UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Cardio-Thoraco-Vascular Sciences and Public Health

Ph.D. COURSE IN: Traslational Specialistic Medicine "G.B. Morgagni"

CURRICULUM: Cardiothoracic and Vascular Sciences

SERIES: XXXII

Idiopathic Pulmonary Fibrosis (IPF): tissue identification of crucial biomarkers by RNA-Sequencing approach.

Coordinator: Ch.mo Prof. Annalisa Angelini

Supervisor: Ch.mo Prof. Fiorella Calabrese

Ph.D. student: Stefania Edith Vuljan

INDEX

ABSTRACT

Background. Idiopathic pulmonary fibrosis (IPF) is one of the most common idiopathic interstitial lung diseases characterized by progressive lung scarring and a very poor overall prognosis with a median survival of 2-3 years. Despite extensive research efforts, the etiopathogenesis and pathophysiology of IPF are still little understood and consequently only slight improvement has been made for appropriate management and effective therapies. Some advances have been made in understanding the multiple interrelated pathogenic pathways underlying IPF and the disease is now considered to result from complex interactions among genetic, epigenetic, transcriptional, post-transcriptional, metabolic and environmental factors. Thus, the discovery and validation of theranostic biomarkers are necessary to enable a more precise and earlier diagnosis of IPF and to improve the prediction of future disease behavior. Usual interstitial pneumonia (UIP) is the histopathological pattern of IPF characterized by spatial/temporal heterogeneous histological lesions. Fibroblastic focus (FF) areas include fibroblasts and myofibroblasts arranged in a linear fashion with a pale staining matrix, with metaplastic epithelial cells lining on top of them. They are usually detected in the transitional area from dense scarring to normal lung and are considered an "active" component in IPF pathogenesis.

Aim of the research. The main goal of the present PhD research project was the identification of crucial biomarkers in IPF pathogenesis by extracting RNA from FF areas (FF plus metaplastic epithelial cells lining FF). The main steps of the study were the following:

1) validation of a protocol for the isolation of FF areas by laser microdissection of formalin-fixed and paraffin-embedded (FFPE) tissues of IPF patients and controls;

2) total RNA extraction, quality and quantity evaluation;

3) creation of cDNA libraries starting from the extracted RNAs;

4) transcriptomic analysis by a Next-Generation Sequencing (NGS) approach (RNA-sequencing; RNA-Seq);

5) validation of the biomarker emerged from the transcriptomic analysis in a more extensive (retrospective) cases series using immunohistochemistry.

Material and Methods. Total RNA was extracted from fibroblastic focus areas isolated with laser microdissection in 10 FFPE IPF lung tissues: 8 from lung transplanted patients and 2 from surgical biopsy. Microdissected fibroblastic focus areas from 2 patients with recurrent pneumothorax were also analyzed and considered as controls. The RNA was extracted using a modified protocol which provides an overnight tissue incubation at 43°C with 10 µl proteinase K. RNA was then preserved by adding RNase inhibitors at the end of the extraction procedure. This was a custom-made protocol (RNeasy® FFPE kit; Qiagen, Hilden, Germany) with additional procedures optimized during my PhD research study. The final RNAs quality and quantity were valuated with an Agilent RNA 6000 Pico Kit using a 2100 Bioanalyzer instrument. Quality was expressed as DV_{200} (percentage of RNA fragments > 200 nucleotides). Libraries were obtained with the SMARTer Stranded Total RNA-Seq Kit pico input mammalian of Takara Bio. DNA library was sequenced with a paired-end sequencing 2x150 bp on a HiSeq 4000 System sequencer of Illumina. MUC5B immunohistochemistry (clone 4A10-H2; 1:200, Novus Biologicals, Centennial, Colorado, USA) was performed in 44 interstitial lung disease (ILD) cases (39 UIP/IPF and 5 ILD with no UIP pattern) and 6

controls (5 spontaneous pneumothorax and 1 emphysema) following the antibody manufacturer's protocol using a Leica Bond-III Autostainer.

Results. Considering the whole population the mean quantity of extracted RNA was 2992.8 pg/µl±2473 (mean ± SD), ranging from 840 pg/µl to 7530 pg/µl. Quality evaluation showed 42% of total cases with a medium/high quality (DV_{200}>50%). In all cases molecular analyses were performed and final libraries were obtained with a concentration ranging from 3.4 to 22.6 ng/ul and a mean cDNA fragment length of 289 nucleotides. RNA-Seq analysis showed that 323 genes were differentially expressed in UIP/IPF cases than controls: 14 of them were up-regulated and 309 down-regulated. The most significant up-regulated gene was *MUC5B*, the other up-regulated genes were those involved in epithelial-to-mesenchymal transition (EMT) and epithelial carcinogenesis process. Gene Ontology Enrichment Analysis (GOEA) was performed to identify the most enriched Gene Ontology (GO) categories for the down-regulated genes. We found that extracellular matrix structure and organization were the principally down-regulated pathways.

The overexpressed gene *MUC5B* was validated by immunohistochemistry. MUC5B was expressed only in IPF/UIP and ILDs, never in control group. The MUC5B expression was mainly detected in metaplastic epithelial cells lining: a) honeycomb areas, b) other alveolar spaces and c) in the metaplastic epithelial cells lining FF. Interestingly, a gradient of MUC5B expression was detected both in IPF/UIP and ILDs samples where MUC5B was overexpressed in lower lobes. Interestingly, MUC5B was overexpressed in upper and middle lobes of IPF/UIP compared with the same lobes of other ILDs.

Conclusion. The principal results obtained from the present research study offer interesting insights into the complex molecular system signature of IPF:

1) adequate quantity and quality of RNA was extracted from microdissected FF areas of FFPE IPF lung tissues. The quantity/quality of RNA was suitable to create cDNA libraries for transcriptomic analysis by RNA-seq: this represents an important step forward in tissue molecular investigation of this disease characterized by high tissue heterogeneity. Only a very few papers in the literature have used lung FFPE tissue for molecular analysis, in particular, this molecular approach on specific affected IPF tissues has not previously used.
2) Comparative analysis performed on selected areas found an overexpression of epithelial proliferation/cancer progression and EMT transcripts: this highlights the crucial role of metaplastic epithelial cells that are the key actors also in the FF areas, considered the active injured lesion in IPF.
3) The up-regulated transcript *MUC5B*, validated also by immunohistochemistry, confirms the crucial role of this mucin in the disease. Indeed previous works, mainly performed in blood samples, had highlighted the importance of this gene in the disease. Selective regulation of *MUC5B* in experimental models could open up an entire line of investigation that could bring us closer to understanding regulation of *MUC5B* and providing novel therapeutic options.

RIASSUNTO

Introduzione. La fibrosi polmonare idiopatica (FPI), la forma più comune di interstziopatia polmonare, è caratterizzata da una fibrosi progressiva e da una prognosi infausta, con una sopravvivenza media di circa 2-3 anni dal momento della diagnosi. Nonostante gli ampi sforzi compiuti negli ultimi anni per comprendere la natura della malattia, l'esatta eziopatogenesi della FPI continua a rimanere ignota. Di conseguenza, solo lievi miglioramenti sono stati apportati per quanto riguarda la gestione dei pazienti e l'impiego di terapie realmente efficaci. Ad oggi, alcuni progressi sono stati computi nella comprensione delle molteplici vie coinvolte nella patogenesi della FPI: si ritiene che la malattia derivi da complesse interazioni tra fattori genetici, epigenetici, trascrizionali, post-trascrizionali, metabolici ed ambientali. Pertanto, la scoperta e la validazione di possibili biomarcatori teranostici risultano necessarie per ottenere una diagnosi più precisa e precoce della FPI e per poterne prevedere l'evoluzione. L'aspetto istologico patognomonico della FPI è la polmonite interstiziale usuale (UIP) caratterizzata dalla presenza di foci di proliferazione fibroblastica e tessuto cicatriziale denso, alternati ad aree di tessuto polmonare normale (eterogeneità). I foci fibroblastici, generalmente localizzati nell'area di transizione tra zone di parenchima normale e di parenchima fibrotico, sono costituiti da fibroblasti e miofibroblasti al di sopra dei quali si colloca un sottile strato di cellule metaplastiche epiteliali, e vengono considerati la "componente attiva" della FPI.

Scopo della ricerca. L'obiettivo primario di questa ricerca è stato quello di individuare possibili biomarcatori cruciali nello sviluppo della FPI mediante estrazione di RNA dai foci fibroblastici di pazienti FPI. Le principali fasi dello studio sono state:

1) validazione di un protocollo per l'isolamento dei foci fibroblastici mediante microdissezione laser di tessuti fissati in formalina e inclusi in paraffina (FFPE);
2) estrazione dell'RNA totale e successiva valutazione della qualità e quantità;
3) creazione di librerie di cDNA a partire dagli RNA estratti;
4) analisi trascrittomica sul RNA estratto mediante metodica di Next-Generation Sequencing (RNA-sequencing; RNA-Seq);
5) validazione del biomarcatore emerso dall'analisi trascrittomica in una casistica più estesa (retrospettiva) mediante immunoistochimica.

Materiali e metodi. L'RNA totale è stato estratto dai foci fibroblastici isolati mediante microdissezione laser da tessuto polmonare FFPE di pazienti con diagnosi di UIP/PFI: 8 da polmoni espiantati da pazienti trapiantati di polmone e 2 biopsie diagnostiche. Come controllo sono state considerati i foci fibroblastici microdissezionati da 2 pazienti affetti da pneumotorace spontaneo ricorrente.

L'RNA è stato estratto utilizzando un kit specifico per tessuti FFPE (RNeasy® FFPE kit; Qiagen, Hilden, Germany), il cui protocollo è stato modificato durante il percorso di dottorato di ricerca: i tessuti microdissezionati sono stati incubati *overnight* a 43°C con 10 µl di proteinasi K, con l'aggiunta finale di inibitori delle RNasi all'eluato. La quantità e qualità degli RNA estratti è stata valutata mediante l'utilizzo del RNA 6000 Pico Kit con lo strumento Agilent Bioanalyzer 2100. La DV_{200} (percentuale di frammenti di RNA > di 200 nucleotidi) è stata impiegata nella valutazione della qualità. Le librerie sono state ottenute con il kit SMARTer Stranded Total RNA-Seq pico input

mammalian della Takara Bio e successivamente sequenziate nello strumento HiSeq 4000 Illumina, mediante un sequenziamento di tipo *paired-end* 2x150 bp.

L'indagine immunoistochimica per MUC5B (clone 4A10-H2; 1:200, Novus Biologicals, Centennial, Colorado, USA) è stata eseguita su 44 casi di interstiziopatia polmonare (ILD) di cui 39 UIP/IPF, 5 ILD senza pattern UIP e 6 controlli (5 pneumotoraci spontanei e 1 enfisema), seguendo il protocollo del produttore, su piattaforma automatizzata Leica Bond-III.

Risultati. La concentrazione dell'RNA estratto variava da 840 pg/µl a 7530 pg/µl con una concentrazione di 2992.8 pg/µl±2473 (media ± SD). Per quanto riguarda la qualità degli RNA estratti, il 42% di campioni presentava una qualità medio/alta con una DV_{200} maggiore del 50%. Tutti i campioni sono stati quindi impiegati per la successiva analisi molecolare che ha fornito delle librerie con una concentrazione compresa tra 3,4 ng/ul e 22,6 ng/ul ed una lunghezza media dei frammenti di cDNA di 289 nucleotidi. In seguito all'analisi di RNA-Seq, sono stati identificati 323 geni differenzialmente espressi di cui 14 sovraespressi e 309 sottoespressi rispetto ai controlli. Il gene maggiormente sovraespresso è risultato essere il *MUC5B,* mentre tutti gli altri geni sovraregolati erano quelli principalmente coinvolti nella transizione epitelio-mesenchima e nei processi oncogenetici a livello delle cellule metaplastiche epiteliali del focus fibroblastico. Per i geni sottoespressi è stata possibile eseguire un'analisi di geneontologia mediante una *Gene Ontology Enrichment Analysis* (GOEA) per l'identificazione dei principali termini ontologici (GO) in grado di caratterizzare i geni individuati. Grazie a quest'analisi, è stato quindi possibile identificare che l'organizzazione e la struttura della matrice extracellulare risultano essere i *pathways* maggiormente regolati dal cluster di geni individuati come sottoespressi rispetto ai controlli.

La validazione dell'over-espressione di *MUC5B* a livello tissutale è stata infine eseguita mediante immunoistochimica. Una forte espressione di MUC5B è stata riscontrata a livello delle cellule epiteliali metaplastiche delle: a) aree di *honeycombing* dei tessuti polmonari, b) negli spazi alveolari e c) nelle cellule epiteliali metaplastiche dei foci fibroblastici sia nei pazienti UIP/FPI sia nei pazienti con altre ILD. Nei controlli non è stata riscontrata nessuna positività per MUC5B. In particolare, mettendo a confronto queste due tipologie di pazienti, è emerso come nei pazienti affetti da FPI l'espressione di MUC5B risultava maggiore rispetto ai pazienti con altre ILD. É stato interessante notare anche una sorta di "gradiente d'espressione" di MUC5B sia nei campioni FPI/UIP sia in quelli ILDs: MUC5B risultava maggiormente espresso nei lobi inferiori. È stata riscontrata anche un'overespressione di MUC5B nei lobi superiori e medi di pazienti FPI/UIP rispetto alle altre ILD.

Conclusione. I risultati principali ottenuti da questo studio, offrono approfondimenti interessanti per quanto riguarda la comprensione del complesso sistema molecolare della FPI:

1) è stato possibile ottenere un'adeguata quantità e qualità di RNA a partire dai foci fibroblastici dei tessuti polmonari FFPE, in modo da poter ottenere delle librerie di cDNA ottimali per la successiva analisi in RNA-Seq. Tutto ciò rappresenta un punto di partenza molto importante per studiare, da un punto di vista molecolare, questa malattia caratterizzata da una così elevata eterogeneità a livello tissutale. A oggi, secondo la letteratura, solamente un numero esiguo di studi ha impiegato tessuti polmonari FFPE per eseguire delle analisi molecolari. In particolare, nessuno ha utilizzato questa metodologia in specifiche aree del tessuto polmonare.

