, 2002). A TNoM <1 with a p value <0.05 was chosen to increase the stringency criteria as few genes reached a FDR <0.05 rate, whereas multiple genes reached a non-stringent p value <0.05. Other studies have used a non-stringent p value <0.05 with ≥2fold-change in gene expression (Huang et al, 2014), however, what change in gene expression is sufficient to have a biological effect is unknown. Therefore, to avoid missing genes with smaller, but potentially important fold-change in expression, no fold-change criteria was applied. One potential limitation of using TNoMs in IPF studies is that IPF fibroblast cell lines are notoriously heterogeneous (Martinez et al, 2005, Habiel and Hogaboam, 2014, DePianto et al, 2015). Therefore, filtering samples over 1 TNoM may be too stringent, however a large number of genes reached this statistical threshold (568 in IPF and 688 in SSc compared to control lung fibroblasts) many of which have previously been linked to fibrosis including IL8 (Ziegenhagen et al, 1998), WNT2B (Zhou et al, 2014), PPARγ (Lakatos et al, 2007), S100A4 (Tomcik et al, 2014), NOTCH3 (Dees et al, 2011), IGFBP7 (Hsu et al, 2011), CCL13 (Yanaba et al, 2010), IL7R (Grigoryev et al, 2008) and TIMP4 (Elias et al, 2008), suggesting this approach was a good compromise.

7.3. Genome-wide methylation in IPF and in SSc compared to control lung fibroblasts.

The results from this study identified 7153 CpGs corresponding to 4563 genes in IPF and 8392 CpGs corresponding to 5294 genes in SSc which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) compared to control lung fibroblasts. In agreement with previous studies examining global DNA methylation in lung tissue and lung fibroblasts (Rabinovich et al, 2012, Huang et al, 2014) the majority of CpGs (69%) had decreased methylation in IPF. In contrast, Sanders et al, reported no global differences in CpG methylation however, the Illumina 27k array they used covered fewer CpGs, was heavily biased towards CGIs in promoter regions and did not take into account methylation at other genomic regions. Furthermore, they used lung tissue as opposed to lung fibroblasts. As previously stated, lung tissue contains multiple cell types which may result in over or under estimation of methylation, thus fibroblast-specific changes in methylation were impossible to determine. In agreement with previously published data from targeted studies using lung fibroblasts, *THY-1*, *PTGER2* and *P14ARF* all had CpG sites with increased methylation in IPF compared to control lung fibroblasts (**Table 7.3.1**), although the increase in methylation did not meet the stringency criteria I chose (Δβ ≥ 0.136; P<0.05)**.**

Gene name	IPF Δβ	P-value
THY1	0.268	0.11644
P^{14}ARF	0.02593	0.039913
PTGER2	0.06889	0.363231

Table 7.3.1. Comparison of microarray data with genes previously identified as having increased methylation in IPF compared to control lung fibroblasts.

The majority (67%) of CpGs which had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in SSc had increased methylation compared to control lung fibroblasts. Furthermore, 7827 CpGs corresponding to 5082 genes had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF compared to SSc lung fibroblasts. These data suggest that global methylation patterns are different between IPF and SSc which may reflect outcomes such as disease progression and sex bias.

Despite global differences in methylation, multiple genes with significantly altered methylation (Δβ ≥ 0.136; P<0.05) overlapped between IPF and SSc lung fibroblasts. These genes included, interleukins, WNTs, collagens, miRs and mucins, all of which have been implicated in the pathogenesis of IPF and/or SSc. Furthermore, multiple biological processes and pathways were enriched in genes that had significantly altered methylation (Δβ ≥ 0.136; P<0.05) in IPF and SSc compared to control lung fibroblasts. Many of these pathways including coagulation (Scotton et al, 2009), apoptosis (Ramos et al, 2001, Thannickal et al, 2006, Fattman, 2008) and ECM interactions (Pardo et al, 2005, Emblom-Callahan et al, 2010, Huang et al, 2014), have all previously been linked to IPF. Wnt signalling was also an enriched biological process identified in both IPF and SSc lung fibroblasts which has previously been associated with the pathogenesis of both IPF and SSc (Chilosi et al, 2003, Konigshoff et al, 2008, Vuga et al, 2009, Wei et al, 2012). In support of this finding, a recent study found two genes involved in Wnt signalling; DKK1 and SFRP1 have decreased expression as a direct result from promoter hypermethylation in SSc fibroblasts (Dees et al, 2013). Pathway analysis is discussed in detail in **section 7.6**.

7.4. Distribution of methylation in lung fibroblasts

Methylation across the genome is typically regarded as bimodal with CpGs having low methylation in CpG islands close to gene promoters and high methylation in other genomic areas (Jones, 2012). CpG islands within promoter regions are classically reported as being unmethylated to allow transcription to occur, however few studies have been able to delineate methylation distribution patterns based on the location of CpGs (with respect to CpG islands) and within different genomic regions due to limited CpG coverage on previous array platform technology. This has important implications particularly when trying to determine relationships between methylation and gene transcription as methylation within different genomic regions can act in opposing ways (Wan et al, 2015). Although there are several studies looking at genome-wide methylation in a variety of tissues or in a disease context, to my knowledge there are no studies which have examined unbiased genome-wide distribution of methylation in primary human lung fibroblasts.

Distribution of methylation in autosomes was bimodal in control, IPF and SSc lung fibroblasts with the highest frequency of CpGs having low methylation (0-15%) and high methylation (85-95%) (**Section 3.2, figure 3.2.1**). A bimodal pattern of CpG methylation was also observed on the X-chromosome with the highest frequency of CpGs having 25-35% methylation. However genome-wide distribution of methylation on the X-chromosome in female, but not male lung fibroblasts, showed a partially

methylated pattern of methylation which was considered consistent with X-inactivation in females (Bell et al, 2011, Johansson et al, 2013, Joo et al, 2014).

Distribution of CpG methylation in CGIs within 1.5kb of their corresponding gene's TSS was unimodal with the majority of CpGs with low methylation (0-15%) consistent with the dogma that the majority of CGIs are unmethylated (Bird et al, 2012). However, CpGs located in CGIs further than 1.5kb from their corresponding gene's TSS had a bimodal distribution of CpG methylation with the highest frequencies of CpGs having 10-15% and 85-90% methylation. These CGIs located distal to promoter regions, coined 'orphans' (Illingworth et al, 2010), have a different distribution pattern of methylation compared to CGIs within promoter region. Some studies have identified CpG methylation in gene bodies positively correlates with gene expression (Lister et al, 2009, Kulis et al, 2012, Banovich et al, 2014), thus the dogma that methylation in CGIs inversely correlates to gene expression (Bird et al, 2002) may not be true for CGIs located in different genomic areas. This also opens up the possibility that methylation could both positively and negatively correlate with gene expression based on the location of CpG methylation. Indeed, out of the 724 genes in lung fibroblasts which had correlation between methylation and expression, 96 had CpGs which both positively and negatively correlated with their respective genes expression.

Distribution of CpG methylation in non-CGIs, which covers shore, shelf and open sea regions were analysed individually. North and south shore regions flank CGIs and are located up to 2kb away. Within 1.5kb of their corresponding gene's TSS, shore regions had a high frequency of CpGs with low methylation, whereas shore regions further than 1.5kb from their gene's TSS had a higher frequency of CpGs with high methylation. North and south shelf regions are located 2-4kb away from CGIs. Within and further than 1.5kb of their gene's corresponding TSS, shelf regions had a high frequency of CpGs with high methylation. Open sea regions which are denoted as being located beyond 4kb of a CGI had the highest frequency of CpGs with high methylation within and further than 1.5kb from their gene's TSS. Open sea regions within 1.5kb from their genes TSS had a bimodal distribution of CpG methylation. The distribution of CpG methylation further than 1.5kb from their gene's TSS was also bimodal but with a greater proportion of CpGs having high methylation compared to low methylation.

In agreement with other studies examining genome-wide distribution of methylation patterns in other cell types, shore regions tended to have a similar methylation pattern to CGIs based on genomic regions, whereas the majority of shelf regions had high methylation (Lokk et al, 2014, Zhang et al, 2015). Also in agreement with these studies, open sea regions had the highest number of CpGs with high methylation (Lokk et al, 2014, Zhang et al, 2015), however, a large number of CpGs (~2000) had low methylation in open sea regions within 1.5kb of their corresponding gene's TSS which may play an important role in regulating gene expression but have been ignored by array platforms which only focus on CGIs.

7.4.1. Distribution of CpGs with altered methylation in IPF/SSc compared to control lung fibroblasts

Of the 7153 CpGs with altered methylation in IPF compared to control lung fibroblasts, 985 (14%) were located in CGIs, whereas 6168 (86%) were located in non-CGIs. This is consistent with the study published in 2014 by Yang et al, which reported the majority of CpGs with altered methylation were located distal to CGIs. Interestingly Yang et al, show that the majority of CpGs with altered methylation in IPF compared to control lung tissue were located in shore regions, however, they defined shore regions as 0-3000bp away from a CGI as opposed to the standard definition of 0-2000bp away and thus incorporate CpGs normally defined as being in shelf regions (2000-4000bp away) into their analysis. They also identified a large number of differentially methylated regions which are >3000bp away from CGIs. This is consistent with our study, although the majority of CpGs with altered methylation in IPF occured in open sea regions (at least 4000bp away from a CGI). Barring the obvious fact that Yang et al, analysed lung tissue as opposed to lung fibroblasts, these differences may be due to the differences in CpG coverage of the arrays. The Illumina 450k is able to analyse ~482,000 CpGs (324,973 after applying filtering) whereas Yang et al, were able to analyse 4.6 million, and thus were potentially able to determine the distribution of CpG methylation more accurately, albeit in lung tissue. Furthermore, Yang et al, did not state what percentage of CpGs were analysed within each region, thus making it impossible to determine whether CpGs in different regions were actually over-represented relative to the number of CpGs studied in each region. Therefore, shore and open sea regions may have been represented in a different ratio compared to our study, which could explain the discussed differences.

To try to determine whether different regions were over-represented in CpG with altered methylation, the total number of CpGs in each location was divided by the total number of CpGs on the array to determine what percentage of CpGs were analysed in each area. In our study 34% of all CpGs analysed were in open sea regions compared to 24% in shore regions, 33% in CGIs and 9% in shelves. Open sea regions consistently had a higher O/E ratio of CpGs with altered methylation (≥13.6%) in IPF compared to control lung fibroblasts. Shore regions had a higher O/E ratio on CpGs with increased methylation, but no significant difference in O/E ratios with CpGs with decreased methylation. However, whilst this analysis confirmed that altered methylation occurred more frequently in open sea regions, it did not take into account that CpGs with low methylation could not go lower than 0% methylation, nor CpGs with high methylation, which could not go over 100%. Thus, a 13.6% difference was mathematically impossible which would bias results for CpGs within regions such as CGIs/shores as they have a high frequency of CpGs with low methylation.

To draw more accurate comparisons, a greater number of CpGs in lung fibroblasts would need to be analysed to fully determine which locations have altered CpG methylation. Nonetheless, this study suggests the majority of CpGs with altered methylation in IPF are located in non-CGI areas.

Furthermore, even though the Illumina 27k array is heavily biased towards CGIs in promoter regions, the majority of CpGs identified by Huang et al, which had altered methylation in IPF compared to control lung fibroblasts, were located outside of CGIs (Huang et al, 2014).

7.5. Validation of Illumina microarrays

TNXB was bisulfite sequenced to validate the methylation microarrays, based on a large number of CpGs on the array with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in IPF compared to SSc and control lung fibroblasts and a thorough literature search suggesting its potential role in collagen deposition, fibril organisation (Elefteriou et al, 2001) and fibrosis (Jing et al, 2011). Bisulfite sequencing confirmed similar changes in methylation at multiple CpGs in two different CGIs (1 located in exon 3 and one located in exon 10), with all having decreased methylation in IPF compared to SSc and control lung fibroblasts and showed strong correlation with microarray data (Pearsons r=0.77). Furthermore, 7 CpGs not on the array also had significantly decreased methylation in IPF compared to control lung fibroblasts. The total methylation observed was different between bisulfite and microarray data, however, this could potentially be explained by different normalisation methods. One unexpected challenge during bisulfite sequencing was designing primers which didn't overlap CpGs in the dense CGI located in exon 3. CpGs should not be included in bisulfite primers to avoid discrimination against methylated or unmethylated DNA (Li and Dahiya, 2002). Bisulfite primers should also be longer than primers used in regular PCR to ensure specificity and the length of the PCR product should ideally be less than 400bp to avoid potential DNA degradation during bisulfite modification (Li and Tollefsbo, 2011). In this region there were few primer choices which fit all the criteria other than the primers I designed (**see Chapter 2: section 2.11.1**). In future studies, to overcome this problem for other dense areas of CpG methylation for which no suitable primers can be made, different techniques such as pyrosequencing could be used.

Gene expression arrays (basal and 5-Aza) were validated using qRT-PCR on multiple genes including, IL8, CADM1, EIF1AY, MMP10 and MMP12. These genes were chosen based on previously being associated with fibrosis and/or had large changes in expression in IPF and/or SSc compared to control lung fibroblasts. qRT-PCR and microarray data showed strong correlation for each gene analysed (IL8: r=0.91, CADM1: r=0.74, EIF1AY: r=0.85), validating the microarray's ability to detect significant differences in gene expression. MMP10 and MMP12 were used to validate the 5-Aza expression arrays with both showing a strong correlation between qRT-PCR and microarray data, however, the fold-changes observed by qRT-PCR were far greater than on the microarray. This is most likely due to the increased sensitivity of qRT-PCR compared to microarray techniques and is a common observation (Chen et al, 2009). TNXB mRNA expression was examined using qRT-PCR as the probe on the array was not detected. This may have been due to a faulty probe or a probe which did not have a low enough P value for detection during quality control. qRT-PCR showed TNXB was expressed in lung fibroblasts and it's expression was significantly increased in IPF compared to SSc and control lung

fibroblasts. IHC in lung tissue samples derived from control, IPF and SSc lungs confirmed increased TNX expression in IPF. Furthermore, other members within our group were examining other genes of interest and showed similar results, which further validated the methylation and expression arrays.

7.6. Biological interpretation and pathway analysis

As previously described (**see Chapter 4: section 4.1**), two different types of enrichment analyses were performed; GO-term and PFAM. GO-term analysis is commonly used to identify enriched functional groups for a given list of genes and can be performed using a variety of online platforms such as DAVID analysis (Huang et al, 2008, Fang and Gough, 2013) or Genecodis (Carmona-Saez et al, 2007, Nogales-Cadenas et al, 2009, Tabas-Madrid et al, 2012). PFAM domain enrichment analysis uses pre-defined protein-level GO annotations to determine if a specific biological process is enriched in given list of PFAMs. Furthermore, each biological process is given a level of specificity ranging from highly general to highly specific (1=highly general, 2; general, 3; specific, 4; highly specific). For example, the general term 'ECM organisation' includes PFAMs associated broadly with this ontology, whereas the specific term 'collagen fibril organisation' only contains PFAM associated with this process. The main difference between GO-term and PFAM enrichment analysis is that GO-term enrichment analysis uses ontology annotations relating to genes associated with a specific biological process, whereas PFAM-domain centric analysis uses ontology annotations relating to the functional units of proteins associated with a specific biological process. PFAM enrichment analysis therefore identifies PFAMs containing genes associated with a specific biological process. This allows the potential to identify genes which share domains which are associated with a specific biological process but have not yet themselves been associated with the specific biological process. Thus, PFAM analysis can potentially aid in the identification of novel genes involved in biological processes associated with fibrosis.

Multiple biological processes enriched in genes with significantly altered methylation and/or expression were consistent with previous literature and with diseases characterised by fibrosis (Yang et al, 2014, Huang et al, 2014). These biological processes included Wnt signalling, ECM organisation, apoptosis, gene expression, integrin signalling and EMT. Of the biological processes enriched in PFAMs containing genes with significantly altered methylation, 86% were the same in both IPF and SSc. 79% of biological processes enriched in PFAMs containing genes with significantly altered expression were the same in both IPF and SSc. Furthermore, 66% of biological processes enriched in PFAMs containing genes with altered methylation in IPF compared to control lung fibroblasts were the same biological processes enriched in PFAMs containing genes with altered expression. Similarly, 63% of biological processes enriched in PFAM containing genes with altered methylation in SSc compared to control lung fibroblasts were the same biological processes enriched in PFAMs containing genes with altered expression. This demonstrates a potential link between methylation and expression and suggests that multiple biological processes overlap between IPF and SSc. GO-term enrichment showed similar biological processes contain genes with altered methylation in IPF and SSc compared to control lung

fibroblasts, but far fewer biological processes were identified containing genes with significantly altered expression in IPF and SSc compared to control lung fibroblasts. This is likely due to the fact that in PFAM enrichment, the number of PFAMs associated with specific biological process are far less than the number of genes and only one gene needs to be associated with a PFAM for it to be linked to a biological process. GO-term enrichment requires the number of genes observed in a given list to exceed the number of genes expected for a given biological process. For example 38 PFAMs are associated with WNT signalling. 14 in IPF and 16 in SSc contained genes with altered expression compared to control fibroblasts. In GO-term enrichment, there are 110 genes associated with Wnt signalling, however, only 2 in IPF and 3 in SSc had significantly altered expression. Combining these two types of enrichment analysis therefore allows one to identify genes which may or may not be involved but share similar domains associated with a specific biological process, such as Wnt signalling, whilst also allowing one to identify genes which have already been associated with Wnt signalling.

Although multiple biological processes overlapped between IPF and SSc, 81% of these processes were enriched in PFAMs containing genes with altered methylation in IPF compared to SSc lung fibroblasts. This data further suggests that multiple biological processes overlap between IPF and SSc but contain different genes which have altered methylation in each disease. This may underlie the different phenotypes of each disease and could potentially provide insights as to why IPF has a much poorer prognosis than SSc. For example, Wnt signalling which has been implicated in both IPF (Königshoff et al, 2008) and SSc (Wei et al, 2012) was enriched in genes with altered methylation in both IPF and SSc compared to control and in IPF compared to SSc. In IPF, the majority of genes within the WNT pathway had decreased methylation compared to control lung fibroblasts, whereas the majority of genes had increased methylation in SSc. Furthermore, multiple genes had significantly altered methylation between IPF and SSc, the majority of which had decreased methylation in IPF. These data suggest genes associated with Wnt signalling have aberrant methylation in both IPF and SSc but multiple different genes are affected. Genecodis was used for GO-term enrichment instead of DAVID analysis, as the DAVID tool has not been updated since early 2010, however, a major limitation of using GO-terms is that many do not overlap with well-defined pathway databases such as KEGG (Mao et al, 2005). Thus, when examining how genes within the Wnt pathway, which had altered methylation/expression could potentially interact with each other, both KEGG and GO-term enrichment was used. This kind of data analysis could potentially be used in the future to identify biomarkers and facilitate drug production targeting specific pathways associated with pulmonary fibrotic diseases. Furthermore, pathway analysis could be used to identify transcipriton factors or other ‘master’ regulators of gene expression and to identify targets that would benefit groups of patients or specific individuals.

7.7. Male compared to female methylation

Despite fundamental differences in male and female biology and with respect to gender-specific susceptibility to disease and responses to drugs, it is surprising in this era that many studies do not adequately, if at all, analyse data based on gender (Flanagan, 2014). Prior to 2014, there were very few studies which had compared methylation patterns between males and females (Sarter et al, 2005). Since 2014, although there are still very few studies which have compared male and female methylation patterns in a tissue-specific or disease context, there is emerging evidence that gender-specific methylation differences do exist which may play important roles in sex-biased diseases. For example, Spiers et al, recently identified gender-specific differences in methylation on multiple autosomes during fetal brain development (Spiers et al, 2015). Hall et al, identified chromosome-wide and gene-specific differences in methylation in human islets contributed to sex-specific metabolic phenotypes (Hall et al, 2014) and Pinto et al, identified gender differences in methylation in familial breast cancer (Pinto et al, 2013). Furthermore, it has recently been identified that leukocytes have an altered methylation profile in males compared to females (Inoshita et al, 2015) and gender can influence saliva methylation on the X chromosome and autosomes, with many sites of methylation also associated with diseases such as cancer (Liu et al, 2010). Furthermore, a sex-biased pattern of methylation in the human pre-frontal cortex has also recently been shown (Xu et al, 2014).

These studies all suggest that gender differences in methylation exist in a variety of tissues and could potentially be important in disease, particular in those which have a sex-bias such as IPF and SSc. The study by Spiers et al, is of particular interest as multiple neurological diseases such as Autism affect males more than females (Croen et al, 2002) which could potentially be in part explained by gender differences in brain methylation as identified in their study. It is therefore logical and important to examine gender differences in methylation for all cell types and tissues and in all sex-bias diseases such as IPF and SSc. To my knowledge, there have been no studies which have examined gender differences in primary human lung fibroblasts or methylation differences between male and female IPF lung fibroblasts.

One of the main limitations of the only previous study examining methylation in IPF lung fibroblasts was that they used mainly male control and female IPF lung fibroblasts (Huang et al, 2014). Thus it is hard to delineate whether CpGs had significantly altered methylation in IPF compared to controls or whether these CpGs had significantly altered methylation between male and females. For the same reason in my studies, it was not possible to analyse male and female differences in SSc lung fibroblasts as 6/7 cell lines were female. However, this work represents the first study to examine methylation differences between male and female human lung fibroblasts (using IPF and control lung fibroblasts) and the first study to compare methylation differences in male controls with male IPF and female controls with female IPF lung fibroblasts separately.

DNA methylation is known to play an important role in X-inactivation, which occurs to silence genes on one of the X-chromosomes in each female cell (Carrel and Willard, 2005). Thus, one would expect differences in methylation between male and females on X-linked genes. Indeed, 50% of genes with significantly altered methylation between male and female lung fibroblasts were located on the X-chromosome. The distribution of methylation was bimodal in male lung fibroblasts, whereas female lung fibroblasts displayed a partially methylated pattern on the X-chromosome which is consistent with X-chromosome inactivation. However, 49% of all genes with significantly altered methylation between males and female lung fibroblasts were located on autosomes. Interestingly, the distribution of CpGs with significantly altered methylation in male compared to female lung fibroblasts was either unimodal or bimodal in male lung fibroblasts (depending of genomic location), whereas in female lung fibroblasts the distribution of CpG methylation was always that of a partially methylated pattern. CGIs had a higher frequency of CpGs with decreased methylation in male compared to female lung fibroblasts, whereas open sea regions had a higher frequency of CpGs with increased methylation in male compared to female lung fibroblasts. Shore regions within 1.5kb of their corresponding genes TSS had a higher frequency of CpGs with low methylation in male compared to female lung fibroblasts whereas CpGs in shelf regions had a higher frequency of CpGs with increased methylation in male compared to female lung fibroblasts. This was consistent with the locations of male/female differences in methylation in the study by Hall et al, who showed CGIs and shore regions had decreased methylation in male compared to female islets, whereas shelf and open had increased methylation. This suggests multiple genes in male lung fibroblasts have a distinct methylation pattern compared to female lung fibroblasts.

Functional analyses identified multiple biological processes with relevance to fibrosis, including, Wnt signalling, EMT, ECM organisation and response to viruses, which were enriched in genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male compared to female lung fibroblasts. Many of these biological processes were also enriched in genes with altered methylation in IPF and/or SSc compared to control lung fibroblasts, suggesting that genes belonging to these processes have disease-specific and gender-specific differences in methylation. This could provide a novel insight as to why IPF predominates in males and potentially be used to target drugs specifically designed for use in male or female patients, based on their respective gene methylation/expression profiles and pathways.

On the discovery that multiple genes had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male compared to female lung fibroblasts, IPF males were compared to control male lung fibroblasts and IPF females were compared to control female lung fibroblasts separately. Although the number of cell lines used were small, the analysis did reveal a number of interesting findings. Firstly, the distribution of methylation on the Y-chromosome was different in male IPF compared to male control lung fibroblasts. Furthermore, multiple Y-linked genes had large fold-changes in expression in male IPF compared to male control lung fibroblasts. Although IPF occurs in females and is increasing, aberrant

methylation and/or expression of Y-linked genes could play a role in why IPF predominates in males as female do not possess the Y-chromosome. Multiple CpGs/genes had significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male IPF compared to male control and female IPF compared to female control lung fibroblasts, but very few CpGs/genes overlapped. Whilst this is preliminary data, it suggests that multiple CpGs/genes have altered methylation in male IPF lung fibroblasts compared to male control lung fibroblasts which are distinct from those which have altered methylation in IPF female lung fibroblasts compared to female control lung fibroblasts. However, multiple biological processes enriched in PFAMs containing genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in male IPF compared to male control overlapped with enriched biological processes in female IPF compared to female control lung fibroblasts. This may explain why there is a sex-bias but similar outcomes in IPF. For example methylation and/or expression of specific genes may induce IPF more rapidly and increase the risk of developing IPF. Further studies into gender-specific differences in methylation in IPF are clearly needed to validate this data.

Analysing methylation patterns in IPF compared to control lung fibroblasts based on gender could provide further insight into the role of methylation in PF and may lead to a more personalised medicine approach in the treatment of IPF (and other sex-bias diseases), such as a gender-based drug program as opposed to a 'one drug for all'. Interestingly, several biological processes enriched in PFAMs containing genes with significantly altered methylation in male IPF compared to male control included those relating to viruses and Actinobacteria. The lung microbiome is an increasingly active area of research in IPF (Han et al, 2014) and recently Actinobacteria levels have been shown to be decreased in IPF (Molyneaux et al, 2013). Viruses have previously been linked with IPF (Egan et al, 1997) and it is well known that males and females differ in response to being infected by certain viruses (Klein, 2012). Viruses have also been shown to play a role in acute exacerbations in many lung diseases including asthma (Busse et al, 2010) and chronic obstructive pulmonary disease (Wedzicha, 2004). Furthermore, in liver fibrosis, men have a worse virological profile than women and significantly more advanced fibrosis (Collazos et al, 2011). Thus, treatment of IPF and/or SSc-PF with vaccines could be designed with male/female differences in mind. Circadian rhythm was also enriched in PFAMs containing genes with altered methylation in male IPF compared to male control lung fibroblasts but not female IPF compared to female control lung fibroblasts. Increasing evidence suggests circadian rhythms are important in disease (Maury et al, 2010, Takeda and Maemura, 2011) and gender-differences have been reported (Wever et al, 1984, Bertossa et al, 2013, Krizo et al, 2014). Furthermore, recent evidence suggests they can regulate anti-oxidant pathways to modulate pulmonary fibrosis (Pekovic-Vaughan et al, 2014).

Biological processes enriched in PFAMs containing genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) in female IPF compared to female control but not in male IPF compared to male control lung fibroblasts included response to hexose, response to corticosteroids and ER stress. Corticosteroids have been trialled in IPF studies, however, there is no evidence showing they have a

beneficial effect (Richeldi et al, 2003, Atkins et al, 2014). Nonetheless, corticosteroids are currently used in the treatment of some IPF patients with acute exacerbations (Atkins et al, 2014). The finding that the biological process 'response to corticosteroids' is enriched in females but not males could potentially explain why corticosteroids have so far failed to treat IPF but may work for a specific minority. Response to hexose was also identified as being enriched in female IPF compared to female control but not male IPF compared to male controls. Hexose is a monosaccharide and increased levels correlate with disease severity in cystic fibrosis (Chace et al, 1983). Cyclophosphamide-induced lung fibrosis also causes hexose levels to increase (Vemkatesan et al, 1998), however, there are no studies examining the role of hexose in IPF. ER stress has previously been linked to IPF (Tanjore et al, 2012) and ER stress can modulate gene expression (Baumeister et al, 2005, Chen et al, 2014). Furthermore, methylation of ER-associated genes can regulate their expression (Han et al, 2013) therefore, genes with altered methylation could potentially explain how ER stress is initiated or how ER stress activates certain genes and contributes to PF potentially specifically in females.

7.8. Moduation of DNA methylation using 5-Aza

5-Aza is a chemical analogue of cytosine which can be incorporated into DNA, bind to DNMTs (inhibiting their activity) and subsequently results in genome-wide demethylation. 5-Aza is already used to treat myelodysplastic syndromes (Sauntharajah, 2013) and previous studies have shown benefits of 5-Aza treatment in murine models of fibrosis in both IPF (Dakhlallah et al, 2013) and SSc (Dees et al, 2009). Furthermore, treatment with 5-Aza can activate the expression of specific genes which are hypermethylated in either IPF (Cisneros et al, 2010, Huang et al, 2014) or SSc (Wang et al, 2006) suggesting it may be useful in the treatment of PF.

5-Aza was used in this study to inhibit methylation and subsequently the effects on gene expression were measured. Originally, analysis using the average change in methylation and expression after 5-Aza treatment compared with basal levels in control, IPF and SSc lung fibroblasts identified multiple changes in CpG methylation but few were ≥13.6%. Furthermore, very few genes had large changes (≥2FC) in expression after treatment with 5-Aza, suggesting that alterations to DNA methylation had limited consequences on transcription. Interestingly the loss of methylation in promoter regions, islands and shores in other diseases which use 5-Aza as a treatment solution, such as acute myeloid leukaemia, show no correlation with changes to gene expression (Lund et al, 2014). However, cluster analysis revealed the effects of 5-Aza treatment were different in each cell line (**see Chapter 6: section 6.2, figure 6.2.2**). Subsequent analyses showed some cell lines had a large response to 5-Aza, where large changes across genome-wide methylation were observed. Other cell lines including all of the SSc cell lines had a much weaker response. Reasons behind this observation are unclear, however it has been shown that some cell lines are resistant to 5-Aza treatment (Flatau et al, 1984, Qin et al, 2009). This might be caused by insufficient incorporation into the DNA (Qin et al, 2009) or through an altered DNA damage response pathway (Palli et al, 2008), although the different population doublings

between control, IPF and SSc cell lines after treatment with 5-Aza did not correlate with the number of CpGs/genes with altered methylation/expression. Deficiency and mutations in deoxycytidine kinase (DCK), a gene responsible for converting 5-Aza into its activate form by phosphorylation, have also been associated with increased resistance to the effects of 5-Aza (Stegmann et al, 1995) although no study has reported mutations or deficiency of this gene in SSc patients.

The number of CpGs/genes with large changes (≥13.6%) in methylation after 5-Aza treatment correlated with an increased number of genes with large changes in expression (≥2FC). However, in SSc lung fibroblast cell lines, many genes had large changes in expression without having large changes in methylation. This suggested that smaller changes in methylation may potentially have a large effect on gene expression. Indeed, when analysing changes to methylation with respect to gene expression across all 18 cell lines, small changes in methylation appeared to correlate with large changes in expression in multiple genes. However, the Illumina 450K array is unable to accurately measure small changes in methylation (Dedeurwaerder et al, 2011), therefore more accurate methods of detecting methylation, such as bisulfite sequencing would be required to confirm these results.

Multiple genes had correlation between methylation and their expression before and after treatment with 5-Aza. This suggests many genes are potentially directly regulated by DNA methylation in lung fibroblasts. This is consistent with a recent study in IPF lung tissue showing multiple differentially methylated regions correlate with changes in gene expression although changes in methylation correlated with expression in an inverse manner (Yang et al, 2014). Furthermore, the Illumina 450k array only covers 99% of all genes, representing ~2% of the entire human methylome. Thus, one can assume that multiple CpGs not covered by the array will have also had changes in methylation after 5-Aza treatment and may be specifically important in regulating their corresponding gene's expression. To fully understand and elucidate the role methylation has in regulating gene expression in lung fibroblasts, the entire methylome would need to be analysed and integrated with other epigenetic mechanisms including miRs and histone modifications.

Whilst multiple genes with altered methylation after 5-Aza treatment correlated with changes to their expression, multiple genes had increased or decreased expression after 5-Aza-2'-deoxycytidine treatment without having alterations to their methylation. This could be explained also by CpGs not covered on the array having an effect on gene expression. Another plausible explanation for this observation is that genes activated or inhibited by 5-Aza treatment could potentially regulate the expression of other genes. Recently published data supports this hypothesis in IPF. Methylation of miR-17-92 silences its expression which leads to a number of genes that are associated with fibroproliferative responses being upregulated (Dakhlallah et al, 2013). This suggests that other genes in IPF lung fibroblasts that are activated or inhibited by demethylation could increase or decrease the expression of multiple other genes. Genes may also be activated in response to the damage 5-Aza can cause to DNA. It has previously been shown that ataxia telangiectasia pathways (ATM and ATR) are

activated upon DNA damage via 5-Aza (Palii et al, 2008) and thus could be responsible for the increased expression of genes related to DNA repair, adding further complexity in finding genes whose expression levels are directly or indirectly regulated by methylation.

Although 5-Aza is a demethylation agent, surprisingly, multiple CpGs had increased methylation after treatment with 5-Aza, a finding which is rarely, if at all described in any literature on 5-Aza studies. The increase in methylation after 5-Aza could potentially be explained by 5-Aza directly or indirectly affecting the expression of other genes or signalling pathway which may be able to specifically remethylate DNA. However, no obvious genes associated with remethylation (such as DNMTs) were identified as having significantly altered mRNA expression after 5-Aza treatment. This is consistent with some previous reports showing 5-Aza does not affect the mRNA expression level of some DNMTs (Ghoshal et al, 2005, Scheider-Stock et al, 2005). It has been shown that DNMT1 (Ghoshal et al, 2005) and DNMT3b (Scheider-Stock et al, 2005) mRNA expression is not affected by 5-Aza and only DNMT1 is rapidly degraded in mammalian cells by 5-Aza, thus other DNMTs may still be able to function (Ghoshal et al, 2005). Furthermore, 5-Aza can induce histone acetylation (Takebayashi et al, 2001, Yang et al, 2010) and modify histone methylation by increasing lys-4 and increasing or decreasing lys-9 depending on the concentration of 5-Aza used (Kondo et al, 2003, Coombes et al, 2003). Increasing evidence shows DNA methylation and histone lysine methylation are highly inter-related (Rose and Klose, 2014) and can influence each other's state (Cedar and Bergman, 2009, Rose and Klose, 2014). Genes marked with specific histone modifications on histone H3 dimethylation on lysine 4 (H3K4me2) and histone H3 trimethylated on lysine 27 (H3K27me3) can gain histone marks which target CGIs for DNA methylation (Ohm et al, 2007). This could potentially explain why multiple genes had increased methylation after 5-Aza or again, why multiple genes had altered expression without any observations in changes to their DNA methylation after 5-Aza treatment.