2) L’analisi omica comparativa, compiuta su aree ben specifiche della malattia, ha messo in luce il ruolo cruciale delle cellule epiteliali metaplastiche dei foci fibroblastici, considerati ad oggi le zone di malattia “attiva” della FPI: abbiamo infatti rilevato una maggiore espressione di trascritti, caratteristici delle cellule epiteliali, coinvolti in processi di proliferazione cellulare/carcinogenesi e nella transizione epitelio-mesenchima.
3) L’overespressione del trascritto *MUC5B*, validata anche in immunoistochimica, ha confermato il ruolo cruciale delle mucine nello sviluppo di questa patologia. Infatti, molti lavori condotti soprattutto su campioni di sangue di pazienti affetti da FPI, hanno messo in luce l’importanza di questo gene nella malattia. Lo studio della regolazione di *MUC5B* in modelli sperimentali potrebbe andare ad aprire una nuova linea di ricerca al fine di comprendere maggiormente il ruolo di *MUC5B* nella FPI e andando così a fornire in futuro nuove possibilità terapeutiche.

1) INTRODUCTION

1.1) Historical findings of Idiopathic Pulmonary Fibrosis

Hippocrates, a Greek physician often referred to as the "Father of Medicine", was the first that described fibrotic changes in the lungs in the fifth century BC[1].

First recognition of the disease in the modern age, is attributed to D.J. Corrigan in 1838, who called it cirrhosis of the lung[2]. Corrigan's description prevailed until 1893, when, in the text *Principles and Practice of Medicine*, William Osler renamed this disease chronic interstitial pneumonia keeping "cirrhosis of the lung" as a subtitle[3]. In his book, Osler described for the first time that ''*the disease is unilateral''*, but what he observed was the chronic evolution of acute infectious pneumonia rather than an idiopathic fibrotic process. Osler largely referred to Jean-Martin Charcot, who had studied chronic pneumonia in detail, reporting the sequence leading from acute pneumonia to those he called ''pneumonic fibrous metamorphosis'' including many ''fusiform cells'' likely corresponding to myofibroblasts, so he was the first to describe this aspect that is now pathognomonic for idiopathic pulmonary fibrosis (IPF)[1].

The term IPF began to appear in the medical literature in the mid-1900s and was initially used by clinicians and radiologists to refer to fibrosing pneumonitis of unknown cause[4,5,6]. However, entities with similar clinical presentations, such as interstitial lung disease (ILD) associated with connective tissue disorders (CTD) or fibrotic hypersensitivity pneumonitis (HP), were often identified as cases of IPF by clinicians due to similar clinical presentation and radiographic appearance. As our knowledge of ILD expanded over the last three decades of the twentieth century, it became clear that histopathologic patterns could be found that paved the way to our current clinic pathologic classification of the Idiopathic Interstitial Pneumonias (IIPs),

with the term "IPF" used exclusively to designate patients with idiopathic usual interstitial pneumonia (UIP).

In the early 1900s pathologists began to recognize that some patients, at the time of autopsy, had bilateral lung disease that was not related to common causes of death such as infection or malignancy. These lungs were scarred, shrunken, and cystic, with these changes appearing to be most prominent peripherally and at the lung bases with these moderate aspects at the apices and more central areas. In some cases smooth muscle cells were visible, and the term "muscular cirrhosis" of the lung was used. Moreover, the presence of cysts was referred to as "honeycomb change" or "honeycomb lung" [7].

In the 1930s, Louis Hamman and Arnold Rich reported a clinical-radiographic-pathologic description of four patients who succumbed for respiratory insufficiency at the Johns Hopkins Hospital in Baltimore, US[8]. Advanced lung fibrosis and honeycombing was found at autopsy, and this was recognized as a disease that became known as the "Hamman–Rich syndrome." They referred that these patients had ''acute diffuse interstitial fibrosis of the lung'', which was considered an entity close to IPF until the 1960s but is now indicated as acute interstitial pneumonia, thus differing from IPF. In the article they wrote:

> "*The symptoms of each case differed somewhat from those of the others and yet bore a striking resemblance so that having seen two cases, the correct diagnosis was suggested when the third appeared. The pathological process in the lungs was identical in all and this lesion was so extraordinary and distinctive that there can be no doubt that the symptoms displayed by each case, different thought they were, arose from the same underlying morbid process, variations in*

symptoms depending upon the stage this process had reached at the time the patients came under observation"[8]

In 1948, Robbins first used the term IPF to describe patients with interstitial opacities on chest radiology that had pulmonary fibrosis but with no identifiable causes. In this article, a review of the literature was done and the authors identified numerous known causes of pulmonary fibrosis, with a reference to a case of fibrosis with unknown etiology. The author identified some possible etiological factors like inflammatory processes (tuberculosis, bronchiectasis, and even syphilis); radiations; asthma, that can lead to the development of severe emphysema with a considerable amount of fibrosis; and pneumoconiosis. Pulmonary fibrosis was also associated with rarer conditions, such as Raynaud's disease, scleroderma, and periarteritis nodosa. A small group of cases remained without any possible etiological cause of fibrosis: these consequently were considered cases of IPF[9].

In the mid and late 1960s, the "modern" era of interstitial lung disease began when the pathologist Averill Liebow proposed a pathologic classification scheme describing five histopathological subgroups of chronic idiopathic interstitial pneumonia without any suspicious of infection or malignancy: one of these was called "usual interstitial pneumonia (UIP)" pattern, considered the result of diffuse alveolar damage with hyaline membranes, further interstitial proliferation and honeycombing[10]. Liebow and colleagues coined the term "UIP" because, it was the most common usual form, among interstitial pneumonias of unknown etiology. The authors observed that UIP pattern tended to occur in older adults and appeared to affect lobes in a highly variable manner. The other four different histopathologic patterns that Liebow described were desquamative interstitial pneumonia (DIP), bronchiolitis obliterans with interstitial

pneumonia (BIP), lymphoid interstitial pneumonia (LIP), and giant cell interstitial pneumonia (GIP)[10].

However, the term IPF was rarely used until a 1976 review article by Crystal and colleagues, that popularized the term. The authors described IPF like a fatal disorder that started as an alveolitis and progressed to interstitial fibrosis. At that time, biochemical studies suggested that the disease was probably related to a collagen rearrangement rather than collagen content increase. Thus, the hypothesis was that peripheral lymphocytes of these patients recognized collagen as "non-self" and produced lymphocytes and cell lysis. The authors thought that the fibrotic process was irreversible, but the inflammatory and immune processes that lead to fibrosis could be responsive to therapy if diagnosed early[11].

In 1978, Charles Carrington reported the results of a prospective longitudinal study on a series of 53 cases of patients with UIP and DIP. UIP was defined as a ‘‘*highly variegated structure often including the entire spectrum from normal alveolar walls to fibrotic, end-stage lesions in the same tissue sample; dense pleomorphic interstitial cellular infiltrate including many lymphocytes and monocytes but relatively few eosinophils*’’. At the end of 5 years, the mortality for UIP was 44%, compared with only 5% in DIP. Fibrosis and honeycombing were considered to be nonspecific features of both UIP and DIP [12].

In the following decade, the prevalent idea was that ‘‘*alveolar macrophages direct the alveolitis associated with IPF*’’ with limited, if any, role attributed to fibroblasts[13].

The introduction of computed tomography (CT)-scan contributed to a better characterization of the different chronic interstitial pneumonias.

Further studies aimed to better describe IPF in the 1990s[14,15,16,17,18] since it was definetively accepted as a distinct entity only in 2008[19].

1.2) Classification of Interstitial lung disease (ILDs)

Interstitial Lung Disease (ILD) are characterized by abnormal collagen accumulation in the lungs due to the proliferation of the interstitial compartments associated to infiltration of inflammatory cells and fibrosis. Interstitial fibrosis is the predominant phenotype in most cases[20]. ILDs are classified on the basis of the known or unknown etiology (figure 1).

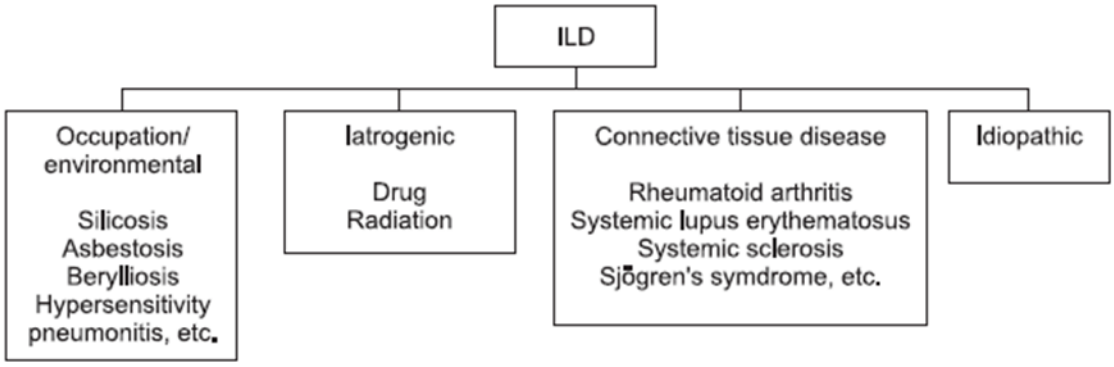

Figure 1*: classification of interstitial lung disease (ILD). From Park SW et al.,Tuberc Respir Dis (Seoul). 2019, page 3*

The classification system based on ILD etiology, show four types:

1) Occupational/environmental ILDs, that includes occupational diseases such as silicosis, asbestosis and Hypersensitivity Pneumonitis (HP);

2) iatrogenic ILDs, that include those induced by radiation or drugs such as chemotherapeutic agents;

3) autoimmune ILDs, that include disorders associated with Connective Tissue Disease (CTD) or autoimmune diseases such as rheumatoid arthritis and lupus;

4) ILDs without a known etiology are defined as idiopathic interstitial pneumonia (IIP). IPF is one specific presentation of IIP[21].

The 2002 American Thoracic Society/European Respiratory Society (ATS/ERS) classification of IIPs was updated in September 2013[22].

In this new classification there are three main categories of idiopathic interstitial pneumonias (IIPs): major IIPs, rare IIPs, and unclassifiable IIPs (figure 2). The major IIPs are grouped into:

- chronic fibrosing IIPs (this includes IPF and non-specific interstitial pneumonia [NSIP]);
- smoking-related (respiratory bronchiolitis–interstitial lung disease [RB-ILD]);
- desquamative interstitial pneumonia (DIP);
- acute/subacute IIPs (cryptogenic organizing pneumonia [COP] and acute interstitial pneumonia [AIP]).

TABLE 1. REVISED AMERICAN THORACIC SOCIETY/EUROPEAN RESPIRATORY SOCIETY CLASSIFICATION OF IDIOPATHIC INTERSTITIAL PNEUMONIAS: MULTIDISCIPLINARY DIAGNOSES

Major idiopathic interstitial pneumonias
Idiopathic pulmonary fibrosis
Idiopathic nonspecific interstitial pneumonia
Respiratory bronchiolitis–interstitial lung disease
Desquamative interstitial pneumonia
Cryptogenic organizing pneumonia
Acute interstitial pneumonia
Rare idiopathic interstitial pneumonias
Idiopathic lymphoid interstitial pneumonia
Idiopathic pleuroparenchymal fibroelastosis
Unclassifiable idiopathic interstitial pneumonias*

***Figure 2**: revision of the IIP classification. From Travis WD et al., Am J Respir Crit Care Med. 2013;188(6):733-48.*

1.3) Epidemiology of Idiopathic Pulmonary Fibrosis

Investigators have used epidemiologic studies to determine the burden of IPF and to identify possible exposures/risk factors for the disease development. The availability of epidemiological data relating to ILDs has increased over recent years. We know that IPF occurs worldwide but determining the real epidemiology of this disease is a difficult duty: a large variability of prevalence and incidence rates is observed in national and international studies due to a series of problems first of all the absence, in the past decades, of a uniform definition and identification of IPF before the 2000 consensus statement[23].

IPF represent above the 20% of all cases of ILD, and is the most frequent and severe form of IIPs[20] (figure 3). IPF is now considered to be a rare disease because occurring in less than 5 per 10,000 person-years[24].

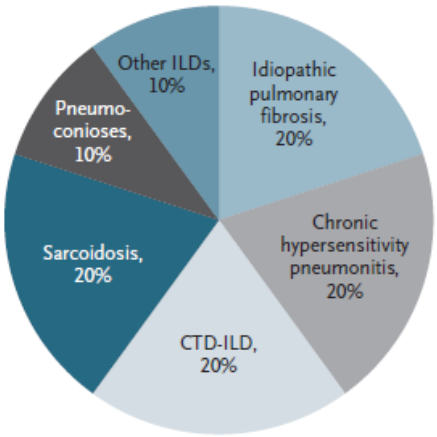

***Figure 3**: estimated Relative Distribution of Specific Interstitial Lung Diseases (ILDs) in the United States. From Lederer DJ et al., N Engl J Med. 2018;378(19):1811-1823.*

The estimated incidence in Europe and North America of IPF has been reported to range between 2.8 and 19 cases per 100,000 people per year[25, 26, 27] and appears to be higher in North America and Europe (3 to 9 cases per 100,000 person-years) than in Asia (South

Korea, Taiwan and Japan), with incidence rates ranging from 1.2 to 4.16 per 100,000 people per year[25]. The highest rates of IPF incidence in Europe are found in the UK, ranging between 4.6 and 8.65 per 100,000 people per year, and 6000 people diagnosed annually[25]. The incidence of IPF appears to be lower in Scandinavia, with incidence reported between 1.3 and 4.3 per 100,000 people per year[25].

In the United States, the prevalence of IPF has been reported to range from 10 to 60 cases per 100,000, although in one study, the prevalence was 494 cases per 100,000 in 2011 among adults over the age of 65 years, which was twice as high as the prevalence recorded 10 years earlier[28]. In Canada, the prevalence of IPF was reported to increase from 25.9 per 100,000 population in patients aged 50–59 years, to 507.0 per 100,000 population at ≥90 years of age. In the UK, 85% of patients diagnosed with IPF are more than 70 years old. It is worth noting that for a small subset of patients with IPF (0.5–3.7%) a familiarity can be recognized and the disease can be earlier[29,30,31].

When stratified by age and gender, the prevalence of IPF increased with age, with the majority of patients aged >50 years at diagnosis, with a higher proportion of males than females[20]. Patients with IPF who are younger than 50 years old are rare and may subsequently manifest features of an underlying CTD or may have familial IPF.

The majority of patients have a history of past cigarette smoking, indeed this is the most important environmental risk factor, even many years after smoking cessation[32]. Other risk factors associated with IPF include gastroesophageal reflux, chronic viral infections such as Epstein-Barr virus, hepatitis C, adenovirus, herpes virus, autoimmunity (lymphocyte aggregates composed of CD3+ T cells and CD20+ B cells and the presence of autoantibodies), environmental exposures (metal dusts, automobile emissions, wood dust) and a family history of ILD[33].

Many patients with IPF also have other comorbidities such as emphysema (combined pulmonary fibrosis and emphysema), lung cancer, pulmonary hypertension, sleep apnea, and coronary artery disease[34].

In some genetic forms, there are also extrapulmonary disease that manifests as bone marrow failure and liver disease. In some patients, biological members of the family (primary relatives) also have IPF. At least 30% of patients who have sporadic or familial pulmonary fibrosis have genetic predisposing factors that are known to increase the risk of pulmonary fibrosis[35].

Only 20 to 30% of subjects were alive 5 years after diagnosis with a mean survival of about 2-5 years from the time of symptoms and diagnosis. Many patients die from progressive, chronic hypoxemic respiratory failure. In public health terms, IPF-related mortality is similar to many malignancies diseases like non-Hodgkin lymphoma, renal cancer and esophageal cancer[36]. Each year, approximately 10 to 20% of patients with IPF have an acute exacerbation, with worsened hypoxemic respiratory failure, bilateral ground-glass opacities, consolidation, or both on high-resolution CT imaging. Most patients with an acute exacerbation die from acute respiratory failure. Exacerbations may be triggered by a clinical event (e.g. infection, aspiration, or drug toxicity) but are frequently idiopathic[20]. In a USA study the most frequent cause of death was respiratory failure (60%), cardiovascular disease (8.5%), and lung cancer (2.9%)[33]. An important complication in IPF patients concerns primarily the development of lung carcinoma, in particular lung adenocarcinoma. Clinico-pathological features of this type of tumor included older age at occurrence, male predominance, smoking history, the develop in the periphery and lower lobe of the lung, large tumor size, high incidence of lymph vessel invasion, pleural invasion, metastasis and poor survival rate with a worse prognosis compared to IPF patients without cancer. Histologically IPF patients with

adenocarcinoma have an invasive mucinous predominant phenotype. From a genetic point of view, this type of tumor has a low incidence of EGFR mutation and high incidence of KRAS mutation. Taking together all these aspects, underline that this type of tumor represent a specific form of lung adenocarcinoma that should not be compared to the conventional one[37].

These studies suggest that mortality for IPF is increasing, and IPF is an important and growing public health concern, particularly in the aging population.

1.4) Clinical features of Idiopathic Pulmonary Fibrosis

In 2000, IPF was defined as a specific form of chronic, progressive, fibrosing interstitial pneumonia of unknown etiology[23]. It occurs primarily in older adults, is limited to the lungs, and is defined by the histopathologic and/or radiologic pattern of Usual Interstitial Pneumonia (UIP).

Clinicians should be considered IPF in the differential diagnosis in all adult patients presenting unexplained chronic exertional dyspnea, dry cough, bibasilar inspiratory crackles, and/or digital clubbing that occur without constitutional or other symptoms that suggest a multisystem disease. Patients with IPF most often present with exertional dyspnea typically progresses over a period of months to years. Chest pain, fatigue, malaise, and weight loss are other nonspecific symptoms. Pulmonary function test results may be normal in patients with mild disease but will show some degree of restriction (i.e., reduced vital capacity and reduced total lung capacity but preserved residual volume). Diffusion capacity is commonly decreased even in patients with mild disease, although this finding is nonspecific. In practice, patients with interstitial lung

disease often initially receive a diagnosis of heart failure or chronic obstructive pulmonary disease, suggesting that clinicians frequently fail to consider interstitial lung disease in patients with dyspnea[20,35,38].

According to the recent Fleischner Society White Paper, a clear focus of a patient's clinical examination should focus on the clinical probability to have IPF, which is particularly increased when the patient is older than 60 years, male, and has a history of cigarette smoking. Other recommendation is that in every patient with fibrosing ILD, identification of exposure to antigens is important because might result in hypersensitivity pneumonitis (lists of such antigens are available)[39]. In general, early recognition and accurate diagnosis are likely to improve outcomes through potentially harmful therapies and prompt initiation of therapies that are effective even in the early stages of disease[40,41].

From a clinical point of view in terms of disease progression, IPF can be distinct into two different types of patterns: the ''slow'' progressors with a long duration of symptoms before diagnosis and with a slowly progressive clinical course and the ''rapid'' progressors with a more rapidly progressive clinical course with a shorter duration of symptoms before diagnosis and progression to death[42].

Selman et al.[43] demonstrated the presence of this distinct clinical phenotype of IPF that differs in clinical course and transcriptional profile. They suggest that during the development of this disease, genetic modifiers (and environmental factors, i.e. smoking) and the up-regulation of several pathways may play an important role in determining the clinical phenotype, inducing a more aggressive IPF phenotype (rapid progressor). These pathways include genes involved in cell motility, myofibroblast differentiation, oxidative stress, coagulation and development. Genes that are increased in rapid progressor where Adenosine A_{2B} receptor gene that is involved in the differentiation of

human lung fibroblasts into myofibroblasts, prominin-1/CD133 which is found in hematopoietic stem cells and embryonic epithelium and MMP-9 that high levels was find in the BAL fluids. Decreased gene in ‘‘rapid’’ progressors included Smad6, disintegrin-like and metalloprotease with thrombospondin type 1 motif, 7 (ADAMTS7), the chemokine receptor CXCR6, and Bcl2-L-10[43].

Morphometric analysis of the explanted lung showed a prominent cellular inflammatory infiltrate. Balestro et al.[44] in their article observed also that two clinical phenotypes can be clearly identified and the differences in their pathological features are present, with the inflammation (innate and adaptive immunity) being the most striking one. Conceivably these differences in lung pathology might be important contributors to the different clinical behavior[44].

1.5) Pathogenesis of Idiopathic Pulmonary Fibrosis

The normal reparative process to tissue injury is characterized by the involvement of different cell types under the influence of a lot of mediators with the goal of reestablishing tissue integrity and barrier function[45].

Fibrosis, characterized by excessive extracellular matrix accumulation and disruption of normal tissue architecture, can occur as a result of dysregulation of this normal reparative process. The wall of the alveolus of the lung is formed by monolayers of alveolar epithelial cells (AECs) and endothelial cells, separated only by their basement membranes[46]. The surrounding interstitial spaces of the lung are composed by of a network of proteins (collagens, fibronectin, elastin) and glycosaminoglycans. In pulmonary fibrosis, there is a disruption of this structure, with expansion of the connective tissue of the lung due to the accumulation of matrix components that causes

the progressive disruption of the normal lung architecture. The current hypothesis about the development of IPF focuses on multiple and focal episodes of epithelial injury followed by a pathological fibrotic repair mechanism[47].

In this process alveolar epithelial cells, mainly type 2 (AEC2s), are primarily affected. In normal condition, AEC2s have surfactant, metabolic and immunological functions and are progenitor cells for alveolar epithelial type 1 (AEC1s). In IPF, AEC2s are characterize by genomic instability, telomere attrition, epigenetic changes, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence and altered intercellular communication.

Indeed, IPF is a complex disorder resulting from interactions between a series of risk factors and the interactions between epithelial cells, mesenchymal cells and extracellular matrix (ECM). However, the exact mechanisms of how these factors interact to cause this disease remain unclear[48].

An approach to understanding IPF pathogenesis is to consider it as a three-stage process composed by a plausible sequence of events that lead to the development of pulmonary fibrosis: the predisposition stage, the initiation/activation stage and the progression stage (figure 4).

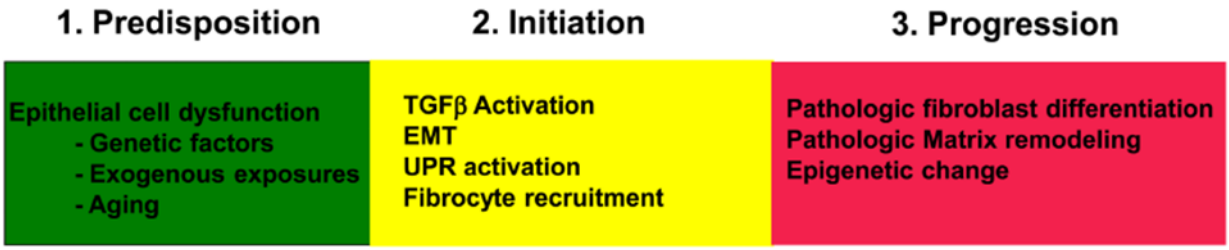

Figure 4: *three-stage description of IPF pathogenesis. In the predisposition stage, recurrent environmental insults lead, in genetically predisposed individuals, to develop lung fibrosis. The second initiation stage includes profibrotic processes, such as TGF-β activation, fibrocyte recruitment, epithelial-to-mesenchymal transition (EMT), endoplasmic reticulum-stress-mediated activation of unfolded protein response (UPR) and apoptosis that accelerate the profibrotic processes.*
The final progression stage includes molecular processes that lead directly to fibrosis, such as pathologic fibroblast differentiation, (the pathological matrix promotes additional differentiation of fibroblasts to myofibroblasts, which deposit more matrix and further activate fibroblasts) matrix deposition and remodeling, increased matrix stiffness, and profibrotic epigenetic changes within fibroblasts and epithelial cells. From Wolters PJ et al., Annu Rev Pathol. 2014;9:157-79.

1.5.1) Predisposition stage

This stage includes factors such as genetic mutations or variations (i.e. SNPs), environmental exposures (i.e. smoking), ageing and epigenetic alterations that predispose individuals to develop lung fibrosis. These factors ultimately lead to epithelial cell dysfunction.

1.5.1.1) Genetics

Occurrence of pulmonary fibrosis in multiple members of the same family, commonly referred to as familial pulmonary fibrosis (FPF), accounts for 2–20% of the overall cases of IPF, suggesting a genetic predisposition to some forms of the disease[49] The common genetic variants in IPF involved principally telomere related genes, surfactant proteins and MUC5B (figure 5).

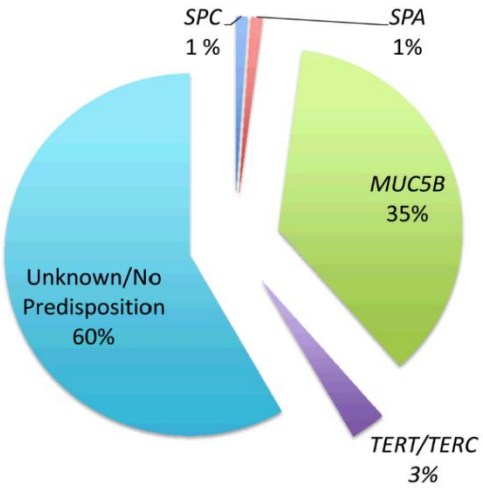

***Figure 5**: estimated frequencies of genetic mutations predisposing to IPF. From Wolters PJ et al., Annu Rev Pathol. 2014;9:157-79.*

Telomere related genes (TRGs)

Rare genetic variants involve mainly TRGs that are involved in telomere maintenance, controlling addition of repeated DNA sequences in the telomere region of chromosomes (TTAGGG repeats bound by a six-protein complex known as shelterin) and protecting the chromosomes from loss of material during mitosis. Among TRG mutations, telomerase reverse transcriptase (*TERT*) mutations are the most frequently observed, being detected in approximately 15% of the cases of familial pulmonary fibrosis, whereas mutations in regulator of telomere length 1 (*RTEL1*), poly(A)-specific ribonuclease (*PARN*), telomerase RNA component (*TERC*), dyskerin pseudouridine synthase 1 (*DKC1*), TERF1 interacting nuclear factor 2 (*TINF2*) and nuclear assembly factor 1 ribonucleoprotein (*NAF1*) are much rarer[50,51,52,53].

Immunohistochemistry for RTEL1 showed that this gene was expressed by bronchial epithelial cells, alveolar macrophages, hyperplastic alveolar epithelial cells and lymphocytes (particularly within lymphoid follicles). RTEL1 was localized in the nucleus and in the cytoplasm in all this cell types, but was not detected in fibroblastic foci or endothelial cells[54]. It is believed that the loss of function of the telomerase complex may influence the turnover and healing of alveolar epithelial cells after a damaging stimulus, thus triggering IPF[49] (figure 6).

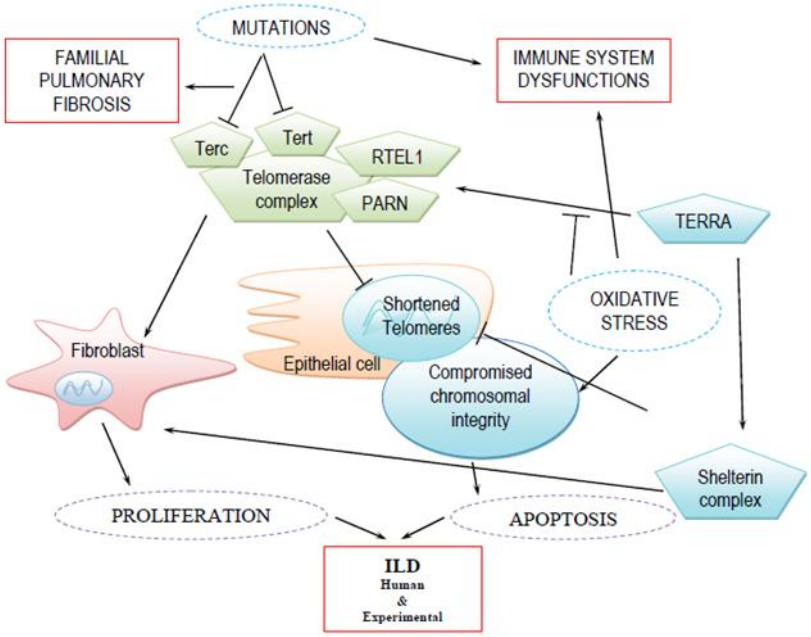

Figure 6: *principal factors involved in the regulation of telomere maintenance in the fibrotic lung. External or internal factors (such as mutations and oxidative stress) may positively (pointed arrow) or negatively (block arrow) act the telomere maintenance through telomerase complex (green pentagon) leading to structural and functional damage of telomeric and chromosomal DNA . The results of these disruptions depend on the cell type and vary from apoptosis in epithelial cells to cellular proliferation in fibroblasts. In lungs, in particular, this may produce a proliferation of fibrotic tissue that can manifest IPF. From Arish N. et al., Int J Mol Sci. 2019;20(12).*

In their article Snetselaar et al.[55] observed that, for sporadic IPF subjects, AEC2s cell telomere length (TL) in non-fibrotic areas was 56% longer than in fibrotic areas. In subjects carrying a telomerase reverse transcriptase (TERT) mutation, AEC2s cell TL was significantly shorter than in sporadic subjects with IPF. Finally, it was determined that IPF subjects with shortest lung TL had a significantly worse survival than patients with long TL[55].