Another potentially interesting explanation as to why 5-Aza increased CpG methylation in lung fibroblasts at multiple CpGs, relates to the effects of 5-Aza on the activation of EBV. 5-Aza can activate the EBV genome (Ben-Sasson and Klein, 1981, Masucci et al, 1989), a virus which has previously been associated with IPF (Lok et al, 2001, Kelly et al, 2002). Recently in immortalised oral keratinocytes, EBV infection resulted in increased DNA methylation at 10676 CpGs (Birdwell et al, 2014). The patient data for the control IPF and SSc lung fibroblasts did not include EBV status, however, EBV can infect lung fibroblasts (Adachi et al, 2001) and is one of the most common human viruses in the world. Thus it is likely these cells line were infected with EBV and could explain why multiple CpGs had increased methylation after 5-Aza treatment. These observations regarding EBV activation by 5-Aza and its ability to increase methylation could have important clinical implications in IPF and SSc, such as 5-Aza treatment having a different effect on methylation in a non-EBV infected patient compared to an EBV infected patient. Future studies could examine methylation in EBV infected compared to non-EBV infected IPF or SSc individuals, which may provide a link between epigenetic and environmental factors associated with these diseases. Furthermore, this could be extremely interesting as males and

females have an altered response and predisposition to viruses and bacteria (Klein, 2012, Giefing-Kröll et al, 2015). The mechanisms of how 5-Aza works are also not fully understood, therefore future studies may shed more information on how and to what extent methylation changes with 5-Aza treatment.

7.8.1. Limitations of 5-Aza

The exact in vivo mechanisms of 5-Aza drugs remain poorly understood, although 5-Aza-2'-deoxycytidine is known to inhibit cell proliferation (Karpf et al, 2001). Enzymatic de-amination of 5-Azacytidine and 5-Aza-2'-deoxycytidine causes increased cytotoxicity (Vesely et al, 1969) although 5-Aza-2'-deoxycytidine is 10 times more cytotoxic to cultured cells than 5-Azacytidine (Flatau et al, 1984). The cytotoxicity of 5-Aza-2'-deoxycytidine and its unknown in vivo mechanisms represent a major concern when trying to analyse the effects of reducing methylation on gene expression. Low concentrations of 5-Aza (1μm) were used to limit toxicity. Other less toxic DNMT inhibitors could be used in future studies such as 1-(β-D-ribofuranosyl)-1,2-dihydropyrimidin-2-one (Zebularine) which has been shown to be able to reactivate epigenetically silenced genes (Cheng et al, 2003). Zebularine inhibits DNMTs and cytidine deaminase which is essential for deamination of cytidine nucleosides (Marquez et al, 1980). Another de-methylating drug called Hydralazine, which is a non-nucleoside analogue, can inhibit DNMTs and subsequently reduce methylation without altering DNMT expression (Cruz-Hernandez et al, 2011). It has been shown that hydralazine can reduce kidney fibrosis by reducing methylation of the RAS protein activator like 1 (RASL1) gene via activation of Tet methylcytosine dioxygenase 3 (TET3) (Tampe et al, 2015). Hydralazine has also been shown to reduce cardiac fibrosis (Qi et al, 2011) and fibrosis in stable ILD patients (Lupi-Herrera et al, 1985). However, it remains unknown whether hydralazine could be beneficial in IPF/SSc-PF, and there are reports of hydralazine causing autoimmunity and frequently causing deleterious effects in patients with pulmonary hypertension (Packer et al, 1982).

As previously discussed, multiple genes had increased expression after 5-Aza treatment. This makes 5-Aza's application in in vivo models somewhat limited as 5-Aza could inadvertently activate multiple genes which could then activate other genes and cause disease. Nonetheless, the use of 5-Aza does provide an insight into which genes can be activated by reducing methylation, which could play a role in the pathogenesis or IPF and other fibrotic diseases.

7.9. The role of TNXB in pulmonary fibrosis

Previous studies have identified deficiency of TNXB as a pathological cause of EDS which is characterised by hyper-mobility and hyper-extensible joints (Zweers et al, 2003). TNXB knockout mice also show reduced collagen density resulting in a 30% reduction in collagen content in skin (Mao et al, 2002). Tenascin-X has also been shown to increase collagen fibril formation in vitro (Egging et al, 2007) and bind to fibril-associated type XII and XIV collagens (Lethias et al, 2006), both of which have been

shown to be upregulated after bleomycin-induced pulmonary fibrosis (Tzortzaki et al, 2003). Furthermore, increased TNXB is associated with fibrous tumours including mesothelioma (Yuan et al, 2009) whereas TNXB knockout rats are protected from cardiac fibrosis, mediated by an increase in PPARγ and a decrease in TGFß (Jing et al, 2011). Whilst tenascin C has been studied in IPF (Estany et al, 2014) and SSc (Brissett et al, 2012, Inoue et al, 2013), no studies prior to this thesis, have examined the role of TNX In pulmonary fibrosis. Furthermore, how TNXB expression is regulated is poorly understood, although, hypomethylation of a CpG island located in exon 3 in muscle corresponds with increased TNXB expression (Rakyan et al, 2004), suggesting a potential regulatory role of methylation on TNXB expression.

The Illumina 450k array identified multiple CpGs with significantly altered methylation in IPF and SSc compared to control and IPF compared to SSc lung fibroblasts. Therefore, based on the array data combined with the previous TNXB literature, the role of TNXB in pulmonary fibrosis was examined. The microarray identified 27 CpG sites in IPF (10 increased and 17 decreased) and 41 CpG sites in SSc (38 increased and 3 decreased) as having a significant change in methylation compared to control fibroblasts. The majority of CpGs with increased methylation in IPF lung fibroblasts were located in north shelf and open sea regions, whereas the majority of CpGs with increased methylation in SSc lung fibroblasts were located in island and open sea regions. The majority of CpGs with decreased methylation in IPF lung fibroblasts were located in open sea and south shelf regions, whereas in SSc lung fibroblasts only 3 CpGs had decreased methylation (1 in the north shore, 1 in the south shelf and 1 in open sea). As previously mentioned, CGIs with increased methylation have been strongly linked with decreased gene expression in many diseases including IPF (Sanders et al, 2008, Sanders et al, 2012, Rabinovich et al, 2012, Cisneros et al, 2010) and SSc (Wang et al, 2006), however, the role of shore/shelf and open sea CpG methylation remains poorly understood in disease (Irizzary et al, 2008).

The microarray covered 3 CpG islands in the TNXB gene (one in exon 3, one in intron 6 and one in exon 10). Although data from the microarray in this study suggested the majority of CpGs in IPF lung fibroblasts had reduced methylation in open sea regions, previously published data has shown hypomethylation of a CGI in exon 3 correlated with high TNXB expression in muscle tissue compared to other tissues (Rakyan et al, 2004). Our microarray data also identified 1 CpG with significantly decreased methylation (13.6%) and multiple other CpG sites in IPF lung fibroblasts, with a small but significant decrease in methylation ($P<0.05$ <0.136) in this CpG island. Furthmore, open sea regions are sparsely populated with CpGs, whereas CGIs have a high frequency of CpGs close together. Therefore, in order to maximise the number of CpGs sequenced to validate the microarray data, a 349bp region located in the exon 3 CGI was bisulfite sequenced. This region covered 7 of the CpGs identified on the microarray, 6 of which had significantly decreased methylation in IPF compared to control lung fibroblasts.

Changes in methylation between control, IPF and SSc lung fibroblasts for all 7 CpGs using bisulfite sequencing were comparable with the microarray data (Pearson r=0.77). The total percentage of methylation however did vary between microarray and bisulfite sequencing. This is most likely to be due to differences in normalising data to different background controls, although the efficiency of the bisulfite conversion could also have an impact on results. Poor bisulfite conversion could reduce the accuracy of the results by underestimating total methylation. However, both methods showed consistent differences in methylation of 7 CpGs between control, IPF and SSc lung fibroblasts. Furthermore, bisulfite sequencing of a 440bp region in the CpG island located in exon 10 of the TNXB gene identified 7 CpGs with significantly decreased methylation in IPF compared to control, 5 of which were also significantly decreased in IPF compared with SSc lung fibroblasts.

For unknown reasons there was no data regarding TNXB expression on the expression microarray, perhaps explained by the probe not working or not being detected. Therefore, the expression of TNXB in lung fibroblasts was examined by qRT-PCR and the expression of TNX, in lung tissue, by IHC to determine whether altered TNXB methylation correlated with expression in control, IPF and SSc lung fibroblasts. qRT-PCR analysis demonstrated TNXB was increased in IPF compared to control and SSc lung fibroblasts, whereas there was no significant difference in TNXB expression between SSc and control lung fibroblasts. The increase in TNX expression in IPF, but not SSc, was confirmed by IHC staining of lung tissue sections.

These data support a role for methylation in regulating TNXB expression. The observation that the majority of CpGs had decreased methylation in IPF and increased methylation in SSc compared to control lung fibroblasts, suggests the location of the methylated CpG sites may be of fundamental importance. However, whether small changes in methylation at one or multiple CpG sites within specific regions or just one CpG site with a change in methylation can have an effect on gene expression is currently unknown. In order to elucidate what change in methylation could have an effect on TNXB expression, TNXB methylation and expression were analysed in control, IPF and SSc lung fibroblasts after treatment with 5-Aza.

7.9.1. 5-Aza treatment and its effects on TNXB expression in control, IPF and SSc lung fibroblasts

Microarray analysis identified 68, 34 and 36 CpG sites in control, IPF and SSc lung fibroblasts, respectively, which had significantly decreased ($P<0.05$) methylation after 5-Aza treatment (**see section 6.6**). qRT-PCR analysis of TNXB expression identified control lung fibroblast cell lines had significantly increased gene expression after 5-Aza treatment in contrast to IPF and SSc lung fibroblasts which had no change in expression. Three out of six control lung fibroblast cell lines had large increases in expression (>5FC) which correlated with large changes in TNXB methylation (≥13.6) after 5-Aza treatment. These control cell lines were the 3 cell lines identified as ‘strong responders’ which had a

greater number of CpGs with altered methylation after 5-Aza treatment compared to the other cell lines. Furthermore, treatment with 5-Aza brought the average methylation of TNXB in control lung fibroblasts down to similar levels of basal TNXB methylation in IPF lung fibroblasts. Interestingly, three IPF cell lines had large changes (≥13.6%) in TNXB methylation after 5-Aza treatment but did not have any change in TNXB expression, whereas no SSc lung fibroblast cell line had large changes (≥13.6%) in TNXB methylation or changes in TNXB expression after 5-Aza treatment. These data potentially suggests that large changes (≥13.6) in TNXB methylation are required to have an effect on TNXB expression and that once the threshold has been met, further demethylation has little or no effect on TNXB expression. However, as previously discussed, multiple genes had large changes (≥2FC) in expression in cell lines which had relatively few CpGs with large changes (≥13.6) in methylation. Therefore, activation of any negative regulators or other regulatory mechanisms of TNXB could mask the effects of small changes in TNXB methylation on TNXB expression. Thus, whilst data in this study suggests large changes in methylation are required to have an effect on TNXB expression, one cannot rule out entirely that small changes in TNXB methylation may also have an effect on TNXB expression.

7.9.2. siRNA knockdown of TNXB in IPF lung fibroblasts and its effects on collagen gel contration

The mechano-properties of the ECM and its ability to modulate cell responses has previously been implicated in the pathogenesis of IPF and other fibrotic diseases (Liu et al, 2010, Marinkovic et al, 2013, Zhou et al, 2013). Fibroblasts in contact with stiff matrices can respond differently to fibroblasts in contact with softer matrices in a number of ways, including proliferation rate, survival rate, collagen synthesis and gene expression (Marinkovic et al, 2013, Zhou et al, 2013). Expression of TGFß, a profibrotic cytokine which has strongly been associated with IPF (Johnston et al, 1990, Khalil et al, 1991, Laurent et al, 2008) can be modulated by matrix stiffness (Wipff et al, 2007, Tatler and Gisli, 2012). Subsequently, TGFß can regulate the expression of other genes such as PPARγ (Wei et al, 2010) and COX2 (Keerthisingam et al, 2001) which may have anti-fibrotic effects in lung fibrosis.

Pathway analysis using Cytoscape 3.2.1, Genecodis 3 and KEGG pathways, identified biological processes and/or pathways including cytoskeletal organisation, focal adhesion and ECM interactions as being in enriched in genes including TNXB. Alterations to these processes can affect cell contractility (Parsons, 2010). Therefore, it was hypothesised that altered methylation and subsequent expression of TNXB, may play an important role in tissue contractility. Collagen gel assays were used to measure fibroblast-mediated collagen gel contraction in control and IPF lung fibroblasts and siRNA targeting TNXB was used to assess the effects of TNXB knockdown on collagen gel contraction in IPF lung fibroblasts. Measuring the effects of knocking down a gene using siRNA is often difficult due to siRNA off-targeting effects, where siRNA binds to unintended mRNA targets (Jackson and Linsley, 2010) and toxic phenotypes being induced, via high levels of siRNA (Fedorov et al, 2006). These adverse effects of siRNA were reduced by using INTERFERin (Polyplus, USA), a transfection reagent which allows a

much lower concentration of siRNA to be used, thus reducing off-target effects and toxicity. Results identified that IPF lung fibroblasts contracted collagen gels to a greater extent than control lung fibroblasts. This could be explained by increased/decreased expression of genes involved in regulating cell contraction (as identified by enrichment analyses) and/or differences in ECM remodelling. Knockdown of TNXB resulted in decreased collagen gel contraction, suggesting TNXB is important in regulating the ability of fibroblasts to contract collagen. However, knockdown of TNXB did not have a significant effect on TGFß or PPARγ which is in contrast to a previous study, which concluded TNXB could initiate myocardial fibrosis via upregulation of TGFß and downregulation of PPARγ (Jing et al, 2011).

7.10. Summary and conclusions

To my knowledge, this thesis represents the largest and most detailed report of DNA methylation and its role in fibrotic lung fibroblasts and extends beyond the only other study examining methylation in IPF lung fibroblasts (Huang et al, 2014) by interrogating more CpG sites, different locations of CpG methylation and comparing methylation with expression using the same cell cultures. Furthermore, the data presented in this thesis are the first to examine gender-differences in methylation in lung fibroblasts, compare male IPF with female IPF methylation and compare global methylation in two diseases characterised by pulmonary fibrosis; IPF and SSc.

In agreement with previous studies examining DNA methylation in IPF lung tissue (Rabinovich et al, 2012, Sanders et al, 2012, Yang et al, 2014) and IPF lung fibroblasts (Huang et al, 2014), data presented in this thesis provide further evidence that altered DNA methylation in lung fibroblasts plays an important role in pulmonary fibrosis. CpGs/genes were predominately hypomethylated in IPF but conversely, hypermethylated in SSc, compared to control lung fibroblasts. Many of these genes overlapped between IPF and SSc, although, fewer CpGs overlapped. This suggests, for genes which overlap, different CpGs have altered methylation in IPF/SSc. Furthermore, multiple genes were distinct to each disease and multiple CpGs/genes had altered methylation and/or expression in IPF compared to SSc lung fibroblasts. This suggests IPF and SSc have distinct methylation and expression profiles which could, in part, explain differences in disease prevalence, progression and manifestations.

Consistent with recent IPF studies (Huang et al, 2014, Yang et al, 2014), the majority of CpGs with altered methylation in IPF compared to control lung fibroblasts were located outside of CGIs. (Huang et al, 2014, Yang et al, 2014). This was also true for SSc lung fibroblasts. In other cell types and diseases, emerging data suggests that non-CGI methylation plays an important regulatory role (Jones, 2012), therefore interrogating CpGs outside of CGIs is important in future methylation studies. The distribution of CpG methylation in control, IPF and SSc lung fibroblasts was bimodal on autosomes. On the X-chromosome the bimodal pattern of methylation was clear but there were more CpGs with a partially methylated. Further analysis identified the X-chromosome had a bimodal pattern of

methylation in male lung fibroblasts, whereas a partially methylated pattern was observed in female lung fibroblasts, most likely caused by the effects of methylation on X-chromosome inactivation (Sharp et al, 2011). Further distribution analysis identified that different genomic locations had specific patterns of methylation suggesting the location of CpG methylation may be important in determining whether a CpG plays an important role in regulating gene expression.

Multiple genes were also identified as having significantly altered expression (TNoM ≤1; P<0.05) in IPF/SSc compared to control lung fibroblasts and correlation analysis identified strong associations between methylation and expression in multiple genes, suggesting methylation plays an important role in both IPF and SSc. Many of these genes have previously been linked to PF whereas others identified, including TNXB, were novel to pulmonary fibrosis, suggesting multiple genes in lung fibroblasts can be regulated by methylation and may be relevant to pulmonary fibrosis.

The data in this thesis represents the first evidence suggesting that TNX plays an important role in IPF. Multiple CpGs within the TNXB gene had significantly decreased methylation in IPF compared to SSc and control lung fibroblasts, which correlated with increased TNXB expression in IPF lung fibroblasts and lung tissue. Furthermore, treatment of control cell lines with 5-Aza resulted in large decreases (≥13.6%) in TNXB methylation which correlated with large increases (≥2FC) in TNXB expression, suggesting that methylation plays an important role in TNXB regulation. Pathway and enrichment analysis identified TNXB as playing an important role in cytoskeletal organisation, focal adhesions and ECM interactions, all of which are important in regulating cell contractility (Tomasek et al, 1992). Subsequent analysis identified IPF fibroblasts could contract collagen gels to a greater extent than control lung fibroblasts and that knockdown of TNXB in IPF fibroblasts significantly reduced collagen gel contraction. Therefore, increased TNX expression via hypomethylation of the TNXB gene, may contribute to the pathogenesis of IPF through altered regulation of ECM deposition and increased tissue contractility.

Multiple CpGs/genes also had significant differences in methylation and expression in male compared to female lung fibroblasts, irrespective of where they were derived from. The majority of CpGs/genes on autosomes (49%), the X-chromosome (50%) and the Y-chromosome (1%) had decreased methylation in male compared to female lung fibroblasts. Multiple CpGs/genes had significantly altered methylation in male IPF compared to male control and female IPF compared to female control lung fibroblasts, however, very few CpGs/genes overlapped, suggesting IPF males have a difference pattern of methylation compared to IPF females. Interestingly, enrichment analysis identified similar biological processes enriched in genes with altered methylation in both sexes, suggesting different genes belonging to the same biological processes are affected. Specific biological processes, including those relating to viruses, were only found in male lung fibroblasts. Viral infections have previously been linked to IPF (Egan et al, 1995), although it is unclear whether viruses play a major role in IPF (Wootton et al, 2011). This data could potentially explain, in part, why IPF predominates in males.

In summary, data presented in this thesis suggests multiple genes are potentially directly regulated by methylation in lung fibroblasts. Multiple CpGs/genes have significantly altered methylation and/or significantly altered expression in IPF/SSc compared to control lung fibroblasts many of which are associated with biological processes/pathways relevant to fibrosis. Multiple genes also overlap between IPF and SSc, although the majority are distinct to each disease suggesting different methylation profiles in each disease. Functional analyses on one gene, TNXB, suggested increased TNX expression via hypomethylation contributes to the pathogenesis of IPF by regulating tissue contractility and could distinguish IPF from other diseases characterised by pulmonary fibrosis, such as SSc. Treatment of fibroblasts with 5-Aza altered methylation at multiple CpGs and the expression of multiple genes confirming the importance of methylation in regulating gene expression. For multiple genes, large changes in methylation (≥13.6%) correlated with large changes in gene expression (≥2FC). However, in SSc lung fibroblasts, small changes in methylation (<13.6) were associated with large changes in gene expression (≥2FC), potentially suggesting that some genes might be sensitive to small changes in methylation. Furthermore, chromosome-wide and gene-specific sex differences in methylation and expression were identified in lung fibroblasts, suggesting a potential role for gender-differences in methylation in sex-biased diseases such as IPF and SSc.

7.11. Future work

7.11.1. Future methylation and expression analysis

The Illumina Infinium HumanMethylation 450 BeadChip array identified multiple CpGs/genes with altered methylation in IPF/SSc compared to control lung fibroblasts and represents a significant upgrade on the only other previous study examining methylation in IPF lung fibroblasts (Huang et al, 2014). However, the 450k array still only covers approximately 2% of the entire human methylome. Therefore, 98% of the human lung fibroblast methylome remains to be determined. CpGs not covered by the Illumina 450k array could potentially play important roles in IPF/SSc. Being able to analyse methylation of each CpG in a gene, particular those genes identified in this thesis with large changes in expression, but with no changes in methylation after 5-Aza treatment, would be of particular interest, as they may be directly regulated by methylation. To extend the current understanding on the role of methylation on gene expression in human lung fibroblasts and potentially reveal more novel genes potentially regulated by methylation, more CpGs would need to be analysed. This could be done by performing whole genome bisulfite sequencing or by using arrays which cover millions of CpGs such as the comprehensive high-throughput arrays for relative methylation (CHARM) arrays previously used to study DNA methylation in lung tissue (Yang et al, 2014).

7.11.2. Comparisons with other fibrotic diseases and analysis of other cell types

Comparing methylation of IPF with another fibrotic disease, SSc, gave a valuable insight into which CpGs/genes overlapped between both diseases and which were specific to each disease. Future

studies could extend this analysis by examining more fibrotic lung fibroblast cell lines and compare IPF with other fibrotic diseases, particular those which are idiopathic. This analysis could potentially reveal more overlapping pathways and biological processes affected by altered methylation and/or expression in idiopathic fibroses. Whilst fibroblasts are key effector cells in pulmonary fibrosis, altered methylation and expression of genes in other cell types, including myofibroblasts, alveolar epithelial cells, macrophages and neutrophils, could also be important. Three studies have examined methylation in lung tissue (Sanders et al, 2012, Rabinovich et al, 2012, Yang et al, 2014) which contains all these cell types. However, it is impossible to determine any cell-specific changes in methylation from this data. Future studies should therefore focus on examining methylation in other specific cell types, which could elucidate more genes whose expression is regulated by methylation. Furthermore, IPF is characterized by areas of fibrosis and normal lung, with fibrotic foci believed to represent areas of active fibrosis. The heterogeneity often associated with IPF could be in part explained by fibroblasts coming from both fibrotic and normal regions of IPF lung. Using a laser capture system, future studies could analyse fibroblasts derived from only fibrotic foci, thus reducing potentially heterogeneity by providing a 'pure' sample of fibrotic lung fibroblasts to analyse. Data could then be compared from multiple different cell types to generate a greater understanding of the role of DNA methylation in pulmonary fibrosis.

7.11.3. Male and female methylation/expression differences

To my knowledge, this is the first study to analyse gender-differences in methylation in primary human lung fibroblasts and compare gender-differences in IPF lung fibroblasts. This study identified multiple CpGs/genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and/or expression (TNoM ≤1; $P<0.05$) in male compared to female lung fibroblasts irrespective of where they were derived. Furthermore, analysis of male IPF compared to male control and female IPF compared to female control lung fibroblasts identified multiple CpGs distinct to each sex which had significantly altered methylation. This is extremely interesting as recently, other studies have identified gender-differences in methylation in other cell types (Liu et al, 2010, Hall et al, 2014, Inoshita et al, 2015) and diseases (Pinto et al, 2013). Furthermore, sex-specific differences in methylation during brain development can affect multiple genes, leading to the speculation that a number of neurological diseases with a sex-bias maybe linked to altered methylation (Spiers et al, 2015). Therefore the identification of gender differences in methylation in IPF lung fibroblasts, could, in part, explain why IPF predominates in males. However, the number of cell lines analysed in this study were small (male: n=4, female: n=7). To confirm gender-difference in methylation in IPF, future studies could examine fibroblasts from a larger cohort of patients and/or other cell types to determine the full extent and role of gender-difference in methylation in IPF. Due to the lack of male SSc lung fibroblast cell lines (n=1), gender-differences in methylation, in SSc, were not examined. However, SSc is also a sex-biased disease that predominates in females. Therefore, it would be of interest to study gender-differences in methylation

in SSc lung fibroblasts and/or other cell types to determine whether gender-differences in methylation exist in other diseases in which pulmonary fibrosis often occurs.

7.11.4. Pathway analysis and data visualisation

STRING 10.0 (Szklarczyk et al, 2015), Cytoscape (Shannon et al, 2003), DC:GO PFAM enrichment (Fang and Gough, 2013) and Genecodis 3 GO-term and KEGG enrichment analyses (Carmona-Saez et al, 2007, Nogales-Cadenas et al, 2009, Tabas-Madrid et al, 2012) were all used to analyse methylation and gene expression data. GO-term enrichment enabled the identification of biological processes significantly enriched in genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and/or expression (TNoM ≤1; $P<0.05$) in IPF/SSc compared to control lung fibroblasts. PFAM enrichment enabled the identification of biological processes enriched in PFAMs belonging to genes with significantly altered methylation ($\Delta\beta \geq 0.136$; $P<0.05$) and/or expression (TNoM ≤1; $P<0.05$) in IPF/SSc compared to control lung fibroblasts. Thus, with this combined approach, genes either previously linked to a biological process or genes which shared domains previously linked to a biological process, but not necessarily directly associated with that biological process, could be identified. This enabled the discovery of novel genes potentially regulated by methylation and involved in biological processes potentially relevant to fibrosis. Future studies could further analyse specific genes or groups of genes associated with specific biological processes/pathways of interest, to determine their role in IPF/SSc and whether methylation plays an important role in their regulation.

STRING 10.0 was used to identify potential protein-protein interactions between genes with altered methylation and/or expression and helped visualise how these genes could potentially interact with each other. The KEGG plugin for Cytoscape was used in combination with GO-term and KEGG pathway enrichment analyses to map genes with significantly altered methylation and/or expression onto known KEGG pathways, thus, allowing further visualisation of how genes in a pathway may interact. However, GO-terms do not match well with KEGG pathways (Mao et al, 2005) and KEGG pathways are not always updated with the latest findings in research. Furthermore, the KEGG plugin for Cytoscape required extensive modifications to work as intended. Future analysis should therefore aim to build on these pathways by manually adding data to specific points in a given pathway using Cytoscape, or with other pathway analysis options, such as IPA (Ingenuity, USA) or Pathway Studio (Ariadne Genomics, USA), both of which can be used to manually create pathways.

Data was visualised using the R (v3.2.0) statistical program and a number of scripts were developed to analyse data. These provided valuable insights into how CpG methylation was distributed within different genomic locations and in relation to CGIs, how different cell lines clustered, and which CpGs/genes had significant correlations between methylation and gene expression. Furthermore, analysis using R helped identify how 5-Aza affected different cells lines to varying extents. Future studies could use or adapt these scripts to any given dataset and extend their capability, such as displaying specific genes or groups of genes associated with different pathways/biological processes

directly on graphs. This could potentially reveal in more detail the extent of altered methylation and/or expression in fibrotic lung fibroblasts.

7.11.5. The role of TNXB in pulmonary fibrosis

In this thesis, I identified multiple CpG sites within the TNXB gene as having decreased methylation in IPF compared to SSc and control lung fibroblasts, which correlated with increased TNXB expression in IPF lung fibroblasts. Furthermore, treatment of control lung fibroblasts with 5-Aza caused large decreases (≥13.6%) in TNXB methylation which correlated with large increases (≥2FC) in TNXB expression, thus suggesting that methylation plays an important role in TNXB regulation. However, it remains unknown which CpGs are important for regulating TNXB. It would therefore be of interest to examine more CpGs within the TNXB to determine the specific region/s which are responsible for regulating TNXB expression. For example, there are potentially important Specificity 1 (SP1) and Specificity 3 (SP3) binding sites within the TNXB promoter region, which may be important in regulating TNXB expression (Wijesuriya et al, 2002). Interestingly, methylation of CpGs adjacent to SP1 binding sites can affect SP1 binding (Zhu et al, 2003). Although no changes in methylation between control, IPF and SSc lung fibroblasts were found covering SP1/SP3 binding sites in the TNXB promoter, altered methylation of CpGs in IPF compared to control lung fibroblasts which surrounded SP1 binding sites, were identified in the exon 10 CGI. Therefore, future studies could examine whether SP1/SP3 binding sites in this, or other regions, are affected by methylation in TNXB and whether changes to methylation at these regions can alter SP1/SP3 binding and subsequently, modulate TNXB expression.

Pathway and enrichment analyses identified biological processes/pathways associated with cell contraction as being enriched in genes including TNXB. Using a model of cell-mediated collagen gel contraction, I showed that IPF fibroblasts contracted collagen to a greater extent compared to control lung fibroblasts and that siRNA knockdown of TNXB reduced collagen gel contraction. However, the exact role of TNXB in pulmonary fibrosis and how it affects contraction remains unknown. A recent study in epithelial cells identified the fibrinogen-like (FBG) domain of TNXB can interact with latent TGFß (Alcaraz et al, 2014). Integrin α 11 (ITGA11) was essential for TNXB FBG-mediated activation of TGFß to occur (Alcaraz et al, 2014). It has previously been shown that multiple integrins can interact with latent TGFß causing the release of active TGFß, however their role in IPF is yet to be fully determined (Tatler and Jenkins, 2012). It would therefore be of interest to study whether ITGA11 and/or other integrins can interact with TNXB in lung fibroblasts and whether this interaction is important for latent TGFß activation and matrix contraction.

Increased TGFß can cause EMT (Xu et al, 2009) which has previously been implicated in ILDs including IPF and may contribute to the increased generation of fibroblasts and myofibroblasts (Kage and Borok, 2012). Interestingly, epithelial cells seeded onto the FBG-domain of TNXB undergo EMT, whereas the EMT response is much weaker when epithelial cells are seeded onto the full length of TNXB, explained by an FNIII-repeat region acting negatively on the FBG-domain (Alcaraz et al, 2014). Methylation can

regulate alternative splicing, the process by which multiple transcripts of the same gene are produced. In theory, altered TNXB methylation could result in aberrant alternative splicing which could produce increased transcripts of TNXB lacking the FNIII-repeat region. Therefore, future studies could compare the relative expression of different TNXB transcripts in IPF and control lung fibroblasts and determine their importance in PF.

7.11.6. Determine the role of other epigenetic mechanisms in IPF lung fibroblasts

DNA methylation is one of three main epigenetic mechanisms which can regulate gene expression, the other two being histone modifications and microRNAs (miRs). MiRs have been extensively studied in IPF with evidence indicating multiple miRs are dysregulated (Hagood, 2014, Pandit and Milosevic, 2015). Furthermore, methylation can regulate microRNA expression in IPF (Dakhlallah et al, 2013) thus, different epigenetic mechanisms can interact with each other to regulate gene expression. Histone modifications have also been shown to affect the expression of individual genes (Coward et al, 2010, Sanders et al, 2011, Hagood et al, 2014) and can affect methylation of genes associated with IPF (Sanders et al, 2011). However, there are no genome-wide studies examining the effects of altered histone modifications in IPF. It would therefore be of interest to study other epigenetic mechanisms including miRs and histone modifications in parallel with DNA methylation to determine how different epigenetic mechanisms interact with each other and together, contribute to altered gene expression in IPF and other diseases characterised by PF.

References

Abraham, D.J. & Varga, J., 2005. Scleroderma: From cell and molecular mechanisms to disease models. *Trends in Immunology*, 26(11), pp.587–595.

Abràmoff, M.D., Magalhães, P.J. & Ram, S.J., 2004. Image processing with ImageJ. *Biophotonics International*, 11(7), pp.36–41.

Adachi, H., Saito, I., Horiuchi, M., Ishii, J., Nagata, Y., Mizuno, F., Nakamura, H., Yagyu, H., Takahashi, K. & Matsuoka, T., 2001. Infection of human lung fibroblasts with Epstein-Barr virus causes increased IL-1beta and bFGF production. *Experimental Lung Research*, 27(2), pp.157–71.

Adams, R.L. & Burdon, R.H., 1982. DNA methylation in eukaryotes. *CRC Critical Reviews in Biochemistry*, 13(4), pp.349-384.

Ädelroth, E., Hedlund, U., Blomberg, A., Helleday, R., Ledin, M.C., Levin, J.O., Pourazar, J., Sandström, T. & Järvholm, B., 2006. Airway inflammation in iron ore miners exposed to dust and diesel exhaust. *European Respiratory Journal*, 27(4), pp.714–719.

Agarwal, S.K., Gourh, P., Shete, S., Paz, G., Divecha, D., Reveille, J.D., Assassi, S., Tan, F.K., Mayes, M.D. & Arnett, F.C., 2009. Association of interleukin 23 receptor polymorphisms with anti-topoisomerase-I positivity and pulmonary hypertension in systemic sclerosis. *Journal of Rheumatology*, 36(12), pp.2715–2723.

Ahluwalia, A., Hurteau, J.A., Bigsby, R.M. & Nephew, K.P., 2001. DNA methylation in ovarian cancer. II. Expression of DNA methyltransferases in ovarian cancer cell lines and normal ovarian epithelial cells. *Gynecologic Oncology*, 82(2), pp.299–304.

Akchiche, N., Bossenmeyer-Pourié, C., Kerek, R., Martin, N., Pourié, G., Koziel, V., Helle, D., Alberto, J.M., Ortiou, S., Camadro, J.M., Léger, T., Guéant, J.L. & Daval, J.L., 2012. Homocysteinylation of neuronal proteins contributes to folate deficiency-associated alterations of differentiation, vesicular transport, and plasticity in hippocampal neuronal cells. *FASEB Journal*, 26(10), pp.3980–3992.

Akiyama, S.K., 1996. Intergrins in cell adhesion and signalling. *Human Cell*, 9(3), pp.181-186.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P., 2002. Molecular Biology of the Cell. 4th edition. New York: Garland Science. The extracellular matrix of animals.

Alcaraz, L.B., Exposito, J.Y., Chuvin, N., Pommier, R.M., Cluzel, C., Martel, S., Sentis, S., Bartholin, L., Lethias, C. & Valcourt, U., 2014. Tenascin-x promotes epithelial-to-mesenchymal transition by activating latent TGF-β. *Journal of Cell Biology*, 205(3), pp.409–428.

Alder, J.K., Chen, J.J.-L., Lancaster, L., Danoff, S., Su, S., Cogan, J.D., Vulto, I., Xie, M., Qi, X., 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(35), pp.13051–13056.

Allanore, Y., Saad, M., Dieudé, P., Avouac, J., Distler, J.H.W., Amouyel, P., Matucci-Cerinic, M., Riemekasten, G., Airo, P., Melchers, I., Hachulla, E., Cusi, D., Wichmann, H.E., Wipff, J., Lambert, J.C., Hunzelmann, N., Tiev, K., Caramaschi, P., Diot, E., Kowal-Bielecka, O., Valentini, G., Mouthon, L., Czirják, L., Damjanov, N., Salvi, E., Conti, C., Müller, M., Müller-Ladner, U., Riccieri, V., Ruiz, B., Cracowski, J.L., Letenneur, L., Dupuy, A.M., Meyer, O., Kahan, A., Munnich, A., Boileau, C. & Martinez, M., 2011. Genome-Wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. *PLoS Genetics*, 7(7), pp.e1002091.