More recently, these findings have been translated into a potential therapeutic strategy by using adeno-associated vectors (AAV) to transiently activate telomerase in adult tissues. Povedano et al.[56], tested this hypothesis by using a TERT based gene therapy using AAV9 vectors in mice with pulmonary fibrosis (treated with low doses of the

lung-damaging agent bleomycin) in the context of short telomeres, a scenario that resembles pulmonary fibrosis in humans associated with short telomeres. Their findings demonstrate that TERT treatment significantly improves pulmonary function, decreases inflammation, and accelerates fiber disappearance in fibrotic lungs. At the molecular level AAV9-treatment, that preferentially targets regenerative alveolar type II cells (ATII), results in telomere elongation and increased proliferation of ATII cells, also significantly decreasing DNA damage, apoptosis, and senescence in these cells[56].

Another recent study, focus on shorter telomere lengths and increased chromosomal damage in the lungs of patients with IPF and compared these measurements to regional disease severity in the lungs based on pathologic structural changes associated with parenchymal collapse and fibrosis. Telomere length was associated with fibrosis, with longer telomere lengths present in regions with increased total collagen. DNA damage was correlated with elastin but not with collagens reflecting an association with parenchymal collapse rather than fibrosis suggesting a role for this process in the early stages of IPF progression[57].

Surfactant proteins

Other genes that are involved in familiar IPF are mutations in genes encoding surfactant proteins— primarily the gene encoding surfactant protein C (*SFTPC*), which is exclusively expressed by AEC2s —*SFTPA2* (Surfactant Protein A2) and *SFTPA1* (Surfactant Protein A1). Familial pulmonary fibrosis-associated variants in SFTPC result in miss folded surfactant protein C, which induces endoplasmic reticulum stress in AEC2s because both mutations predict instability of the translated protein, so the mutated proteins are abnormally retained in the endoplasmic reticulum. This subsequent

dysfunctional folding and processing of surfactant, involves dysregulated proteostasis and endoplasmic reticulum stress with the consequent epithelial dedifferentiation (that is epithelial to mesenchymal transition, EMT)[58].

These mutations reportedly occur in only 1% of sporadic cases of pulmonary fibrosis[59,60].

MUC5B

MUC5B is a major gel forming mucin in the lung that plays a key role in mucociliary clearance and host defense. This protein is secreted from proximal submucosal glands and distal airway secretory cells. In patients with IPF, excess MUC5B protein is especially observed in epithelial cells in the respiratory bronchiole and honeycomb cyst, regions of lung involved in lung fibrosis. However, to date, there are only some hypothesis how MUC5B leads to the development of IPF[61].

This localization of MUC5B in IPF lungs suggests a prominent role for injury and abnormal repair to the distal airway epithelium cells in IPF, inconsistent with the traditional view that IPF is a disease of the alveolar epithelium. It suggests that *MUC5B* overexpression in the distal airway may play a role in the development of IPF[62].

In 2011, a study research[63] used a genomewide linkage scan that detected linkage between idiopathic interstitial pneumonia and a 3.4-Mb region of chromosome 11p15 in 82 families. The minor allele of the single-nucleotide polymorphism (SNP) rs35705950, located 3 kb upstream of the MUC5B transcription start site, was present at a frequency of 38% among subjects with IPF. In this study, MUC5B expression in the lung was 14.1 times higher in subjects with IPF than non IPF patients. Authors also found that MUC5B protein was expressed in tissue lesions of IPF patients, suggesting that dysregulated MUC5B expression in the lung may be involved in the pathogenesis of

pulmonary fibrosis assuming, at least, three possible mechanisms. First, they hypothesized that high MUC5B levels impair mucosal host defense, resulting in excessive lung injury from inhaled substances, and over time leading to the development of IPF. Second, the over-expression of MUC5B can alter normal alveolar repair, interfering either the interaction between the type II alveolar epithelial cells and the underlying matrix or the surface-tension properties of surfactant. Third, tissue IPF lesions were spatially heterogeneous, suggesting that the disorder was multifocal, originating in individual bronchoalveolar units. Since the rs35705950 SNP occurred in the putative promoter region of MUC5B and was predicted to disrupt transcription-factor binding sites, one must considered ectopic production of MUC5B in cells or locations that caused injury to the bronchoalveolar unit[63].

Hancock et al.[61] show that *MUC5B*, in human IPF lung tissue, is co-expressed with surfactant protein C in columnar epithelial cells lining honeycomb cysts and in type 2 alveolar epithelia, indicating that cell types involved in lung fibrosis in the distal airspace also express *MUC5B* so this cells are involved in the lung remodeling that is characteristic of IPF. In mice, they demonstrate that MUC5B concentration in bronchoalveolar epithelia is related to impaired mucociliary clearance (MCC) and to the extent and persistence of bleomycin-induced lung fibrosis. Their findings suggest that mucociliary dysfunction might play a causative role in bleomycin-induced pulmonary fibrosis in mice overexpressing *MUC5B*, and that MUC5B in distal airspaces is a potential therapeutic target in humans with IPF. These findings suggest that overexpression of MUC5B in distal airspaces, which is known to occur in IPF, disrupts the equilibrium necessary to sustain effective mucociliary transport thereby impairing mucus function. One potential consequence of mucociliary dysfunction is retention of inhaled substances (air pollutants, cigarette smoke, microorganisms, etc.) and

endogenous inflammatory debris that, over time, results in temporally and spatially distinct areas of microscopic scaring and progressive fibroproliferation in the lung leading to the development of IPF. Alternatively, reduced clearance or enhanced viscosity of MUC5B may initiate a reactive or regenerative fibrotic response localized to the bronchoalveolar region of the lung that eventually leads to the development of IPF[61] (figure 7).

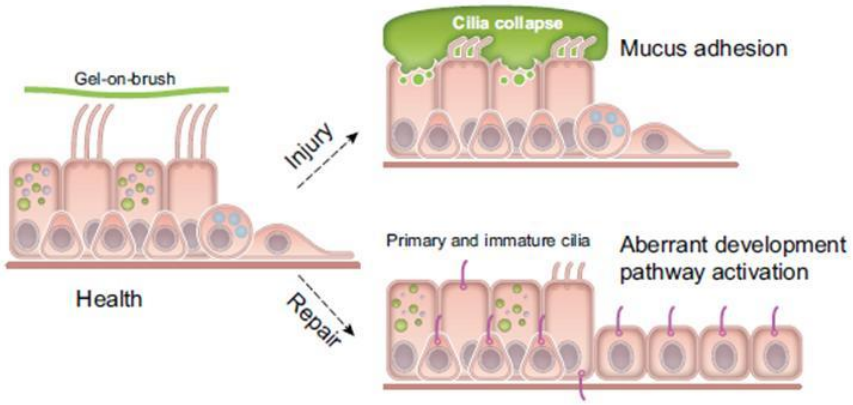

Figure 7*: Regulation of motile and non-motile cilia in the lungs. Under conditions of health, multiciliated cells maintain the normal structure and function of apically localized motile cilia. In addition, membrane mucins present along cilia adsorb water and maintain a stable "grafted brush" structure that supports an overlying gel. Normal repair programs following injury activate progenitor basal-like cells to differentiate fully to secretory and ciliated cells. Primary nonmotile cilia precede motile cilia in this normal repair process. Under conditions where mucins are overproduced or hypersecreted (such as the MUC5B promoter variant rs35705950), or when electrolyte homeostasis is disrupted (e.g., through excessive Na absorption), cilia collapse and aggregates of mucus adhere to airway surfaces, potentially worsening injury. Under injurious conditions, resulting from excess mucus or other forms of lung injury, aberrant repair could lead to abnormal activation of progenitor basal-like cells, resulting in partially differentiated ciliated cells that retain primary cilia, have poorly developed apical motile cilia, and developmental programs that are aberrantly activated. From Evans C.M. et al., Physiol Rev. 2016;96(4):1567-91.*

In the study of Yang et al[64], *MUC5B* promoter variant was associated with a 34.1-fold increase in *MUC5B* expression in lung tissue among unaffected subjects. MUC5B immunohistochemical staining showed dense accumulation of this protein in terminal bronchioles as well as the pseudostratified bronchial epithelium and lumen of honeycomb cysts but not in fibroblastic foci. The authors speculated on the possible role

that MUC5B plays in the pathogenesis of IPF: they proposed at least two distinct pathogenetic mechanisms that are not mutually exclusive, and that they could both contribute to the pathogenesis of IPF. One possible mechanism is that MUC5B overexpression results in chronic mucus hypersecretion and accumulation in the peripheral airspace, which in turn impairs mucociliary transport (also observed in the other research by Hancock et al.), resulting in mucus adhesion in the bronchoalveolar region, and consequently chronic inflammation and injury. Indeed, MUC5B accumulates in terminal bronchioles and areas of microscopic honeycombing in IPF lung. The other distinct mechanism is that MUC5B overexpression results in an aberrant repair process after injury to the bronchoalveolar regions of the lung. In this case, honeycomb cysts in IPF lung would represent a failed regenerative process after injury[64].

In the review of Zhang Q. et al[62], the authors highlight that although the mechanisms how MUC5B have a pathological effects on IPF remain unclear, two hypotheses are possibilities and consistent. First, excessive MUC5B compromises the mucosal host defense and reduces lung clearance of inhaled particles, dissolved chemicals, and microorganisms. Over time, reduced clearance may lead to scar tissue formation and persistent fibroproliferation that expands and displaces normal lung tissue. Given that cigarette smoking is a strong risk factor for the development of IPF, it is logical to speculate that the inhaled particles associated with cigarette smoking might cause defects in mucosal host defense and subsequent interstitial injury. Second, excessive MUC5B in the respiratory bronchioles may interfere with alveolar repair. It has been established that local expansion of type II alveolar epithelial cells following lung injury may repopulate denuded alveolar basement membranes. One possibility is that MUC5B impedes alveolar repair either by interfering with the interaction between type II

alveolar epithelial cells and the underlying matrix or by interfering with the surface-tension properties of the surfactant. The failure to reepithelialize damaged alveoli could enhance the collapse and fibrosis of bronchoalveolar units and eventually lead to the development of idiopathic pulmonary fibrosis. The changes of MUC5B in the distal conducting airways potentially enhance injury or disrupt repair responses in alveoli.

Besides the promoter variant rs35705950, which definitely causes an overexpression of *MUC5B* and appears to be predictive of IPF, little is known about the influencing factors of gene expression. Identifying the associated influencing factors of the *MUC5B* gene in pulmonary diseases will contribute to the exploration of explicit signaling pathways in MUC5B overproduction. At present, genetic mutation, epigenetic changes, effects of some transcriptional factors as well as inflammatory mediators, and the associated signaling pathways may be involved in the regulation of MUC5B production[62] (figure 8).

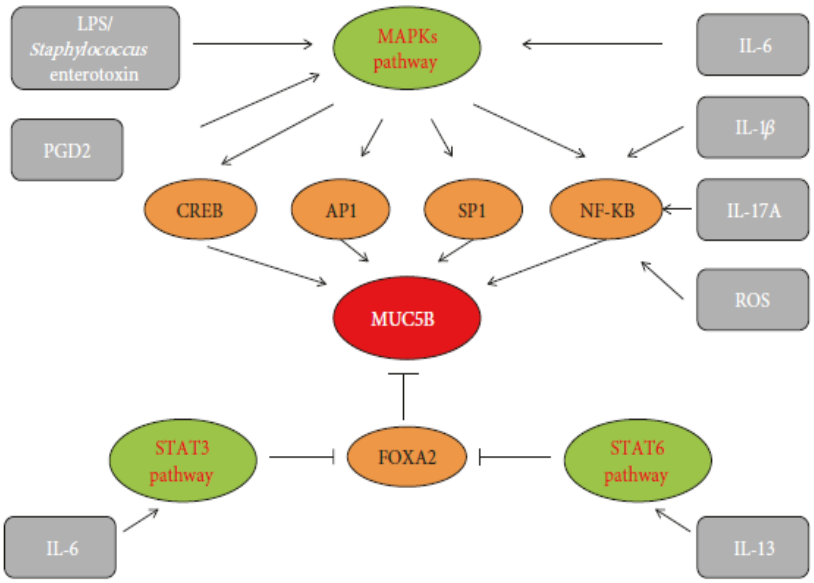

Figure 8: *regulation of MUC5B expression in airway epithelial cells in IPF. Different stimuli, including some interleukins, ROS, PGD2, and certain bacterial components may induce MUC5B overexpression by MAPKs, STAT3, or STAT6 pathways. The important transcriptional factors including FOXA2, CREB, AP-1, SP1, and NF-κB also directly or indirectly participate in MUC5B overexpression. From Zhang Q. et al., Biomed Res Int. 2019;2019:9712464.*

Several study have also demonstrated that aberrant expression of mucins molecules play a major role in various cancers and has been associated with proliferation, altered cellular adhesion, invasion and metastasis properties of cancer cells (figure 9). Among the various membrane bound mucins, MUC1 (aberrantly expressed in several epithelial cancers such as breast, ovarian and pancreatic cancer), MUC4 (involved in malignant transformation, metastasis and altered cellular signaling in the tissues of pancreatic, ovarian, endometrial and breast cancer) and MUC16 (involved in the EMT process by altering cell surface proteins as well as intracellular downstream signaling molecules in fact loss of cell surface expression of MUC16 induces mesenchymal features such as longer and fibroblast like structures) are so far identified to be associated with the EMT process in various cancers. Also the gel forming mucin MUC5AC appears to be expressed in many cancers such as pancreas, colon and lung etc. A study demonstrated that MUC5AC is overexpressed in pancreatic cancer and significantly correlated with proliferation, invasion and motility of cancer cells. MUC5B was also shown to be involved in tumor cell invasion and metastasis in luminal type of breast cancer cells by affecting cell-cell and cell-matrix interaction[65].

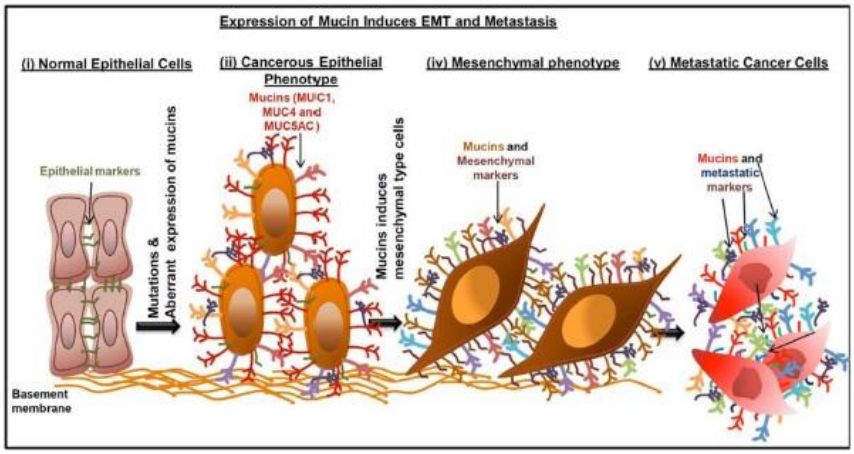

Figure 9: *aberrant expression of mucins induce EMT phenotype. Normal epithelial cells carry many epithelial markers and junctional proteins that anchor epithelial cells to the basement membrane. Many of the cancerous cells apparently express mucin molecules and upregulates the EMT markers. Mucin molecules involved in the induction of mesenchymal cell type. Metastatic process initiated by mesenchymal cell types in relation to mucin expression. From Ponnusamy MP. et al., Curr Cancer Drug Targets. 2013;13(9):945-56.*

1.5.1.2) Exogenous exposures

Several environmental exposures increase the risk of IPF, with the most consistent being the cigarette smoking. IPF patients with cigarette smoking history have a poorer survival compared to non-smokers. Occupational exposures such as agriculture and farming, livestock, wood dust, metal dust, stone dust and silica has also been linked to an increased risk[66].