Allcock, R.J., Forrest, I., Corris, P.A., Crook, P.R. & Griffiths, I.D., 2004. A study of the prevalence of systemic sclerosis in northeast England. *Rheumatology (Oxford, England)*, 43(5), pp.596–602.

Allen, J.T. & Spiteri, M.A., 2002. Growth factors in idiopathic pulmonary fibrosis: relative roles. *Respiratory Research*, 3, p.13.

Allen, J.T., Knight, R.A., Bloor, C.A. & Spiteri, M.A., 1999. Enhanced insulin-like growth factor binding protein-related protein 2 (Connective tissue growth factor) expression in patients with idiopathic pulmonary fibrosis and pulmonary sarcoidosis. *American Journal of Respiratory Cell and Molecular Biology*, 21(6), pp.693–700.

American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias American Journal of Respiratory and Critical Care Medicine, 2002. *The American Journal of Respiratory and Critical Care Medicine*, 165(2), pp.277-304.

Antequera, F. & Bird, A., 1993. Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 90(24), pp.11995–11999.

Aranda, A. & Pascual, A., 2001. Nuclear hormone receptors and gene expression. *Physiological Reviews*, 81(3), pp.1269–1304.

Arase, Y., Suzuki, F., Suzuki, Y., Akuta, N., Kobayashi, M., Kawamura, Y., Yatsuji, H., Sezaki, H., Hosaka, T., Hirakawa, M., Saito, S., Ikeda, K. & Kumada, H., 2008. Hepatitis C virus enhances incidence of idiopathic pulmonary fibrosis. *World Journal of Gastroenterology*, 14(38), pp.5880–5886.

Arcucci, A., Ruocco, M.R., Amatruda, N., Riccio, A., Tarantino, G., Albano, F., Mele, V. & Montagnani, S., 2011. Analysis of extracellular superoxide dismutase in fibroblasts from patients with systemic sclerosis. *Journal of Biological Regulators and Homeostatic Agents*, 25(4), pp.647–654.

Arias-Nuñez, M.C., Llorca, J., Vazquez-Rodriguez, T.R., Gomez-Acebo, I., Miranda-Filloy, J.A., Martin, J., Gonzalez-Juanatey, C. & Gonzalez-Gay, M.A., 2008. Systemic sclerosis in northwestern Spain: a 19-year epidemiologic study. *Medicine*, 87(5), pp.272–280.

Armanios, M.Y., Chen, J.J.-L., Cogan, J.D., Alder, J.K., Ingersoll, R.G., Markin, C., Lawson, W.E., Xie, M., Vulto, I., Phillips, J.A., Lansdorp, P.M., Greider, C.W. & Loyd, J.E., 2007. Telomerase mutations in families with idiopathic pulmonary fibrosis. *The New England Journal of Medicine*, 356(13), pp.1317–1326.

Arnett, F.C., Cho, M., Chatterjee, S., Aguilar, M.B., Reveille, J.D. & Mayes, M.D., 2001. Familial occurrence frequencies and relative risks for systemic sclerosis (Scleroderma) in three United States cohorts. *Arthritis and Rheumatism*, 44(6), pp.1359–1362.

Arnett, F.C., Howard, R.F., Tan, F., Moulds, J.M., Bias, W.B., Durban, E., Cameron, H.D., Paxton, G., Hodge, T.J., Weathers, P.E. & Reveille, J.D., 1996. Increased prevalence of systemic sclerosis in a Native American tribe in Oklahoma. Association with an Amerindian HLA haplotype. *Arthritis and Rheumatism*, 39(8), pp.1362–1370.

Aryal, B.K., Khuder, S.A. & Schaub, E.A., 2001. Meta-analysis of systemic sclerosis and exposure to solvents. *American Journal of Industrial Medicine*, 40(3), pp.271–274.

Ashworth, J.L., Murphy, G., Rock, M.J., Sherratt, M.J., Shapiro, S.D., Shuttleworth, C.A. & Kielty, C.M., 1999. Fibrillin degradation by matrix metalloproteinases: implications for connective tissue remodelling. *The Biochemical Journal*, 340 (Pt 1), pp.171–181.

Assassi, S. & Tan, F.K., 2005. Genetics of scleroderma: update on single nucleotide polymorphism analysis and microarrays. *Current Opinion in Rheumatology*, 17(6), pp.761–767.

Atkins, C.P., Loke, Y.K. & Wilson, a. M., 2014. Outcomes in idiopathic pulmonary fibrosis: A meta-analysis from placebo controlled trials. *Respiratory Medicine*, 108(2), pp.376–387.

Azuma, A., Nukiwa, T., Tsuboi, E., Suga, M., Abe, S., Nakata, K., Taguchi, Y., Nagai, S., Itoh, H., Ohi, M., Sato, A., Kudoh, S. & Raghu, G., 2005. Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 171(9), pp.1040–1047.

Bando, M., Takahashi, M., Ohno, S., Hosono, T., Hironaka, M., Okamoto, H. & Sugiyama, Y., 2008. Torque teno virus DNA titre elevated in idiopathic pulmonary fibrosis with primary lung cancer. *Respirology*, 13(2), pp.263–269.

Bansil, R. & Turner, B.S., 2006. Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid and Interface Science*, 11(2-3), pp.164–170.

Barak, A.J., Beckenhauer, H.C., Junnila, M. & Tuma, D.J., 1993. Dietary betaine promotes generation of hepatic S-adenosylmethionine and protects the liver from ethanol-induced fatty infiltration. *Alcoholism, Clinical and Experimental Research*, 17(3), pp.552–555.

Barak, A.J., Beckenhauer, H.C., Tuma, D.J. & Badakhsh, S., 1987. Effects of prolonged ethanol feeding on methionine metabolism in rat liver. *Biochemistry and Cell Biology,* 65(3), pp.230–233.

Barlo, N.P., Van Moorsel, C.H.M., Korthagen, N.M., Heron, M., Rijkers, G.T., Ruven, H.J.T., Van Den Bosch, J.M.M. & Grutters, J.C., 2011. Genetic variability in the IL1RN gene and the balance between interleukin (IL)-1 receptor agonist and IL-1β in idiopathic pulmonary fibrosis. *Clinical and Experimental Immunology*, 166(3), pp.346–351.

Barlo, N.P., Van Moorsel, C.H.M., Van Den Bosch, J.M.M. & Grutters, J.C., 2010. Predicting prognosis in idiopathic pulmonary fibrosis. *Sarcoidosis Vasculitis and Diffuse Lung Diseases*, 27(2), pp.85–95.

Bartel, D.P., 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, 116(2), pp.281–297.

Bateman, E.D., Turner-Warwick, M. & Adelmann-Grill, B.C., 1981. Immunohistochemical study of collagen types in human foetal lung and fibrotic lung disease. *Thorax*, 36(9), pp.645–653.

Baumeister, P., Luo, S., Skarnes, W.C., Sui, G., Seto, E., Shi, Y. & Lee, A.S., 2005. Endoplasmic reticulum stress induction of the Grp78/BiP promoter: activating mechanisms mediated by YY1 and its interactive chromatin modifiers. *Molecular and Cellular Biology*, 25(11), pp.4529–40.

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. *American Journal of Epidemiology*, 152(4), pp.307–315.

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. *American Journal of Respiratory and Critical Care Medicine*, 155(1), pp.242–248.

Bayle, J., Fitch, J., Jacobsen, K., Kumar, R., Lafyatis, R. & Lemaire, R., 2008. Increased expression of Wnt2 and SFRP4 in Tsk mouse skin: role of Wnt signaling in altered dermal fibrillin deposition and systemic sclerosis. *The Journal of Investigative Dermatology*, 128(4), pp.871–881.

Bechtel, W., McGoohan, S., Zeisberg, E.M., Müller, G.A., Kalbacher, H., Salant, D.J., Müller, C.A., Kalluri, R. & Zeisberg, M., 2010. Methylation determines fibroblast activation and fibrogenesis in the kidney. *Nature Medicine*, 16(5), pp.544–550.

Bell, E., Ivarsson, B. & Merrill, C., 1979. Production of a tissue-like structure by contraction of collagen lattices by human fibroblasts of different proliferative potential in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 76(3), pp.1274–1278.

Bell, J.T., Tsai, P.C., Yang, T.P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., Shin, S.Y., Dempster, E.L., Murray, R.M., Grundberg, E., Hedman, A.K., Nica, A., Small, K.S., Dermitzakis, E.T., McCarthy, M.I., Mill, J., Spector, T.D. & Deloukas, P., 2012. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genetics*, 8(4).

Bensadoun, E.S., Burke, A.K., Hogg, J.C. & Roberts, C.R., 1996. Proteoglycan deposition in pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 154(6), pp.1819–1828.

Ben-Sasson, S.A. & Klein, G., 1981. Activation of the Epstein-Barr virus genome by 5-Aza-cytidine in latently infected human lymphoid lines. *International Journal of Cancer*, 28(2), pp.131–5.

Bérezné, A., Ranque, B., Valeyre, D., Brauner, M., Allanore, Y., Launay, D., Le Guern, V., Kahn, J.E., Couderc, L.J., Constans, J., Cohen, P., Mahr, A., Pagnoux, C., Hachulla, E., Kahan, A., Cabane, J., Guillevin, L. & Mouthon, L., 2008. Therapeutic strategy combining intravenous cyclophosphamide followed by oral azathioprine to treat worsening interstitial lung disease associated with systemic sclerosis: A retrospective multicenter open-label study. *Journal of Rheumatology*, 35(6), pp.1064–1072.

Berezne, A., Valeyre, D., Ranque, B., Guillevin, L. & Mouthon, L., 2007. Interstitial lung disease associated with systemic sclerosis: what is the evidence for efficacy of cyclophosphamide? *Annals of the New York Academy of Sciences*, 1110, pp.271–284.

Besingi, W. & Johansson, Å., 2014. Smoke-related DNA methylation changes in the etiology of human disease. *Human Molecular Genetics,* 23(9), pp.2290–2297.

Bestor, T.H., 2000. The DNA methyltransferases of mammals. *Human Molecular Genetics,* 9(16), pp.2395–2402.

Bhattacharyya, S., Kelley, K., Melichian, D.S., Tamaki, Z., Fang, F., Su, Y., Feng, G., Pope, R.M., Budinger, G.R.S., Mutlu, G.M., Lafyatis, R., Radstake, T., Feghali-Bostwick, C. & Varga, J., 2013. Toll-like receptor 4 signaling augments transforming growth factor-β responses: A novel mechanism for maintaining and amplifying fibrosis in scleroderma. *American Journal of Pathology*, 182(1), pp.192–205.

Bhattacharyya, S.N., Passero, M.A., DiAugustine, R.P. & Lynn, W.S., 1975. Isolation and characterization of two hydroxyproline containing glycoproteins from normal animal lung lavage and lamellar bodies. *Journal of Clinical Investigation*, 55(5), pp.914–920.

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R. & Gunderson, K.L., 2009. Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics*, 1(1), pp.177–200.

Bienkowski, R.S. & Gotkin, M.G., 1995. Control of collagen deposition in mammalian lung. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, 209(2), pp.118–140.

Billings, C.G. & Howard, P., 1994. Hypothesis: exposure to solvents may cause fibrosing alveolitis. *European Respiratory Journal*, 7(6), pp.1172–1176.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. *Genes and Development*, 16(1), pp.6–21.

Bird, A., 2007. Perceptions of epigenetics. *Nature*, 447(7143), pp.396–398.

Bird, A., Taggart, M., Frommer, M., Miller, O.J. & Macleod, D., 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1), pp.91–99.

Bird, A.P., 1978. Use of restriction enzymes to study eukaryotic DNA methylation. *Journal of Molecular Biology*, 118(1), pp.49–60.

Birdwell, C.E., Queen, K.J., Kilgore, P., Rollyson, P., Trutschl, M., Cvek, U. & Scott, R.S., 2014. Genome-wide DNA methylation as an epigenetic consequence of Epstein-Barr virus infection of immortalized keratinocytes. *Journal of Virology*, 88, pp.11442–11458.

Bjoraker, J.A., Ryu, J.H., Edwin, M.K., Myers, J.L., Tazelaar, H.D., Schroeder, D.R. & Offord, K.P., 1998. Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 157(1), pp.199–203.

Blaauboer, M.E., Boeijen, F.R., Emson, C.L., Turner, S.M., Zandieh-Doulabi, B., Hanemaaijer, R., Smit, T.H., Stoop, R. & Everts, V., 2013. Extracellular matrix proteins: A positive feedback loop in lung fibrosis? *Matrix biology: Journal of the International Society for Matrix Biology*, pp.2–10.

Black, C.M., Silman, A.J., Herrick, A.I., Denton, C.P., Wilson, H., Newman, J., Pompon, L. & Shi-Wen, X., 1999. Interferon-alpha does not improve outcome at one year in patients with diffuse cutaneous scleroderma: results of a randomized, double-blind, placebo-controlled trial. *Arthritis and Rheumatism*, 42(2), pp.299-305.

Blair, J.D. & Price, E.M., 2012. Illuminating Potential Technical Artifacts of DNA-Methylation Array Probes. The *American Journal of Human Genetics*, 91(4), pp.760–762.

Bockmuhl, Y., Patchev, A. V, Madejska, A., Hoffmann, A., Sousa, J.C., Sousa, N., Holsboer, F., Almeida, O.F.X. & Spengler, D., 2015. Methylation at the CpG island shore region upregulates Nr3c1 promoter activity after early-life stress. *Epigenetics: Official Journal of the DNA Methylation Society*, 10(3), pp.247–257.

Boks, M.P., Derks, E.M., Weisenberger, D.J., Strengman, E., Janson, E., Sommer, I.E., Kahn, R.S. & Ophoff, R.A., 2009. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS ONE*, 4(8), e6767.

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. *American Journal of Respiratory and Critical Care Medicine*, 186(9), pp.866–876.

Bossini-Castillo, L., Broen, J.C.A., Simeon, C.P., Beretta, L., Vonk, M.C., Ortego-Centeno, N., Espinosa, G., Carreira, P., Camps, M.T., Navarrete, N., González-Escribano, M.F., Vicente-Rabaneda, E., Rodríguez, L., Tolosa, C., Román-Ivorra, J.A., Gómez-Gracia, I., García-Hernández, F.J., Castellví, I., Gallego, M., Fernández-Nebro, A., García-Portales, R., Egurbide, M.V., Fonollosa, V., de la Peña, P.G., Pros, A., González-Gay, M.A., Hesselstrand, R., Riemekasten, G., Witte, T., Coenen, M.J.H., Koeleman, B.P., Houssiau, F., Smith, V., de Keyser, F., Westhovens, R., De Langhe, E., Voskuyl, A.E., Schuerwegh, A.J., Chee, M.M., Madhok, R., Shiels, P., Fonseca, C., Denton, C., Claes, K., Padykov, L., Nordin, A., Palm, O., Lie, B.A., Airó, P., Scorza, R., van Laar, J.M., Hunzelmann, N., Kreuter, A., Herrick, A., Worthington, J., Radstake, T.R.D.J., Martín, J. & Rueda, B., 2011. A replication study confirms the association of TNFSF4 (OX40L) polymorphisms with systemic sclerosis in a large European cohort. *Annals of the Rheumatic Diseases*, 70(4), pp.638–641.

Bouros, D., Wells, A.U., Nicholson, A.G., Colby, T. V., Polychronopoulos, V., Pantelidis, P., Haslam, P.L., Vassilakis, D.A., Black, C.M. & Du Bois, R.M., 2002. Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *American Journal of Respiratory and Critical Care Medicine*, 165(12), pp.1581–1586.

Boute, N., Exposito, J.Y., Boury-Esnault, N., Vacelet, J., Noro, N., Miyazaki, K., Yoshizato, K. & Garrone, R., 1997. Type IV collagen in sponges, the missing link in basement membrane ubiquity. *Biology of the Cell*, 88(1-2), pp.37–44.

Bradley, K., McConnell-Breul, S. & Crystal, R.G., 1974. Lung collagen heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America*, 71(7), pp.2828–2832.

Bramwell, B., 1914. Diffuse scleroderma: its frequency, its occurrence in stonemasons, its treatment by fibrinolysin, and elevation of temperature do to fibrinolysin injections. *Edinburgh Medical Journal*, 12, pp.387–40.

Breitling, L.P., Yang, R., Korn, B., Burwinkel, B. & Brenner, H., 2011. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *American Journal of Human Genetics*, 88(4), pp.450–457.

Brissett, M., Veraldi, K.L., Pilewski, J.M., Medsger, T.A. & Feghali-Bostwick, C.A., 2012. Localized expression of tenascin in systemic sclerosis-associated pulmonary fibrosis and its regulation by insulin-like growth factor binding protein 3. *Arthritis and Rheumatism*, 64(1), pp.272–280.

Broen, J.C., Radstake, T.R. & Rossato, M., 2014. The role of genetics and epigenetics in the pathogenesis of systemic sclerosis. *Nature Reviews. Rheumatology*, pp.1–11.

Brown, J.M., Pfau, J.C. & Holian, A., 2004. Immunoglobulin and lymphocyte responses following silica exposure in New Zealand mixed mice. *Inhalation Toxicology*, 16(3), pp.133–139.

Burch, G.H., Gong, Y., Liu, W., Dettman, R.W., Curry, C.J., Smith, L., Miller, W.L. & Bristow, J., 1997. Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nature Genetics*, 17(1), pp.104-108.

Busse, W.W., Lemanske, R.F.J. & Gern, J.E., 2010. Role of viral respiratory infections in asthma and asthma exacerbations. *Lancet*, 376(9743), pp.826–834.

Bussone, G. & Mouthon, L., 2011. Interstitial lung disease in systemic sclerosis. *Autoimmunity Reviews*, 10(5), pp.248–255.

Calabrese, F., Kipar, A., Lunardi, F., Balestro, E., Perissinotto, E., Rossi, E., Nannini, N., Marulli, G., Stewart, J.P. & Rea, F., 2013. Herpes virus infection is associated with vascular remodeling and pulmonary hypertension in idiopathic pulmonary fibrosis. *PLoS ONE*, 8(2), e55715.

Calderwood, C.J., Jones, M.G., Hoile, L., Havelock, T., Maher, T.M., O'Reilly, K.M.A. & Davies, D., 2012. Mechanisms of chronic lung disease: p113 secreted lysyl oxidase is elevated in the bronchoalveolar lavage fluid of patients with idiopathic pulmonary fibrosis. *Thorax*, 67(Supplement 2), A111.

Calvanese, V., Lara, E., Kahn, A. & Fraga, M.F., 2009. The role of epigenetics in aging and age-related diseases. *Ageing Research Reviews*, 8(4), pp.268–276.

Campo, I., Zorzetto, M., Mariani, F., Kadija, Z., Morbini, P., Dore, R., Kaltenborn, E., Frixel, S., Zarbock, R., Liebisch, G., Hegermann, J., Wrede, C., Griese, M. & Luisetti, M., 2014. A large kindred of pulmonary fibrosis associated with a novel ABCA3 gene variant. *Respiratory Research*, 15(43), pp.1–15.

Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J.M. & Pascual-Montano, A., 2007. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biology*, 8(1), p.R3.

Carr, P.A., Erickson, H.P. & Palmer, A.G., 1997. Backbone dynamics of homologous fibronectin type III cell adhesion domains from fibronectin and tenascin. *Structure (London, England : 1993)*, 5(7), pp.949–959.

Carrel, L. & Willard, H.F., 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, 434(7031), pp.400–404.

Carwile LeRoy, E., Black, C., Fleischmajer, R., Jablonska, S., Krieg, T., Medsger, T.A. & Wollheim, F., 1988. Scleroderma (systemic sclerosis): Classification, subsets and pathogenesis. *Journal of Rheumatology*, 15(2), pp.202–205.

Cedar, H. & Bergman, Y., 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews. Genetics*, 10(5), pp.295–304.

Chace, K. V, Leahy, D.S., Martin, R., Carubelli, R., Flux, M. & Sachdev, G.P., 1983. Respiratory mucous secretions in patients with cystic fibrosis: relationship between levels of highly sulfated mucin component and severity of the disease. *Clinica Chimica Acta*, 132(2), pp.143–155.

Chambers, R.C. & Scotton, C.J., 2012. Coagulation cascade proteinases in lung injury and fibrosis. *Proceedings of the American Thoracic Society*, 9(3), pp.96–101.

Chambers, R.C., 2008. Procoagulant signalling mechanisms in lung inflammation and fibrosis: novel opportunities for pharmacological intervention? *British Journal of Pharmacology*, 153(Supplement 1), pp.S367–S378.

Chambers, R.C., Leoni, P., Kaminski, N., Laurent, G.J. & Heller, R.A., 2003. Global expression profiling of fibroblast responses to transforming growth factor-beta1 reveals the induction of inhibitor of differentiation-1 and provides evidence of smooth muscle cell phenotypic switching. *The American Journal of Pathology*, 162(2), pp.533–546.

Chang, W., Wei, K., Jacobs, S.S., Upadhyay, D., Weill, D. & Rosen, G.D., 2010. SPARC suppresses apoptosis of idiopathic pulmonary fibrosis fibroblasts through constitutive activation of β-catenin. *Journal of Biological Chemistry*, 285(11), pp.8196–8206.

Chen, H. & Boutros, P.C., 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, 12(1), p.35.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E., 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, 23(16), pp.5594–5605.

Chen, Y., Choufani, S., Ferreira, J.C., Grafodatskaya, D., Butcher, D.T. & Weksberg, R., 2011. Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray. *Genomics*, 97(4), pp.214–222.

Chen, Y., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J. & Weksberg, R., 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics: Official Journal of the DNA Methylation Society*, 8(2), pp.203–9.

Cheng, J.C., Matsen, C.B., Gonzales, F. a, Ye, W., Greer, S., Marquez, V.E., Jones, P. & Selker, E.U., 2003. Inhibition of DNA methylation and reactivation of silenced genes by zebularine. *Journal of the National Cancer Institute*, 95(5), pp.399–409.

Cheresh, P., Kim, S.-J., Tulasiram, S. & Kamp, D.W., 2013. Oxidative stress and pulmonary fibrosis. *Biochimica et Biophysica Acta*, 1832(7), pp.1028–40.

Chiang, J.Y.L., 2009. Bile acids: regulation of synthesis. *Journal of Lipid Research*, 50(10), pp.1955–1966.

Chifflot, H., Fautrel, B., Sordet, C., Chatelus, E. & Sibilia, J., 2008. Incidence and Prevalence of Systemic Sclerosis: A Systematic Literature Review. *Seminars in Arthritis and Rheumatism*, 37(4), pp.223–235.

Chilosi, M., Poletti, V., Zamò, A., Lestani, M., Montagna, L., Piccoli, P., Pedron, S., Bertaso, M., Scarpa, A., Murer, B., Cancellieri, A., Maestro, R., Semenzato, G. & Doglioni, C., 2003. Aberrant Wnt/beta-catenin pathway activation in idiopathic pulmonary fibrosis. *The American Journal of Pathology*, 162(5), pp.1495–1502.

Chouliaras, L., Mastroeni, D., Delvaux, E., Grover, A., Kenis, G., Hof, P.R., Steinbusch, H.W.M., Coleman, P.D., Rutten, B.P.F. & van den Hove, D.L.A., 2013. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiology of Aging*, 34(9), pp.2091–2099.

Christman, J.K., 2002. 5-Azacytidine and 5-Aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*, 21(35), pp.5483–5495.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G. & Li, B.F., 1997. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science (New York, N.Y.)*, 277(5334), pp.1996–2000.

Chung, M.J., Liu, T., Ullenbruch, M. & Phan, S.H., 2007. Antiapoptotic effect of found in inflammatory zone (FIZZ)1 on mouse lung fibroblasts. *Journal of Pathology*, 212(2), pp.180–187.

Ciechomska, M., O'Reilly, S., Suwara, M., Bogunia-Kubik, K. & van Laar, J.M., 2014. Mir-29a reduces timp-1 production by dermal fibroblasts via targeting tgf-β activated kinase 1 binding protein 1, implications for systemic sclerosis. *PLoS ONE*, 9(12), e115596.

Cisneros, J., Hagood, J., Checa, M., Ortiz-Quintero, B., Negreros, M., Herrera, I., Ramos, C., Pardo, A. & Selman, M., 2012. Hypermethylation-mediated silencing of p14ARF in fibroblasts from idiopathic pulmonary fibrosis. *The American Journal of Physiology: Lung Cellular and Molecular Physiology*, 303(4), pp.L295–L303.

Clark, J.G., Kostal, K.M. & Marino, B. A, 1982. Modulation of collagen production following bleomycin-induced pulmonary fibrosis in hamsters. *The Journal of Biologicl Chemistry*, 257(14), pp.8098–8105.

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(1), pp.1–9.

Clements, P.J., Seibold, J.R., Furst, D.E., Mayes, M., White, B., Wigley, F., Weisman, M.D., Barr, W., Moreland, L., Medsger, T.A., Steen, V., Martin, R.W., Collier, D., Weinstein, A., Lally, E., Varga, J., Weiner, S.R., Andrews, B., Abeles, M. & Wong, W.K., 2004. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis trial: lessons learned. *Seminars in Arthritis and Rheumatism*, 33(4), pp.249–263.

Collard, H.R., Moore, B.B., Flaherty, K.R., Brown, K.K., Kaner, R.J., King, T.E., Lasky, J.A., Loyd, J.E., Noth, I., Olman, M.A., Raghu, G., Roman, J., Ryu, J.H., Zisman, D.A., Hunninghake, G.W., Colby, T. V., Egan, J.J., Hansell, D.M., Johkoh, T., Kaminski, N., Dong, S.K., Kondoh, Y., Lynch, D.A., Müller-Quernheim, J., Myers, J.L., Nicholson, A.G., Selman, M., Toews, G.B., Wells, A.U. & Martinez, F.J., 2007. Acute exacerbations of idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 176(7), pp.636–643.

Collard, H.R., Ryu, J.H., Douglas, W.W., Schwarz, M.I., Curran-Everett, D., King, T.E. & Brown, K.K., 2004. Combined corticosteroid and cyclophosphamide therapy does not alter survival in idiopathic pulmonary fibrosis. *Chest*, 125(6), pp.2169–2174.

Collazos, J., Carton, J.A. & Asensi, V., 2011. Gender differences in liver fibrosis and hepatitis C virus-related parameters in patients co-infected with human immunodeficiency virus. *Current HIV Research*, 9(5), pp.339–345.

Coombes, M.M., Briggs, K.L., Bone, J.R., Clayman, G.L., El-Naggar, A.K. & Dent, S.Y.R., 2003. Resetting the histone code at CDKN2A in HNSCC by inhibition of DNA methylation. *Oncogene*, 22(55), pp.8902–11.

Costabel, U. & Bonella, F., 2011. Treatment of pulmonary fibrosis. New substances and new interventions. *The Internist*, 52(12), pp.1422–8.

Cottin, V., 2013. The role of pirfenidone in the treatment of idiopathic pulmonary fibrosis. *Respiratory Research*, 14(Supplement 1), p.S5.

Couillard, J., Demers, M., Lavoie, G. & St-Pierre, Y., 2006. The role of DNA hypomethylation in the control of stromelysin gene expression. *Biochemical and Biophysical Research Communications*, 342(4), pp.1233–1239.

Coultas, D.B., Zumwalt, R.E., Black, W.C. & Sobonya, R.E., 1994. The epidemiology of interstitial lung diseases. *American Journal of Respiratory and Critical Care Medicine*, 150(4), pp.967–72.

Coward, W.R., Watts, K., Feghali-Bostwick, C.A., Jenkins, G. & Pang, L., 2010. Repression of IP-10 by interactions between histone deacetylation and hypermethylation in idiopathic pulmonary fibrosis. *Molecular and Cellular Biology*, 30(12), pp.2874–86.

Cox, T.R. & Erler, J.T., 2011. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Disease Models & Mechanisms*, 4(2), pp.165–178.

Cox, T.R., Bird, D., Baker, A.M., Barker, H.E., Ho, M.W.Y., Lang, G. & Erler, J.T., 2013. LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis. *Cancer Research*, 73(6), pp.1721–1732.

Crapo, J.D., Barry, B.E., Gehr, P., Bachofen, M. & Weibel, E.R., 1982. Cell number and cell characteristics of the normal human lung. *The American Review of Respiratory Disease*, 126(2), pp.332–337.

Croen, L. A., Grether, J.K. & Selvin, S., 2002. Descriptive epidemiology of autism in a California population: Who Is at Risk? *Journal of Autism and Developmental Disorders*, 32(3), pp.217–224.

Cruz-hernandez, E. De, Trujillo, J. & Medina-franco, J.L., 2011. DNA Demethylating Activity of Hydralazine in Cancer Cell Lines. *Life Sciences and Medicine Research*, 2011, pp.1–8.

Crystal, R.G., Fulmer, J.D., Roberts, W.C., Moss, M.L., Line, B.R. & Reynolds, H.Y., 1976. Idiopathic pulmonary fibrosis. Clinical, histologic, radiographic, physiologic, scintigraphic, cytologic, and biochemical aspects. *Annals of Internal Medicine*, 85(6), pp.769–788.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Kähäri, A.K., Keenan, S., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Aken, B.L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S.M.J., Spudich, G., Trevanion, S.J., Yates, A., Zerbino, D.R. & Flicek, P., 2014. Ensembl 2015. *Nucleic Acids Research*, 43, pp.D662–9.

Cuozzo, C., Porcellini, A., Angrisano, T., Morano, A., Lee, B., Di Pardo, A., Messina, S., Iuliano, R., Fusco, A., Santillo, M.R., Muller, M.T., Chiariotti, L., Gottesman, M.E. & Avvedimento, E. V., 2007. DNA damage, homology-directed repair, and DNA methylation. *PLoS Genetics*, 3(7), pp.1144–1162.

Cushing, L., Kuang, P.P., Qian, J., Shao, F., Wu, J., Little, F., Thannickal, V.J., Cardoso, W. V. & Lü, J., 2011. miR-29 is a major regulator of genes associated with pulmonary fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, 45(2), pp.287–294.

D'Alessio, S., Fibbi, G., Cinelli, M., Guiducci, S., Del Rosso, A., Margheri, F., Serratì, S., Pucci, M., Kahaleh, B., Fan, P., Annunziato, F., Cosmi, L., Liotta, F., Matucci-Cerinic, M. & Del Rosso, M., 2004. Matrix metalloproteinase 12-dependent cleavage of urokinase receptor in systemic sclerosis microvascular endothelial cells results in impaired angiogenesis. *Arthritis and Rheumatism*, 50(10), pp.3275–3285.

Dakhlallah, D., Batte, K., Wang, Y., Cantemir-Stone, C.Z., Yan, P., Nuovo, G., Mikhail, A., Hitchcock, C.L., Wright, V.P., Nana-Sinkam, S.P., Piper, M.G. & Marsh, C.B., 2013. Epigenetic regulation of miR-17-92 contributes to the pathogenesis of pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 187(4), pp.397–405.

De Leon, A.D., Cronkhite, J.T., Katzenstein, A.L.A., Godwin, J.D., Raghu, G., Glazer, C.S., Rosenblatt, R.L., Girod, C.E., Garrity, E.R., Xing, C. & Garcia, C.K., 2010. Telomere lengths, pulmonary fibrosis and telomerase (TERT) Mutations. *PLoS ONE*, 5(5).

De Vuyst, P., Dumortier, P., Rickaert, F., Van de Weyer, R., Lenclud, C. & Yernault, J.C., 1986. Occupational lung fibrosis in an aluminium polisher. *European Journal of Respiratory Diseases*, 68(2), pp.131-140.

De Wever, O., Demetter, P., Mareel, M. & Bracke, M., 2008. Stromal myofibroblasts are drivers of invasive cancer growth. *International Journal of Cancer*, 123(10), pp.2229–2238.

Deaton, A.M. & Bird, A., 2011. CpG islands and the regulation of transcription. *Genes & Development*, 25(10), pp.1010–1022.

Decaris, M.L., Gatmaitan, M., FlorCruz, S., Luo, F., Li, K., Holmes, W.E., Hellerstein, M.K., Turner, S.M. & Emson, C.L., 2014. Proteomic analysis of altered extracellular matrix turnover in bleomycin-induced pulmonary fibrosis. *Molecular & Cellular Proteomics*, 13(7), pp.1741–52.

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C. & Fuks, F., 2011. Evaluation of the Infinium Methylation 450K technology. *Epigenomics*, 3(6), pp.771–784.

Dees, C., Schlottmann, I., Funke, R., Distler, A., Palumbo-Zerr, K., Zerr, P., Lin, N.-Y., Beyer, C., Distler, O., Schett, G. & Distler, J.H.W., 2014. The Wnt antagonists DKK1 and SFRP1 are downregulated by promoter hypermethylation in systemic sclerosis. *Annals of the Rheumatic Diseases*, 73(6), pp.1232–9.

Dees, C., Tomcik, M., Zerr, P., Akhmetshina, A., Horn, A., Palumbo, K., Beyer, C., Zwerina, J., Distler, O., Schett, G. & Distler, J.H.W., 2011. Notch signalling regulates fibroblast activation and collagen release in systemic sclerosis. *Annals of the Rheumatic Diseases*, 70(7), pp.1304–1310.

DePianto, D.J., Chandriani, S., Abbas, A.R., Jia, G., N'Diaye, E.N., Caplazi, P., Kauder, S.E., Biswas, S., Karnik, S.K., Ha, C., Modrusan, Z., Matthay, M.A., Kukreja, J., Collard, H.R., Egen, J.G., Wolters, P.J. & Arron, J.R., 2014. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. *Thorax*, 70(1), pp.48–56.

Desmoulière, A., Badid, C., Bochaton-Piallat, M.L. & Gabbiani, G., 1997. Apoptosis during wound healing, fibrocontractive diseases and vascular wall injury. *International Journal of Biochemistry and Cell Biology*, 29(1), pp.19–30.

Desplats, P., Spencer, B., Coffee, E., Patel, P., Michael, S., Patrick, C., Adame, A., Rockenstein, E. & Masliah, E., 2011. α-Synuclein sequesters Dnmt1 from the nucleus: A novel mechanism for epigenetic alterations in Lewy body diseases. *Journal of Biological Chemistry*, 286(11), pp.9031–9037.

Di Sario, A., Bendia, E., Svegliati Baroni, G., Ridolfi, F., Casini, A., Ceni, E., Saccomanno, S., Marzioni, M., Trozzi, L., Sterpetti, P., Taffetani, S. & Benedetti, A., 2002. Effect of pirfenidone on rat hepatic stellate cell proliferation and collagen production. *Journal of Hepatology*, 37(5), pp.584–591.