Several studies have also suggested that viruses might play a part in the pathogenesis of IPF with Epstein–Barr virus (EBV) and latent human herpesviruses (such as cytomegalovirus and Kaposi sarcoma-associated herpesvirus)[67,68].

Also bacteria have a possible rule in IPF pathogenesis with Staphylococcus spp. and Streptococcus spp. An imbalance in bacterial composition has also been observed in patients with interstitial lung disease, when compared with healthy lungs[69,70].

Studies investigating drugs as antivirals, antibiotics, and antifungals have shown a great promise for IPF treatment, consolidating the link between microbiome and IPF[71].

Microaspirations of gastric content have been suggested as another potential cause of injury to the lung epithelium in IPF with an elevated prevalence of gastro-esophageal reflux in IPF patients compared with other patients[72].

1.5.1.3) Ageing

Aging can be defined as the inevitable time-dependent functional decline, characterized by progressive loss of physiological integrity, reduced homeostatic control and increased vulnerability to death[73].

Lopez-Otin et al.[74], proposed nine "hallmarks of aging" including genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient

sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication.

Since aging confers a risk for IPF, some have suggested that IPF may represent a form of "accelerated" lung aging[75]. Alveolar epithelial cells and fibroblasts have been shown to assume senescent identities in IPF, contributing to development of fibrosis[76,77,78].

Established senescent markers such as p21, p16 and senescence-associated β-galactosidase activity (SA-β-gal) are increased in both fibroblasts and epithelial cells in human IPF lung tissue as compared with controls[78,79].

IPF derived fibroblasts exhibit accelerated replicative cellular senescence and increased resistance to oxidative stress induced cytotoxicity as compared to normal lung fibroblasts. A recent paper suggests that fibroblasts from lungs of old mice express a fibrogenic phenotype that leads to resistance to apoptosis and increased susceptibility to fibrotic response after injury. These findings have been partially associated with an increased expression of plasminogen activator inhibitor 1 (PAI-1), which is an effector of Transforming Growth Factor-Beta 1 (TGF-β1), a key factor in the development of senescence through the induction of p21[80].

1.5.1.4) Epigenetic alterations

Any process that modifies gene activity without changing the underline genetic code is defined as epigenetic alteration. Traditionally, epigenetic modifications refer to DNA methylation and histone modifications. The leading mechanisms of DNA methylation and histone modifications seem to mediate both genetic and environmental influence on gene expression and disease features, especially with age. Increasing evidences support a central role for epigenetic alterations in IPF[81,82]. DNA methylation changes consist of

both hyper- and hypo-methylation of cytosine residues in different genes, with accidental errors in methylation[83].

In a 2014 study, a genome-wide DNA methylation analysis of lung tissue involving 94 patients with IPF and 67 controls, recognized 2130 genome-wide differentially methylated regions of which about a third were associated with significant changes in gene expression[84].

Other studies in patients with IPF have also identified significant modifications in miRNAs with potential pathogenetic roles. Evidences have identified significant changes in the levels of regulatory miRNAs in IPF patients when compared with healthy subjects[85].

Few studies have also examined the role of structural changes of chromatin in IPF. One report showed that almost all class I and II histone deacetylases are upregulated in IPF, mainly in myofibroblasts and in abnormal bronchiolar basal cells[86]. However, their exact targets and effects in the setting of IPF are currently unknown[48].

Cigarette smoking and ageing are the main effectors of epigenetic modifications, given their association with IPF and the relationship between them and DNA methylation[87].

1.5.2) Initiation/progression stage

In the progression phase, the normal alveolar structure of the lung is lost and replaced by remodeled fibrotic tissue characterized by bronchiolised cystic airspaces, which might include continuous proliferation of bronchiolar epithelium to honeycomb cysts[32]. This stage is characterized by the activation of some pro-fibrotic process that are discussed above.

1.5.2.1) Endoplasmic reticulum stress

the activation of endoplasmic reticulum (ER)-stress-mediated activation of unfolded protein response (UPR). The ER is the cellular compartment where secreted and membrane proteins are made, folded, and matured for packaging and trafficking via the Golgi complex. ER stress occurs when there is an imbalance between cellular demand for protein synthesis and the ER's capacity to synthesize, process, and package proteins. In response to this stress, the cell activates a cellular response termed the unfolded protein response (UPR) that leads to activation of biochemical pathways designed to match the protein production capacity of the ER. If the UPR cannot match the demand, a terminal UPR is activated and the cell sacrifices itself through apoptotic pathways. Studies have reported that markers of UPR activation are elevated in alveolar type II cells of patients with IPF[88].

UPR stimulates also the production of profibrotic mediators, as TGF-β1, PDGF, CXCL12 (C-X-C motif chemokine 12), CCL2 (chemokine C-C motif ligand 2)[85].

1.5.2.2) TGF-β activation

TGF-β1 is probably the most important mediator involved in IPF pathogenesis. Levels of transforming growth factor β (TGF-β) are increased in the lungs of patients with IPF. AEC2s may produce it as a consequence of actin/myosin-mediated cytoskeletal contraction induced by UPR, through αvβ6 integrin activation. All three isoforms (TGF-β1, TGF-β2, and TGF-β3) of inactive TGF-β are synthesized and secreted bound to the latency-associated peptide (LAP). Under normal conditions, TGF-β is bound in its inactive form to LAP cross-linked to the extracellular matrix. During the development of lung fibrosis, AECs express increased levels of the integrin αvβ6, which can bind to

the arginine-glycine-aspartate (RGD) sequence of LAP. Activation of epithelial cells expressing αvβ6 through binding of mediators to their receptors induces actin/myosin-mediated contraction of the epithelial cell. This contraction pulls on the latent TGF-β tethered to the epithelial cell via binding of LAP to the integrin αvβ6. This retraction induces TGF-β activation and signaling in localized, because the active TGF-β does not appear to be released into a soluble form. The αvβ6 integrin/TGF-β1 pathway is a fundamental biological process: the molecules are constitutively bound, suggesting that the system is primed to detect injurious stimuli[88].

TGFβ1 is a strong pro-fibrotic mediator: it promotes epithelial cell apoptosis, epithelial mesenchymal transition (EMT), epithelial cell migration, production of other profibrotic mediators, circulating fibrocytes recruitment and fibroblasts activation, proliferation and transformation into myofibroblasts, production of VEGF, CTGF (connective-tissue growth factor) and other proangiogenic mediators and several other pathways[85].

1.5.2.3) Epithelial-to-mesenchymal transition

Epithelial-to-mesenchymal transition (EMT) is the process by which epithelial cells acquire molecular and cell physiologic features commonly associated with mesenchymal cells following activation by specific growth factors, of which TGF-β is the prototype[88].

A typical epithelium is a sheet of cells, often one cell thick, with individual epithelial cells bound each other in a uniform array. Spaced cell–cell junctions and adhesions between epithelial cells hold them tightly together and inhibit the movement of individual cells away from the epithelial monolayer. The epithelial sheet is polarized, meaning that the apical and basal surfaces are likely to be visually different.

On the other hand, adhesions between mesenchymal cells are less strong than in their epithelial counterparts, allowing for increased migratory capacity. These cells also have a more extended and elongated shape and, unlike epithelia, the irregular structure of mesenchyme does not allow for rigid to pological specialization. Epithelial cells move as a sheet en block, whereas mesenchymal migration is considerably more dynamic. Mesenchymal cells move individually and can leave part of the trailing region behind.

Turning an epithelial cell into a mesenchymal cell requires alterations in morphology, cellular architecture, adhesion, and migration capacity (figure 10). Commonly used molecular markers for EMT include increased expression of N-cadherin and vimentin, αSMA, nuclear localization of β-catenin, and increased production of the transcription factors such as Snail1 (Snail), Snail2 (Slug), Twist, EF1/ZEB1, SIP1/ZEB2, and/or E47 that inhibit E-cadherin production. Phenotypic markers for an EMT include an increased capacity for migration and three-dimensional invasion, as well as resistance to apoptosis[89].

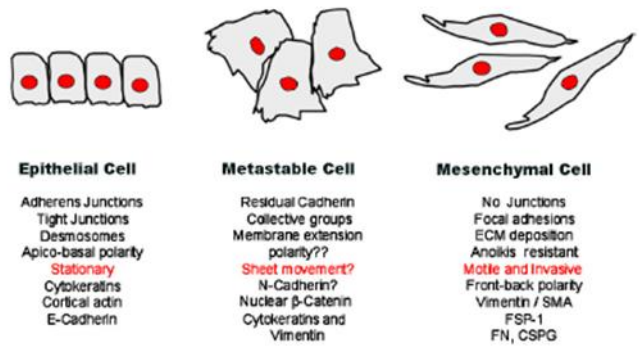

Figure 10: *different phenotypes of epithelial, metastable cell (hybrid cell showing both epithelial and mesenchymal traits) and mesenchimal cells. From Lee JM et al., J Cell Biol. 2006;172(7):973-81.*

This molecular reprogramming of EMT, which causes epithelial cells to express mesenchymal cell–associated genes, occurs in three biologic contexts: development, cancer, and fibrosis. These changes in protein expression cause epithelial cells to lose their polarity and tight junctions and to become more mobile. The evidence that alveolar

cells exhibit EMT in IPF patients is based on immunohistochemical studies colocalizing epithelial cell– and mesenchymal cell-associated proteins within IPF lungs[88].

In their study Morbini et al.[90] observed that only bronchiolar basal metaplastic cell in IPF, rather than hyperplastic pneumocytes, showed evidence of complete EMT with loss of E-cadherin and the expression of vimentin and fibronectin. So, only bronchial basal cells seems to be specific of UIP and might have a rule in the development of the disease[90].

The UPR, TGF-β, and EMT are activated in patients with IPF, but how this activation occurs is still widely undefined. Genetic predisposition may explain the activation of these processes in some patients (figure 11).

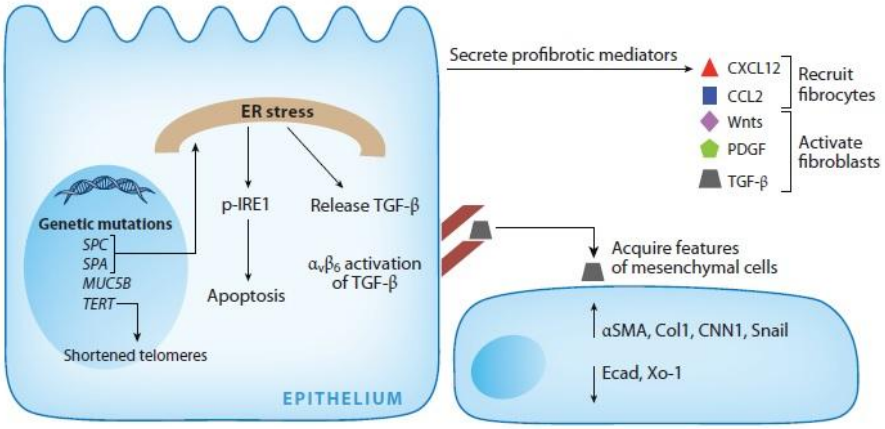

Figure 11: *profibrotic attributes of epithelial cells in IPF lungs. Genetic mutations in epithelial cell–associated proteins predispose to the development of lung fibrosis by leading to the development of short telomeres or endoplasmic reticulum (ER) stress. ER stress may be profibrotic by causing apoptosis or release of TGF-β. Epithelial cells release profibrotic mediators such as Wnts, PDGF, and TGF-β, which activate fibroblasts, or CXCL12 and CCL2, which recruit fibrocytes to the lung. Epithelial cells produce the integrin αvβ6, which activates TGF-P and may thereby cause epithelial cells to undergo epithelial-to-mesenchymal transition. From Wolters PJ et al., Annu Rev Pathol. 2014;9:157-79.*

1.5.3) Progression stage

During this stage, pathologic mesenchymal cells release abnormal types and quantities of matrix proteins, which remodel and scar the lung. The pathological matrix promotes additional differentiation of fibroblasts to myofibroblasts, which deposit more matrix and further activate fibroblasts in a feed-forward loop of lung remodeling. The development of lung remodeling and fibrosis depends on an interaction between (a) an epithelium that is dysregulated by the underlying genetic mutations, UPR activation, and/or EMT and (b) fibroblasts, the collagen-secreting cells of the lung.

Fibrocytes are bone marrow-derived mesenchymal cells that can be identified in the circulation or in tissues. These cells contribute to development of lung fibrosis by directly producing extracellular matrix proteins such as type I and type III collagen, or by differentiating into fibroblasts or myofibroblasts, or by producing cytokines that induce collagen deposition. Fibrocytes have been found both in circulation and in the lung parenchyma of patients with IPF[88].

Heukels et al.[91] demonstrate that lung fibrocytes have different characteristics than circulating fibrocytes and represent an intermediate cell population between circulating fibrocytes and lung fibroblast. They found also increased percentages of circulating fibrocytes in patients with IPF compared to the control population[91].

Also another study research found increased counts of fibrocytes in the blood of patients with IPF compared with healthy control subjects. Fibrocytes did not correlate with clinical parameters but were predictive of higher mortality[92].

The alveolar epithelium may play a role in recruiting fibrocytes to IPF lungs. Fibrocytes express the chemokine receptors CXCR4 and CCR2, similarly the alveolar epithelium of IPF patients expresses CCL2 , the ligand for CCR2. These observations suggest that

circulating fibrocytes may be recruited to IPF lungs through a CXCR4/CXCL12 or CCR2/CCL2 axis and that expansion of fibrocytes in the lungs may contribute to IPF.