Dieudé, P., Guedj, M., Wipff, J., Avouac, J., Fajardy, I., Diot, E., Granel, B., Sibilia, J., Cabane, J., Mouthon, L., Cracowski, J.L., Carpentier, P.H., Hachulla, E., Meyer, O., Kahan, A., Boileau, C. & Allanore, Y., 2009. Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: A new perspective for pulmonary fibrosis. *Arthritis and Rheumatism*, 60(1), pp.225–233.

Dieudé, P., Guedj, M., Wipff, J., Ruiz, B., Hachulla, E., Diot, E., Granel, B., Sibilia, J., Tiev, K., Mouthon, L., Cracowski, J.L., Carpentier, P.H., Amoura, Z., Fajardy, I., Avouac, J., Meyer, O., Kahan, A., Boileau, C. & Allanore, Y., 2009. STAT4 is a genetic risk factor for systemic sclerosis having additive effects with IRF5 on disease susceptibility and related pulmonary fibrosis. *Arthritis and Rheumatism*, 60(8), pp.2472–2479.

Dieudé, P., Wipff, J., Guedj, M., Ruiz, B., Melchers, I., Hachulla, E., Riemekasten, G., Diot, E., Hunzelmann, N., Sibilia, J., Tiev, K., Mouthon, L., Cracowski, J.L., Carpentier, P.H., Distler, J., Amoura, Z., Tarner, I., Avouac, J., Meyer, O., Kahan, A., Boileau, C. & Allanore, Y., 2009. BANK1 is a genetic risk factor for diffuse cutaneous systemic sclerosis and has additive effects with IRF5 and STAT4. *Arthritis and Rheumatism*, 60(11), pp.3447–3454.

Doan, M.L., Guillerman, R.P., Dishop, M.K., Nogee, L.M., Langston, C., Mallory, G.B., Sockrider, M.M. & Fan, L.L., 2008. Clinical, radiological and pathological features of ABCA3 mutations in children. *Thorax*, 63(4), pp.366–373.

Dodge, J.E., Ramsahoye, B.H., Wo, Z.G., Okano, M. & Li, E., 2002. De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene*, 289(1-2), pp.41–48.

Dogan, M. V, Shields, B., Cutrona, C., Gao, L., Gibbons, F.X., Simons, R., Monick, M., Brody, G.H., Tan, K., Beach, S.R. & Philibert, R. a, 2014. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics*, 15(1), p.151.

Dominguez-Salas, P., Moore, S.E., Baker, M.S., Bergen, A.W., Cox, S.E., Dyer, R. A, Fulford, A.J., Guan, Y., Laritsky, E., Silver, M.J., Swan, G.E., Zeisel, S.H., Innis, S.M., Waterland, R. A, Prentice, A.M. & Hennig, B.J., 2014. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nature Communications*, 5, p.3746.

Downing, T.E., Sporn, T.A., Bollinger, R.R., Davis, R.D., Parker, W. & Lin, S.S., 2008. Pulmonary histopathology in an experimental model of chronic aspiration is independent of acidity. *Experimental Biology and Medicine*, 233(10), pp.1202–1212.

Du, P., Kibbe, W. A. & Lin, S.M., 2008. Lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, 24(13), pp.1547–8.

Duisters, R.F., Tijsen, A.J., Schroen, B., Leenders, J.J., Lentink, V., Van Der Made, I., Herias, V., Van Leeuwen, R.E., Schellings, M.W., Barenbrug, P., Maessen, J.G., Heymans, S., Pinto, Y.M. & Creemers, E.E., 2009. MiR-133 and miR-30 regulate connective tissue growth factor: Implications for a role of micrornas in myocardial matrix remodeling. *Circulation Research*, 104(2), pp.170–178.

Eastman, Q. & Grosschedl, R., 1999. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Current Opinion in Cell Biology*, 11(2), pp.233–240.

Egan, J.J., Stewart, J.P., Hasleton, P.S., Arrand, J.R., Carroll, K.B. & Woodcock, A.A., 1995. Epstein-Barr virus replication within pulmonary epithelial cells in cryptogenic fibrosing alveolitis. *Thorax*, 50(12), pp.1234–1239.

Egan, J.J., Woodcock, A.A. & Stewart, J.P., 1997. Viruses and idiopathic pulmonary fibrosis. *European Respiratory Journal*, 10(7), pp.1433–1437.

Egger, G., Liang, G., Aparicio, A. & Jones, P.A., 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, 429(6990), pp.457–463.

Egging, D., Van Den Berkmortel, F., Taylor, G., Bristow, J. & Schalkwijk, J., 2007. Interactions of human tenascin-X domains with dermal extracellular matrix molecules. *Archives of Dermatological Research*, 298(8), pp.389–396.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., Mccune, R.A. & Gehrke, C., 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, 10(8), pp.2709–2721.

Eisenberg, H., DuBois, E.L., Sherwin, R.P. & Balchum, O.J., 1973. Diffuse interstitial lung disease in systemic lupus erythematosus. *Annals of Internal Medicine*, 79, pp.37-45.

Elefteriou, F., Exposito, J.Y., Garrone, R. & Lethias, C., 2001. Binding of tenascin-X to decorin. *FEBS Letters*, 495(1-2), pp.44–47.

Elias, G.J., Ioannis, M., Theodora, P., Dimitrios, P.P., Despoina, P., Kostantinos, V., Charalampos, K., Vassilios, V. & Petros, S.P., 2008. Circulating tissue inhibitor of matrix metalloproteinase-4 (TIMP-4) in systemic sclerosis patients with elevated pulmonary arterial pressure. *Mediators of Inflammation*, 2008, e164134.

Elkington, P.T.G. & Friedland, J.S., 2006. Matrix metalloproteinases in destructive pulmonary pathology. *Thorax*, 61(3), pp.259–266.

Elliott, H.R., Tillin, T., McArdle, W.L., Ho, K., Duggirala, A., Frayling, T.M., Davey Smith, G., Hughes, A.D., Chaturvedi, N. & Relton, C.L., 2014. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clinical Epigenetics*, 6(1), p.4.

Englert, H., Small-McMahon, J., Chambers, P., O'Connor, H., Davis, K., Manolios, N., White, R., Dracos, G. & Brooks, P., 1999. Familial risk estimation in systemic sclerosis. *Australian and New Zealand Journal of Medicine*, 29(1), pp.36–41.

Erasmum, L.D., 1957. Scleroderma in goldminers on the Witwatersrand with particular reference to pulmonary manifestations. *South African Journal of Laboratory and Clinical Medicine*, 3(3), pp.209-231.

Erfurth, F.E., Popovic, R., Grembecka, J., Cierpicki, T., Theisler, C., Xia, Z.-B., Stuart, T., Diaz, M.O., Bushweller, J.H. & Zeleznik-Le, N.J., 2008. MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), pp.7517–22.

Estany, S., Vicens-Zygmunt, V., Llatjós, R., Montes, A., Penín, R., Escobar, I., Xaubet, A., Santos, S., Manresa, F., Dorca, J. & Molina-Molina, M., 2014. Lung fibrotic tenascin-C upregulation is associated with other extracellular matrix proteins and induced by TGFβ1. *BMC Pulmonary Medicine*, 14(1), pp.120–142.

Esteller, M. & Herman, J.G., 2002. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *Journal of Pathology*, 196(1), pp.1–7.

Esteller, M., Corn, P.G., Baylin, S.B. & Herman, J.G., 2001. A gene hypermethylation profile of human cancer. *Cancer Research*, 61(8), pp.3225–3229.

Exposito, J.Y., Valcourt, U., Cluzel, C. & Lethias, C., 2010. The fibrillar collagen family. *International Journal of Molecular Sciences*, 11(2), pp.407–426.

Fang, H. & Gough, J., 2013. DcGO: database of domain-centric ontologies on functions, phenotypes, diseases and more. *Nucleic Acids Research*, 41, pp.D536–44.

Farina, A., Cirone, M., York, M., Lenna, S., Padilla, C., McLaughlin, S., Faggioni, A., Lafyatis, R., Trojanowska, M. & Farina, G. A, 2014. Epstein-Barr Virus infection induces aberrant tlr activation pathway and fibroblast-myofibroblast conversion in scleroderma. *The Journal of Investigative Dermatology*, 134(4), pp.954–64.

Fatini, C., Mannini, L., Sticchi, E., Rogai, V., Guiducci, S., Conforti, M.L., Cinelli, M., Pignone, A.M., Bolli, P., Abbate, R. & Cerinic, M.M., 2006. Hemorheologic profile in systemic sclerosis: Role of NOS3 -786T>C and 894G>T polymorphisms in modulating both the hemorheologic parameters and the susceptibility to the disease. *Arthritis and Rheumatism*, 54(7), pp.2263–2270.

Fattman, C.L., 2008. Apoptosis in pulmonary fibrosis: too much or not enough? *Antioxidants and Redox Signaling*, 10(2), pp.379–385.

Feghali-Bostwick, C., Medsger, T.A. & Wright, T.M., 2003. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. *Arthritis and Rheumatism*, 48(7), pp.1956–1963.

Feinberg, A.P. & Tycko, B., 2004. The history of cancer epigenetics. *Nature Reviews. Cancer*, 4(2), pp.143–153.

Feinberg, A.P. & Vogelstein, B., 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301(5895), pp.89–92.

Feinberg, A.P., 2007. Phenotypic plasticity and the epigenetics of human disease. *Nature*, 447(7143), pp.433–440.

Ferri, C., Giuggioli, D., Sebastiani, M., Panfilo, S., Abatangelo, G., Zakrzewska, K. & Azzi, A., 2002. Parvovirus B19 infection of cultured skin fibroblasts from systemic sclerosis patients: Comment on the article by Ray et al. *Arthritis and Rheumatism*, 46(8), pp.2262–2263.

Figueroa, S., Gerstenhaber, B., Welch, L., Klimstra, D., Smith, G.J.W. & Beckett, W., 1992. Hard metal interstitial pulmonary disease associated with a form of welding in a metal parts coating plant. *American Journal of Industrial Medicine*, 21(3), pp.363–373.

Fingerlin, T.E., Murphy, E., Zhang, W., Peljto, A.L., Brown, K.K., Steele, M.P., Loyd, J.E., Cosgrove, G.P., Lynch, D., Groshong, S., Collard, H.R., Wolters, P.J., Bradford, W.Z., Kossen, K., Seiwert, S.D., du Bois, R.M., Garcia, C.K., Devine, M.S., Gudmundsson, G., Isaksson, H.J., Kaminski, N., Zhang, Y., Gibson, K.F., Lancaster, L.H., Cogan, J.D., Mason, W.R., Maher, T.M., Molyneaux, P.L., Wells, A.U., Moffatt, M.F., Selman, M., Pardo, A., Kim, D.S., Crapo, J.D., Make, B.J., Regan, E. A, Walek, D.S., Daniel, J.J., Kamatani, Y., Zelenika, D., Smith, K., McKean, D., Pedersen, B.S., Talbert, J., Kidd, R.N., Markin, C.R., Beckman, K.B., Lathrop, M., Schwarz, M.I. & Schwartz, D. A, 2013. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nature Genetics*, 45(6), pp.613–20.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J. & Punta, M., 2014. Pfam: The protein families database. *Nucleic Acids Research*, 42(D1), pp. 222-230.

Flaherty, K.R., Toews, G.B., Lynch, J.P., Kazerooni, E.A., Gross, B.H., Strawderman, R.L., Hariharan, K., Flint, A. & Martinez, F.J., 2001. Steroids in idiopathic pulmonary fibrosis: a prospective assessment of adverse reactions, response to therapy, and survival. *The American Journal of Medicine*, 110(4), pp.278–282.

Flanagan, K.L., 2014. Sexual dimorphism in biomedical research: A call to analyse by sex. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 108(7), pp.385–387.

Flatau, E., Gonzales, F. A., Michalowsky, L. A. & Jones, P. A., 1984. DNA methylation in 5-Aza-2'-deoxycytidine-resistant variants of C3H 10T1/2 C18 cells. *Molecular and Cellular Biology*, 4(10), pp.2098–2102.

Fonseca, C., Lindahl, G.E., Ponticos, M., Sestini, P., Renzoni, E.A., Holmes, A.M., Spagnolo, P., Pantelidis, P., Leoni, P., McHugh, N., Stock, C.J., Shi-Wen, X., Denton, C.P., Black, C.M., Welsh, K.I., du Bois, R.M. & Abraham, D.J., 2007. A polymorphism in the CTGF promoter region associated with systemic sclerosis. *The New England Journal of Medicine*, 357(12), pp.1210–1220.

Fonseca, C., Renzoni, E., Sestini, P., Pantelidis, P., Lagan, A., Bunn, C., McHugh, N., Welsh, K.I., Du Bois, R.M., Denton, C.P., Black, C. & Abraham, D., 2006. Endothelin axis polymorphisms in patients with scleroderma. *Arthritis and Rheumatism*, 54(9), pp.3034–3042.

Fortin, J., 2014. Functional normalization of 450k methylation array data improves replication in large cancer studies. *BioRxiv*, pp.0–42.

Fraga, M.F. & Esteller, M., 2007. Epigenetics and aging: the targets and the marks. *Trends in Genetics*, 23(8), pp.413–418.

Frankel, S.K., Cosgrove, G.P., Cha, S.I., Cool, C.D., Wynes, M.W., Edelman, B.L., Brown, K.K. & Riches, D.W.H., 2006. TNF-alpha sensitizes normal and fibrotic human lung fibroblasts to Fas-induced apoptosis. *American Journal of Respiratory Cell and Molecular Biology*, 34(3), pp.293–304.

Frantz, C., Stewart, K.M. & Weaver, V.M., 2010. The extracellular matrix at a glance. *Journal of Cell Science*, 123(Pt 24), pp.4195–4200.

Friedman, R.C., Farh, K.K.H., Burge, C.B. & Bartel, D.P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1), pp.92–105.

Fu, L., Ma, C., Cong, B., Li, S., Chen, H. & Zhang, J., 2011. Hypomethylation of proximal CpG motif of interleukin-10 promoter regulates its expression in human rheumatoid arthritis. *Acta Pharmacologica Sinica*, 32(11), pp.1373–1380.

Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., Miyazaki, T., Ogura, C., Okazaki, Y. & Jinno, Y., 2004. Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: An HPLC-based study. *Annals of Human Genetics*, 68(3), pp.196–204.

Fuks, F., 2005. DNA methylation and histone modifications: Teaming up to silence genes. *Current Opinion in Genetics and Development*, 15(5 SPEC. ISS.), pp.490–495.

Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L. & Kouzarides, T., 2000. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature Genetics*, 24(1), pp.88–91.

Fuks, F., Burgers, W.A., Godin, N., Kasai, M. & Kouzarides, T., 2001. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO Journal*, 20(10), pp.2536–2544.

Gao, F., Koenitzer, J.R., Tobolewski, J.M., Jiang, D., Liang, J., Noble, P.W. & Oury, T.D., 2008. Extracellular superoxide dismutase inhibits inflammation by preventing oxidative fragmentation of hyaluronan. *Journal of Biological Chemistry*, 283(10), pp.6058–6066.

García-de-Alba, C., Becerril, C., Ruiz, V., González, Y., Reyes, S., García-Alvarez, J., Selman, M. & Pardo, A., 2010. Expression of matrix metalloproteases by fibrocytes: Possible role in migration and homing. *American Journal of Respiratory and Critical Care Medicine*, 182(9), pp.1144–1152.

Garcia-Sancho Figueroa, M.C., Carrillo, G., Perez-Padilla, R., Fernandez-Plata, M.R., Buendia-Roldan, I., Vargas, M.H. & Selman, M., 2010. Risk factors for idiopathic pulmonary fibrosis in a Mexican population. A case-control study. *Respiratory Medicine*, 104(2), pp.305–309.

García-Sancho, C., Buendía-Roldán, I., Fernández-Plata, M.R., Navarro, C., Pérez-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(12), pp.1902–1907.

Garro, A.J., McBeth, D.L., Lima, V. & Lieber, C.S., 1991. Ethanol consumption inhibits fetal DNA methylation in mice: implications for the fetal alcohol syndrome. *Alcoholism, Clinical and Experimental Research*, 15(3), pp.395–398.

Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T. & Jacob, S.T., 2005. 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the ken box, bromo-adjacent homology domain, and nuclear localization signal. *Molecular and Cellular Biology*, 25(11), pp.4727–4741.

Giefing-Kröll, C., Berger, P., Lepperdinger, G. & Grubeck-Loebenstein, B., 2015. How sex and age affect immune responses, susceptibility to infections, and response to vaccination. *Aging Cell*, 14(3), pp.309–21.

Gilchrist, F.C., Bunn, C., Foley, P.J., Lympany, P.A., Black, C.M., Welsh, K.I. & du Bois, R.M., 2001. Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. *Genes and Immunity*, 2(2), pp.76–81.

Gladman, D.D., Kung, T.N., Siannis, F., Pellett, F., Farewell, V.T. & Lee, P., 2005. HLA markers for susceptibility and expression in scleroderma. *Journal of Rheumatology*, 32(8), pp.1481–1487.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E. & Bestor, T.H., 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science*, 311(5759), pp.395–398.

Gomez, D.E., Alonso, D.F., Yoshiji, H. & Thorgeirsson, U.P., 1997. Tissue inhibitors of metalloproteinases: structure, regulation and biological functions. *European Journal of Cell Biology*, 74(2), pp.111–122.

Gordon, M.K. & Hahn, R.A., 2010. Collagens. *Cell and Tissue Research*, 339(1), pp.247–257.

Gosain, A. & DiPietro, L.A., 2004. Aging and Wound Healing. *World Journal of Surgery*, 28(3), pp.321–326.

Gourh, P., Agarwal, S.K., Divecha, D., Assassi, S., Paz, G., Arora-Singh, R.K., Reveille, J.D., Shete, S., Mayes, M.D., Arnett, F.C. & Tan, F.K., 2009. Polymorphisms in TBX21 and STAT4 increase the risk of systemic sclerosis: Evidence of possible gene-gene interaction and alterations in Th1/Th2 cytokines. *Arthritis and Rheumatism*, 60(12), pp.3794–3806.

Gourh, P., Arnett, F.C., Tan, F.K., Assassi, S., Divecha, D., Paz, G., McNearney, T., Draeger, H., Reveille, J.D., Mayes, M.D. & Agarwal, S.K., 2010. Association of TNFSF4 (OX40L) polymorphisms with susceptibility to systemic sclerosis. *Annals of the Rheumatic Diseases*, 69(3), pp.550–555.

Grassegger, A., Schuler, G., Hessenberger, G., Walder-Hantich, B., Jabkowski, J., MacHeiner, W., Salmhofer, W., Zahel, B., Pinter, G., Herold, M., Klein, G. & Fritsch, P.O., 1998. Interferon-gamma in the treatment of systemic sclerosis: a randomized controlled multicentre trial. *The British Journal of Dermatology*, 139(4), pp.639-648.

Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C., Tsykin, A., Farshid, G., Vadas, M.A., Khew-Goodall, Y. & Goodall, G.J., 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biology*, 10(5), pp.593–601.

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(11), pp.980–985.

Grigoryev, D.N., Mathai, S.C., Fisher, M.R., Girgis, R.E., Zaiman, A.L., Housten-Harris, T., Cheadle, C., Gao, L., Hummers, L.K., Champion, H.C., Garcia, J.G.N., Wigley, F.M., Tuder, R.M., Barnes, K.C. & Hassoun, P.M., 2008. Identification of candidate genes in scleroderma-related pulmonary arterial hypertension. *Translational Research: The Journal of Laboratory and Clinical Medicine*, 151(4), pp.197–207.

Guéant, J.L., Namour, F., Guéant-Rodriguez, R.M. & Daval, J.L., 2013. Folate and fetal programming: A play in epigenomics? *Trends in Endocrinology and Metabolism*, 24(6), pp.279–289.

Guida, F., Sandanger, T.M., Castagné, R., Campanella, G., Polidoro, S., Palli, D., Krogh, V., Tumino, R., Sacerdote, C., Panico, S., Severi, G., Kyrtopoulos, S.A., Georgiadis, P., Vermeulen, R.C.H., Lund, E., Vineis, P. & Chadeau-Hyam, M., 2015. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Human Molecular Genetics,* 24 (8), pp.2349–2359.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L. & Gu, T., 2014. Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell*, 15, pp.1–12.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., Fan, X., Yong, J., Wen, L., Xie, S.X., Tang, F. & Qiao, J., 2014. The DNA methylation landscape of human early embryos. *Nature*, 511(7511), pp.606–610.

Guo, S. & Dipietro, L.A., 2010. Factors affecting wound healing. *Journal of Dental Research*, 89(3), pp.219–229.

Gustafson, T., Dahlman-Höglund, A., Nilsson, K., Ström, K., Tornling, G. & Torén, K., 2007. Occupational exposure and severe pulmonary fibrosis. *Respiratory Medicine*, 101(10), pp.2207–2212.

Habiel, D. & Hogaboam, C., 2014. Heterogeneity in fibroblast proliferation and survival in idiopathic pulmonary fibrosis. *Frontiers in Pharmacology*, 5, pp.5–10.

Hagimoto, N., Kuwano, K., Inoshima, I., Yoshimi, M., Nakamura, N., Fujita, M., Maeyama, T. & Hara, N., 2002. TGF-beta 1 as an enhancer of Fas-mediated apoptosis of lung epithelial cells. *Journal of Immunology*, 168(12), pp.6470–6478.

Hagood, J., 2014. Beyond the genome: epigenetic mechanisms in lung remodeling. *Physiology*, 29(3), pp.177–85.

Hagood, J.S., Prabhakaran, P., Kumbla, P., Salazar, L., MacEwen, M.W., Barker, T.H., Ortiz, L.A., Schoeb, T., Siegal, G.P., Alexander, C.B., Pardo, A. & Selman, M., 2005. Loss of fibroblast Thy-1 expression correlates with lung fibrogenesis. *The American Journal of Pathology*, 167(2), pp.365–379.

Haines, T.R., Rodenhiser, D.I. & Ainsworth, P.J., 2001. Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. *Developmental Biology*, 240(2), pp.585–598.

Hall, E., Volkov, P., Dayeh, T., Esguerra, J., Salö, S., Eliasson, L., Rönn, T., Bacos, K. & Ling, C., 2014. Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biology*, 15(12), p.522.

Halliday, N.L. & Tomasek, J.J., 1995. Mechanical properties of the extracellular matrix influence fibronectin fibril assembly in vitro. *Experimental Cell Research*, 217(1), pp.109–117.

Halsted, C.H., Villanueva, J., Chandler, C.J., Stabler, S.P., Allen, R.H., Muskhelishvili, L., James, S.J. & Poirier, L., 1996. Ethanol feeding of micropigs alters methionine metabolism and increases hepatocellular apoptosis and proliferation. *Hepatology,* 23(3), pp.497–505.

Halsted, C.H., Villanueva, J.A., Devlin, A.M. & Chandler, C.J., 2002. Metabolic interactions of alcohol and folate. *The Journal of Nutrition*, 132(8 Supplement), pp.2367S–2372S.

Hamman, L. & Rich, A.R., 1935. Fulminating diffuse interstitial fibrosis of the lungs. *Transactions of the American Clinical and Climatological Association*, 51, pp.154–163.

Hamman, L. & Rich, A.R., 1944. Acute diffuse interstitial fibrosis of the lungs. *Bulletin of the Johns Hopkins Hospital Journal*, 74, pp.177-212.

Han, H., Hu, J., Lau, M.Y., Feng, M., Petrovic, L.M. & Ji, C., 2013. Altered methylation and expression of ER-associated degradation factors in long-term alcohol and constitutive ER stress-induced murine hepatic tumors. *Frontiers in Genetics*, 4, p.224.

Han, M.K., Murray, S., Fell, C.D., Flaherty, K.R., Toews, G.B., Myers, J., Colby, T. V., Travis, W.D., Kazerooni, E.A., Gross, B.H. & Martinez, F.J., 2008. Sex differences in physiological progression of idiopathic pulmonary fibrosis. *European Respiratory Journal*, 31(6), pp.1183–1188.

Han, M.K., Zhou, Y., Murray, S., Tayob, N., Noth, I., Lama, V.N., Moore, B.B., White, E.S., Flaherty, K.R., Huffnagle, G.B. & Martinez, F.J., 2014. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: An analysis of the COMET study. *The Lancet Respiratory Medicine*, 2(7), pp.448–456.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T. & Zhang, K., 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49(2), pp.359–367.

Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M. & Gartler, S.M., 1999. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 96(25), pp.14412–14417.

Harlid, S., Xu, Z., Panduri, V., Sandler, D.P. & Taylor, J.A., 2014. CpG sites associated with cigarette smoking: Analysis of epigenome-wide data from the sister study. *Environmental Health Perspectives*, 122(7), pp.673–678.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C. & Xu, G.-L., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, N.Y.)*, 333(6047), pp.1303–1307.

Heard, E. & Avner, P., 1994. Role play in X-inactivation. *Human Molecular Genetics,* 3(1), pp.1481–1485.

Heino, J., 2007. The collagen family members as cell adhesion proteins. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 29(10), pp.1001–1010.

Henry, M.T., McMahon, K., Mackarel, A.J., Prikk, K., Sorsa, T., Maisi, P., Sepper, R., Fitzgerald, M.X. & O'Connor, C.M., 2002. Matrix metalloproteinases and tissue inhibitor of metalloproteinase-1 in sarcoidosis and IPF. *The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology*, 20(5), pp.1220–1227.

Hernando-Herraez, I., Prado-Martinez, J., Garg, P., Fernandez-Callejo, M., Heyn, H., Hvilsom, C., Navarro, A., Esteller, M., Sharp, A.J. & Marques-Bonet, T., 2013. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genetics*, 9(9), e1003763.

Herrera, I., Cisneros, J., Maldonado, M., Ramírez, R., Ortiz-Quintero, B., Anso, E., Chandel, N.S., Selman, M. & Pardo, A., 2013. Matrix metalloproteinase (MMP)-1 induces lung alveolar epithelial cell migration and proliferation, protects from apoptosis, and represses mitochondrial oxygen consumption. *Journal of Biological Chemistry*, 288(36), pp.25964–25975.

Herzog, E. L., Mathur, A., Tager, A. M., Feghali-Bostwick, C., Schneider, F., & Varga, J., 2014. Interstitial lung disease associated with systemic sclerosis and idiopathic pulmonary fibrosis: how similar and distinct? *Arthritis and Rheumatology*, 66(8), pp.1967–1978.

Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J. V., Setien, F., Carmona, F.J., Puca, A.A., Sayols, S., Pujana, M.A., Serra-Musach, J., Iglesias-Platas, I., Formiga, F., Fernandez, A.F., Fraga, M.F., Heath, S.C., Valencia, A., Gut, I.G., Wang, J. & Esteller, M., 2012. Distinct DNA methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences USA*, 109(26), pp.10522–10527.

Hilberg, F., Roth, G.J., Krssak, M., Kautschitsch, S., Sommergruber, W., Tontsch-Grunt, U., Garin-Chesa, P., Bader, G., Zoephel, A., Quant, J., Heckel, A. & Rettig, W.J., 2008. BIBF 1120: Triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy. *Cancer Research*, 68(12), pp.4774–4782.

Hisatomi, K., Sakamoto, N., Mukae, H., Hayashi, T., Amenomori, M., Ishimoto, H., Fujita, H., Ishii, H., Nakayama, S., Ishimatsu, Y. & Kohno, S., 2009. Elevated levels of tenascin-C in patients with cryptogenic organizing pneumonia. *Internal Medicine,* 48(17), pp.1501–1507.

Hissaria, P., Lester, S., Hakendorf, P., Woodman, R., Patterson, K., Hill, C., Ahern, M.J., Smith, M.D., Walker, J.G. & Roberts-Thomson, P.J., 2011. Survival in scleroderma: Results from the population-based South Australian Register. *Internal Medicine Journal*, 41(5), pp.381–390.

Ho, Y.Y., Lagares, D., Tager, A.M. & Kapoor, M., 2014. Fibrosis-a lethal component of systemic sclerosis. *Nature Reviews. Rheumatology*, 10(7), pp.390–402.

Hoff, C.R., Perkins, D.R. & Davidson, J.M., 1999. Elastin gene expression is upregulated during pulmonary fibrosis. *Connective Tissue Research*, 40(2), pp.145–153.

Holliday, R., 2006. Epigenetics: A historical overview. *Epigenetics*, 1(2), pp.76–80.

Homolka, J., 1987. Idiopathic pulmonary fibrosis: a historical review. *Canadian Medical Association Journal*, 137(11), pp.1003–1005.

Honda, N., Jinnin, M., Kajihara, I., Makino, T., Makino, K., Masuguchi, S., Fukushima, S., Okamoto, Y., Hasegawa, M., Fujimoto, M. & Ihn, H., 2012. TGF-β-mediated downregulation of microRNA-196a contributes to the constitutive upregulated type I collagen expression in scleroderma dermal fibroblasts. *Journal of Immunology*, 188(7), pp.3323–31.

Honda, N., Jinnin, M., Kira-Etoh, T., Makino, K., Kajihara, I., Makino, T., Fukushima, S., Inoue, Y., Okamoto, Y., Hasegawa, M., Fujimoto, M. & Ihn, H., 2013. MiR-150 down-regulation contributes to the constitutive type I collagen overexpression in scleroderma dermal fibroblasts via the induction of integrin β3. *American Journal of Pathology*, 182(1), pp.206–216.

Hotchkiss, D., 1948. The quantitative separation of purines, pyrimidines and nucleosides by paper chromatography. *The Journal of Biological Chemistry*, 175, pp.315-332.

Hoyles, R.K., Ellis, R.W., Wellsbury, J., Lees, B., Newlands, P., Goh, N.S.L., Roberts, C., Desai, S., Herrick, A.L., McHugh, N.J., Foley, N.M., Pearson, S.B., Emery, P., Veale, D.J., Denton, C.P., Wells, A.U., Black, C.M. & Du Bois, R.M., 2006. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis and Rheumatism*, 54(12), pp.3962–3970.

Hsu, E., Shi, H., Jordan, R.M., Lyons-Weiler, J., Pilewski, J.M. & Feghali-Bostwick, C.A., 2011. Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. *Arthritis and Rheumatism*, 63(3), pp.783–794.

Huang, D.W., Lempicki, R.A. & Sherman, B.T., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), pp.44–57.

Huang, J., Okuka, M., Lu, W., Tsibris, J.C.M., McLean, M.P., Keefe, D.L. & Liu, L., 2013. Telomere shortening and DNA damage of embryonic stem cells induced by cigarette smoke. *Reproductive Toxicology,* 35, pp.89–95.

Huang, S.K., Fisher, A.S., Scruggs, A.M., White, E.S., Hogaboam, C.M., Richardson, B.C. & Peters-Golden, M., 2010. Hypermethylation of PTGER2 confers prostaglandin E2 resistance in fibrotic fibroblasts from humans and mice. *The American Journal of Pathology*, 177(5), pp.2245–2255.

Huang, S.K., Scruggs, a M., Donaghy, J., Horowitz, J.C., Zaslona, Z., Przybranowski, S., White, E.S. & Peters-Golden, M., 2013. Histone modifications are responsible for decreased Fas expression and apoptosis resistance in fibrotic lung fibroblasts. *Cell Death & Disease*, 4(5), p.e621.

Huang, S.K., Scruggs, A.M., Mceachin, R.C., White, E.S. & Peters-golden, M., 2014. Lung fibroblasts from patients with idiopathic pulmonary fibrosis exhibit genome-wide differences in DNA methylation compared to fibroblasts from nonfibrotic lung. *PLoS ONE*, 9(9), e107055.

Huang, X., Yang, N., Fiore, V.F., Barker, T.H., Sun, Y., Morris, S.W., Ding, Q., Thannickal, V.J. & Zhou, Y., 2012. Matrix stiffness-induced myofibroblast differentiation is mediated by intrinsic mechanotransduction. *American Journal of Respiratory Cell and Molecular Biology*, 47(3), pp.340–348.

Huaux, F., Lasfargues, G., Lauwerys, R. & Lison, D., 1995. Lung toxicity of hard metal particles and production of interleukin-1, tumor necrosis factor-alpha, fibronectin, and cystatin-c by lung phagocytes. *Toxicology and Applied Pharmacology*, 132(1), pp.53–62.

Hubbard, R., Cooper, M., Antoniak, M., Venn, A., Khan, S., Johnston, I., Lewis, S. & Britton, J., 2000. Risk of cryptogenic fibrosing alveolitis in metal workers. *Lancet*, 355(9202), pp.466–467.

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(8997), pp.284–289.

Hudson, M., Lo, E., Lu, Y., Hercz, D., Baron, M. & Steele, R., 2011. Cigarette smoking in patients with systemic sclerosis. *Arthritis and Rheumatism*, 63(1), pp.230–238.

Hutchinson, J., Fogarty, A., Hubbard, R. & McKeever T., 2015. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *European Respiratory Journal*, 46(3), pp.795-806.

Hutyrová, B., Pantelidis, P., Drábek, J., Žůrková, M., Kolek, V., Lenhart, K., Welsh, K.I., Du Bois, R.M. & Petřek, M., 2002. Interleukin-1 gene cluster polymorphisms in sarcoidosis and idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 165(2), pp.148–151.

Huxley-Jones, J., Clarke, T.-K., Beck, C., Toubaris, G., Robertson, D.L. & Boot-Handford, R.P., 2007. The evolution of the vertebrate metzincins; insights from Ciona intestinalis and Danio rerio. *BMC Evolutionary Biology*, 7, p.63.

Huxley-Jones, J., Robertson, D.L. & Boot-Handford, R.P., 2007. On the origins of the extracellular matrix in vertebrates. *Matrix Biology*, 26(1), pp.2–11.

Hwang, H.-W. & Mendell, J.T., 2006. MicroRNAs in cell proliferation, cell death, and tumorigenesis. *British Journal of Cancer*, 94(6), pp.776–780.

Hynes, R.O. & Naba, A., 2012. Overview of the matrisome-An inventory of extracellular matrix constituents and functions. *Cold Spring Harbor Perspectives in Biology*, 4(1), a004903.

Ide, M., Ishii, H., Mukae, H., Iwata, A., Sakamoto, N., Kadota, J. ichi & Kohno, S., 2008. High serum levels of thrombospondin-1 in patients with idiopathic interstitial pneumonia. *Respiratory Medicine*, 102(11), pp.1625–1630.

Ideraabdullah, F.Y., Vigneau, S. & Bartolomei, M.S., 2008. Genomic imprinting mechanisms in mammals. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 647(1-2), pp.77–85.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R. & Bird, A.P., 2010. Orphan CpG Islands identify numerous conserved promoters in the mammalian genome. *PLoS Genetics*, 6(9), e1001134.