Epithelial cell dysfunction and aberrant epithelial-mesenchymal signaling lead to the activation of fibroblasts and the deposition and remodeling of matrix. Several mediators, including TGF-β, can induce the differentiation of fibroblasts to myofibroblasts that are the classic pathologic fibroblast phenotype described in IPF lungs. Compared with resident lung fibroblasts, myofibroblasts secrete excessive amounts of matrix, including type I collagen. This excess matrix deposition may lead to pathologic lung fibrosis and remodeling.

IPF has features similar to those of cancer. Most notable is the relentless progression of IPF, which is similar to that of cancer, and the appearance of foci of fibroblasts, which are reminiscent of tumorlets. Although the foci of fibroblasts are more accurately described as reticula of fibroblasts, their anatomic appearance suggests the possibility that fibroblasts isolated from IPF patients may be able to invade the extracellular matrix much like metastatic cancer cells do. Significantly, IPF fibroblasts invade artificial basement membranes more readily than control fibroblasts do. The mechanisms for this enhanced invasion are poorly understood[88].

Overall, when the pathogenesis of IPF is considered as a continuum of predisposition, activation, and progression, distal bronchiolar and alveolar epithelial lung cells are shown to be the pathologically abnormal cells in IPF lungs, and fibrosis to be the consequence of epithelial-cell dysfunction. Therefore, IPF is a disease of lung epithelial cells that manifests as fibrosis rather than being an intrinsically fibrotic disease[32].

1.6) Histopathology of Idiopathic Pulmonary Fibrosis

The appearance of lungs with IPF shows a characteristic fibrosis that is distributed along the inferior portions of the lobes with sub-pleural accentuation, that causing remodeling of lung architecture. The pleural surface has a cobblestone appearance with areas of air space enlargement, and fibrotic reaction that resembles cirrhotic liver. This pattern of fibrosis has been called microscopic honeycombing[88]. These areas are frequently lined by bronchiolar epithelium and filled with mucus and inflammatory cells[35].

The typical microscopic appearance of IPF has been termed as Usual Interstitial Pneumonia (UIP) pattern with spatial and temporal heterogeneity (figure 12).

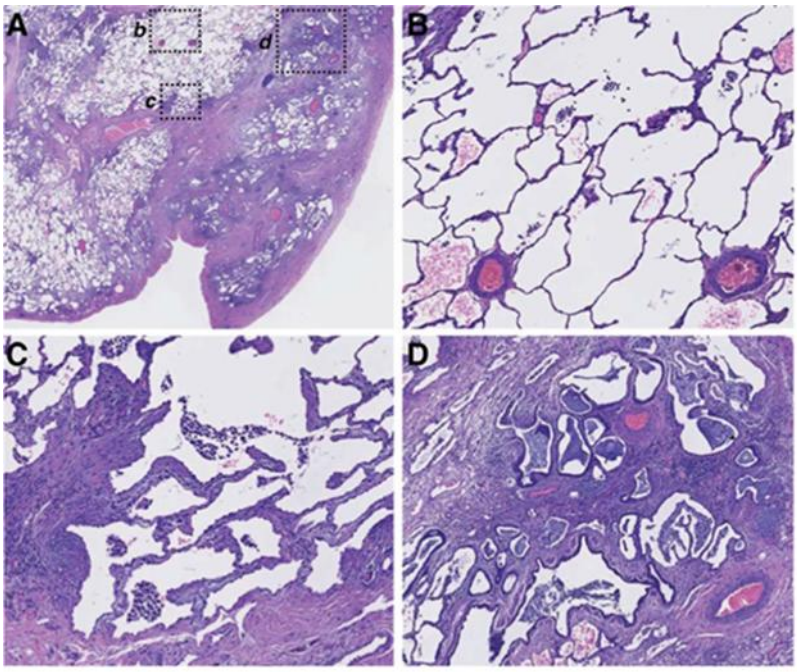

Figure 12: *spatial heterogeneity of pathophysiology in a single representative IPF explanted lung tissue. (A) Low-power magnification of a H&E-stained section. b, c, and d indicate regions shown at higher magnification in: (B), normal-appearing alveolar structure, (C) transition zone between the normal and the abnormal fibrotic parenchyma, (D) advanced scar tissue with microscopic honeycombing and bronchiolisation. From De Pianto DJ et al., Thorax. 2015 Jan;70(1):48-56*

The spatial heterogeneity of the lesion result in patchy fibrosis composed mainly of dense collagen that is prevalent in the peripheral portion of the secondary pulmonary lobule and spares the central portion of the lobule. As one moves from the peripheral to the central portion of the pulmonary lobule, the lung injury patterns changes from chronic to acute to absent (also reflect the temporal heterogeneity)[88]. These areas of fibrotic appearance are then alternates with areas of less affected parenchyma.

A characteristic finding of UIP pattern is the presence of fibroblast foci, foci of proliferating fibroblast and myofibroblast that are arranged in a linear fashion within a pale staining matrix, at the transitional area from dense scarring to non-fibrotic regions of the lobule (figure 13). Overlying epithelium consists of hyperplastic pneumocytes or columnar non-ciliated bronchiolar cells[88]. Fibroblast foci are characteristic of UIP and represent an important diagnostic feature when seen in the context of patchy fibrosis and honeycomb change. The presence of these microscopic zones that are considered to be the sites of acute lung injury set against a backdrop of chronic scarring accounts for the temporal heterogeneity typical of UIP. Fibroblastic foci have also prognostic significant because their numbers are associated with poor survival[93].

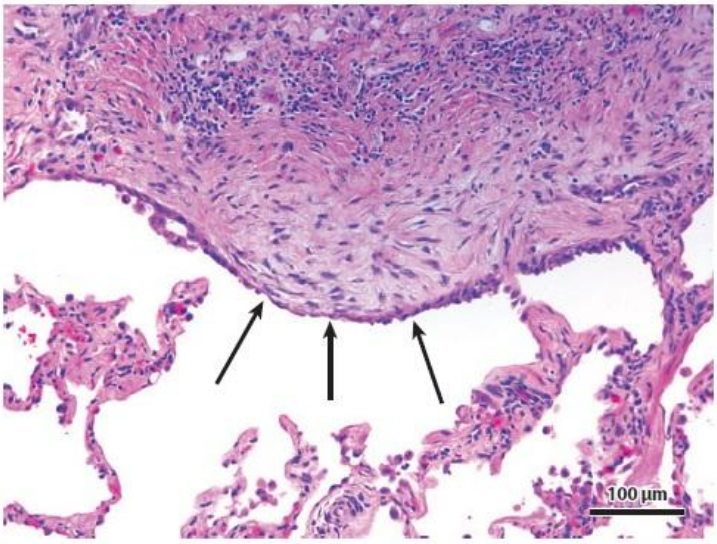

Figure 13: *fibroblast foci. The fibroblast focus shows a proliferation of spindled fibroblasts within a myxoid matrix, producing a bulge of tissue into the airspace. The overlying epithelium often shows plump reactive-appearing cells (arrows). From Wolters PJ et al., Annu Rev Pathol. 2014;9:157-79.*

Inflammatory component is usually mild, and consists of a patchy interstitial infiltrate of lymphocytes and plasma cells associated with hyperplasia of type 2 pneumocytes and bronchiolar epithelium[35]. Inflammation occurs mainly in areas of collagen deposition or honeycomb change, and rarely involves unaltered alveolar septa, so it is not a prominent histopathologic finding in usual interstitial pneumonia[47]. Smooth muscle metaplasia in the interstitium is commonly seen in areas of fibrosis and honeycombing[35].

To date, there are two guidelines that classified IPF from a pathological point of view: the Fleischner Society Guideline 2018[39] and the ATS/ERS/JRS/ALAT Clinical Practice Guideline 2018[35].

According to the new ATS/ERS/JRS/ALAT (American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Latin American Thoracic Association) Clinical Practice Guideline 2018 for the diagnosis of IPF, the histopathologic findings of biopsies can be categorizing into "UIP," "probable UIP," "indeterminate for UIP" and "alternative diagnosis"[35]. Characteristic of each histopathology are summarized in figure 14.

UIP	Probable UIP	Indeterminate for UIP	Alternative Diagnosis
• Dense fibrosis with architectural distortion (i.e., destructive scarring and/or honeycombing) • Predominant subpleural and/or paraseptal distribution of fibrosis • Patchy involvement of lung parenchyma by fibrosis • Fibroblast foci • Absence of features to suggest an alternate diagnosis	• Some histologic features from column 1 are present but to an extent that precludes a definite diagnosis of UIP/IPF *And* • Absence of features to suggest an alternative diagnosis *Or* • Honeycombing only	• Fibrosis with or without architectural distortion, with features favoring either a pattern other than UIP or features favoring UIP secondary to another cause* • Some histologic features from column 1, but with other features suggesting an alternative diagnosis†	• Features of other histologic patterns of IIPs (e.g., absence of fibroblast foci or loose fibrosis) in all biopsies • Histologic findings indicative of other diseases (e.g., hypersensitivity pneumonitis, Langerhans cell histiocytosis, sarcoidosis, LAM)

Figure 14: *histopathology patterns by the new ATS/ERS/JRS/ALAT Clinical Practice Guideline 2018. From Raghu G et al., Am J Respir Crit Care Med. 2018;198(5):e44-e68.*

The most common conditions that should be distinguished histologically from IPF include chronic hypersensitivity pneumonitis, idiopathic non-specific interstitial pneumonia, connective tissue disease, and pleuroparenchymal fibroelastosis[39].

1.7) Multidisciplinary diagnosis of Idiopathic Pulmonary Fibrosis

A diagnosis of IPF requires exclusion of alternative causes of fibrosing ILD that requires a high level of capacity to understand all the clinical, radiological and pathological features[39].

In 2018, a multidisciplinary team composed by members of the Fleischner Society provides an updated approach to the diagnosis of IPF based on a systematic search of the medical literature and expert opinions. This guideline, focus on the important of multidisciplinary team for the assessment of patient with fibrotic ILD with the ability of the specialist to interpret and communicate clinical data and the ability to integrate conflicting information. The multidisciplinary diagnosis consisted of the integration of views from radiologists, pathologists, and pulmonary specialists with an interactive multidisciplinary discussion. A major advantage of multidisciplinary diagnosis is that it reduces diagnostic imprecision due to recognized limitations in each of the three domains (ie, clinical, radiological, and pathological) by combining information from all three; however, the accuracy of each domain is influenced by the individual experience of the clinician, radiologist, and pathologist involved[39].

1.7.1) Diagnostic flowchart criteria for IPF

A correct diagnosis of IPF requires the following steps:

1. Exclusion of other known causes of ILD (e.g., domestic and occupational environmental exposures, connective tissue disease, drug toxicity), and either #2 or #3;

2. The presence of the HRCT (High Resolution Computer Tomography) pattern of UIP;

3. Specific combinations of HRCT patterns and histopathology patterns in patients subjected to lung tissue sampling.

Patients with suspected IPF are initially evaluated for identifiable causes of ILD. If a potential cause is identified, the patient undergoes an evaluation to confirm or exclude hypersensitivity pneumonitis, connective tissue disease, pneumoconiosis, and iatrogenic causes. If a specific diagnosis is not made or no potential cause for ILD is identified, then clinical findings and HRCT are considered during multidisciplinary discussion to either ascertain or exclude the diagnosis of IPF.

IPF is diagnosed if the appropriate combination of HRCT patterns and histopathological patterns are present[35] (figure 15).

IPF suspected*		Histopathology pattern			
		UIP	Probable UIP	Indeterminate for UIP	Alternative diagnosis
HRCT pattern	UIP	IPF	IPF	IPF	Non-IPF dx
	Probable UIP	IPF	IPF	IPF (Likely)**	Non-IPF dx
	Indeterminate for UIP	IPF	IPF (Likely)**	Indeterminate for IPF***	Non-IPF dx
	Alternative diagnosis	IPF (Likely)** /non-IPF dx	Non-IPF dx	Non-IPF dx	Non-IPF dx

Figure 15: *IPF diagnosis based upon the combination of HRCT and biopsy patterns. From Raghu G et al., Am J Respir Crit Care Med. 2018;198(5):e44-e68.*

In this guideline is also present an algorithm for the diagnosis of IPF (figure 16).

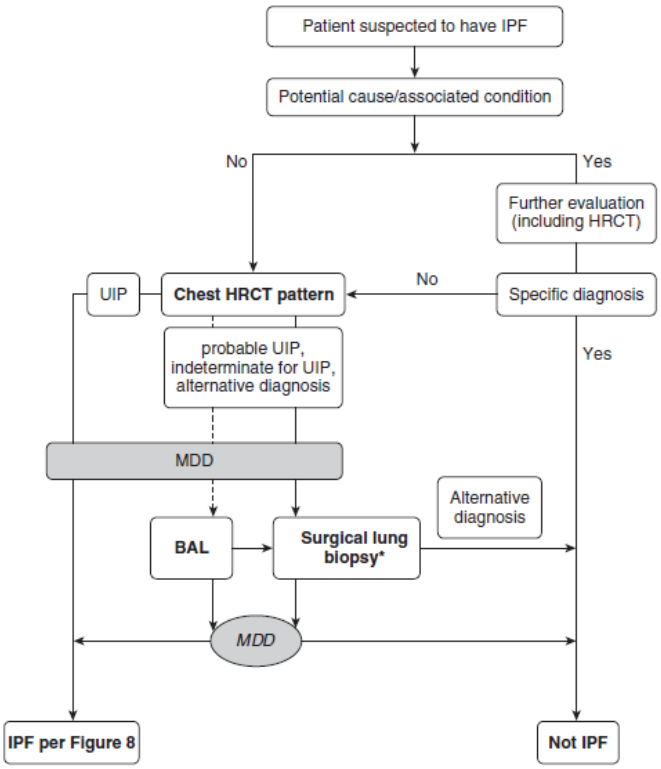

Figure 16: *diagnostic algorithm for idiopathic pulmonary fibrosis (IPF). Patients with suspected IPF (i.e., unexplained symptomatic or asymptomatic bilateral pulmonary infiltrates on a chest radiograph or chest computed tomography [CT] scan, bibasilar inspiratory crackles, and age older than 60 yr), unexplained dyspnea on exertion, and/or cough with evidence of interstitial lung disease (ILD) should be carefully evaluated for potential and/or identifiable causes of ILD, such as domestic and occupational environmental exposures, connective tissue disease (CTD), or drug toxicity. Middle aged adults (40 yr and 60 yr), especially patients with risks for familial pulmonary fibrosis, can rarely present with the otherwise same clinical scenario as the typical patient older than 60 years. If a potential cause for ILD is identified, the patient should undergo a thorough evaluation to confirm or exclude other known causes, such as hypersensitivity pneumonitis, CTD, pneumoconiosis, an iatrogenic causes (e.g., drug toxicity, irradiation). If a specific diagnosis is not made or no potential cause for ILD is identified, further evaluation is influenced by the patterns of high-resolution CT (HRCT) images of the chest and supportive clinical findings surfaced in the course of multidisciplinary discussion to ascertain or exclude the diagnosis of IPF. IPF is diagnosed if the appropriate combination of HRCT patterns and histopathological patterns are present. From Raghu G et al., Am J Respir Crit Care Med. 2018;198(5):e44-e68.*

1.8) Idiopathic Pulmonary Fibrosis and treatments

1.8.1) Nonpharmacologic Management

Nonpharmacologic management strategies help patients with IPF live healthier and more normal. Smoking cessation should be a priority for patients who are actively using tobacco because smoking worsens the disease. Influenza, pneumococcal, and other age-appropriate vaccines should be administered.