Inkinen, K., Soots, A., Krogerus, L., Loginov, R., Bruggeman, C. & Lautenschlager, I., 2005. Cytomegalovirus enhance expression of growth factors during the development of chronic allograft nephropathy in rats. *Transplant International: Official Journal of the European Society for Organ Transplantation*, 18(6), pp.743–749.

Inoshita, M., Numata, S., Tajima, A., Kinoshita, M., Umehara, H., Yamamori, H., Hashimoto, R., Imoto, I. & Ohmori, T., 2015. Sex differences of leukocytes DNA methylation adjusted for estimated cellular proportions. *Biology of Sex Differences*, 6(1), p.11.

Iozzo, R. V & Murdoch, A.D., 1996. Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 10(5), pp.598–614.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J.B., Sabunciyan, S. & Feinberg, A.P., 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, 41(2), pp.178–186.

Irving, W.L., Day, S. & Johnston, I.D., 1993. Idiopathic pulmonary fibrosis and hepatitis C virus infection. *The American Review of Respiratory Disease*, 148(6 Pt 1), pp.1683–1684.

Ishikawa, N., Hattori, N., Yokoyama, A. & Kohno, N., 2012. Utility of KL-6/MUC1 in the clinical management of interstitial lung diseases. *Respiratory Investigation*, 50(1), pp.3–13.

Isler, J.A., Skalet, A.H. & Alwine, J.C., 2005. Human cytomegalovirus infection activates and regulates the unfolded protein response. *Journal of Virology*, 79(11), pp.6890–6899.

Ito, I., Kawaguchi, Y., Kawasaki, A., Hasegawa, M., Ohashi, J., Hikami, K., Kawamoto, M., Fujimoto, M., Takehara, K., Sato, S., Hara, M. & Tsuchiya, N., 2009. Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. *Arthritis and Rheumatism*, 60(6), pp.1845–1850.

Ito, S., D'Alessio, A.C., Taranova, O. V., Hong, K., Sowers, L.C. & Zhang, Y., 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310), pp.1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. & Zhang, Y., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 333(6047), pp.1300–1303.

Iwai, K., Mori, T., Yamada, N., Yamaguchi, M. & Hosoda, Y., 1994. Idiopathic pulmonary fibrosis: Epidemiologic approaches to occupational exposure. *American Journal of Respiratory and Critical Care Medicine*, 150(3), pp.670–675.

Jackson, R.M., Glassberg, M.K., Ramos, C.F., Bejarano, P.A., Butrous, G. & Orlando Gómez-Marín, O., 2010. Sildenafil therapy and exercise tolerance in idiopathic pulmonary fibrosis. *Lung*, 188(2), pp.115–123.

Jacobsen, G., Schlünssen, V., Schaumburg, I., Taudorf, E. & Sigsgaard, T., 2008. Longitudinal lung function decline and wood dust exposure in the furniture industry. *European Respiratory Journal*, 31(2), pp.334–342.

Janoštiak, R., Pataki, A.C., Brábek, J. & Rösel, D., 2014. Mechanosensors in integrin signaling: The emerging role of p130Cas. *European Journal of Cell Biology*, 93, pp.445–454.

Järveläinen, H., Sainio, A., Koulu, M., Wight, T.N. & Penttinen, R., 2009. Extracellular matrix molecules: potential targets in pharmacotherapy. *Pharmacological Reviews*, 61(2), pp.198–223.

Javierre, B.M., Fernandez, A.F., Richter, J., Al-Shahrour, F., Ignacio Martin-Subero, J., Rodriguez-Ubreva, J., Berdasco, M., Fraga, M.F., O'Hanlon, T.P., Rider, L.G., Jacinto, F. V., Javier Lopez-Longo, F., Dopazo, J., Forn, M., Peinado, M.A., Carreño, L., Sawalha, A.H., Harley, J.B., Siebert, R., Esteller, M., Miller, F.W. & Ballestar, E., 2010. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Research*, 20(2), pp.170–179.

Jederlinic, P.J., Abraham, J.L., Churg, A., Himmelstein, J.S., Epler, G.R. & Gaensler, E.A., 1990. Pulmonary fibrosis in aluminum oxide workers. Investigation of nine workers, with pathologic examination and microanalysis in three of them. *The American Review of Respiratory Disease*, 142(5), pp.1179–1184.

Jenkins, G., 2013. Pirfenidone should be prescribed for patients with idiopathic pulmonary fibrosis. *Thorax*, 68(7), pp.603–5.

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S. & Jones, P.A., 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Molecular and Cellular Biology*, 29(19), pp.5366–5376.

Jiang, H., Xiao, R., Lian, X., Kanekura, T., Luo, Y., Yin, Y., Zhang, G., Yang, Y., Wang, Y., Zhao, M. & Lu, Q., 2012. Demethylation of TNFSF7 contributes to CD70 overexpression in CD4+ T cells from patients with systemic sclerosis. *Clinical Immunology*, 143(1), pp.39–44.

Jiang, Y., Liu, S., Chen, X., Cao, Y. & Tao, Y., 2013. Genome-wide distribution of DNA methylation and DNA demethylation and related chromatin regulators in cancer. *Biochimica et Biophysica Acta - Reviews on Cancer*, 1835(2), pp.155–163.

Jing, L., Zhou, L.J., Zhang, F.M., Li, W.M. & Sang, Y., 2011. Tenascin-x facilitates myocardial fibrosis and cardiac remodeling through transforming growth factor-β1 and peroxisome proliferator activated receptor γ in alcoholic cardiomyopathy. *Chinese Medical Journal*, 124(3), pp.390–395.

Johansen, J.S., Milman, N., Hansen, M., Garbarsch, C., Price, P.A. & Graudal, N., 2005. Increased serum YKL-40 in patients with pulmonary sarcoidosis--a potential marker of disease activity? *Respiratory Medicine*, 99(4), pp.396–402.

Johnston, I.D., Prescott, R.J., Chalmers, J.C. & Rudd, R.M., 1997. British Thoracic Society study of cryptogenic fibrosing alveolitis: current presentation and initial management. Fibrosing Alveolitis Subcommittee of the Research Committee of the British Thoracic Society. *Thorax*, 52(1), pp.38–44.

Jones, F.S. & Jones, P.L., 2000. The tenascin family of ECM glycoproteins: Structure, function, and regulation during embryonic development and tissue remodeling. *Developmental Dynamics*, 218(2), pp.235–259.

Jones, P.A. & Baylin, S.B., 2002. The fundamental role of epigenetic events in cancer. *Nature Reviews. Genetics*, 3(6), pp.415–428.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews. Genetics*, 13(7), pp.484–492.

Jowaed, A., Schmitt, I., Kaut, O. & Wüllner, U., 2010. Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains. *The Journal of Neuroscience*, 30(18), pp.6355–6359.

Jung, M. & Pfeifer, G.P., 2015. Aging and DNA methylation. *BMC Biology*, 13(1), pp.1–8.

Jüngel, A., Distler, J.H.W., Gay, S. & Distler, O., 2011. Epigenetic modifications: novel therapeutic strategies for systemic sclerosis? *Expert Review of Clinical Immunology*, 7(4), pp.475–480.

Kadler, K.E., Baldock, C., Bella, J. & Boot-Handford, R.P., 2007. Collagens at a glance. *Journal of Cell Science*, 120(Pt 12), pp.1955-1958.

Kadler, K.E., Hill, A. & Canty-Laird, E.G., 2008. Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators. *Current Opinion in Cell Biology*, 20(5), pp.495–501.

Kage, H. & Borok, Z., 2012. EMT and interstitial lung disease: a mysterious relationship. *Current Opinion in Pulmonary Medicine*, 18(5), pp.517–23.

Kahaleh, B., 2009. The microvascular endothelium in scleroderma. *Rheumatology*, 47(Supplement 5), pp.14-15.

Kähäri, V.M., Olsen, D.R., Rhudy, R.W., Carrillo, P., Chen, Y.Q. & Uitto, J., 1992. Transforming growth factor-beta up-regulates elastin gene expression in human skin fibroblasts. Evidence for post-transcriptional modulation. *Laboratory Investigation*, 66(5), pp.580–588.

Kajihara, I., Jinnin, M., Honda, N., Makino, K., Makino, T., Masuguchi, S., Sakai, K., Fukushima, S., Inoue, Y. & Ihn, H., 2013. Scleroderma dermal fibroblasts overexpress vascular endothelial growth factor due to autocrine transforming growth factor β signaling. *Modern Rheumatology / the Japan Rheumatism Association*, 23(3), pp.516–24.

Kajihara, I., Jinnin, M., Yamane, K., Makino, T., Honda, N., Igata, T., Masuguchi, S., Fukushima, S., Okamoto, Y., Hasegawa, M., Fujimoto, M. & Ihn, H., 2012. Increased accumulation of extracellular thrombospondin-2 due to low degradation activity stimulates type I collagen expression in scleroderma fibroblasts. *American Journal of Pathology*, 180(2), pp.703–714.

Kalabay, L., Fekete, B., Czirják, L., Horváth, L., Daha, M.R., Veres, A., Fónyad, G., Horváth, A., Viczián, Á., Singh, M., Hoffer, I., Füst, G., Romics, L. & Prohászka, Z., 2002. Helicobacter pylori infection in connective tissue disorders is associated with high levels of antibodies to mycobacterial hsp65 but not to human hsp60. *Helicobacter*, 7(4), pp.250–256.

Karpf, A.R., Moore, B.C., Ririe, T.O. & Jones, D.A., 2001. Activation of the p53 DNA damage response pathway after inhibition of DNA methyltransferase by 5-Aza-2'-deoxycytidine. *Molecular Pharmacology*, 59(4), pp.751–757.

Karrer, S., Bosserhoff, A.K., Weiderer, P., Distler, O., Landthaler, M., Szeimies, R.M., Müller-Ladner, U., Schölmerich, J. & Hellerbrand, C., 2005. The -2518 promotor polymorphism in the MCP-1 gene is associated with systemic sclerosis. *Journal of Investigative Dermatology*, 124(1), pp.92–98.

Kelly, B.G., Lok, S.S., Hasleton, P.S., Egan, J.J. & Stewart, J.P., 2002. A rearranged form of Epstein-Barr virus DNA is associated with idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 166(4), pp.510–513.

Kelsey, G., 2007. Genomic imprinting - roles and regulation in development. *Endocrine Development*, 12, pp.99–112.

Keshet, I., Lieman-Hurwitz, J. & Cedar, H., 1986. DNA methylation affects the formation of active chromatin. *Cell*, 44(4), pp.535–543.

Khalil, N., Churg, A., Muller, N. & O'Connor, R., 2007. Environmental, inhaled and ingested causes of pulmonary fibrosis. *Toxicologic Pathology*, 35(1), pp.86–96.

Khalil, N., O'Connor, R.N., Unruh, H.W., Warren, P.W., Flanders, K.C., Kemp, A., Bereznay, O.H. & Greenberg, A.H., 1991. Increased production and immunohistochemical localization of transforming growth factor-beta in idiopathic pulmonary fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, 5(2), pp.155–162.

Kiani, J., Grandjean, V., Liebers, R., Tuorto, F., Ghanbarian, H., Lyko, F., Cuzin, F. & Rassoulzadegan, M., 2013. RNA-mediated epigenetic heredity requires the cytosine methyltransferase DNMT2. *PLoS Genetics*, 9(5), e1003498.

Kilduff, C.E., Counter, M.J., Thomas, G. A., Harrison, N.K. & Hope-Gill, B.D., 2014. Effect of acid suppression therapy on gastroesophageal reflux and cough in idiopathic pulmonary fibrosis: an intervention study. *Cough*, 10(1), p.4.

Kim, S.H., Turnbull, J. & Guimond, S., 2011. Extracellular matrix and cell signalling: The dynamic cooperation of integrin, proteoglycan and growth factor receptor. *Journal of Endocrinology*, 209(2), pp.139–151.

Kim, Y.I., Logan, J.W., Mason, J.B. & Roubenoff, R., 1996. DNA hypomethylation in inflammatory arthritis: reversal with methotrexate. *The Journal of Laboratory and Clinical Medicine*, 128(2), pp.165–172.

Kimura, H. & Shiota, K., 2003. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *Journal of Biological Chemistry*, 278(7), pp.4806–4812.

King, T.E., 2005. Clinical advances in the diagnosis and therapy of the interstitial lung diseases. *American Journal of Respiratory and Critical Care Medicine*, 172(3), pp.268–279.

King, T.E., Albera, C., Bradford, W.Z., Costabel, U., Hormel, P., Lancaster, L., Noble, P.W., Sahn, S.A., Szwarcberg, J., Thomeer, M., Valeyre, D. & du Bois, R.M., 2009. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet*, 374(9685), pp.222–228.

King, T.E., Behr, J., Brown, K.K., Du Bois, R.M., Lancaster, L., De Andrade, J.A., Stähler, G., Leconte, I., Roux, S. & Raghu, G., 2008. BUILD-1: A randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 177(1), pp.75–81.

King, T.E., Brown, K.K., Raghu, G., Du Bois, R.M., Lynch, D.A., Martinez, F., Valeyre, D., Leconte, I., Morganti, A., Roux, S. & Behr, J., 2011. BUILD-3: A randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 184(1), pp.92–99.

King, T.E., Schwarz, M.I., Brown, K., Tooze, J.A., Colby, T. V., Waldron, J.A., Flint, A., Thurlbeck, W. & Cherniack, R.M., 2001. Idiopathic pulmonary fibrosis: Relationship between histopathologic features and mortality. *American Journal of Respiratory and Critical Care Medicine*, 164(6), pp.1025–1032.

Kinnula, V.L., Fattman, C.L., Tan, R.J. & Oury, T.D., 2005. Oxidative stress in pulmonary fibrosis: A possible role for redox modulatory therapy. *American Journal of Respiratory and Critical Care Medicine*, 172(4), pp.417–422.

Kirk, J.M., Heard, B.E., Kerr, I., Turner-Warwick, M. & Laurent, G.J., 1984. Quantitation of types I and III collagen in biopsy lung samples from patients with cryptogenic fibrosing alveolitis. *Collagen and Related Research*, 4(3), pp.169–182.

Klein, C.J., Botuyan, M.-V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C., Simon, M., Lander, C., Boardman, L.A., Cunningham, J.M., Smith, G.E., Litchy, W.J., Boes, B., Atkinson, E.J., Middha, S., B Dyck, P.J., Parisi, J.E., Mer, G., Smith, D.I. & Dyck, P.J., 2011. Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nature Genetics*, 43(6), pp.595–600.

Klein, S.L., 2012. Sex influences immune responses to viruses, and efficacy of prophylaxis and treatments for viral diseases. *Bioessays*, 34(12), pp.1050–1059.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T., Matsui, Y. & Kono, T., 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research*, 23(4), pp.616–627.

Kodama, T., Takehara, T., Hikita, H., Shimizu, S., Shigekawa, M., Tsunematsu, H., Li, W., Miyagi, T., Hosui, A., Tatsumi, T., Ishida, H., Kanto, T., Hiramatsu, N., Kubota, S., Takigawa, M., Tomimaru, Y., Tomokuni, A., Nagano, H., Doki, Y., Mori, M. & Hayashi, N., 2011. Increases in p53 expression induce CTGF synthesis by mouse and human hepatocytes and result in liver fibrosis in mice. *Journal of Clinical Investigation*, 121(8), pp.3343–3356.

Koenig, M., Joyal, F., Fritzler, M.J., Roussin, A., Abrahamowicz, M., Boire, G., Goulet, J.R., Rich, É., Grodzicky, T., Raymond, Y. & Senécal, J.L., 2008. Autoantibodies and microvascular damage are independent predictive factors for the progression of Raynaud's phenomenon to systemic sclerosis: A twenty-year prospective study of 586 patients, with validation of proposed criteria for early systemic sclerosi. *Arthritis and Rheumatism*, 58(12), pp.3902–3912.

Kohli, R.M. & Zhang, Y., 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, 502(7472), pp.472–9.

Kohno, N., Awaya, Y., Oyama, T., Yamakido, M., Akiyama, M., Inoue, Y., Yokoyama, A., Hamada, H., Fujioka, S. & Hiwada, K., 1993. KL-6, a mucin-like glycoprotein, in bronchoalveolar lavage fluid from patients with interstitial lung disease. *The American Review of Respiratory Disease*, 148(3), pp.637–642.

Komatsu, Y., Waku, T., Iwasaki, N., Ono, W., Yamaguchi, C. & Yanagisawa, J., 2012. Global analysis of DNA methylation in early-stage liver fibrosis. *BMC Medical Genomics*, 5, p.5.

Kondo, Y., Shen, L. & Issa, J.-P.J., 2003. Critical role of histone methylation in tumor suppressor gene silencing in colorectal cancer. *Molecular and Cellular Biology*, 23(1), pp.206–215.

Königshoff, M. & Eickelberg, O., 2010. WNT signaling in lung disease: A failure or a regeneration signal? *American Journal of Respiratory Cell and Molecular Biology*, 42(1), pp.21–31.

Königshoff, M., Kramer, M., Balsara, N., Wilhelm, J., Amarie, O.V., Jahn, A., Rose, F., Fink, L., Seeger, W., Schaefer, L., Günther, A. & Eickelberg, O., 2009. WNT1-inducible signaling protein-1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis. *Journal of Clinical Investigation*, 119(4), pp.772–787.

Konishi, K., Gibson, K.F., Lindell, K.O., Richards, T.J., Zhang, Y., Dhir, R., Bisceglia, M., Gilbert, S., Yousem, S.A., Jin, W.S., Dong, S.K. & Kaminski, N., 2009. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 180(2), pp.167–175.

Kono, M., Nakamura, Y., Suda, T., Kato, M., Kaida, Y., Hashimoto, D., Inui, N., Hamada, E., Miyazaki, O., Kurashita, S., Fukamachi, I., Endo, K., Ng, P.S., Takehara, K., Nakamura, H., Maekawa, M. & Chida, K., 2011. Plasma CCN2 (connective tissue growth factor; CTGF) is a potential biomarker in idiopathic pulmonary fibrosis (IPF). *Clinica Chimica Acta*, 412(23-24), pp.2211–2215.

Korfei, M., Ruppert, C., Mahavadi, P., Henneke, I., Markart, P., Koch, M., Lang, G., Fink, L., Bohle, R.M., Seeger, W., Weaver, T.E. & Guenther, A., 2008. Epithelial endoplasmic reticulum stress and apoptosis in sporadic idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 178(8), pp.838–846.

Korthagen, N.M., Van Moorsel, C.H.M., Barlo, N.P., Ruven, H.J.T., Kruit, A., Heron, M., Van Den Bosch, J.M.M. & Grutters, J.C., 2011. Serum and BALF YKL-40 levels are predictors of survival in idiopathic pulmonary fibrosis. *Respiratory Medicine*, 105(1), pp.106–113.

Korthagen, N.M., Van Moorsel, C.H.M., Kazemier, K.M., Ruven, H.J.T. & Grutters, J.C., 2012. IL1RN genetic variations and risk of IPF: A meta-analysis and mRNA expression study. *Immunogenetics*, 64(5), pp.371–377.

Kowal-Bielecka, O., Landewé, R., Avouac, J., Chwiesko, S., Miniati, I., Czirjak, L., Clements, P., Denton, C., Farge, D., Fligelstone, K., Földvari, I., Furst, D.E., Müller-Ladner, U., Seibold, J., Silver, R.M., Takehara, K., Toth, B.G., Tyndall, A., Valentini, G., van den Hoogen, F., Wigley, F., Zulian, F. & Matucci-Cerinic, M., 2009. EULAR recommendations for the treatment of systemic sclerosis: a report from the EULAR Scleroderma Trials and Research group (EUSTAR). *Annals of the Rheumatic Diseases*, 68(5), pp. 620–628.

Kozomara, A. & Griffiths-Jones, S., 2011. MiRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, 39(1), pp.D152-D157.

Kristensen, J.H., Karsdal, M. A., Sand, J.M., Willumsen, N., Diefenbach, C., Svensson, B., Hägglund, P. & Oersnes-Leeming, D.J., 2015. Serological assessment of neutrophil elastase activity on elastin during lung ECM remodeling. *BMC Pulmonary Medicine*, 15(1), pp.1–7.

Kropski, J. A., Lawson, W.E., Young, L.R. & Blackwell, T.S., 2013. Genetic studies provide clues on the pathogenesis of idiopathic pulmonary fibrosis. *Disease Models and Mechanisms*, 6(1), pp.9–17.

Kuhn, C. & Mason, R.J., 1995. Immunolocalization of SPARC, tenascin, and thrombospondin in pulmonary fibrosis. *The American Journal of Pathology*, 147(6), pp.1759–1769.

Kuhn, C. & McDonald, J.A., 1991. The roles of the myofibroblast in idiopathic pulmonary fibrosis. Ultrastructural and immunohistochemical features of sites of active extracellular matrix synthesis. *The American Journal of Pathology*, 138(5), pp.1257–1265.

Kulis, M., Queirós, A.C., Beekman, R. & Martín-Subero, J.I., 2013. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(11), pp.1161–1174.

Kuwano, K., Nomoto, Y., Kunitake, R., Hagimoto, N., Matsuba, T., Nakanishi, Y. & Hara, N., 1997. Detection of adenovirus E1A DNA in pulmonary fibrosis using nested polymerase chain reaction. *The European Respiratory Journal*, 10(7), pp.1445–1449.

Lakatos, H.F., Thatcher, T.H., Kottmann, R.M., Garcia, T.M., Phipps, R.P. & Sime, P.J., 2007. The role of PPARs in lung fibrosis. *PPAR Research*, 71323.

Lakos, G., Takagawa, S. & Varga, J., 2004. Animal models of scleroderma. *Methods in Molecular Medicine*, 102, pp.377–393.

Lam, A.P. & Gottardi, C.J., 2011. β-catenin signaling: a novel mediator of fibrosis and potential therapeutic target. *Current Opinion in Rheumatology*, 23(6), pp.562-567.

Lambert, N.C., Distler, O., Müller-Ladner, U., Tylee, T.S., Furst, D.E. & Nelson, J.L., 2000. HLA-DQA1*0501 is associated with diffuse systemic sclerosis in Caucasian men. *Arthritis and Rheumatism*, 43(9), pp.2005–2010.

Lamouille, S., Xu, J. & Derynck, R., 2014. Molecular mechanisms of epithelial-mesenchymal transition. *Nature Reviews. Molecular Cell Biology*, 15(3), pp.178–96.

Larue, L. & Bellacosa, A., 2005. Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways. *Oncogene*, 24(50), pp.7443–7454.

Laurent, G.J & Tetley, T.D., 1984. Pulmonary fibrosis and emphysema: connective tissue disorders of the lung. *European Journal of Clinical Investigation*, 14(6), pp.411-413.

Laurent, G.J. & McAnulty, R.J., 1983. Protein metabolism during bleomycin-induced pulmonary fibrosis in rabbits. In vivo evidence for collagen accumulation because of increased synthesis and decreased degradation of the newly synthesized collagen. *The American Review of Respiratory Disease*, 128(1), pp.82–88.

Laurent, G.J., 1982. Rates of collagen synthesis in lung, skin and muscle obtained in vivo by a simplified method using ^{3H} proline. *The Biochemical Journal*, 206(3), pp.535–544.

Laurent, G.J., Harrison, N.K. & McAnulty, R.J., 1988. The regulation of collagen production in normal lung and during interstitial lung disease. *Postgraduate Medical Journal*, 64(Supplement 4), pp.26-34.

Laurila, J.P., Laatikainen, L.E., Castellone, M.D. & Laukkanen, M.O., 2009. SOD3 reduces inflammatory cell migration by regulating adhesion molecule and cytokine expression. *PLoS ONE*, 4(6), pp.e5786

Lawson, W.E. & Loyd, J.E., 2006. The genetic approach in pulmonary fibrosis: can it provide clues to this complex disease? *Proceedings of the American Thoracic Society*, 3(4), pp.345–349.

Lawson, W.E., Cheng, D.-S., Degryse, A.L., Tanjore, H., Polosukhin, V. V., Xu, X.C., Newcomb, D.C., Jones, B.R., Roldan, J., Lane, K.B., Morrisey, E.E., Beers, M.F., Yull, F.E. & Blackwell, T.S., 2011. Endoplasmic reticulum stress enhances fibrotic remodeling in the lungs. *Proceedings of the National Academy of Sciences of the United States of America*, 108(26), pp.10562–10567.

Lawson, W.E., Crossno, P.F., Polosukhin, V. V, Roldan, J., Cheng, D.-S., Lane, K.B., Blackwell, T.R., Xu, C., Markin, C., Ware, L.B., Miller, G.G., Loyd, J.E. & Blackwell, T.S., 2008. Endoplasmic reticulum stress in alveolar epithelial cells is prominent in IPF: association with altered surfactant protein processing and herpes virus infection. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 294(6), pp.L1119–L1126.

Le Guern, V., Mahr, A., Mouthon, L., Jeanneret, D., Carzon, M. & Guillevin, L., 2004. Prevalence of systemic sclerosis in a French multi-ethnic county. *Rheumatology*, 43(9), pp.1129–1137.

Leask, A. & Abraham, D.J., 2004. TGF-beta signaling and the fibrotic response. T*he Federation of American Societies for Experimental Biology*, 18(7), pp.816–827.

Leask, A. & Abraham, D.J., 2006. All in the CCN family: essential matricellular signaling modulators emerge from the bunker. *Journal of Cell Science*, 119(Pt 23), pp.4803–4810.

Lederer, D.J., Caplan-Shaw, C.E., O'Shea, M.K., Wilt, J.S., Basner, R.C., Bartels, M.N., Sonett, J.R., Arcasoy, S.M. & Kawut, S.M., 2006. Racial and ethnic disparities in survival in lung transplant candidates with idiopathic pulmonary fibrosis. *American Journal of Transplantation*, 6(2), pp.398–403.

Lee, J.S., Collard, H.R., Raghu, G., Sweet, M.P., Hays, S.R., Campos, G.M., Golden, J.A. & King, T.E., 2010. Does chronic microaspiration cause idiopathic pulmonary fibrosis? *American Journal of Medicine*, 123(4), pp.304–311.

Lee, K.W.K. & Pausova, Z., 2013. Cigarette smoking and DNA methylation. *Frontiers in Epigenomics and Epigenetics*, 4, p.132.

Lei, W., Luo, Y., Yan, K., Zhao, S., Li, Y., Qiu, X., Zhou, Y., Long, H., Zhao, M., Liang, Y., Su, Y. & Lu, Q., 2009. Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. *Scandinavian Journal of Rheumatology*, 38(5), pp.369–74.

Lemjabbar, H., Gosset, P., Lechapt-Zalcman, E., Franco-Montoya, M.L., Wallaert, B., Harf, A. & Lafuma, C., 1999. Overexpression of alveolar macrophage gelatinase B (MMP-9) in patients with idiopathic pulmonary fibrosis effects of steroid and immunosuppressive treatment. *American Journal of Respiratory Cell and Molecular Biology*, 20(5), pp.903–913.

LeRoy, E.C. & Medsger, J., 2001. Criteria for the classification of early systemic sclerosis. *Journal of Rheumatology*, 28(7), pp.1573–1576.

LeRoy, E.C., 1996. Systemic sclerosis. A vascular perspective. *Rheumatic Diseases Clinics of North America*, 22(4), pp.675–694.

LeRoy, E.C., Smith, E.A., Kahaleh, M.B., Trojanowska, M. & Silver, R.M., 1989. A strategy for determining the pathogenesis of systemic sclerosis. Is transforming growth factor beta the answer? *Arthritis and Rheumatism*, 32(7), pp.817-825.

Lethias, C., Carisey, A., Comte, J., Cluzel, C. & Exposito, J.Y., 2006. A model of tenascin-X integration within the collagenous network. *FEBS Letters*, 580(26), pp.6281–6285.

Lewis, T., 1929. Experiments relating to the peripheral mechanism involved in spasmodic arrest of the circulation in the fingers. A variety of Raynaud's disease. *Heart*, 15, pp.7–101.

Ley, B., & Collard, H. R., 2013. Epidemiology of idiopathic pulmonary fibrosis. *Clinical Epidemiology*, 5, pp.483–492.

Li, E., 2002. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews. Genetics*, 3(9), pp.662–673.

Li, E., Beard, C. & Jaenisch, R., 1993. Role for DNA methylation in genomic imprinting. *Nature*, 366(6453), pp.362–365.

Li, E., Bestor, T.H. & Jaenisch, R., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6), pp.915–926.

Li, H., Yang, R., Fan, X., Gu, T., Zhao, Z., Chang, D., Wang, W. & Wang, C., 2012. MicroRNA array analysis of microRNAs related to systemic scleroderma. *Rheumatology International*, 32(2), pp.307–313.

Li, P., Zhao, G.-Q., Chen, T.-F., Chang, J.-X., Wang, H.-Q., Chen, S.-S. & Zhang, G.-J., 2013. Serum miR-21 and miR-155 expression in idiopathic pulmonary fibrosis. *The Journal of Asthma*, 50(9), pp.960–964.

Lian, X., Xiao, R., Hu, X., Kanekura, T., Jiang, H., Li, Y., Wang, Y., Yang, Y., Zhao, M. & Lu, Q., 2012. DNA demethylation of CD40L in CD4+ T cells from women with systemic sclerosis: A possible explanation for female susceptibility. *Arthritis and Rheumatism*, 64(7), pp.2338–2345.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W. & Jones, P.A., 2002. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Molecular and Cellular Biology*, 22(2), pp.480–491.

Liang, H., Xu, C., Pan, Z., Zhang, Y., Xu, Z., Chen, Y., Li, T., Li, X., Liu, Y., Huangfu, L., Lu, Y., Zhang, Z., Yang, B., Gitau, S., Lu, Y., Shan, H. & Du, Z., 2014. The antifibrotic effects and mechanisms of microRNA-26a action in idiopathic pulmonary fibrosis. *Molecular Therapy*, 22(6), pp.1122–33.

Lin, S.M., Du, P., Huber, W. & Kibbe, W. a., 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research*, 36(2), pp.1–9.

Lindahl, G.E., Chambers, R.C., Papakrivopoulou, J., Dawson, S.J., Jacobsen, M.C., Bishop, J.E. & Laurent, G.J., 2002. Activation of fibroblast procollagen alpha 1(I) transcription by mechanical strain is transforming growth factor-beta-dependent and involves increased binding of CCAAT-binding factor (CBF/NF-Y) at the proximal promoter. *Journal of Biological Chemistry*, 277(8), pp.6153–6161.

Lino Cardenas, C.L., Henaoui, I.S., Courcot, E., Roderburg, C., Cauffiez, C., Aubert, S., Copin, M.C., Wallaert, B., Glowacki, F., Dewaeles, E., Milosevic, J., Maurizio, J., Tedrow, J., Marcet, B., Lo-Guidice, J.M., Kaminski, N., Barbry, P., Luedde, T., Perrais, M., Mari, B. & Pottier, N., 2013. miR-199a-5p Is Upregulated during Fibrogenic Response to Tissue Injury and Mediates TGFbeta-Induced Lung Fibroblast Activation by Targeting Caveolin-1. *PLoS Genetics*, 9(2), e10003291.

Lison, D., Lauwerys, R., Demedts, M. & Nemery, B., 1996. Experimental research into the pathogenesis of cobalt/hard metal lung disease. *European Respiratory Journal*, 9(5), pp.1024–1028.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B. & Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), pp.315–322.

Liu, C., Li, Y., Semenov, M., Han, C., Baeg, G.H., Tan, Y., Zhang, Z., Lin, X. & He, X., 2002. Control of β-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell*, 108(6), pp.837–847.

Liu, F., Lagares, D., Choi, K.M., Stopfer, L., Marinković, A., Vrbanac, V., Probst, C.K., Hiemer, S.E., Sisson, T.H., Horowitz, J.C., Rosas, I.O., Fredenburgh, L.E., Feghali-Bostwick, C., Varelas, X., Tager, A.M. & Tschumperlin, D.J., 2015. Mechanosignaling through YAP and TAZ drives fibroblast activation and fibrosis. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 308(4), pp.L344–L357.

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. *Journal of Cell Biology*, 190(4), pp.693–706.

Liu, G., Friggeri, A., Yang, Y., Milosevic, J., Ding, Q., Thannickal, V.J., Kaminski, N. & Abraham, E., 2010. miR-21 mediates fibrogenic activation of pulmonary fibroblasts and lung fibrosis. *The Journal of Experimental Medicine*, 207(8), pp.1589–1597.

Liu, J., Morgan, M., Hutchison, K. & Calhoun, V.D., 2010. A study of the influence of sex on genome wide methylation. *PLoS One*, 5(4), p.e10028.

Liu, Q., Liu, L., Zhao, Y., Zhang, J., Wang, D., Chen, J., He, Y., Wu, J., Zhang, Z. & Liu, Z., 2011. Hypoxia induces genomic DNA demethylation through the activation of HIF-1α and transcriptional upregulation of MAT2A in hepatoma cells. *Molecular Cancer Therapeutics*, 10(6), pp.1113–1123.

Liu, T., Dhanasekaran, S.M., Jin, H., Hu, B., Tomlins, S.A., Chinnaiyan, A.M. & Phan, S.H., 2004. FIZZ1 stimulation of myofibroblast differentiation. *The American Journal of Pathology*, 164(4), pp.1315–1326.

Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4), pp.402–408.

Lok, S.S., Stewart, J.P., Kelly, B.G., Hasleton, P.S. & Egan, J.J., 2001. Epstein-Barr virus and wild p53 in idiopathic pulmonary fibrosis. *Respiratory Medicine*, 95(10), pp.787–91.

Lokk, K., Vooder, T., Kolde, R., Välk, K., Võsa, U., Roosipuu, R., Milani, L., Fischer, K., Koltsina, M., Urgard, E., Annilo, T., Metspalu, A. & Tõnisson, N., 2012. Methylation markers of early-stage non-small cell lung cancer. *PLoS ONE*, 7(6), e39813.

López-Isac, E., Bossini-Castillo, L., Simeon, C.P., Egurbide, M.V., Alegre-Sancho, J.J., Callejas, J.L., Roman-Ivorra, J.A., Freire, M., Beretta, L., Santaniello, A., Airó, P., Lunardi, C., Hunzelmann, N., Riemekasten, G., Witte, T., Kreuter, A., Distler, J.H.W., Schuerwegh, A.J., Vonk, M.C., Voskuyl, A.E., Shiels, P.G., van Laar, J.M., Fonseca, C., Denton, C., Herrick, A., Worthington, J., Assassi, S., Koeleman, B.P., Mayes, M.D., Radstake, T.R.D.J. & Martin, J., 2014. A genome-wide association study follow-up suggests a possible role for PPARG in systemic sclerosis susceptibility. *Arthritis Research and Therapy*, 16(1), pp.R6–R6.

Loyd, J.E., 2003. Pulmonary fibrosis in families. *American Journal of Respiratory Cell and Molecular Biology*, 29(Supplement 3), S47-50.