Clinical practice guidelines also strongly recommend supplemental oxygen for patients with IPF that reduces exertional dyspnea and improves exercise tolerance[35]. Pulmonary rehabilitation, a structured exercise program designed for adults with advanced lung disease, has been shown to improve exercise capacity and health-related quality of life for patients with IPF[94].

To date, lung transplantation remains the only definitive therapeutic approach. Unless specific contraindications exist, patients with severe functional impairment, oxygen dependency and a deteriorating course of the disease, should be listed for lung transplantation. Relative contraindications to lung transplantation include unstable or inadequate psychosocial profile/stability or significant extrapulmonary disorders (e.g., liver, renal, or cardiac dysfunction) that may negatively influence survival. Many centers limit lung transplantation candidates to 60 years of age[95].

Lung transplantation can improve survival and quality of life in highly selected candidates; however, among recipients 2010-2012, 1-year survival was 85.8%; 68.8% of transplant recipients surviving to 3 years and 56.1% to 5 years[96]. Frequent complications include primary graft dysfunction, acute and chronic forms of allograft rejection, cytomegalovirus or other infections, and cancer and are mainly responsible for the high mortality rates after lung transplantation[96].

1.8.2) Pharmacologic Management

During the past 5 years, notable advances have been made in pharmaco-therapeutic approaches to IPF. The 2015 ATS/ERS/JRS/ALAT guideline update was revolutionary in conditionally recommending treatment with several pharmacotherapies on the basis of safety and efficacy showed in clinical trials[97].

Two recently approved antifibrotic treatments, nintedanib and pirfenidone, have been shown to be safe and effective in the treatment of IPF by slowing disease progression (figure 17). Both have shown some efficacy in reducing severe respiratory events, such as acute exacerbations, and hospitalization for respiratory events[20].

Variable	Nintedanib	Pirfenidone
Mechanism of action	Tyrosine kinase inhibition	Inhibition of TGF-β production and downstream signaling, collagen synthesis, and fibroblast proliferation (selected list)
Efficacy	Slows FVC decline by 50%	Slows FVC decline by 50%
FDA-approved dose	150 mg by mouth twice daily	801 mg by mouth thrice daily
Common side effects	Diarrhea	Anorexia, nausea, photosensitivity
Enzyme metabolism	Ester cleavage (major), CYP 3A4 (minor)	CYP 1A2 (major), other CYP enzymes (minor)
Cautions	Risks of both bleeding and arterial thrombosis; risk of gastrointestinal perforation (rare); anticoagulant and prothrombotic drugs should be avoided	CYP 1A2 inhibitors (e.g., fluvoxamine and ciprofloxacin) can raise pirfenidone levels; CYP 1A2 inducers (e.g., omeprazole and smoking) can lower pirfenidone levels
Need for liver-function monitoring	Yes†	Yes‡
Clinical strategies to minimize side effects	Use of antidiarrheal agents, temporary dose reduction to 100 mg twice daily	Slow dose increase over 14-day period, medication to be taken with food, use of antacids, use of antiemetic agents, sun avoidance

Figure 17: *pharmacologic management of IPF. From Lederer DJ et al., N Engl J Med. 2018;379(8):797-798.*

Nintedanib is a multiple tyrosine kinase inhibitor that was approved for therapeutic use in European Union and in the United States in 2014. Nintedanib targets growth factor pathways, including vascular endothelial growth factor (VEGF) receptors 1, 2, and 3, fibroblast growth factor (FGF) and planted-derived growth factor (PDGF) receptors[98]. Adverse effects of this drug include diarrhea, weight loss, liver toxicity and potential

bleeding risks; despite these effects, nintedanib was generally well tolerated in the clinical trials. Monitoring of liver function before and periodically during treatment is suggested. Arterial thromboembolic events occurred in numerically more nintedanib-treated patients than placebo-treated patients; myocardial infarction was the most common event. Caution has been advised in patients with known coronary artery disease and myocardial ischaemia[48].

Pirfenidone, an oral pyridine derivate bioavailable synthetic compound, was approved for the treatment of IPF in 2011 in the European Union and in 2014 in the United States. To date, randomized placebo-controlled phase III studies have demonstrated that pirfenidone significantly slows disease progression[99,100,101,102].

Pirfenidone has a number of anti-inflammatory, antioxidant and antifibrotic effects, including inhibition of collagen synthesis, down-regulation of TGF-β and tumor necrosis factor alpha, and a reduction in fibroblast proliferation[98]. Adverse effects include upper gastrointestinal symptoms, photosensitivity, skin rash, anorexia and liver toxicity; these effects were well tolerated in patients participating in clinical trials, especially when the dose was decreased. Monitoring of liver function before and periodically during treatment is suggested[48].

Both these drugs have shown therapeutic benefits in early IPF stages, thus early diagnosis and treatment is crucial for reducing functional decline, slowing disease progression and improving quality of life. To date, the exact mechanism and pharmacodynamics of both drugs remain not well defined[98].

It is difficult to recommend one agent over the other, since there have been no head-to-head comparisons. Metwork meta-analysis concluded that pirfenidone and nintedanib provide similar benefits[103,104].

Recent data on treatment that combines these agents suggest clinically significant gastrointestinal side effects[105].

1.9) Gene expression profiling in IPF

Our understanding of IPF has dramatically changed in the last two decades. This happened because of an increased availability of human tissues and the introduction of high throughput molecular technologies.

Indeed, the history of transcriptomics in pulmonary fibrosis is characterized by an increased technological throughput and enhanced data analysis. Gene expression microarrays which allowed the parallel analysis of hundreds and, later, thousands of genes, emerged in the second half of 90's. The first study reporting microarray analysis of an experimental model of pulmonary fibrosis was published in 2000 and ~6,000 transcripts were profiled[106].

The first study on human lung tissues was published in 2002 and included eight samples[107]. Even several years later, studies that aimed at profiling this disease included a numbers of samples from 12 to 88[108,109,43,110,111,112].

RNA sequencing (RNA-Seq) is a Next-Generation-Sequencing (NGS) approach that permits a deeper sequencing than microarray platforms, resulting in profiling the whole transcriptome including coding and non-coding RNAs. Thus, RNA-Seq has the advantage to detect larger dynamic ranges of transcripts, and identify novel transcripts and variants. Another important aspects related to this approach is the possibility to analyze also low-input and degraded RNA samples, enabling research on lung microenvironments archival tissues, typically formalin-fixed-paraffin-embedded

(FFPE)[113]. Transcriptomics studies revealed numerous novel molecules and pathways that seen relevant in IPF pathogenesis and are summarized in figure 18.

Gene ID[a]	Gene name	Direction of expression	Tissue localization	Relevant pathway
Expressed in lung epithelium in IPF				
MMP7	Matrix metallopeptidase 7	Increased	Lung (alveolar epithelial cells and fibroblasts), peripheral blood and BAL	Extracellular matrix degradation, defensins, SPP1, and WNT/β-catenin pathway
MMP3	Matrix metallopeptidase 3	Increased	Lung, epithelial cells	Extracellular matrix degradation, β-catenin pathway
MMP19	Matrix metallopeptidase 19	Increased	Lung, epithelial cells	Extracellular matrix degradation, PTGS2 pathway
MMP1	Matrix metallopeptidase 1	Increased	Lung, epithelial cells	Extracellular matrix degradation, mitochondrial function/HIF-1-alpha pathway
SPP1	Osteopontin	Increased	Lung (epithelial cells)	Extracellular matrix degradation
IGFBP-4	Insulin-like growth factor binding protein 4	Increased	Lung (alveolar and basal cells)	IGF1 pathway
CCNA2	Cyclin A2	Increased	Lung (alveolar epithelial cells)	Cell cycle regulation
HIF1A	Hypoxia-inducible factor-1 alpha	Increased	Lung (alveolar epithelial cells)	Hypoxia, p53/VEGF pathways
CAV1	Caveolin-1	Decreased	Lung	Cell cycle regulation, TGF-b/JNK pathway
SYN-2	Syndecan-2	Increased	Lung, alveolar macrophages	TGF-b pathway
TAGLN	Transgelin	Increased	Lung, ATII cells	TGF-b pathway
CRLF 1	Cytokine receptor-like factor 1	Increased	Lung, ATII	Th1 cells inflammatory response
EGFR	Epidermal growth factor receptor	Increased	Lung, epithelial cells	Reepithelization
LYCAT	Lysocardiolipin acyltransferase	Decreased	Lung (epithelial cells), peripheral blood mononuclear cell (PBMC)	Mitochondrial membrane potential
SERPINF1 (PEDF)	Pigment epithelium-derived factor	Increased	Lung	Angiogenesis
Fibroblasts related gene expression in IPF				
FOXF1	Forkhead box F1	Increased	Lung	COL1/ARPC1 pathway
VCAM-1	Vascular cell adhesion molecule 1	Increased	Lung, fibroblast foci and blood vessels	TGF-b/ERK/Cyclin D pathway
FKBP10	FK506-binding protein 10	Increased	Lung, fibroblasts, and CD68 (+) macrophages	TGF-b/Col I synthesis
RXFP1	Relaxin/insulin-like family peptide receptor 1	Decreased	Lung	TGF-b
TAZ	Transcriptional coactivator with PDZ-binding motif	Increased	Lung	CTGF and Col1 pathways
IGFBP3, IGFBP5	Insulin-like growth factor binding proteins 3 and 5	Increased	Lung	IGF pathway
WNT pathway in IPF				
WNT1, 3a, 5a, 7b, 10b, Fzd2 and 3, β-catenin, Lef1, Gsk-3β	Wingless and others	Increased	Lung, fibroblasts, alveolar and bronchial epithelium	Wnt signaling
LRP5	Wnt co-receptor	Increased	Lung, PBMC	Wnt and TGF-b pathway
WISP1	Wnt1-inducible signaling protein-1	Increased	Lung	Wnt signaling
Apoptotic response in IPF				
TWIST1	Twist basic helix–loop–helix transcription factor 1	Increased	Lung—fibroblastic foci	Apoptosis/PDGF pathway
CXCL12	Chemokine ligand 12	Increased	Lung	Inflammation
TNSF10, BAX, CASP6	Apoptotic regulators	Altered expression	Lung	Apoptosis
SHP2 (PTPN11)	SH2 domain-containing tyrosine phosphatase-2	Decreased	Lung	Apoptosis/Tyr and Ser/Thr kinase pathways
Host defense implicated in IPF				
DEFA3–4	Defensin alpha 3 and 4	Increased	Lung and peripheral blood	Host defense
AGER (RAGE)	Advanced glycosylation end product-specific receptor	Decreased	Lung and peripheral blood	Inflammation
Mitochondria-related genes in IPF				
PINK1	PTEN-induced putative kinase 1	Decreased	Lung	Dysfunction of mitochondria
DIO2	Iodothyronine deiodinase 2	Increased	Lung	TH pathway/mitochondrial biogenesis

Figure 18: *summary of relevant idiopathic pulmonary fibrosis (IPF) genes identified by transcriptome profiling. From Vukmirovic M et al., Front Med (Lausanne). 2018;5:87.*

Gangwar et al.[114], in 2017, attempts to unify the best available gene expression data for IPF towards generating a computational model of IPF pathology. In this manner, they try to unify the large volume of high throughput genomics data to derive the most common molecular signatures in IPF. A set of 39 differentially expressed genes (DEGs) were found critical and most of these genes were found associated with the process of lung development, maintenance, immune system signaling, collagen metabolism, extracellular matrix (ECM) deposition, lipid metabolism, and cell-cell interactions. miRNAs are one of the most important regulatory components of cell system. Clusters of [miR-181a, miR-181b], [miR-93, miR-106b], [miR-17, miR-18a, miR-92a] and [miR-30b, miR-30d] were differentially downregulated whereas cluster of miR-133a and miR-1 was upregulated in IPF. Down regulated miRNA clusters were associated with genes involved in Wnt, p53, Jak-STAT, PI3K-Akt, Prolactin signaling and cell adhesion molecules (CAMs) whereas upregulated miRNA cluster was found targeting genes of fructose and mannose metabolism, cAMP, Hedgehog, PPAR, AMPK signaling, sphingolipid and fatty acid metabolism.

A total of 56 pathways crosstalks were also found with the involvement of four crucial biological pathways (Hedgehog signaling, Wnt signaling, TGFβ signaling and Cytokine-chemokine signaling) that causing increased cell proliferation, adhesion, reduced differentiation, altered apoptosis and EMT. In their work Gangwar et al. identified so a system of five serially connected downregulated genes (ACADL-HMGCR-FLT1-FZD5-ARRB1), potentially targeted by nine different miRNAs and under control of eight different transcription factors stands out as a crucially IPF specific spot for promising therapeutic interventions for IPF[114].

All this transcriptomic profiling in IPF has largely been performed by microarrays using RNA obtained from whole lung lysates from fresh frozen tissues, so one research group

focus on RNA-Seq analysis from the whole lung FFPE tissues. They compared their results with those obtain from gene expression microarrays from RNA extract from fresh frozen tissues (FF). At the end they observed a high concordance between RNA-Seq (FFPE) and microarray (FF) expression profile because the overlap was statistically significant. The common genes were enriched for signaling pathways relevant to IPF such as: ECM remodeling process, WNT, TGF-β, NFAT, IL-8 in angiogenesis, CCL2 signaling and PEDF signaling[113].

A 2019 study[115], used total RNA to performed RNA sequencing from lung tissue of patients with advanced IPF that underwent lung transplantation at the University of Pennsylvania compared to donor lungs. The aim of the study was to identify the pathways and mechanisms that contribute to severe/end-stage IPF. They identified a strong up-regulation of pathways associated with tumor cell infiltration and development of cancer in addition to expected alterations in extracellular matrix (ECM) remodeling pathways and epithelial to mesenchymal transition (EMT). Notably, T-cell activation and survival pathways (CD4, CD8 and regulatory T-cells) were strongly up-regulated concomitant with a robust increase in the expression of the checkpoint effectors programmed cell death (PD)1, LAG3 and CTLA4, the T-cell co-stimulatory receptor CD28 as well as chemokine/chemokine receptors including CCR5, CCR6, CXCR3 and CXCR5. Genes involved in inflammatory signalling leading to myeloid-derived suppressor cells or M2 macrophages were strongly down-regulated. They also found that the cholesterol homeostasis pathway was strongly and significantly down-regulated in IPF compared with controls[115].