Lu, P., Weaver, V.M. & Werb, Z., 2012. The extracellular matrix: A dynamic niche in cancer progression. *Journal of Cell Biology*, 196(4), pp.395–406.

Lu, Q., Kaplan, M., Ray, D., Ray, D., Zacharek, S., Gutsch, D. & Richardson, B., 2002. Demethylation of ITGAL (CD11a) regulatory sequences in systemic lupus erythematosus. *Arthritis and Rheumatism*, 46(5), pp.1282–1291.

Lu, Q., Wu, A., Tesmer, L., Ray, D., Yousif, N. & Richardson, B., 2007. Demethylation of CD40LG on the inactive X in T cells from women with lupus. *Journal of Immunology*, 179(9), pp.6352–6358.

Lu, S.C., Huang, Z.Z., Yang, H., Mato, J.M., Avila, M.A. & Tsukamoto, H., 2000. Changes in methionine adenosyltransferase and S-adenosylmethionine homeostasis in alcoholic rat liver. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, 279(1), pp.G178–G185.

Lunardi, C., Bason, C., Navone, R., Millo, E., Damonte, G., Corrocher, R. & Puccetti, A., 2000. Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells. *Nature Medicine*, 6(10), pp.1183–1186.

Lunardi, C., Dolcino, M., Peterlana, D., Bason, C., Navone, R., Tamassia, N., Beri, R., Corrocher, R. & Puccetti, A., 2006. Antibodies against human cytomegalovirus in the pathogenesis of systemic sclerosis: A gene array approach. *PLoS Medicine*, 3(1), pp.94–108.

Lund, K., Cole, J.J., VanderKraats, N.D., McBryan, T., Pchelintsev, N. a, Clark, W., Copland, M., Edwards, J.R. & Adams, P.D., 2014. DNMT inhibitors reverse a specific signature of aberrant promoter DNA methylation and associated gene silencing in AML. *Genome Biology*, 15(8), p.406.

Lupi-Herrera, E., Seoane, M., Verdejo, J., Gomez, A., Sandoval, J., Barrios, R. & Martinez, W., 1985. Hemodynamic effect of hydralazine in interstitial lung disease patients with cor pulmonale. Immediate and short-term evaluation at rest and during exercise. *Chest*, 87(5), pp.564–573.

Lyon, M.F., 1962. Sex chromatin and gene action in the mammalian X-chromosome. *American Journal of Human Genetics*, 14, pp.135–148.

MacNee, W., 2005. Pathogenesis of chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 2(4), pp.258–266; discussion 290–291.

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. *American Journal of Respiratory and Critical Care Medicine*, 182(1), pp.73–82.

Maher, T.M., Wells, A.U. & Laurent, G.J., 2007. Idiopathic pulmonary fibrosis: multiple causes and multiple mechanisms? *The European Respiratory Journal*, 30(5), pp.835–9.

Maitra, M., Wang, Y., Gerard, R.D., Mendelson, C.R. & Garcia, C.K., 2010. Surfactant protein A2 mutations associated with pulmonary fibrosis lead to protein instability and endoplasmic reticulum stress. *Journal of Biological Chemistry*, 285(29), pp.22103–22113.

Makino, K., Jinnin, M., Hirano, A., Yamane, K., Eto, M., Kusano, T., Honda, N., Kajihara, I., Makino, T., Sakai, K., Masuguchi, S., Fukushima, S. & Ihn, H., 2013. The downregulation of microRNA let-7a contributes to the excessive expression of type I collagen in systemic and localized scleroderma. *Journal of Immunology*, 190(8), pp.3905–15.

Manika, K., Alexiou-Daniel, S., Papakosta, D., Papa, A., Kontakiotis, T., Patakas, D. & Antoniadis, A., 2007. Epstein-Barr virus DNA in bronchoalveolar lavage fluid from patients with idiopathic pulmonary fibrosis. *Sarcoidosis, Vasculitis, and Diffuse Lung Diseases,* 24(2), pp.134–140.

Mann, J., Chu, D.C.K., Maxwell, A., Oakley, F., Zhu, N.L., Tsukamoto, H. & Mann, D.A., 2010. MeCP2 controls an epigenetic pathway that promotes myofibroblast transdifferentiation and fibrosis. *Gastroenterology*, 138(2), pp.705-714.

Mannino, D.M., Etzel, R.A. & Gibson Parrish, R., 1996. Pulmonary fibrosis deaths in the United States, 1979-1991: An analysis of multiple-cause mortality data. *American Journal of Respiratory and Critical Care Medicine*, 153(5), pp.1548–1552.

Mao, J.R., Taylor, G., Dean, W.B., Wagner, D.R., Afzal, V., Lotz, J.C., Rubin, E.M. & Bristow, J., 2002. Tenascin-X deficiency mimics Ehlers-Danlos syndrome in mice through alteration of collagen deposition. *Nature Genetics*, 30(4), pp.421–425.

Mao, X., Cai, T., Olyarchuk, J.G. & Wei, L., 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*, 21(19), pp.3787–3793.

Margaron, Y., Bostan, L., Exposito, J.Y., Malbouyres, M., Trunfio-Sfarghiu, A.M., Berthier, Y. & Lethias, C., 2010. Tenascin-X increases the stiffness of collagen gels without affecting fibrillogenesis. *Biophysical Chemistry*, 147(1-2), pp.87–91.

Maricq, H.R., Weinrich, M.C., Keil, J.E., Smith, E.A., Harper, F.E., Nussbaum, A.I., LeRoy, E.C., McGregor, A.R., Diat, F. & Rosal, E.J., 1989. Prevalence of scleroderma spectrum disorders in the general population of South Carolina. *Arthritis Rheumatism*, 32, pp.998–1006.

Marinković, A., Liu, F. & Tschumperlin, D.J., 2013. Matrices of physiologic stiffness potently inactivate idiopathic pulmonary fibrosis fibroblasts. *American Journal of Respiratory Cell and Molecular Biology*, 48(4), pp.422–430.

Markiewicz, M., Smith, E.A., Rubinchik, S., Dong, J.Y., Trojanowska, M. & LeRoy, E.C., 2004. The 72-kilodalton IE-1 protein of human cytomegalovirus (HCMV) is a potent inducer of connective tissue growth factor (CTGF) in human dermal fibroblasts. *Clinical and Experimental Rheumatology*, 22(Supplement 33).

Marquez, V.E., Liu, P.S., Kelley, J.A., Driscoll, J.S. & McCormack, J.J., 1980. Synthesis of 1,3-diazepin-2-one nucleosides as transition-state inhibitors of cytidine deaminase. *Journal of Medicinal Chemistry*, 23(7), pp.713–715.

Marshall, R.P., Puddicombe, A., Cookson, W.O. & Laurent, G.J., 2000. Adult familial cryptogenic fibrosing alveolitis in the United Kingdom. *Thorax*, 55(2), pp.143–146.

Martinez, F.J., Safrin, S., Weycker, D., Starko, K.M., Bradford, W.Z., King Jr., T.E., Flaherty, K.R., Schwartz, D.A., Noble, P.W., Raghu, G., Brown, K.K. & Group, I.P.F.S., 2005. The clinical course of patients with idiopathic pulmonary fibrosis. *Annals of Internal Medicine*, 142, pp.963–967.

Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P.D. & Rogers, J., 2010. Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation. *Neurobiology of Aging*, 31(12), pp.2025–2037.

Mastronardi, F.G., Noor, A., Wood, D.D., Paton, T. & Moscarello, M.A., 2007. Peptidyl argininedeiminase 2 CpG island in multiple sclerosis white matter is hypomethylated. *Journal of Neuroscience Research*, 85(9), pp.2006–2016.

Mastruzzo, C., Crimi, N. & Vancheri, C., 2002. Role of oxidative stress in pulmonary fibrosis. *Monaldi Archives for Chest Disease*, 57(3-4), pp.173-176.

Masucci, M.G., Contreras-salazar, B., Ragnar, E.V. a, Falk, K., Minarovits, J., Ernberg, I. & Klein, G., 1989. 5-Azacytidine up-regulates the expression of Epstein-Barr Virus nuclear antigen 2 (EBNA-2) through EBNA-6 and latent membrane protein in the Burkitt's lymphoma line rael. *Journal of Virology*, 2(7), pp.3135–3141.

Matsumoto, L., Takuma, H., Tamaoka, A., Kurisaki, H., Date, H., Tsuji, S. & Iwata, A., 2010. CpG demethylation enhances alpha-synuclein expression and affects the pathogenesis of Parkinson's disease. *PLoS ONE*, 5(11), pp.e15522.

Mattuzzi, S., Barbi, S., Carletto, A., Ravagnani, V., Moore, P.S., Bambara, L.M. & Scarpa, A., 2007. Association of polymorphisms in the IL1B and IL2 genes with susceptibility and severity of systemic sclerosis. *Journal of Rheumatology*, 34(5), pp.997–1004.

Matute-Bello, G., Wurfel, M.M., Lee, J.S., Park, D.R., Frevert, C.W., Madtes, D.K., Shapiro, S.D. & Martin, T.R., 2007. Essential role of MMP-12 in fas-induced lung fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, 37(2), pp.210–221.

Maurer, B., Stanczyk, J., Jüngel, A., Akhmetshina, A., Trenkmann, M., Brock, M., Kowal-Bielecka, O., Gay, R.E., Michel, B.A., Distler, J.H.W., Gay, S. & Distler, O., 2010. MicroRNA-29, a key regulator of collagen expression in systemic sclerosis. *Arthritis and Rheumatism*, 62(6), pp.1733–1743.

Maury, E., Ramsey, K.M. & Bass, J., 2010. Circadian rhythms and metabolic syndrome: From experimental genetics to human disease. *Circulation Research*, 106(3), pp.447–462.

Mayes, M.D., Lacey, J. V., Beebe-Dimmer, J., Gillespie, B.W., Cooper, B., Laing, T.J. & Schottenfeld, D., 2003. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis and Rheumatism*, 48(8), pp.2246–2255.

Mays, P.K., McAnulty, R.J. & Laurent, G.J., 1989. Age-related changes in lung collagen metabolism. A role for degradation in regulating lung collagen production. *The American Review of Respiratory Disease*, 140(2), pp.410–416.

Mc Cormic, Z.D., Khuder, S.S., Aryal, B.K., Ames, A.L. & Khuder, S.A., 2010. Occupational silica exposure as a risk factor for scleroderma: A meta-analysis. *International Archives of Occupational and Environmental Health*, 83(7), pp.763–769.

McAnulty, R.J. & Laurent, G.J., 1987. Collagen synthesis and degradation in vivo. Evidence for rapid rates of collagen turnover with extensive degradation of newly synthesized collagen in tissues of the adult rat. *Collagen and Related Research*, 7(2), pp.93–104.

McAnulty, R.J. & Laurent, G.J., 1995. Pathogenesis of lung fibrosis and potential new therapeutic strategies. Experimental Nephrology, 3(2), pp.96-107.

McAnulty, R.J. & Laurent, G.J., 2002. Fibroblasts. In: Barnes, P., Drazen, J., Rennard, S., Thomson, N. & editors. Asthma and COPD: basic mechanisms and clinical management. *London: Academic Press*; pp.139–144.

McAnulty, R.J., 2007. Fibroblasts and myofibroblasts: Their source, function and role in disease. *International Journal of Biochemistry and Cell Biology*, 39(4), pp.666–671.

McAnulty, R.J., Campa, J.S., Cambrey, A.D. & Laurent, G.J., 1991. The effect of transforming growth factor beta on rates of procollagen synthesis and degradation in vitro. *Biochimica et Biophys Acta*, 1091(2), pp.231–235.

McAnulty, R.J., Hernández-Rodriguez, N.A., Mutsaers, S.E., Coker, R.K. & Laurent, G.J., 1997. Indomethacin suppresses the anti-proliferative effects of transforming growth factor-beta isoforms on fibroblast cell cultures. *Biochemical Journal*, 321(Pt 3), pp.639–643.

McCarthy, D.S., Baragar, F.D., Dhingra, S., Sigurdson, M., Sutherland, J.B., Rigby, M. & Martin, L., 1988. The lungs in systemic sclerosis (scleroderma): a review and new information. *Seminars in Arthritis and Rheumatism*, 17(4), pp.271-283.

McKeown, S., Richter, A.G., O'Kane, C., McAuley, D.F. & Thickett, D.R., 2009. MMP expression and abnormal lung permeability are important determinants of outcome in IPF. *European Respiratory Journal*, 33(1), pp.77–84.

McKleroy, W., Lee, T.-H. & Atabai, K., 2013. Always cleave up your mess: targeting collagen degradation to treat tissue fibrosis. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 304(11), pp.L709–21.

Mecham. R.P., Stenmark, K.R. & Parks, W.C., 1991. Connective tissue production by vascular smooth muscle in development and disease. *Chest*, 99(3), s43-47.

Merline, R., Schaefer, R.M. & Schaefer, L., 2009. The matricellular functions of small leucine-rich proteoglycans (SLRPs). *Journal of Cell Communication and Signaling*, 3(3-4), pp.323–335.

Milosevic, J., Pandit, K., Magister, M., Rabinovich, E., Ellwanger, D.C., Yu, G., Vuga, L.J., Weksler, B., Benos, P. V., Gibson, K.F., McMillan, M., Kahn, M. & Kaminski, N., 2012. Profibrotic role of miR-154 in pulmonary fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, 47(6), pp.879–887.

Minamitani, T., Ikuta, T., Saito, Y., Takebe, G., Sato, M., Sawa, H., Nishimura, T., Nakamura, F., Takahashi, K., Ariga, H. & Matsumoto, K.I., 2004. Modulation of collagen fibrillogenesis by tenascin-X and type VI collagen. *Experimental Cell Research*, 298(1), pp.305–315.

Miura, Y., Saito, T., Fujita, K., Tsunoda, Y., Tanaka, T., Takoi, H., Rin, S., Sekine, A., Hayashihara, K., Nei, T. & Azuma, A., 2014. Clinical experience with pirfenidone in five patients with scleroderma-related interstitial lung disease. *Sarcoidosis Vasculitis and Diffuse Lung Disease*, 31(3), pp.235–238.

Moeller, A., Gilpin, S.E., Ask, K., Cox, G., Cook, D., Gauldie, J., Margetts, P.J., Farkas, L., Dobranowski, J., Boylan, C., O'Byrne, P.M., Strieter, R.M. & Kolb, M., 2009. Circulating fibrocytes are an indicator of poor prognosis in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 179(7), pp.588–594.

Mohandas, T., Sparkes, R.S. & Shapiro, L.J., 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science,* 211(4480), pp.393–396.

Molyneaux, P.L. & Maher, T.M., 2013. The role of infection in the pathogenesis of idiopathic pulmonary fibrosis. *European Respiratory Review*, 22(129), pp.376–81.

Monaghan, H., Wells, A.U., Colby, T. V., Du Bois, R.M., Hansell, D.M. & Nicholson, A.G., 2004. Prognostic implications of histologic patterns in multiple surgical lung biopsies from patients with idiopathic interstitial pneumonias. *Chest*, 125(2), pp.522–526.

Monick, M.M., Beach, S.R.H., Plume, J., Sears, R., Gerrard, M., Brody, G.H. & Philibert, R.A., 2012. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 159(2), pp.141–151.

Monso, E., Tura, J.M., Marsal, M., Morell, F., Pujadas, J. & Morera, J., 1990. Mineralogical microanalysis of idiopathic pulmonary fibrosis. *Archives of Environmental Health*, 45(3), pp.185-188.

Moodley, Y.P., Caterina, P., Scaffidi, A.K., Misso, N.L., Papadimitriou, J.M., McAnulty, R.J., Laurent, G.J., Thompson, P.J. & Knight, D.A., 2004. Comparison of the morphological and biochemical changes in normal human lung fibroblasts and fibroblasts derived from lungs of patients with idiopathic pulmonary fibrosis during FasL-induced apoptosis. *The Journal of Pathology*, 202(4), pp.486–495.

Moodley, Y.P., Misso, N.L.A., Scaffidi, A.K., Fogel-Petrovic, M., McAnulty, R.J., Laurent, G.J., Thompson, P.J. & Knight, D. a., 2003. Inverse effects of interleukin-6 on apoptosis of fibroblasts from pulmonary fibrosis and normal lungs. *American Journal of Respiratory Cell and Molecular Biology*, 29(4), pp.490–498.

Moore, B.B. & Moore, T.A., 2015. Viruses in idiopathic pulmonary fibrosis. Etiology and exacerbation. *Annals of the American Thoracic Society*, 12(2), pp.S186–92.

Moore, B.B., Coffey, M.J., Christensen, P., Sitterding, S., Ngan, R., Wilke, C.A., McDonald, R., Phare, S.M., Peters-Golden, M., Paine, R. & Toews, G.B., 2000. GM-CSF regulates bleomycin-induced pulmonary fibrosis via a prostaglandin-dependent mechanism. *Journal of Immunology*, 165(7), pp.4032–4039.

Morlá, M., Busquets, X., Pons, J., Sauleda, J., MacNee, W. & Agustí, A.G.N., 2006. Telomere shortening in smokers with and without COPD. *European Respiratory Journal*, 27(3), pp.525–528.

Morrisey, E.E., 2003. Wnt signaling and pulmonary fibrosis. *American Journal of Physical Anthropology*, 162(5), pp.1393–1397.

Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M.C. & Leonhardt, H., 2005. Recruitment of DNA methyltransferase I to DNA repair sites. *Proceedings of the National Academy of Sciences of the United States of America*, 102(25), pp.8905–8909.

Mosser, D.M. & Edwards, J.P., 2008. Exploring the full spectrum of macrophage activation. *Nature Reviews. Immunology*, 8(12), pp.958–969.

Mulugeta, S., Nguyen, V., Russo, S.J., Muniswamy, M. & Beers, M.F., 2005. A surfactant protein C precursor protein BRICHOS domain mutation causes endoplasmic reticulum stress, proteasome dysfunction, and caspase 3 activation. *American Journal of Respiratory Cell and Molecular Biology*, 32(6), pp.521–530.

Munger, J.S., Huang, X., Kawakatsu, H., Griffiths, M.J., Dalton, S.L., Wu, J., Pittet, J.F., Kaminski, N., Garat, C., Matthay, M.A., Rifkin, D.B. & Sheppard, D., 1999. The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis. *Cell*, 96(3), pp.319–328.

Murray, L.A., Chen, Q., Kramer, M.S., Hesson, D.P., Argentieri, R.L., Peng, X., Gulati, M., Homer, R.J., Russell, T., Van Rooijen, N., Elias, J.A., Hogaboam, C.M. & Herzog, E.L., 2011. TGF-beta driven lung fibrosis is macrophage dependent and blocked by Serum amyloid P. *International Journal of Biochemistry and Cell Biology*, 43(1), pp.154–162.

Mushiroda, T., Wattanapokayakit, S., Takahashi, A., Nukiwa, T., Kudoh, S., Ogura, T., Taniguchi, H., Kubo, M., Kamatani, N. & Nakamura, Y., 2008. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. *Journal of Medical Genetics*, 45(10), pp.654–656.

Muth, M., Hussein, K., Jacobi, C., Kreipe, H. & Bock, O., 2011. Hypoxia-induced down-regulation of microRNA-449a/b impairs control over targeted SERPINE1 (PAI-1) mRNA - a mechanism involved in SERPINE1 (PAI-1) overexpression. *Journal of Translational Medicine*, 4(9), pp.24–33.

Nagai, S., Kitaichi, M., Hamada, K., Nagao, T., Hoshino, Y., Miki, H. & Izumi, T., 1999. Hospital-based historical cohort study of 234 histologically proven Japanese patients with IPF. *Sarcoidosis Vasculitis and Diffuse Lung Disease*, 16(2), pp.209–214.

Nakamura, N. & Takenaga, K., 1998. Hypomethylation of the metastasis-associated S100A4 gene correlates with gene activation in human colon adenocarcinoma cell lines. *Clinical & Experimental Metastasis*, 16(5), pp.471–479.

Nakamura, T., Hamada, F., Ishidate, T., Anai, K., Kawahara, K., Toyoshima, K. & Akiyama, T., 1998. Axin, an inhibitor of the Wnt signalling pathway, interacts with beta-catenin, GSK-3beta and APC and reduces the beta-catenin level. *Genes to Cells: Devoted to Molecular and Cellular Mechanisms*, 3(6), pp.395–403.

Nakashima, T., Jinnin, M., Yamane, K., Honda, N., Kajihara, I., Makino, T., Masuguchi, S., Fukushima, S., Okamoto, Y., Hasegawa, M., Fujimoto, M. & Ihn, H., 2012. Impaired IL-17 signaling pathway contributes to the increased collagen expression in scleroderma fibroblasts. *Journal of Immunology*, 188(8), pp.3573–83.

Nalysnyk, L., Cid-Ruzafa, J., Rotella, P. & Esser, D., 2012. Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. *European Respiratory Review*, 21(126), pp.355–361.

Namboodiri, A.M., Rocca, K.M., Kuwana, M. & Pandey, J.P., 2006. Antibodies to human cytomegalovirus protein UL83 in systemic sclerosis. *Clinical and Experimental Rheumatology*, 24(2), pp.174–178.

Nannini, C., West, C.P., Erwin, P.J. & Matteson, E.L., 2008. Effects of cyclophosphamide on pulmonary function in patients with scleroderma and interstitial lung disease: a systematic review and meta-analysis of randomized controlled trials and observational prospective cohort studies. *Arthritis Research and Therapy*, 10(5), p.R124.

Nath, S. & Mukherjee, P., 2014. MUC1: A multifaceted oncoprotein with a key role in cancer progression. *Trends in Molecular Medicine*, 20(6), pp.332–342.

Navaratnam, V., Fleming, K.M., West, J., Smith, C.J.P., Jenkins, R.G., Fogarty, A. & Hubbard, R.B., 2011. The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax*, 66(6), pp.462–467.

Nebert, D.W. & Russell, D.W., 2002. Clinical importance of the cytochromes P450. *Lancet*, 360(9340), pp.1155–1162.

Ngalamika, O., Zhang, Y., Yin, H., Zhao, M., Gershwin, M.E. & Lu, Q., 2012. Epigenetics, autoimmunity and hematologic malignancies: A comprehensive review. *Journal of Autoimmunity*, 39(4), pp.451–465.

Nho, R.S. & Hergert, P., 2014. IPF fibroblasts are desensitized to type i collagen matrix-induced cell death by suppressing low autophagy via aberrant AKT/MTOR kinases. *PLoS ONE*, 9(4), e94616.

Nicholls, R.D., Knoll, J.H., Butler, M.G., Karam, S. & Lalande, M., 1989. Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. *Nature*, 342(6247), pp.281–285.

Nicholson, A.G. & Wells, A.U., 2001. Nonspecific interstitial pneumonia: nobody said it's perfect. *The American Journal of Respiratory and Critical Care Medicine*, 164, pp.1553-4.

Nicholson, A.G., Fulford, L.G., Colby, T. V., Du Bois, R.M., Hansell, D.M. & Wells, A.U., 2002. The relationship between individual histologic features and disease progression in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 166(2), pp.173–177.

Nielsen, D.A., Yuferov, V., Hamon, S., Jackson, C., Ho, A., Ott, J. & Kreek, M.J., 2009. Increased OPRM1 DNA methylation in lymphocytes of methadone-maintained former heroin addicts. *Neuropsychopharmacology*, 34(4), pp.867–873.

Nietert, P.J. & Silver, R.M., 2000. Systemic sclerosis: environmental and occupational risk factors. *Current Opinion in Rheumatology*, 12(6), pp.520–526.

Nieto, M.A., 2002. The snail superfamily of zinc-finger transcription factors. *Nature Reviews. Molecular Cell Biology*, 3(3), pp.155–166.

Nihtyanova, S.I. & Denton, C.P., 2010. Autoantibodies as predictive tools in systemic sclerosis. *Nature Reviews. Rheumatology*, 6(2), pp.112–116.

Nishida, K., Ono, K., Kanaya, S. & Takahashi, K., 2014. KEGGscape: a Cytoscape app for pathway data integration. *F1000 Research*, 3, p.144.

Noble, P.W., Albera, C., Bradford, W.Z., Costabel, U., Glassberg, M.K., Kardatzke, D., King, T.E., Lancaster, L., Sahn, S.A., Szwarcberg, J., Valeyre, D. & du Bois, R.M., 2011. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet*, 377(9779), pp.1760–1769.

Nogales-Cadenas, R., Carmona-Saez, P., Vazquez, M., Vicente, C., Yang, X., Tirado, F., Carazo, J.M. & Pascual-Montano, A., 2009. GeneCodis: Interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Research*, 37(2), pp.317–322.

Nogee, L.M., Dunbar, A.E. 3rd., Wert, S.E., Askin, F., Hamvas, A. & Whitsett, J.A., 2001. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *The New England Journal of Medicine*, 344(8), pp.573-579.

Nordenbaek, C., Johansen, J.S., Halberg, P., Wiik, A., Garbarsch, C., Ullman, S., Price, P.A. & Jacobsen, S., 2005. High serum levels of YKL-40 in patients with systemic sclerosis are associated with pulmonary involvement. *Scandinavian Journal of Rheumatology*, 34(4), pp.293–7.

Norlin, M. & Wikvall, K., 2007. Enzymes in the conversion of cholesterol into bile acids. *Current Molecular Medicine*, 7(2), pp.199–218.

Noth, I., Anstrom, K.J., Calvert, S.B., De Andrade, J., Flaherty, K.R., Glazer, C., Kaner, R.J. & Olman, M.A., 2012. A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 186(1), pp.88–95.

Noth, I., Zhang, Y., Ma, S.-F., Flores, C., Barber, M., Huang, Y., Broderick, S.M., Wade, M.S., Hysi, P., Scuirba, J., Richards, T.J., Juan-Guardela, B.M., Vij, R., Han, M.K., Martinez, F.J., Kossen, K., Seiwert, S.D., Christie, J.D., Nicolae, D., Kaminski, N. & Garcia, J.G.N., 2013. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *The Lancet. Respiratory Medicine*, 1(4), pp.309–17.

O'Donoghue, R.J.J., Knight, D.A., Richards, C.D., Prêle, C.M., Lau, H.L., Jarnicki, A.G., Jones, J., Bozinovski, S., Vlahos, R., Thiem, S., McKenzie, B.S., Wang, B., Stumbles, P., Laurent, G.J., McAnulty, R.J., Rose-John, S., Zhu, H.J., Anderson, G.P., Ernst, M.R. & Mutsaers, S.E., 2012. Genetic partitioning of interleukin-6 signalling in mice dissociates Stat3 from Smad3-mediated lung fibrosis. *EMBO Molecular Medicine*, 4(9), pp.939–951.

Ochs, M., Nyengaard, J.R., Jung, A., Knudsen, L., Voigt, M., Wahlers, T., Richter, J. & Gundersen, H.J.G., 2004. The number of alveoli in the human lung. *American Journal of Respiratory and Critical Care Medicine*, 169(1), pp.120–124.

Oh, C.K., Murray, L.A. & Molfino, N.A., 2012. Smoking and idiopathic pulmonary fibrosis. *Pulmonary Medicine*, 2012, pp.808260.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., Berman, D.M., Jenuwein, T., Pruitt, K., Sharkis, S.J., Watkins, D.N., Herman, J.G. & Baylin, S.B., 2007. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genetics*, 39(2), pp.237–42.

Ohnishi, H., Yokoyama, A., Yasuhara, Y., Watanabe, A., Naka, T., Hamada, H., Abe, M., Nishimura, K., Higaki, J., Ikezoe, J. & Kohno, N., 2003. Circulating KL-6 levels in patients with drug induced pneumonitis. *Thorax*, 58(10), pp.872–875.

Okano, M., Bell, D.W., Haber, D.A. & Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3), pp.247–257.

Oku, H., Nakazato, H., Horikawa, T., Tsuruta, Y. & Suzuki, R., 2002. Pirfenidone suppresses tumor necrosis factor-alpha, enhances interleukin-10 and protects mice from endotoxic shock. *European Journal of Pharmacology*, 446(1-3), pp.167–176.

Olsen, K.C., Sapinoro, R.E., Kottmann, R.M., Kulkarni, A.A., Iismaa, S.E., Johnson, G.V.W., Thatcher, T.H., Phipps, R.P. & Sime, P.J., 2011. Transglutaminase 2 and its role in pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 184(6), pp.699–707.

Olson, A.L., Swigris, J.J., Lezotte, D.C., Norris, J.M., Wilson, C.G. & Brown, K.K., 2007. Mortality from pulmonary fibrosis increased in the United States from 1992 to 2003. *American Journal of Respiratory and Critical Care Medicine*, 176(3), pp.277–284.

Orozco, L.D., Rubbi, L., Martin, L.J., Fang, F., Hormozdiari, F., Che, N., Smith, A.D., Lusis, A.J. & Pellegrini, M., 2014. Intergenerational genomic DNA methylation patterns in mouse hybrid strains. *Genome Biology*, 15(5), p.R68.

Ostlere, L.S., Harris, D., Buckley, C., Black, C. & Rustin, M.H.A., 1992. Atypical systemic sclerosis following exposure to vinyl chloride monomer. A case report and review of the cutaneous aspects of vinyl chloride disease. *Clinical and Experimental Dermatology*, 17(3), pp.208–210.

Otero, N.K.H., Thomas, J.D., Saski, C.A., Xia, X. & Kelly, S.J., 2012. Choline supplementation and DNA methylation in the hippocampus and prefrontal cortex of rats exposed to alcohol during development. *Alcoholism: Clinical and Experimental Research*, 36(10), pp.1701–1709.

Ouko, L.A., Shantikumar, K., Knezovich, J., Haycock, P., Schnugh, D.J. & Ramsay, M., 2009. Effect of alcohol consumption on CpG methylation in the differentially methylated regions of H19 and IG-DMR in male gametes - Implications for fetal alcohol spectrum disorders. *Alcoholism: Clinical and Experimental Research*, 33(9), pp.1615–1627.

Pacaud, R., Sery, Q., Oliver, L., Vallette, F.M., Tost, J. & Cartron, P.F., 2014. DNMT3L interacts with transcription factors to target DNMT3L/DNMT3B to specific DNA sequences: Role of the DNMT3L/DNMT3B/p65-NFkB complex in the (de-)methylation of TRAF1. *Biochimie*, 104(1), pp.36–49.

Packer, M., Greenberg, B., Massie, B. & Dash, H., 1982. Deleterious effects of hydralazine in patients with pulmonary hypertension. *The New England Journal of Medicine*, 306(22), pp.1326–1331.

Palii, S.S., Van Emburgh, B.O., Sankpal, U.T., Brown, K.D. & Robertson, K.D., 2008. DNA methylation inhibitor 5-Aza-2'-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 and 3B. *Molecular and Cellular Biology*, 28(2), pp.752–71.

Pan, L.H., Yamauchi, K., Uzuki, M., Nakanishi, T., Takigawa, M., Inoue, H. & Sawai, T., 2001. Type II alveolar epithelial cells and interstitial fibroblasts express connective tissue growth factor in IPF. *European Respiratory Journal*, 17(6), pp.1220–1227.

Pandey, J.P. & LeRoy, E.C., 1998. Current Comment: Human cytomegalovirus and the vasculopathies of autoimmune diseases (especially scleroderma), allograft rejection, and coronary restenosis. *Arthritis and Rheumatism*, 41(1), pp.10–15.

Pandit, K. V & Milosevic, J., 2015. MicroRNA regulatory networks in idiopathic pulmonary fibrosis. *Biochemistry and Cell Biology*, 93(2), pp.129–37.

Pandit, K. V., Corcoran, D., Yousef, H., Yarlagadda, M., Tzouvelekis, A., Gibson, K.F., Konishi, K., Yousem, S.A., Singh, M., Handley, D., Richards, T., Selman, M., Watkins, S.C., Pardo, A., Ben-Yehudah, A., Bouros, D., Eickelberg, O., Ray, P., Benos, P. V. & Kaminski, N., 2010. Inhibition and role of let-7d in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 182(2), pp.220–229.

Pandit, K. V., Milosevic, J. & Kaminski, N., 2011. MicroRNAs in idiopathic pulmonary fibrosis. *Translational research : The Journal of Laboratory and Clinical Medicine*, 157(4), pp.191–199.

Paone, C., Chiarolanza, I., Cuomo, G., Ruocco, L., Vettori, S., Menegozzo, M., La Montagna, G. & Valentini, G., 2007. Twelve-month azathioprine as maintenance therapy in early diffuse systemic sclerosis patients treated for 1-year with low dose cyclophosphamide pulse therapy. *Clinical and Experimental Rheumatology*, 25(4), pp.613–616.

Pardo, A. & Selman, M., 2006. Matrix metalloproteases in aberrant fibrotic tissue remodeling. *Proceedings of the American Thoracic Society*, 3(4), pp.383–388.

Pardo, A. & Selman, M., 2012. Role of matrix metalloproteases in idiopathic pulmonary fibrosis. *Fibrogenesis & Tissue Repair*, 5(Supplement 1), p.S9.

Pardo, A., Gibson, K., Cisneros, J., Richards, T.J., Yang, Y., Becerril, C., Yousem, S., Herrera, I., Ruiz, V., Selman, M. & Kaminski, N., 2005. Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. *PLoS Medicine*, 2(9), pp.0891–0903.

Parker, M.W., Rossi, D., Peterson, M., Smith, K., Sikstrom̈, K., White, E.S., Connett, J.E., Henke, C.A., Larsson, O. & Bitterman, P.B., 2014. Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *Journal of Clinical Investigation*, 124(4), pp.1622–1635.

Parks, C.G., Cooper, G.S., Nylander-French, L.A., Sanderson, W.T., Dement, J.M., Cohen, P.L., Dooley, M.A., Treadwell, E.L., St.Clair, E.W., Gilkeson, G.S., Hoppin, J.A. & Savitz, D.A., 2002. Occupational exposure to crystalline silica and risk of systemic lupus erythematosus: A population-based, case-control study in the southeastern United States. *Arthritis and Rheumatism*, 46(7), pp.1840–1850.

Parry, L. & Clarke, A.R., 2011. The roles of the methyl-CpG binding proteins in cancer. *Genes and Cancer*, 2(6), pp.618–630.

Payer, B. & Lee, J.T., 2008. X chromosome dosage compensation: how mammals keep the balance. *Annual Review of Genetics*, 42, pp.733–772.

Pekovic-Vaughan, V., Gibbs, J., Yoshitane, H., Yang, N., Pathiranage, D., Guo, B., Sagami, A., Taguchi, K., Bechtold, D., Loudon, A., Yamamoto, M., Chan, J., van der Horst, G.T.J., Fukada, Y. & Meng, Q.J., 2014. The circadian clock regulates rhythmic activation of the NRF2/glutathionemediated antioxidant defense pathway to modulate pulmonary fibrosis. *Genes and Development*, 28(6), pp.548–560.

Philibert, R.A., Plume, J.M., Gibbons, F.X., Brody, G.H. & Beach, S.R.H., 2012. The impact of recent alcohol use on genome wide DNA methylation signatures. *Frontiers in Genetics*, 3(54), e3389.

Pinto, R., Pilato, B., Ottini, L., Lambo, R., Simone, G., Paradiso, A. & Tommasi, S., 2013. Different methylation and MicroRNA expression pattern in male and female familial breast cancer. *Journal of Cellular Physiology*, 228(6), pp.1264–1269.

Plow, E.F., Haas, T.A., Zhang, L., Loftus, J. & Smith, J.W., 2000. Ligand binding to integrins. *Journal of Biological Chemistry*, 275(29), pp.21785–21788.

Price, M.E., Cotton, A.M., Lam, L.L., Farré, P., Emberly, E., Brown, C.J., Robinson, W.P. & Kobor, M.S., 2013. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics & Chromatin*, 6(1), p.4.

Qi, G., Jia, L., Li, Y., Bian, Y., Cheng, J., Li, H., Xiao, C. & Du, J., 2011. Angiotensin II infusion-induced inflammation, monocytic fibroblast precursor infiltration, and cardiac fibrosis are pressure dependent. *Cardiovascular Toxicology*, 11(2), pp.157–167.

Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M.J. & Zhou, M.M., 2008. Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *Journal of Biological Chemistry*, 283(50), pp.34490–34494.

Qin, T., Jelinek, J., Si, J., Shu, J. & Issa, J.P.J., 2009. Mechanisms of resistance to 5-Aza-2'-deoxycytidine in human cancer cell lines. *Blood*, 113(3), pp.659–667.

Quintero-Ronderos, P. & Montoya-Ortix, G., 2012. Epigenetics and autoimmune diseases. *Autoimmune Diseases*, 2012, pp.1-16.

Rabinovich, E.I., Kapetanaki, M.G., Steinfeld, I., Gibson, K.F., Pandit, K. V., Yu, G., Yakhini, Z. & Kaminski, N., 2012. Global methylation patterns in idiopathic pulmonary fibrosis. *PLoS ONE*, 7(4), e33770.

Radić, M., Kaliterna, D.M. & Radić, J., 2011. Helicobacter pylori infection and systemic sclerosis-is there a link? *Joint Bone Spine*, 78(4), pp.337–340.

Radisky, E.S. & Radisky, D.C., 2010. Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, 15(2), pp.201–212.

Radstake, T.R.D.J., Gorlova, O., Rueda, B., Martin, J.-E., Alizadeh, B.Z., Palomino-Morales, R., Coenen, M.J., Vonk, M.C., Voskuyl, A.E., Schuerwegh, A.J., Broen, J.C., van Riel, P.L.C.M., van 't Slot, R., Italiaander, A., Ophoff, R.A., Riemekasten, G., Hunzelmann, N., Simeon, C.P., Ortego-Centeno, N., González-Gay, M.A., González-Escribano, M.F., Airo, P., van Laar, J., Herrick, A., Worthington, J., Hesselstrand, R., Smith, V., de Keyser, F., Houssiau, F., Chee, M.M., Madhok, R., Shiels, P., Westhovens, R., Kreuter, A., Kiener, H., de Baere, E., Witte, T., Padykov, L., Klareskog, L., Beretta, L., Scorza, R., Lie, B.A., Hoffmann-Vold, A.-M., Carreira, P., Varga, J., Hinchcliff, M., Gregersen, P.K., Lee, A.T., Ying, J., Han, Y., Weng, S.-F., Amos, C.I., Wigley, F.M., Hummers, L., Nelson, J.L., Agarwal, S.K., Assassi, S., Gourh, P., Tan, F.K., Koeleman, B.P.C., Arnett, F.C., Martin, J. & Mayes, M.D., 2010. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nature Genetics*, 42(5), pp.426–429.

Raghu, G. & Brown, K.K., 2004. Interstitial lung disease: Clinical evaluation and keys to an accurate diagnosis. *Clinics in Chest Medicine*, 25(3), pp.409–419.

Raghu, G., Anstrom, K.J., King, T.E., Lasky, J.A. & Martinez, F.J., 2012. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *The New England Journal of Medicine*, 366(21), pp.1968–77.

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., Müller, N.L., Nicholson, A.G., Richeldi, L., Selman, M., Dudden, R.F., Griss, B.S., Protzko, S.L. & Schünemann, H.J., 2011a. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine*, 183(6), pp.788–824.

Raghu, G., Weycker, D., Edelsberg, J., Bradford, W.Z. & Oster, G., 2006. Incidence and prevalence of idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 174(7), pp.810–816.

Rakyan, V.K., Beyan, H., Down, T. a., Hawa, M.I., Maslau, S., Aden, D., Daunay, A., Busato, F., Mein, C. a., Manfras, B., Dias, K.R.M., Bell, C.G., Tost, J., Boehm, B.O., Beck, S. & Leslie, R.D., 2011. Identification of type 1 Diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genetics*, 7(9), pp.1–9.

Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A. V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A., Fischer, J., Gut, I.G., Berlin, K. & Beck, S., 2004. DNA methylation profiling of the human major histocompatibility complex: A pilot study for the Human Epigenome Project. *PLoS Biology*, 2(12), pp.e405.

Ramage, J.E., Roggli, V.L., Bell, D.Y. & Piantadosi, C.A., 1988. Interstitial lung disease and domestic wood burning. *The American Review of Respiratory Diseases*, 137(5), pp.1229-1232.

Ramos, C., Montaño, M., García-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. *American Journal of Respiratory Cell and Molecular Biology*, 24(5), pp.591–598.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P. & Jaenisch, R., 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), pp.5237–5242.

Rao, X., Evans, J., Chae, H., Pilrose, J., Kim, S., Yan, P., Huang, R.-L., Lai, H.-C., Lin, H., Liu, Y., Miller, D., Rhee, J.-K., Huang, Y.-W., Gu, F., Gray, J.W., Huang, T.-M. & Nephew, K.P., 2013. CpG island shore methylation regulates caveolin-1 expression in breast cancer. *Oncogene*, 32(38), pp.4519–4528.

Rauch, T. & Wu, X., 2009. A human B cell methylome at 100- base pair resolution. *Proceedings of the National Academy of Sciences of the USA*, 106(3), pp.671–678.

Recklies, A.D., White, C. & Ling, H., 2002. The chitinase 3-like protein human cartilage glycoprotein 39 (HC-gp39) stimulates proliferation of human connective-tissue cells and activates both extracellular signal-regulated kinase- and protein kinase B-mediated signalling pathways. *The Biochemical Journal*, 365(Pt 1), pp.119–126.

Reik, W., Constância, M., Fowden, A., Anderson, N., Dean, W., Ferguson-Smith, A., Tycko, B. & Sibley, C., 2003. Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. *The Journal of Physiology*, 547(Pt 1), pp.35–44.

Reik, W., Dean, W. & Walter, J., 2001. Epigenetic reprogramming in mammalian development. *Science,* 293(5532), pp.1089–1093.

Richardson, B., 2003. Impact of aging on DNA methylation. *Ageing Research Reviews*, 2(3), pp.245–261.

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., Klüglich, M. & du Bois, R.M., 2011. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *The New England Journal of Medicine*, 365(12), pp.1079–87.

Richeldi, L., Davies, H.R., Ferrara, G. & Franco, F., 2003. Corticosteroids for idiopathic pulmonary fibrosis. *The Cochrane Database of Systematic Reviews*, (3), p.CD002880.

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., 2014. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *The New England Journal of Medicine*, 370(22), pp.2071–82.

Richter, A.G., McKeown, S., Rathinam, S., Harper, L., Rajesh, P., McAuley, D.F., Heljasvaara, R. & Thickett, D.R., 2009. Soluble endostatin is a novel inhibitor of epithelial repair in idiopathic pulmonary fibrosis. *Thorax*, 64(2), pp.156–161.

Richter, A.G., Stockley, R.A., Harper, L. & Thickett, D.R., 2009. Pulmonary infection in Wegener granulomatosis and idiopathic pulmonary fibrosis. *Thorax*, 64(8), pp.692–697.

Ries, C., 2014. Cytokine functions of TIMP-1. *Cellular and Molecular Life Sciences*, 71(4), pp.659–672.

Riha, R.L., Yang, I.A., Rabnott, G.C., Tunnicliffe, A.M., Fong, K.M. & Zimmerman, P. V., 2004. Cytokine gene polymorphisms in idiopathic pulmonary fibrosis. *Internal Medicine Journal*, 34(3), pp.126–129.

Rio, D.C., Ares, M., Hannon, G.J. & Nilsen, T.W., 2010. Purification of RNA using TRIzol (TRI Reagent). *Cold Spring Harbor Protocols*, 5(6), pp.1–4.

Robertson, K.D., 2005. DNA methylation and human disease. *Nature Reviews. Genetics*, 6(8), pp.597–610.

Robinson, C.M., Neary, R., Levendale, A., Watson, C.J. & Baugh, J.A., 2012. Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is ssociated with thy-1 promoter methylation and the development of a pro-fibrotic Phenotype. *Respiratory Research*, 13(1), p.74.

Rock, J.R., Barkauskas, C.E., Cronce, M.J., Xue, Y., Harris, J.R., Liang, J., Noble, P.W. & Hogan, B.L.M., 2011. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proceedings of the National Academy of Sciences*, 108(52), pp.E1475–E1483.

Rodan, G.P., Benedek, T.G., Medsger, T.A. Jr. & Cammarata, R.J., 1967. The association of progressive systemic sclerosis (scleroderma) with coal miners' pneumoconiosis and other forms of silicosis. *Annals of Internal Medicine*, 66(2), 323-334.

Rosas, I.O., Richards, T.J., Konishi, K., Zhang, Y., Gibson, K., Lokshin, A.E., Lindell, K.O., Cisneros, J., MacDonald, S.D., Pardo, A., Sciurba, F., Dauber, J., Selman, M., Gochuico, B.R. & Kaminski, N., 2008. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Medicine*, 5(4), pp.0623–0633.

Rose, N.R. & Klose, R.J., 2014. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta*, 1839(12), pp.1362–1372.

Roth, T.L., Lubin, F.D., Funk, A.J. & Sweatt, J.D., 2009. Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biological Psychiatry*, 65(9), pp.760–769.

Rountree, M.R., Bachman, K.E. & Baylin, S.B., 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nature Genetics*, 25(3), pp.269–277.

Rozario, T. & DeSimone, D.W., 2010. The extracellular matrix in development and morphogenesis: A dynamic view. *Developmental Biology*, 341(1), pp.126–140.

Rozen, S. & Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, 132, pp.365–386.

Rucker, R.B. & Dubick, M.A., 1984. Elastin metabolism and chemistry: Potential roles in lung development and structure. *Environmental Health Perspectives*, 55, pp.179–191.

Rueda, B., Gourh, P., Broen, J., Agarwal, S.K., Simeon, C., Ortego-Centeno, N., Vonk, M.C., Coenen, M., Riemekasten, G., Hunzelmann, N., Hesselstrand, R., Tan, F.K., Reveille, J.D., Assassi, S., Garcia-Hernandez, F.J., Carreira, P., Camps, M., Fernandez-Nebro, A., Garcia de la Peña, P., Nearney, T., Hilda, D., Gónzalez-Gay, M.A., Airo, P., Beretta, L., Scorza, R., Radstake, T.R.D.J., Mayes, M.D., Arnett, F.C. & Martin, J., 2010. BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians. *Annals of the Rheumatic Diseases*, 69(4), pp.700–705.

Ruiz, X.D., Mlakar, L.R., Yamaguchi, Y., Su, Y., Larregina, A.T., Pilewski, J.M. & Feghali-Bostwick, C.A., 2012. Syndecan-2 is a novel target of insulin-like growth factor binding protein-3 and is over-expressed in fibrosis. *PLoS ONE*, 7(8), pp.e43049.

Sado, T., Wang, Z., Sasaki, H. & Li, E., 2001. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development,* 128(8), pp.1275–1286.

Sanders, Y.Y., Ambalavanan, N., Halloran, B., Zhang, X., Liu, H., Crossman, D.K., Bray, M., Zhang, K., Thannickal, V.J. & Hagood, J.S., 2012. Altered DNA methylation profile in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 186(6), pp.525–535.

Sanders, Y.Y., Pardo, A., Selman, M., Nuovo, G.J., Tollefsbol, T.O., Siegal, G.P. & Hagood, J.S., 2008. Thy-1 promoter hypermethylation: A novel epigenetic pathogenic mechanism in pulmonary fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, 39(5), pp.610–618.

Sanders, Y.Y., Tollefsbol, T.O., Varisco, B.M. & Hagood, J.S., 2011. Epigenetic regulation of Thy-1 by histone deacetylase inhibitor in rat lung fibroblasts. *American Journal of Respiratory Cell and Molecular Biology*, 45(1), pp.16–23.

Sandoval, J., Heyn, H.A., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M. & Esteller, M., 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*, 6, pp.692–702.

Sarkar, S., Abujamra, A.L., Loew, J.E., Forman, L.W., Perrine, S.P. & Faller, D. V., 2011. Histone deacetylase inhibitors reverse CpG methylation by regulating DNMT1 through ERK signaling. *Anticancer Research*, 31(9), pp.2723–2732.

Sarter, B., Long, T.I., Tsong, W.H., Koh, W.P., Yu, M.C. & Laird, P.W., 2005. Sex differential in methylation patterns of selected genes in Singapore Chinese. *Human Genetics*, 117(4), pp.402–403.

Sato, H., Lagan, A.L., Alexopoulou, C., Vassilakis, D.A., Ahmad, T., Pantelidis, P., Veeraraghavan, S., Renzoni, E., Denton, C., Black, C., Wells, A.U., Du Bois, R.M. & Welsh, K.I., 2004. The TNF-863A allele strongly associates with anti-centromere antibody positivity in scleroderma. *Arthritis and Rheumatism*, 50(2), pp.558–564.

Sato, S., Hayakawa, I., Hasegawa, M., Fujimoto, M. & Takehara, K., 2003. Function blocking autoantibodies against matrix metalloproteinase-1 in patients with systemic sclerosis. *Journal of Investigative Dermatology*, 120(4), pp.542–547.

Sato, S., Nagaoka, T., Hasegawa, M., Nishijima, C. & Takehara, K., 2000. Elevated serum KL-6 levels in patients with systemic sclerosis: association with the severity of pulmonary fibrosis. *Dermatology (Basel, Switzerland)*, 200(3), pp.196–201.

Sauntharajah, Y., 2013. Key clinical observations after 5-Azacytidine and decitabine treatment of myelodysplastic syndromes suggest practical solutions for better outcomes. Hematology / the Education Program of the American Society of Hematology. *American Society of Hematology*, 1, pp.511–521.

Savagner, P., Yamada, K.M. & Thiery, J.P., 1997. The zinc-finger protein slug causes desmosome dissociation, an initial and necessary step for growth factor-induced epithelial-mesenchymal transition. *Journal of Cell Biology*, 137(6), pp.1403–1419.

Savarino, E., Zentilin, P., Furnari, M., Bodini, G., Marabotto, E. & Savarino, V., 2014. Not all patients with non-erosive reflux disease share psychological distress as main mechanism of disease. *Journal of Neurogastroenterology and Motility*, 20(1), pp.129-130.

Saxonov, S., Berg, P. & Brutlag, D.L., 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), pp.1412–1417.

Scadding, J.G. & Hinson, K.F., 1967. Diffuse fibrosing alveolitis (diffuse interstitial fibrosis of the lungs). Correlation of histology at biopsy with prognosis. *Thorax*, 22(4), pp.291–304.

Scadding, J.G., 1960. Chronic diffuse interstitial fibrosis of the lungs. *British Medical Journal*, 1(5171), pp.443–450.

Schaefer, C.J., Ruhrmund, D.W., Pan, L., Seiwert, S.D. & Kossen, K., 2011. Antifibrotic activities of pirfenidone in animal models. *European Respiratory Review*, 20(120), pp.85–97.

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M. & Lyko, F., 2010. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes and Development*, 24(15), pp.1590–1595.

Schneider-Stock, R., Diab-Assef, M., Rohrbeck, A., Foltzer-Jourdainne, C., Boltze, C., Hartig, R., Schönfeld, P., Roessner, A. & Gali-Muhtasib, H., 2005. 5-Aza-cytidine is a potent inhibitor of DNA methyltransferase 3A and induces apoptosis in HCT-116 colon cancer cells via Gadd45- and p53-dependent mechanisms. *The Journal of Pharmacology and Experimental Therapeutics*, 312(2), pp.525–536.

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(6759), pp.1015–1017.

Scotton, C.J. & Chambers, R.C., 2007. Molecular targets in pulmonary fibrosis: The myofibroblast in focus. *Chest*, 132(4), pp.1311–1321.

Scotton, C.J. & Chambers, R.C., 2010. Bleomycin revisited: towards a more representative model of IPF? *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 299(4), pp.L439–L441.

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. *The New England Journal of Medicine*, 364(16), pp.1503–1512.

Selman, M., King T.E., J. & Pardo, A., 2001. Idiopathic pulmonary fibrosis: Prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Annals of Internal Medicine*, 134(2), pp.136–151.

Selman, M., Pardo, A. & Kaminski, N., 2008. Idiopathic pulmonary fibrosis: Aberrant recapitulation of developmental programs? *PLoS Medicine*, 5(3), pp.0373–0380.

Selman, M., Pardo, A., Barrera, L., Estrada, A., Watson, S.R., Wilson, K., Aziz, N., Kaminski, N. & Zlotnik, A., 2006. Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *American Journal of Respiratory and Critical Care Medicine*, 173(2), pp.188–198.

Selman, M., Ruiz, V., Cabrera, S., Segura, L., Ramírez, R., Barrios, R. & Pardo, A., 2000. TIMP-1, -2, -3, and -4 in idiopathic pulmonary fibrosis. A prevailing nondegradative lung microenvironment? *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 279(3), pp.L562–L574.

Selmi, C., Feghali-Bostwick, C.A., Lleo, A., Lombardi, S.A., De Santis, M., Cavaciocchi, F., Zammataro, L., Mitchell, M.M., Lasalle, J.M., Medsger, T. & Gershwin, M.E., 2012. X chromosome gene methylation in peripheral lymphocytes from monozygotic twins discordant for scleroderma. *Clinical and Experimental Immunology*, 169(3), pp.253–262.

Sgonc, R., Gruschwitz, M.S., Dietrich, H., Recheis, H., Gershwin, M.E. & Wick, G., 1996. Endothelial cell apoptosis is a primary pathogenetic event underlying skin lesions in avian and human scleroderma. *Journal of Clinical Investigation*, 98(3), pp.785–792.

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(11), pp.2498–504.

Shapiro, S.D., Endicott, S.K., Province, M.A., Pierce, J.A. & Campbell, E.J., 1991. Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *Journal of Clinical Investigation*, 87(5), pp.1828–1834.

Sharif, R., Mayes, M.D., Tan, F.K., Gorlova, O.Y., Hummers, L.K., Shah, A.A., Furst, D.E., Khanna, D., Martin, J., Bossini-Castillo, L., Gonzalez, E.B., Ying, J., Draeger, H.T., Agarwal, S.K., Reveille, J.D., Arnett, F.C., Wigley, F.M. & Assassi, S., 2012. IRF5 polymorphism predicts prognosis in patients with systemic sclerosis. *Annals of the Rheumatic Diseases*, 71(7), pp.1197–1202.

Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y. & Antonarakis, S.E., 2011. DNA methylation profiles of human active and inactive X chromosomes. *Genome Research*, 21(10), pp.1592–1600.

Shen, J., Chen, X., Hendershot, L. & Prywes, R., 2002. ER stress regulation of ATF6 localization by dissociation of BiP/GRP78 binding and unmasking of golgi localization signals. *Developmental Cell*, 3(1), pp.99–111.

Shenker, N.S., Polidoro, S., van Veldhoven, K., Sacerdote, C., Ricceri, F., Birrell, M.A., Belvisi, M.G., Brown, R., Vineis, P. & Flanagan, J.M., 2013. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Human Molecular Genetics,* 22(5), pp.843–851.

Shiel, W.C. & Prete, P.E., 1984. Pleuropulmonary manifestations of rheumatoid arthritis. *Seminars in Arthritis and Rheumatism*, 13(3), pp.235-43.

Shyy, J.Y.J. & Chien, S., 2002. Role of integrins in endothelial mechanosensing of shear stress. *Circulation Research*, 91(9), pp.769–775.

Siegmund, K., Connor, C., Campan, M. & Long, T., 2007. DNA Methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS ONE*, 2(9), e895.

Silman, A., Jannini, S., Symmons, D. & Bacon, P., 1988. An epidemiological study of scleroderma in the West Midlands. *British Journal of Rheumatology*, 27(4), pp.286–290.

Sing, T., Jinnin, M., Yamane, K., Honda, N., Makino, K., Kajihara, I., Makino, T., Sakai, K., Masuguchi, S., Fukushima, S. & Ihn, H., 2012. MicroRNA-92a expression in the sera and dermal fibroblasts increases in patients with scleroderma. *Rheumatology (United Kingdom)*, 51(9), pp.1550–1556.

Sisson, T.H., 2012. Increased survivin expression contributes to apoptosis-resistance in IPF fibroblasts. *Advances in Bioscience and Biotechnology*, 03(06), pp.657–664.

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M.H., Baldock, R., Barbiera, G., Bardou, P., Beck, T., Blake, A., Bonierbale, M., Brookes, A.J., Bucci, G., Buetti, I., Burge, S., Cabau, C., Carlson, J.W., Chelala, C., Chrysostomou, C., Cittaro, D., Collin, O., Cordova, R., Cutts, R.J., Dassi, E., Genova, A. Di, Djari, A., Esposito, A., Estrella, H., Eyras, E., Fernandez-Banet, J., Forbes, S., Free, R.C., Fujisawa, T., Gadaleta, E., Garcia-Manteiga, J.M., Goodstein, D., Gray, K., Guerra-Assunção, J.A., Haggarty, B., Han, D.-J., Han, B.W., Harris, T., Harshbarger, J., Hastings, R.K., Hayes, R.D., Hoede, C., Hu, S., Hu, Z.-L., Hutchins, L., Kan, Z., Kawaji, H., Keliet, A., Kerhornou, A., Kim, S., Kinsella, R., Klopp, C., Kong, L., Lawson, D., Lazarevic, D., Lee, J.-H., Letellier, T., Li, C.-Y., Lio, P., Liu, C.-J., Luo, J., Maass, A., Mariette, J., Maurel, T., Merella, S., Mohamed, A.M., Moreews, F., Nabihoudine, I., Ndegwa, N., Noirot, C., Perez-Llamas, C., Primig, M., Quattrone, A., Quesneville, H., Rambaldi, D., Reecy, J., Riba, M., Rosanoff, S., Saddiq, A.A., Salas, E., Sallou, O., Shepherd, R., Simon, R., Sperling, L., Spooner, W., Staines, D.M., Steinbach, D., Stone, K., Stupka, E., Teague, J.W., Dayem Ullah, A.Z., Wang, J., Ware, D., Wong-Erasmus, M., Youens-Clark, K., Zadissa, A., Zhang, S.-J. & Kasprzyk, A., 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research*, 43(W1), pp.W589–W598.

Smith, G.K., 2005. limma: Linear Models for Microarray Data. *In Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. pp. 397–420.

Son, J.Y., Kim, S.Y., Cho, S.H., Shim, H.S., Jung, J.Y., Kim, E.Y., Lim, J.E., Park, B.H., Kang, Y.A., Kim, Y.S., Kim, S.K., Chang, J. & Park, M.S., 2013. TGF-β1 T869C polymorphism may affect susceptibility to idiopathic pulmonary fibrosis and disease severity. *Lung*, 191(2), pp.199–205.

Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H. & Held, W.A., 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9), pp.3336–3341.

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. *The European Respiratory Journal*, 37(2), pp.356–63.

Spagnolo, P., Del Giovane, C., Luppi, F., Cerri, S., Balduzzi, S., Walters, E.H., D'Amico, R. & Richeldi, L., 2010. Non-steroid agents for idiopathic pulmonary fibrosis. *Cochrane Database of Systematic Reviews*, (9), p.CD003134.

Spencer-Green, G., Alter, D. & Welch, H.G., 1997. Test performance in systemic sclerosis: Anti-centromere and anti-Scl-70 antibodies. *American Journal of Medicine*, 103(3), pp.242–248.

Spiers, H., Hannon, E., Schalkwyk, L.C., Smith, R., Wong, C.C.Y., Donovan, M.C.O., Bray, N.J. & Mill, J., 2015. Methylomic trajectories across human fetal brain development. *Genome Research*, 25(3), pp.338-352.

Starcher, B.C., 1986. Elastin and the lung. *Thorax*, 41, pp.577-585.

Steegers-Theunissen, R.P., Obermann-Borst, S.A., Kremer, D., Lindemans, J., Siebel, C., Steegers, E.A., Slagboom, P.E. & Heijman, B.T., 2009. Periconceptional maternal folic acid use of 400 µg per day is related to increased methylation of the IGF2 gene in the very young child. *PLoS ONE*, 4(11), pp.e7845.

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., Sporn, T.A., McAdams, H.P., Schwarz, M.I. & Schwartz, D.A., 2005. Clinical and pathologic features of

familial interstitial pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 172(9), pp.1146–1152.

Steen, V. & Medsger, T.A., 2003. Predictors of isolated pulmonary hypertension in patients with systemic sclerosis and limited cutaneous involvement. *Arthritis and Rheumatism*, 48(2), pp.516–522.

Steen, V.D. & Medsger, T.A., 2000. Severe organ involvement in systemic sclerosis with diffuse scleroderma. *Arthritis and Rheumatism*, 43(11), pp.2437–2444.

Steen, V.D. & Medsger, T.A., 2007. Changes in causes of death in systemic sclerosis, 1972-2002. *Annals of the Rheumatic Diseases*, 66(7), pp.940–944.

Steen, V.D., Powell, D.L. & Medsger, T.A., 1988. Clinical correlations and prognosis based on serum autoantibodies in patients with systemic sclerosis. *Arthritis and Rheumatism*, 31(2), pp.196–203.

Stegmann, A.P., Honders, M.W., Willemze, R. & Landegent, J.E., 1995. De novo induced mutations in the deoxycytidine kinase (dck) gene in rat leukemic clonal cell lines confer resistance to cytarabine (AraC) and 5-Aza-2'-deoxycytidine (DAC). *Leukemia*, 9(6), pp.1032–1038.

Stewart, J.P., Egan, J.J., Ross, A.J., Kelly, B.G., Lok, S.S., Hasleton, P.S. & Woodcock, A.A., 1999. The detection of Epstein-Barr virus DNA in lung tissue from patients with idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 159(4 Pt 1), pp.1336–1341.

Stuckenholz, C., Lu, L., Thakur, P.C., Choi, T.Y., Shin, D. & Bahary, N., 2013. Sfrp5 Modulates Both Wnt and BMP Signaling and Regulates Gastrointestinal Organogensis in the Zebrafish, Danio rerio. *PLoS ONE*, 8(4), pp.e62470.

Suga, M., Iyonaga, K., Okamoto, T., Gushima, Y., Miyakawa, H., Akaike, T. & Ando, M., 2000. Characteristic elevation of matrix metalloproteinase activity in idiopathic interstitial pneumonias. *American Journal of Respiratory and Critical Care Medicine*, 162(5), pp.1949–1956.

Sun, Y. V, Lazarus, A., Smith, J.A., Chuang, Y.-H., Zhao, W., Turner, S.T. & Kardia, S.L.R., 2013. Gene-specific DNA methylation association with serum levels of C-reactive protein in African americans. *PloS One*, 8(8), p.e73480.

Suter, M., Ma, J., Harris, A., Patterson, L., Brown, K.A., Shope, C., Showalter, L., Abramovici, A. & Aagaard-Tillery, K.M., 2011. Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. *Epigenetics*, 6(11), pp.1284–1294.

Swigris, J.J., Olson, A.L., Huie, T.J., Fernandez-Perez, E.R., Solomon, J., Sprunger, D. & Brown, K.K., 2012. Ethnic and racial differences in the presence of idiopathic pulmonary fibrosis at death. *Respiratory Medicine*, 106(4), pp.588–593.

Szamosi, S., Szekanecz, Z. & Szucs, G., 2006. Gastrointestinal manifestations in Hungarian scleroderma patients. *Rheumatology International*, 26(12), pp.1120–1124.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J. & von Mering, C., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 43, pp.D447–52.

Tabas-Madrid, D., Nogales-Cadenas, R. & Pascual-Montano, A., 2012. GeneCodis3: A non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Research*, 40(W1), pp.W478–W483.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. & Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324(5929), pp.930–935.

Takebayashi, S., Nakao, M., Fujita, N., Sado, T., Tanaka, M., Taguchi, H. & Okumura, K., 2001. 5-Aza-2'-deoxycytidine induces histone hyperacetylation of mouse centromeric heterochromatin by a mechanism

independent of DNA demethylation. *Biochemical and Biophysical Research Communications*, 288, pp.921–926.

Takeda, N. & Maemura, K., 2011. Circadian clock and cardiovascular disease. *Journal of Cardiology*, 57(3), pp.249–256.

Tampe, B. & Zeisberg, M., 2013. Contribution of genetics and epigenetics to progression of kidney fibrosis. *Nephrology, Dialysis, Transplantation, 29(*4), pp.72-79.

Tampe, B., Tampe, D., Zeisberg, E.M., Müller, G.A., Bechtel-Walz, W., Koziolek, M., Kalluri, R. & Zeisberg, M., 2015. Induction of Tet3-dependent Epigenetic Remodeling by Low-dose Hydralazine Attenuates Progression of Chronic Kidney Disease. *EBioMedicine*, 2(1), pp.19–36.

Tan, A., Denton, C.P., Mikhailidis, D.P. & Seifalian, A.M., 2011. Recent advances in the diagnosis and treatment of interstitial lung disease in systemic sclerosis (scleroderma): a review. *Clinical and Experimental Rheumatology*, 29(2 Supplement 65), pp.S66–74.

Tanaka, S., Suto, A., Ikeda, K., Sanayama, Y., Nakagomi, D., Iwamoto, T., Suzuki, K., Kambe, N., Matsue, H., Matsumura, R., Kashiwakuma, D., Iwamoto, I. & Nakajima, H., 2013. Alteration of circulating miRNAs in SSc: MiR-30b regulates the expression of PDGF receptor β. *Rheumatology,* 52(11), pp.1963–1972.

Tang, Y.W., Johnson, J.E., Browning, P.J., Cruz-Gervis, R.A., Davis, A., Graham, B.S., Brigham, K.L., Oates, J.A., Loyd, J.E. & Stecenko, A.A., 2003. Herpesvirus DNA is consistently detected in lungs of patients with idiopathic pulmonary fibrosis. *Journal of Clinical Microbiology*, 41(6), pp.2633–2640.

Tanjore, H., Blackwell, T.S. & Lawson, W.E., 2012. Emerging evidence for endoplasmic reticulum stress in the pathogenesis of idiopathic pulmonary fibrosis. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 302(8), pp.L721–9.

Tanjore, H., Cheng, D.-S., Degryse, A.L., Zoz, D.F., Abdolrasulnia, R., Lawson, W.E. & Blackwell, T.S., 2011. Alveolar epithelial cells undergo epithelial-to-mesenchymal transition in response to endoplasmic reticulum stress. *The Journal of Biological Chemistry*, 286(35), pp.30972–30980.

Tao, M.H., Marian, C., Shields, P.G., Nie, J., McCann, S.E., Millen, A., Ambrosone, C., Hutson, A., Edge, S.B., Krishnan, S.S., Xie, B., Winston, J., Vito, D., Russell, M., Nochajski, T.H., Trevisan, M. & Freudenheim, J.L., 2011. Alcohol consumption in relation to aberrant DNA methylation in breast tumors. *Alcohol,* 45(7), pp.689–99.

Tapp, H.S., Commane, D.M., Bradburn, D.M., Arasaradnam, R., Mathers, J.C., Johnson, I.T. & Belshaw, N.J., 2013. Nutritional factors and gender influence age-related DNA methylation in the human rectal mucosa. *Aging Cell*, 12(1), pp.148–155.

Tarin, D. & Croft, C.B., 1970. Ultrastructural studies of wound healing in mouse skin. II. Dermo-epidermal interrelationships. *Journal of Anatomy*, 106(Pt 1), pp.79–91.

Tashkin, D.P., Elashoff, R., Clements, P.J., Roth, M.D., Furst, D.E., Silver, R.M., Goldin, J., Arriola, E., Strange, C., Bolster, M.B., Seibold, J.R., Riley, D.J., Hsu, V.M., Varga, J., Schraufnagel, D., Theodore, A., Simms, R., Wise, R., Wigley, F., White, B., Steen, V., Read, C., Mayes, M., Parsley, E., Mubarak, K., Connolly, M.K., Golden, J., Olman, M., Fessler, B., Rothfield, N., Metersky, M., Khanna, D., Li, N. & Li, G., 2007. Effects of 1-year treatment with cyclophosphamide on outcomes at 2 years in scleroderma lung disease. *American Journal of Respiratory and Critical Care Medicine*, 176(10), pp.1026–1034.

Taskar, V.S. & Coultas, D.B., 2006. Is idiopathic pulmonary fibrosis an environmental disease? *Proceedings of the American Thoracic Society*, 3(4), pp.293–298.

Tatler, A.L. & Jenkins, G., 2012. TGF-ß Activation and Lung Fibrosis. *Proceedings of the American Thoracic Society*, 9(3), pp.130–136.

Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Weisenberger, D.J., Shen, H., Campan, M., Noushmehr, H., Bell, C.G., Maxwell, A.P., Savage, D.A., Mueller-Holzner, E., Marth, C., Kocjan, G., Gayther, S.A., Jones, A., Beck, S., Wagner, W., Laird, P.W., Jacobs, I.J. & Widschwendter, M., 2010. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Research*, 20(4), pp.440–446.

Thannickal, V.J. & Horowitz, J.C., 2006. Evolving concepts of apoptosis in idiopathic pulmonary fibrosis. *Proceedings of the American Thoracic Society*, 3(4), pp.350–356.

Thomas, B., Matson, S., Chopra, V., Sun, L., Sharma, S., Hersch, S., Diana Rosas, H., Scherzer, C., Ferrante, R. & Matson, W., 2013. A novel method for detecting 7-methyl guanine reveals aberrant methylation levels in Huntington disease. *Analytical Biochemistry*, 436(2), pp.112–120.

Thompson, A.E. & Pope, J.E., 2002. Increased prevalence of scleroderma in Southwestern Ontario: A cluster analysis. *Journal of Rheumatology*, 29(9), pp.1867–1873.

Thoua, N.M., Bunce, C., Brough, G., Forbes, A., Emmanuel, A. V & Denton, C.P., 2010. Assessment of gastrointestinal symptoms in patients with systemic sclerosis in a UK tertiary referral centre. *Rheumatology*, 49(9), pp.1770–1775.

Ting, A.H., McGarvey, K.M. & Baylin, S.B., 2006. The cancer epigenome--components and functional correlates. *Genes and Development*, 20(23), pp.3215–3231.

Tingstrom, A., Heldin, C.H. & Rubin, K., 1992. Regulation of fibroblast-mediated collagen gel contraction by platelet- derived growth factor, interleukin-1 alpha and transforming growth factor-beta 1. *Journal of Cell Science*, 102(Pt 2), pp.315–322.

Tomasek, J.J., Gabbiani, G., Hinz, B., Chaponnier, C. & Brown, R.A., 2002. Myofibroblasts and mechano-regulation of connective tissue remodelling. *Nature Reviews. Molecular Cell Biology*, 3(5), pp.349–363.

Tomasek, J.J., Haaksma, C.J., Eddy, R.J. & Vaughan, M.B., 1992. Fibroblast contraction occurs on release of tension in attached collagen lattices: dependency on an organized actin cytoskeleton and serum. *The Anatomical Record*, 232(3), pp.359–368.

Tomcik, M., Palumbo-Zerr, K., Zerr, P., Avouac, J., Dees, C., Sumova, B., Distler, A., Beyer, C., Cerezo, L.A., Becvar, R., Distler, O., Grigorian, M., Schett, G., Senolt, L. & Distler, J.H.W., 2014. S100A4 amplifies TGF-β-induced fibroblast activation in systemic sclerosis. *Annals of the Rheumatic Diseases*, 74(9), pp.1748-1755.

Trang, G., Steele, R., Baron, M. & Hudson, M., 2012. Corticosteroids and the risk of scleroderma renal crisis: A systematic review. *Rheumatology International*, 32(3), pp.645–653.

Travis, W.D., Costabel, U., Hansell, D.M., King, T.E., 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., 2013. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *American Journal of Respiratory and Critical Care Medicine*, 188(6), pp.733–748.

Trebaul, A., Chan, E.K. & Midwood, K.S., 2007. Regulation of fibroblast migration by tenascin-C. *Biochemical Society Transactions*, 35(Pt 4), pp.695–697.

Tsakiri, K.D., Cronkhite, J.T., Kuan, P.J., Xing, C., Raghu, G., Weissler, J.C., Rosenblatt, R.L., Shay, J.W. & Garcia, C.K., 2007. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18), pp.7552–7557.

Tsuchiya, N., Ito, I. & Kawasaki, A., 2010. Association of IRF5, STAT4 and BLK with systemic lupus erythematosus and other rheumatic diseases. *Japanese Journal of Clinical Immunology*, 33(2), pp.57–65.

Turner, R., Lipshutz, W., Miller, W., Rittenberg, G., Schumacher, H.R. & Cohen, S., 1973. Esophageal dysfunction in collagen disease. *The American Journal of the Medical Sciences*, 265(3), pp.191-199.

Uber, C.L., McReynolds, R.A. & Kuhn, C., 1982. Immunotoxicology of Silica. *CRC Critical Reviews in Toxicology*, 10(4), pp.303–319.

Udwadia, Z.F., Mullerpattan, J.B., Balakrishnan, C. & Richeldi, L., 2015. Improved pulmonary function following pirfenidone treatment in a patient with progressive interstitial lung disease associated with systemic sclerosis. *Lung India*, 32(1), pp.50–52.

Ueda, T., Ohta, K., Suzuki, N., Yamaguchi, M., Hirai, K., Horiuchi, T., Watanabe, J., Miyamoto, T. & Ito, K., 1992. Idiopathic pulmonary fibrosis and high prevalence of serum antibodies to hepatitis C virus. *The American Review of Respiratory Disease*, 146(1), pp.266–268.

Vallyathan, V., Bergeron, W.N., Robichaux, P.A. & Craighead, J.E., 1982. Pulmonary fibrosis in an aluminium arc welder. *Chest*, 81(3), pp.372-374.

Van Almen, G.C., Verhesen, W., van Leeuwen, R.E.W., van de Vrie, M., Eurlings, C., Schellings, M.W.M., Swinnen, M., Cleutjens, J.P.M., van Zandvoort, M.A.M.J., Heymans, S. & Schroen, B., 2011. MicroRNA-18 and microRNA-19 regulate CTGF and TSP-1 expression in age-related heart failure. *Aging Cell*, 10(5), pp.769–779.

Van der Rest, M. & Garrone, R., 1991. Collagen family of proteins. T*he Federation of American Societies for Experimental Biology Journal*, 5(13), pp.2814–2823.

Vanyushin, B.F., Nemirovsky, L.E., Klimenko, V.V., Vasiliev, V.K. & Belozersky, A.N., 1973. The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontologia*, 19(3), pp.138-152.

Varga, J. & Abraham, D., 2007. Systemic sclerosis: A prototypic multisystem fibrotic disorder. *Journal of Clinical Investigation*, 117(3), pp.557–567.

Varga, J. & Pasche, B., 2009. Transforming growth factor beta as a therapeutic target in systemic sclerosis. *Nature Reviews. Rheumatology*, 5(4), pp.200–206.

Vaughan, J.H., Shaw, P.X., Nguyen, M.D., Medsger, T.A. Jr., Wright, T.M., Metcalf, J.S. & Leroy, E.C., 2000. Evidence of activation of 2 herpesviruses, Epstein-Barr virus and cytomegalovirus, in systemic sclerosis and normal skins. *The Journal of Rheumatology*, 27(3), pp.821-3.

Veeck, J., Geisler, C., Noetzel, E., Alkaya, S., Hartmann, A., Knüchel, R. & Dahl, E., 2008. Epigenetic inactivation of the secreted frizzled-related protein-5 (SFRP5) gene in human breast cancer is associated with unfavorable prognosis. *Carcinogenesis*, 29(5), pp.991–998.

Venkatesan, N., Punithavathi, D. & Chandrakasan, G., 1998. Glycoprotein composition in cyclophosphamide-induced lung fibrosis. *Biochimica et Biophysica Acta*, 1407(2), pp.125–134.

Vergnon, J.M., Weynants, P., Vincent, M., de Thé, G., Mornex, J.F. & Brune, J., 1984. Cryptogenic fibrosing alveolitis and Epstein-Barr virus: An association? *Lancet*, 2(8406), pp.768–770.

Vernon, R.B. & Gooden, M.D., 2002. An improved method for the collagen gel contraction assay. *In Vitro Cellular and Developmental Biology*, 38(2), p.97.

Vettori, S., 2012. Role of MicroRNAs in Fibrosis. *The Open Rheumatology Journal*, 6(1), pp.130–139.

Voelkel, N.F., Vandivier, R.W. & Tuder, R.M., 2006. Vascular endothelial growth factor in the lung. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 290(2), pp.L209–L221.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A. & Kinzler, K.W., 2013. Cancer genome landscapes. *Science*, 339(6127), pp.1546–58.

Von Zglinicki, T., 2002. Oxidative stress shortens telomeres. *Trends in Biochemical Sciences*, 27(7), pp.339–344.

Vuga, L.J., Ben-Yehudah, A., Kovkarova-Naumovski, E., Oriss, T., Gibson, K.F., Feghali-Bostwick, C. & Kaminski, N., 2009. WNT5A is a regulator of fibroblast proliferation and resistance to apoptosis. *American Journal of Respiratory Cell and Molecular Biology*, 41(5), pp.583–589.

Wagner, J.R., Busche, S., Ge, B., Kwan, T., Pastinen, T. & Blanchette, M., 2014. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biology*, 15(2), p.R37.

Walsh, C.P. & Bestor, T.H., 1999. Cytosine methylation and mammalian development. *Genes & Development*, 13(1), pp.26–34.

Walter, N., Collard, H.R. & King, T.E., 2006. Current perspectives on the treatment of idiopathic pulmonary fibrosis. *Proceedings of the American Thoracic Society*, 3(4), pp.330–338.

Wan, J., Oliver, V.F., Wang, G., Zhu, H., Zack, D.J., Merbs, S.L. & Qian, J., 2015. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics*, 16(1), p.49.

Wang, B., Koh, P., Winbanks, C., Coughlan, M.T., McClelland, A., Watson, A., Jandeleit-Dahm, K., Burns, W.C., Thomas, M.C., Cooper, M.E. & Kantharidis, P., 2011. MiR-200a prevents renal fibrogenesis through repression of TGF-beta2 expression. *Diabetes*, 60(1), pp.280–287.

Wang, Y. & Kahaleh, B., 2013. Epigenetic repression of bone morphogenetic protein receptor II expression in scleroderma. *Journal of Cellular and Molecular Medicine*, 17(10), pp.1291–1299.

Wang, Y. & Zhou, B.P., 2011. Epithelial-mesenchymal transition in breast cancer progression and metastasis. *Chinese Journal of Cancer*, 30(9), pp.603–611.

Wang, Y., Fan, P.S. & Kahaleh, B., 2006. Association between enhanced type I collagen expression and epigenetic repression of the FLI1 gene in scleroderma fibroblasts. *Arthritis and Rheumatism*, 54(7), pp.2271–2279.

Wang, Y., Kuan, P.J., Xing, C., Cronkhite, J.T., Torres, F., Rosenblatt, R.L., DiMaio, J.M., Kinch, L.N., Grishin, N. V. & Garcia, C.K., 2009. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *American Journal of Human Genetics*, 84(1), pp.52–59.

Wang, Z., Jinnin, M., Kudo, H., Inoue, K., Nakayama, W., Honda, N., Makino, K., Kajihara, I., Fukushima, S., Inoue, Y. & Ihn, H., 2013. Detection of hair-microRNAs as the novel potent biomarker: Evaluation of the usefulness for the diagnosis of scleroderma. *Journal of Dermatological Science*, 72(2), pp.134–141.

Watson, C.J., Collier, P., Tea, I., Neary, R., Watson, J.A., Robinson, C., Phelan, D., Ledwidge, M.T., Mcdonald, K.M., Mccann, A., Sharaf, O. & Baugh, J.A., 2014. Hypoxia-induced epigenetic modifications are associated with cardiac tissue fibrosis and the development of a myofibroblast-like phenotype. *Human Molecular Genetics,* 23(8), pp.2176–2188.

Wedzicha, J.A., 2004. Role of viruses in exacerbations of chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 1(2), pp.115–120.

Wei, J., Fang, F., Lam, A.P., Sargent, J.L., Hamburg, E., Hinchcliff, M.E., Gottardi, C.J., Atit, R., Whitfield, M.L. & Varga, J., 2012. Wnt/β-catenin signaling is hyperactivated in systemic sclerosis and induces Smad-dependent fibrotic responses in mesenchymal cells. *Arthritis and Rheumatism*, 64(8), pp.2734–2745.

Weichert, N., Kaltenborn, E., Hector, A., Woischnik, M., Schams, A., Holzinger, A., Kern, S. & Griese, M., 2011. Some ABCA3 mutations elevate ER stress and initiate apoptosis of lung epithelial cells. *Respiratory Research*, 12, p.4.

Wells, A.U. & Hirani, N., 2008. Interstitial lung disease guideline. *Thorax*, 63(5), p.1029.

Wells, A.U., Hansell, D.M., Rubens, M.B., Cailes, J.B., Black, C.M. & Du Bois, R.M., 1997. Functional impairment in lone cryptogenic fibrosing alveolitis and fibrosing alveolitis associated with systemic sclerosis: A comparison. *American Journal of Respiratory and Critical Care Medicine*, 155(5), pp.1657–1664.

Wells, A.U., Hansell, D.M., Rubens, M.B., Cullinan, P., Haslam, P.L., Black, C.M. & Du Bois, R.M., 1994. Fibrosing alveolitis in systemic sclerosis: Bronchoalveolar lavage findings in relation to computed tomographic appearance. *American Journal of Respiratory and Critical Care Medicine*, 150(2), pp.462–468.

Wells, R.G., 2008. The role of matrix stiffness in regulating cell behavior. *Hepatology*, 47(4), pp.1394–1400.

West, J.B. & Mathieu-Costello, O., 1999. Structure, strength and failure of the pulmonary blood-gas barrier. *European Respiratory Monograph*, 4(12), pp.171–202.

White, B., 1996. Immunopathogenesis of systemic sclerosis. *Rheumatic Disease Clinics*, 22(4), pp.695–708.

Whyte, M., Hubbard, R., Meliconi, R., Whidborne, M., Eaton, V., Bingle, C., Timms, J., Duff, G., Facchini, A., Pacilli, A., Fabbri, M., Hall, I., Britton, J., Johnston, I. & Di, G.F., 2000. Increased risk of fibrosing alveolitis associated with interleukin-1 receptor antagonist and tumor necrosis factor-alpha gene polymorphisms. *American Journal of Respir Critical Care Medicine*, 162(2 Pt1), pp.755–758.

Wickham, H., 2010. A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*, 19(1), pp.3–28.

Wight, T.N., Kinsella, M.G. & Qwarnström, E.E., 1992. The role of proteoglycans in cell adhesion, migration and proliferation. *Current Opinion in Cell Biology*, 4(5), pp.793–801.

Wijermans, P., Lübbert, M., Verhoef, G., Bosly, A., Ravoet, C., Andre, M. & Ferrant, A., 2000. Low-dose 5-Aza-2'-deoxycytidine, a DNA hypomethylating agent, for the treatment of high-risk myelodysplastic syndrome: a multicenter phase II study in elderly patients. *Journal of Clinical Oncology*, 18(5), pp.956-962.

Wijesuriya, S.D., Bristow, J. & Miller, W.L., 2002. Localization and analysis of the principal promoter for human tenascin-X. *Genomics*, 80(4), pp.443–452.

Wilson, V.L., Smith, R.A., Ma, S. & Cutler, R.G., 1987. Genomic 5-methyldeoxycytidine decreases with age. *Journal of Biological Chemistry*, 262(21), pp.9948–9951.

Winkelmann, J., Lin, L., Schormair, B., Kornum, B.R., Faraco, J., Plazzi, G., Melberg, A., Cornelio, F., Urban, A.E., Pizza, F., Poli, F., Grubert, F., Wieland, T., Graf, E., Hallmayer, J., Strom, T.M. & Mignot, E., 2012. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Human Molecular Genetics,* 21(10), pp.2205–2210.

Wolff, G.L., Kodell, R.L., Moore, S.R. & Cooney, C.A., 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *The Federation of American Societies for Experimental Biology Journal*, 12(11), pp.949–957.

Wootton, S.C., Kim, D.S., Kondoh, Y., Chen, E., Lee, J.S., Song, J.W., Huh, J.W., Taniguchi, H., Chiu, C., Boushey, H., Lancaster, L.H., Wolters, P.J., DeRisi, J., Ganem, D. & Collard, H.R., 2011. Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 183(12), pp.1698–1702.

Wu, Y.J., La Pierre, D.P., Wu, J., Yee, A.J. & Yang, B.B., 2005. The interaction of versican with its binding partners. *Cell Research*, 15(7), pp.483–494.

Wynn, T.A., 2007. Common and unique mechanisms regulate fibrosis in various fibroproliferative diseases. *Journal of Clinical Investigation*, 117(3), pp.524–529.

Xaubet, A., Marin-Arguedas, A., Lario, S., Ancochea, J., Morell, F., Ruiz-Manzano, J., Rodriguez-Becerra, E., Rodriguez-Arias, J.M., Inigo, P., Sanz, S., Campistol, J.M., Mullol, J. & Picado, C., 2003. Transforming growth factor-beta1 gene polymorphisms are associated with disease progression in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 168(4), pp.431–435.

Xiao, Y., Word, B., Starlard-Davenport, A., Haefele, A., Lyn-Cook, B.D. & Hammons, G., 2008. Age and gender affect DNMT3a and DNMT3b expression in human liver. *Cell Biology and Toxicology*, 24(3), pp.265–272.

Xu, H., Wang, F., Liu, Y., Yu, Y., Gelernter, J. & Zhang, H., 2014. Sex-biased methylome and transcriptome in human prefrontal cortex. *Human Molecular Genetics,* 23(5), pp.1260–1270.

Xu, J., Lamouille, S. & Derynck, R., 2009. TGF-beta-induced epithelial to mesenchymal transition. *Cell Research*, 19(2), pp.156–172.

Yamakage, A., Ishikawa, H., Saito, Y. & Hattori, A., 1980. Occupational scleroderma-like disorder occurring in men engaged in the polymerization of epoxy resins. *Dermatologica*, 161(1), pp.33-44.

Yamane, K., Ihn, H., Kubo, M. & Tamaki, K., 2002. Increased transcriptional activities of transforming growth factor β receptors in scleroderma fibroblasts. *Arthritis and Rheumatism*, 46(9), pp.2421–2428.

Yamashita, C.M., Dolgonos, L., Zemans, R.L., Young, S.K., Robertson, J., Briones, N., Suzuki, T., Campbell, M.N., Gauldie, J., Radisky, D.C., Riches, D.W.H., Yu, G., Kaminski, N., McCulloch, C.A.G. & Downey, G.P., 2011. Matrix metalloproteinase 3 is a mediator of pulmonary fibrosis. *American Journal of Pathology*, 179(4), pp.1733–1745.

Yanaba, K., Yoshizaki, A., Muroi, E., Hara, T., Ogawa, F., Shimizu, K., Hasegawa, M., Fujimoto, M., Takehara, K. & Sato, S., 2010. CCL13 is a promising diagnostic marker for systemic sclerosis. *British Journal of Dermatology*, 162(2), pp.332–336.

Yanagishita, M., 1993. Function of proteoglycans in the extracellular matrix. *Acta Pathologica Japonica*, 43(6), pp.283–293.

Yang, F., Zhang, L., Li, J., Huang, J., Wen, R., Ma, L., Zhou, D. & Li, L., 2010. Trichostatin A and 5-Azacytidine both cause an increase in global histone H4 acetylation and a decrease in global DNA and H3K9 methylation during mitosis in maize. *BMC Plant Biology*, 10, p.178.

Yang, I. V., Pedersen, B.S., Rabinovich, E., Hennessy, C.E., Davidson, E.J., Murphy, E., Guardela, B.J., Tedrow, J.R., Zhang, Y., Singh, M.K., Correll, M., Schwarz, M.I., Geraci, M., Sciurba, F.C., Quackenbush, J., Spira, A., Kaminski, N. & Schwartz, D.A., 2014. Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 190(11), pp.1263–1272.

Yang, Z., Mu, Z., Dabovic, B., Jurukovski, V., Yu, D., Sung, J., Xiong, X. & Munger, J.S., 2007. Absence of integrin-mediated TGFbeta1 activation in vivo recapitulates the phenotype of TGFbeta1-null mice. *The Journal of Cell Biology*, 176(6), pp.787–793.

Yazawa, N., Fujimoto, M., Kikuchi, K., Kubo, M., Ihn, H., Sato, S., Tamaki, T. & Tamaki, K., 1998. High seroprevalence of Helicobacter pylori infection in patients with systemic sclerosis: Association with esophageal involvement. *Journal of Rheumatology*, 25(4), pp.650–653.

Yi, L., Wang, J.C., Guo, X.J., Gu, Y.H., Tu, W.Z., Guo, G., Yang, L., Xiao, R., Yu, L., Mayes, M.D., Assassi, S., Jin, L., Zou, H.J. & Zhou, X.D., 2013. STAT4 is a genetic risk factor for systemic sclerosis in a Chinese population. *International Journal of Immunopathology and Pharmacol*ogy, 26(2), pp.473–478.

Yokoyama, A., Kondo, K., Nakajima, M., Matsushima, T., Takahashi, T., Nishimura, M., Bando, M., Sugiyama, Y., Totani, Y., Ishizaki, T., Ichiyasu, H., Suga, M., Hamada, H. & Kohno, N., 2006. Prognostic value of circulating KL-6 in idiopathic pulmonary fibrosis. *Respirology*, 11(2), pp.164–168.

Young, L.R., Nogee, L.M., Barnett, B., Panos, R.J., Colby, T. V. & Deutsch, G.H., 2008. Usual interstitial pneumonia in an adolescent with ABCA3 mutations. *Chest*, 134(1), pp.192–195.

Yu, W.H. & Woessner, J.F., 2000. Heparan sulfate proteoglycans as extracellular docking molecules for matrilysin (matrix metalloproteinase 7). *Journal of Biological Chemistry*, 275(6), pp.4183–4191.

Yuan, Y., Nymoen, D.A., Stavnes, H.T., Rosnes, A.K., Bjorang, O., Wu, C., Nesland, J.M. & Davidson, B., 2009. Tenascin-X is a novel diagnostic marker of malignant mesothelioma. *American Journal of Surgical Pathology*, 33(11), pp.1673–1682.

Zakrzewska, K., Corcioli, F., Carlsen, K.M., Giuggioli, D., Fanci, R., Rinieri, A., Ferri, C. & Azzi, A., 2009. Human parvovirus B19 (B19V) infection in systemic sclerosis patients. *Intervirology*, 52(5), pp.279–282.

Zanelli, R., Barbic, F., Migliori, M. & Michetti, G., 1994. Uncommon evolution of fibrosing alveolitis in a hard metal grinder exposed to cobalt dusts. *Science of the Total Environment*, 150(1-3), pp. 225–229.

Zhang, F.F., Cardarelli, R., Carroll, J., Fulda, K.G., Kaur, M., Gonzalez, K., Vishwanatha, J.K., Santella, R.M. & Morabia, A., 2011. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics*, 6(5), pp.623–629.

Zhang, Y., Lee, T.C., Guillemin, B., Yu, M.C. & Rom, W.N., 1993. Enhanced IL-1 beta and tumor necrosis factor-alpha release and messenger RNA expression in macrophages from idiopathic pulmonary fibrosis or after asbestos exposure. *Journal of Immunology*, 150(9), pp.4188–4196.

Zhao, Q., Qin, C.Y., Zhao, Z.H., Fan, Y.C. & Wang, K., 2013. Epigenetic modifications in hepatic stellate cells contribute to liver fibrosis. *The Tohoku Journal of Experimental Medicine*, 229(1), pp.35–43.

Zhong, Q., Zhou, B., Ann, D.K., Minoo, P., Liu, Y., Banfalvi, A., Krishnaveni, M.S., Dubourd, M., Demaio, L., Willis, B.C., Kim, K.J., duBois, R.M., Crandall, E.D., Beers, M.F. & Borok, Z., 2011. Role of endoplasmic reticulum stress in epithelial-mesenchymal transition of alveolar epithelial cells: effects of misfolded surfactant protein. *American Journal of Respiratory Cell and Molecular Biology*, 45(3), pp.498–509.

Zhou, L., Jiang, L., Li, Z. & Kang, J., 2010. Change of matrix metalloproteinase-1 and matrix metalloproteinase-7 in serum and bronchoalveolar lavage fluid of patients with idiopathic pulmonary fibrosis and sarcoidosis. *Chinese Journal of Tuberculosis and Respiratory Diseases*, 33(6), pp.441–444.

Zhou, X., Tan, F.K., Reveille, J.D., Wallis, D., Milewicz, D.M., Ahn, C., Wang, A. & Arnett, F.C., 2002. Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. *Arthritis and Rheumatism*, 46(11), pp.2990–2999.

Zhou, X.D., Xiong, M.M., Tan, F.K., Guo, X.J. & Arnett, F.C., 2006. SPARC, an upstream regulator of connective tissue growth factor in response to transforming growth factor beta stimulation. *Arthritis and Rheumatism*, 54(12), pp.3885–3889.

Zhu, H., Li, Y., Qu, S., Luo, H., Zhou, Y., Wang, Y., Zhao, H., You, Y., Xiao, X. & Zuo, X., 2012. MicroRNA expression abnormalities in limited cutaneous scleroderma and diffuse cutaneous scleroderma. *Journal of Clinical Immunology*, 32(3), pp.514–522.

Zhu, H., Luo, H. & Zuo, X., 2013. MicroRNAs: their involvement in fibrosis pathogenesis and use as diagnostic biomarkers in scleroderma. *Experimental and Molecular Medicine*, 45(9), p.e41.

Zhu, J.-K., 2009. Active DNA demethylation mediated by DNA glycosylases. *Annual Review of Genetics*, 43, pp.143–166.

Zhu, W., Srinivasan, K., Dai, Z., Duan, W., Druhan, L.J., Ding, H., Yee, L., Villalona-Calero, M.A., Plass, C. & Otterson, G.A., 2003. Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the p21(Cip1) promoter. *Molecular and Cellular Biology*, 23(12), pp.4056–65.

Zibrak, J.D. & Price, D., 2014. Interstitial lung disease: raising the index of suspicion in primary care. *Primary Care Respiratory Medicine*, 2454(10).

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. & Henikoff, S., 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, 39(1), pp.61–69.

Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., Gnirke, A. & Meissner, A., 2011. Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types. *PLoS Genetics*, 7(12), pp.e1002389.

Zöchbauer-Müller, S., Minna, J.D. & Gazdar, A.F., 2002. Aberrant DNA methylation in lung cancer: biological and clinical implications. *The Oncologist*, 7(5), pp.451–457.

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. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), pp.6292–6297.

Zweers, M.C., Bristow, J., Steijlen, P.M., Dean, W.B., Hamel, B.C., Otero, M. & Schalkwijk, J., 2003. Haploinsufficiency of TNXB is associated with hypermobility type of ehlers-danlos syndrome. *American Journal of Human Genetics*, *73*(1), pp.214–217.

Appendices

Appendice A.

Genes with both positive and negative correlation between methylation and expression.

Symbol	Name	Symbol	Name
ACVRL1	activin A receptor type II-like 1	MGMT	O-6-methylguanine-DNA methyltransferase
ADAM15	ADAM metallopeptidase domain 15	MYOM2	myomesin (M-protein) 2, 165kDa
ADAMTS2	ADAM metallopeptidase with thrombospondin type 1 motif, 2	NETO2	neuropilin (NRP) and tolloid (TLL)-like 2
ALDH3A1	aldehyde dehydrogenase 3 family, memberA1	NLGN4Y	neuroligin 4, Y-linked
ANO1	anoctamin 1, calcium activated chloride channel	NPTX1	neuronal pentraxin I
C13orf15	chromosome 13 open reading frame 15	PAX8	paired box 8
C1orf159	chromosome 1 open reading frame 159	PGM3	phosphoglucomutase 3
CA12	carbonic anhydrase XII	PLA2G5	phospholipase A2, group V
CDH13	cadherin 13, H-cadherin (heart)	PLAG1	pleiomorphic adenoma gene 1
CHST15	carbohydrate (N-acetylgalactosamine 4-sulfate 6-O) sulfotransferase 15	PLAGL1	pleiomorphic adenoma gene-like 1
CLEC14A	C-type lectin domain family 14, member A	PPP3CA	protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform
CPNE8	copine VIII	PRKY	protein kinase, Y-linked
CRIPAK	cysteine-rich PAK1 inhibitor	RAMP1	receptor (G protein-coupled) activity modifying protein 1
DLL1	delta-like 1 (Drosophila)	RPS4Y1	ribosomal protein S4, Y-linked 1
EIF1AY	eukaryotic translation initiation factor 1A, Y-linked	SAMD14	sterile alpha motif domain containing 14
FAM13A	family with sequence similarity 13, member A	SASH1	SAM and SH3 domain containing 1
FST	follistatin	SPON2	spondin 2, extracellular matrix protein
GPER	G protein-coupled estrogen receptor 1	STX18	syntaxin 18
GSTT1	glutathione S-transferase theta 1	TANC1	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1
ICMT	isoprenylcysteine carboxyl methyltransferase	TNFAIP8L3	tumor necrosis factor, alpha-induced protein 8-like 3
IL16	interleukin 16 (lymphocyte chemoattractant factor)	TRIM56	tripartite motif-containing 56
MACF1	microtubule-actin crosslinking factor 1	ZFHX4	zinc finger homeobox 4
MAPRE1	microtubule-associated protein, RP/EB family, member 1	ZFY	zinc finger protein, Y-linked

Appendice B.

Differentially methylated genes in fibrotic compared to control lung fibroblasts involved in the Wnt signalling pathway as determined by GO-term and KEGG enrichment.

Symbol	Name	Symbol	Name
APC	Adenomatous polyposis coli	**MCC**	Mutated in colorectal cancers
ARL6	ADP-ribosylation factor-like 6	**MITF**	Microphthalmia-associated transcription factor
AXIN2	Axin 2	**NDRG2**	NDRG family member 2
BAMBI	Hypothetical LOC729590; BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	**NFATC1**	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1
BCL9	B-cell CLL/lymphoma 9	**NFATC2**	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2
BRD7	Bromodomain containing 7; bromodomain containing 7 pseudogene 2	**NKD2**	Naked cuticle homolog 2 (Drosophila)
BTRC	Beta-transducin repeat containing	**NXN**	Nucleoredoxin
C1orf187	Chromosome 1 open reading frame 187	**PLCB1**	Phospholipase C, beta 1 (phosphoinositide-specific)
CACYBP	Similar to calcyclin binding protein; calcyclin binding protein	**PLCB2**	Phospholipase C, beta 2
CALCOCO1	Calcium binding and coiled-coil domain 1	**PLCB3**	Phospholipase C, beta 3 (phosphatidylinositol-specific)
CAMK2A	Calcium/calmodulin-dependent protein kinase II alpha	**PPP2R5A**	Protein phosphatase 2, regulatory subunit B', alpha isoform
CAMK2B	Calcium/calmodulin-dependent protein kinase II beta	**PPP2R5C**	Protein phosphatase 2, regulatory subunit B', gamma isoform
CAMK2G	Calcium/calmodulin-dependent protein kinase II gamma	**PPP3CA**	Protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform
CCDC88C	Coiled-coil domain containing 88C	**PRICKLE1**	Prickle homolog 1 (Drosophila)
CCND3	Cyclin D3	**PRICKLE2**	Prickle homolog 2 (Drosophila)
CCNY	Cyclin Y	**PRKCA**	Protein kinase C, alpha
CD44	CD44 molecule (Indian blood group)	**PRKCB**	Protein kinase C, beta
CDK14	PFTAIRE protein kinase 1	**PRKCG**	Protein kinase C, gamma
CELSR2	Cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)	**RNF146**	Ring finger protein 146
CPZ	Carboxypeptidase Z	**RSPO2**	R-spondin 2 homolog (Xenopus laevis)
CREBBP	CREB binding protein	**SENP2**	SUMO1/sentrin/SMT3 specific peptidase 2
CSNK1A1L	Casein kinase 1, alpha 1-like	**SFRP1**	Secreted frizzled-related protein 1
CSNK1D	Casein kinase 1, delta	**SFRP2**	Secreted frizzled-related protein 2
CSNK1G1	Casein kinase 1, gamma 1	**SFRP5**	Secreted frizzled-related protein 5
CSNK1G3	Casein kinase 1, gamma 3	**SIAH1**	Seven in absentia homolog 1 (Drosophila)
CSNK2A1	Casein kinase 2, alpha 1 polypeptide pseudogene; casein kinase 2, alpha 1 polypeptide	**SKP1**	S-phase kinase-associated protein 1
CTBP2	C-terminal binding protein 2	**SMAD2**	SMAD family member 2
CTNNB1	Catenin (cadherin-associated protein), beta 1, 88kda	**SMAD3**	SMAD family member 3
CTNNBIP1	Catenin, beta interacting protein 1	**SOST**	Sclerosteosis
CUL1	Cullin 1	**SOSTDC1**	Sclerostin domain containing 1
CXXC4	CXXC finger 4	**SOX17**	SRY (sex determining region Y)-box 17
CYLD	Cylindromatosis (turban tumor syndrome)	**TBL1XR1**	Transducin (beta)-like 1 X-linked receptor 1
DAAM2	Dishevelled associated activator of morphogenesis 2	**TBL1Y**	Transducin (beta)-like 1Y-linked
DACT1	Dapper, antagonist of beta-catenin, homolog 1 (Xenopus laevis)	**TCF7L1**	Transcription factor 7-like 1 (T-cell specific, HMG-box)

DKK4	Dickkopf homolog 4 (Xenopus laevis)	**TCF7L2**	Transcription factor 7-like 2 (T-cell specific, HMG-box)
DRD2	Dopamine receptor D2	**TLE2**	Transducin-like enhancer of split 2 (E(sp1) homolog, Drosophila)
FAM123B	Family with sequence similarity 123B	**TLE3**	Transducin-like enhancer of split 3 (E(sp1) homolog, Drosophila)
FBXW11	F-box and WD repeat domain containing 11	**TNIK**	TRAF2 and NCK interacting kinase
FBXW4	F-box and WD repeat domain containing 4	**TNKS**	Tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase
FOSL1	FOS-like antigen 1	**VANGL2**	Vang-like 2 (van gogh, Drosophila)
FZD1	Frizzled homolog 1 (Drosophila)	**WIF1**	WNT inhibitory factor 1
FZD10	Frizzled homolog 10 (Drosophila)	**WISP1**	WNT1 inducible signaling pathway protein 1
FZD5	Frizzled homolog 5 (Drosophila)	**WNT10A**	Wingless-type MMTV integration site family, member 10A
FZD6	Frizzled homolog 6 (Drosophila)	**WNT10B**	Wingless-type MMTV integration site family, member 10B
FZD7	Frizzled homolog 7 (Drosophila)	**WNT11**	Wingless-type MMTV integration site family, member 11
FZD9	Frizzled homolog 9 (Drosophila)	**WNT16**	Wingless-type MMTV integration site family, member 16
GPC4	Glypican 4	**WNT3**	Wingless-type MMTV integration site family, member 3
GRK5	G protein-coupled receptor kinase 5	**WNT4**	Wingless-type MMTV integration site family, member 4
GRK6	G protein-coupled receptor kinase 6	**WNT5A**	Wingless-type MMTV integration site family, member 5A
GSK3A	Glycogen synthase kinase 3 alpha	**WNT6**	Wingless-type MMTV integration site family, member 6
JUN	Jun oncogene	**WNT7A**	Wingless-type MMTV integration site family, member 7A
KREMEN2	Kringle containing transmembrane protein 2	**WNT7B**	Wingless-type MMTV integration site family, member 7B
LEF1	Lymphoid enhancer-binding factor 1	**WNT8A**	Wingless-type MMTV integration site family, member 8A
LRP5	Low density lipoprotein receptor-related protein 5	**WNT8B**	Wingless-type MMTV integration site family, member 8B
LRRFIP2	Leucine rich repeat (in FLII) interacting protein 2	**WNT9A**	Wingless-type MMTV integration site family, member 9A
MAPK10	Mitogen-activated protein kinase 10		