In a recent research conduct by Spek CA et al[116]. they would investigate the molecular profiling of IPF on the basis that IPF has been proposed to be a disease that appear like a malignant disorder of the lung, so they explored the similarity between IPF and cancer

at the transcriptome level, using microarray analysis, by comparing gene expression datasets of IPF and non-small cell lung cancer (NSCLC) patients. They found that only 123 genes were upregulated in both IPF and NSCLC that contains collagens and metalloproteases grouped into gene ontology categories like cell adhesion, extracellular matrix organization and collagen catabolism. 257 genes were instead downregulated in both IPF and NSCLC were only significantly enriched in the angiogenesis gene ontology category whereas a larger proportion of genes upregulated in IPF patients (n=127) was actually downregulated in NSCLC patients grouped into gene ontology categories related to ciliogenesis (i.e. cilium movement and cilium organization)[116].

One limit of this type of researches is that they don't take into account the changes in cellular compositions or the heterogeneity within IPF such that areas of disease are interspersed with normal areas. The use of whole lung tissue is so a limitation of transcriptomic studies because transcriptomic changes are cell-type specific. For this reason several studies focus on the transcriptomics analysis of specific cellular component of the disease such as epithelial cell, fibroblast, macrophages, etc.

One study extract RNA from lung fibroblast cultured from the surgical specimens of 8 patients with IPF and compared it with 4 normal lungs. With a microarray analysis they found that *CCL8* expression was 22.8-fold higher in IPF fibroblast compared with control fibroblast. This differential expression was validated using an additional number of fibroblast and bronchoalveolar lavage fluid samples where they found a CCL8 protein concentration significantly higher in patients with IPF[117].

In their study research, Sheu et al.[118] used diseased human lung fibroblasts and normal human lung fibroblasts purchased from a biotech company. They tried to identify the dysregulated genes in IPF fibroblasts, elucidate their functions and explore potential micro RNA (miRNA)‐mRNA interactions using a next‐generation sequencing

platform, and bioinformatics analyses. A total of 42 dysregulated genes were identified with the up-regulation of neurotrimin (*NTM*), paired box 8 (*PAX8*) and mesoderm development LRP chaperone. The genes that are down-regulation included ITPR interacting domain containing 2 and Inka box actin regulator 2 (*INKA2*). Gene Ontology analysis indicated that the most significant function of these 42 dysregulated genes was associated with the composition and function of the extracellular matrix (ECM). A total of 60 dysregulated miRNAs were also identified. The integrated analysis of mRNA and miRNA expression data, identified (hsa)-miR-1254-INKA2 and hsa-miR-766-3p-INKA2 as the potential miRNA-mRNA interactions hat may promote proliferation and survival of IPF fibroblasts[118].

Emblom-Callahan et al.[119] comparing the genomic phenotype of non-cultured pulmonary fibroblasts (fibroblasts that were isolated immediately post-lung explantation and were not subjected to long-term in vitro propagation) from advanced IPF patients and compared to non-cultured normal pulmonary fibroblasts, using an oligo-microarray revealed novel genes, biological processes and concomitant pathways previously unreported in IPF fibroblasts. They demonstrate altered expression in proteasomal constituents, ubiquitination-mediators, Wnt, apoptosis and vitamin metabolic pathways and cell cycle regulators, suggestive of loss of cellular homeostasis. Specifically, *FBXO32*, *CXCL14*, *BDKRB1* and *NMNAT1* were up-regulated, while *RARA* and *CDKN2D* were down-regulated. Paradoxically, pro-apoptotic inducers *TNFSF10*, *BAX* and *CASP6* were also found to be increased[119].

To interrogate mechanisms by which pathological activation of fibroblasts/myofibroblasts is regulated in IPF, Mullenbrock and colleagues performed genome-wide analyses of both mRNA and miRNA in these cells (primary fibroblasts

isolated from lung tissues of normal donors or IPF patients who underwent lung transplantation) using RNA-seq and miRNA-seq, and characterized their ECM deposition properties by proteomic analysis using mass spectrometry. RNA-seq analysis of the primary lung fibroblasts identified 168 differentially expressed genes (DEGs). The Gene Ontology (GO) enrichment analysis revealed that IPF fibroblast signatures are associated with the activation of several profibrotic signaling pathways such as WNT, TGF-β, NOTCH1 and HIF1A as well as inhibition of the anti-fibrotic PPARG pathway. To further probe potential downstream functions of genes altered in IPF fibroblasts, they used GO analysis and found that the disease fibroblast gene signatures are enriched for genes associated with "cell proliferation" (upregulated genes), "muscle contraction" (upregulated genes), "response to wounding" (downregulated genes), and "metallopeptidase activity" (downregulated genes). In addition, GO terms associated with ECM were the most significant and frequently observed terms for the up-regulated and down-regulated gene sets for IPF.

Compared to control fibroblasts, IPF fibroblasts exhibited 3 up-regulated and 16 down-regulated miRNAs. In particular, the miRNA signature includes many miRNAs previously linked to fibrosis. They also demonstrated experimentally that several "fibrotic miRNAs" they identified in the study (miR-29b-3p, miR-138-5P and miR-146b-5p) regulate the expression of fibrotic/ECM genes[120].

Few novel technologies have been welcomed with more excitement by the scientific community than single-cell RNA sequencing (scRNA-seq). Very recently, the report by Reyfman and colleagues provides important insight into pathogenic cell types in lung fibrosis on an unprecedented scale, making IPF the first chronic lung disease to be analyzed using scRNA-seq. Their study included 8 human subjects with pulmonary fibrosis and 8 donor lung samples. The authors analyzed and visualized their data as one

integrated set and also as individual data sets for each patient, revealing patient-specific cell types. The authors analyzed differential gene expression between donor and fibrotic lungs in macrophages, alveolar type II cells, and fibroblasts, and confirmed the results by sequencing "bulk" RNA of pooled, flow cytometry–sorted cells. Importantly, they identified a distinct, novel population of profibrotic alveolar macrophages exclusively in patients with fibrosis. They saw up-regulation of Notch ligands and Notch target gene expression in alveolar type II cells and club cells, with down-regulation of Notch target gene expression in endothelial cells. They detected low-level expression of several Wnt ligands in epithelial cell populations and fibroblasts in the normal and fibrotic lungs. They also identified rare cell populations including airway stem cells and senescent cells emerging during pulmonary fibrosis[121].

Another study focus on scRNA-seq to identify gene expression patterns of epithelial cells and associated biological processes involved in the pathogenesis of IPF. Transcriptomic analysis of normal human lung epithelial cells defined gene expression patterns associated with highly differentiated alveolar type 2 (AT2) cells with the identification of 3 distinct subsets of epithelial cell types with characteristics of conducting airway basal and goblet/club cells and an additional atypical transitional cell that contributes to pathological processes in IPF. Individual IPF cells frequently co-expressed alveolar type 1 (AT1), AT2, and selective markers, demonstrating "indeterminate" states of differentiation not seen in normal lung development. Pathway analysis predicted aberrant activation of canonical signaling via TGF-β, HIPPO/YAP, P53, WNT, and AKT/PI3K[122].

A recent study, utilized also this technique to confirm the origin cell identities for MMP7, CCL18 and MUC1 with the examination of 30,865 cells. They found that MMP7 was up-regulated in club cells, CCL18 was up-regulated in macrophages while

MUC1 was up-regulated in type I pneumocytes. The number of AT2 cells decrease dramatically in IPF lower and upper lobes compared to control. In contrast, fibroblasts increased from 2.08% of cells in normal to 4.50% in IPF upper lobes and 13.54% in IPF lower lobes. They concluded that AT2 cells, macrophage and fibroblast all play important roles in pathogenesis [123].

2) AIM OF THE RESEARCH

The UIP pattern is characterized by spatial temporal heterogeneity of histological lesions (see paragraph 1.6; page 46), with extensive areas represented by end stage fibrosis with honeycombing lesions of alveolar space.

The sandwich of fibroblastic focus/metaplastic epithelial cells is considered a manifestation of active lung injury and several studies have demonstrated a strict correlation with physiologic decline and mortality [93,124-127].

RESEARCH QUESTIONS

1. Will it be possible to select the fibroblastic foci (FF) sandwich area from formalin-fixed and paraffin-embedded (FFPE) tissue of UIP lung and to extract adequate RNA from high-throughput sequencing technologies to survey, characterize and quantify the transcriptome of a genome?
2. Will it be possible to detect the crucial transcripts (up or down regulated) and consequently the key actors in the most representative injured lesion of IPF?
3. In order to identify some potential biomarkers, will we able to validate some transcripts in lung tissue of larger case series?

Thus the aims and objectives of the present research were the following:

AIMS

1. Evaluation the possibility to extract RNA from FF sandwich selected from archival IPF tissues and control group (FF of pneumothorax);
2. Assessment of the up and down regulated RNA transcripts in FF sandwich;

3. Analysis and validation of the most significant up or down regulated transcripts.

RESEARCH OBJECTIVES

1. Optimization of a protocol for microdissection of FF sandwiches from FFPE tissues;
2. Identification of crucial transcripts and key cell actors in the active injured lesion of IPF;
3. Validation of biomarkers to better understand the pathogenesis and for preliminary novel therapeutic options.

3) MATERIALS AND METHODS

3.1) Study population

The University Hospital of Padua is one of the most important Italian centers for lung transplantation, in particular for the end-stage IPF patients. Indeed, 171 patients were transplanted for IPF/UIP (from 1995 to 2017), and 40 of them (from 2000 to 2008) had a diagnostic surgical biopsy (VATS, Video Assisted Thoracic Surgery).

In order to perform molecular analyses in the areas of fibroblastic foci + metaplastic epithelial cells (FF sandwich), 10 patients with a diagnosis of IPF [IPF CASES] according to the ATS/ERS/JRS/ALAT Clinical Practice Guideline 2011[128] were included (at that time the new 2018 guideline was not available).

Two patients who had primary spontaneous pneumothorax (PSP) in absence of other concomitant lung diseases [CONTROLS] were also investigated (Table 1).

SAMPLE NUMBER	PATIENT	TIME OF DIAGNOSIS (years)	SEX	AGE AT TRANSPLANTATION (years)	STUDY GROUP	PHARMACOLOGICAL TREATMENT	DISEASE PROGRESSION	SAMPLE TYPE
1	B.R.	2015	M	59	IPF case	pirfenidone	not known	explanted lung
2	P.S.	2015	M	55	IPF case	pirfenidone	slow	explanted lung
3	T.C.	2016	M	60	IPF case	nintedanib	rapid	explanted lung
4	B.R.	2015	M	55	IPF case	steroid	rapid	explanted lung
5	P.A.	2014	M	36	IPF case	steroid	slow	explanted lung
6	V.G.	2006	M	58	IPF case	/	not known	surgical lung biopsy
7	P.S.	2014	M	62	IPF case	steroid	slow	explanted lung
8	G.G.	2011	M	68	IPF case	steroid	rapid	explanted lung
9	F.N.	2017	M	56	IPF case	mycophenolate	slow	explanted lung
10	S.M.	2017	M	19	control	/	/	surgical lung biopsy
11	C.R.	2017	M	18	control	/	/	surgical lung biopsy
12	B.G.	2003	M	66	IPF case	/	not known	surgical lung biopsy

Table 1: *characteristics of patients selected for the molecular analysis. Sex: M=male, F=female; IPF= Idiopathic Pulmonary Fibrosis.*

The choice of these control patients is due to the presence of histological features similar to FF sandwiches detected in the UIP pattern (figure 19).

A PSP is traditionally defined as a pneumothorax which presents itself without an external event in absence of another lung disease. PSP is usually related to the rupture of small apical subpleural blebs or bulla into the pleural cavity. FF sandwiches are nonspecific and non-pathological features that can often be detected also in surgical specimens from patients with PSP, indicative of a repair mechanism. They are found within a myxoid stroma at the pleural-parenchymal interface or leading edge. These lesions often exhibit a wedge-shaped configuration, with the broad base at the pleural surface and the apex towards the lung parenchyma.

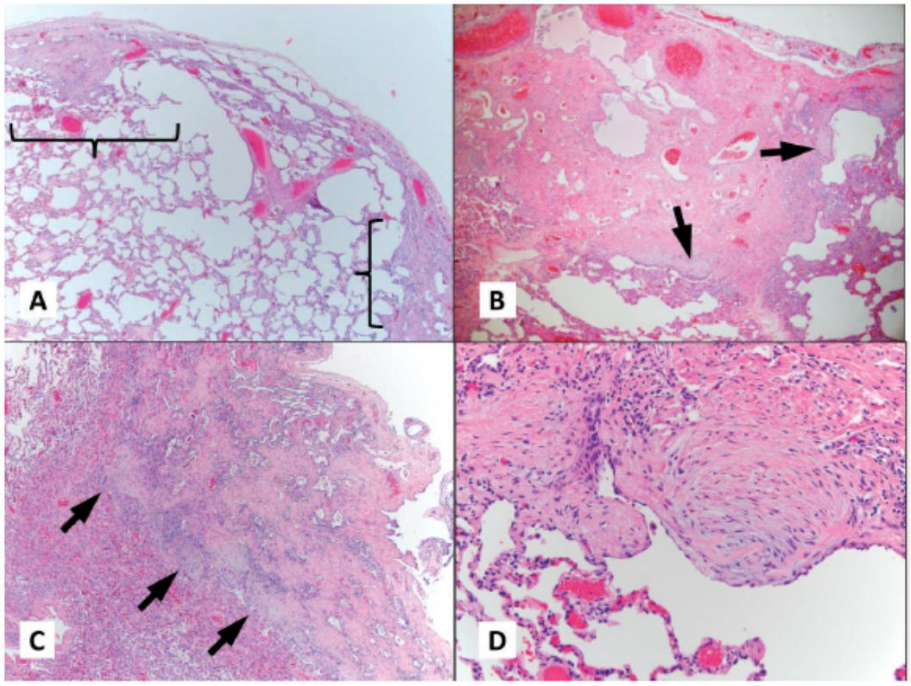

***Figure 19**: histopathologic features of pneumothorax with fibroblastic foci. A: wedge-shaped area of zonation demonstrated by peripheral collagenous fibrosis and intervening normal pleura. B and C: higher-power view illustrating the zonation and fibroblastic foci (arrowheads) at the leading edge. D: fibroblastic focus. From Belchis DA, et al., Arch Pathol Lab Med. 2012;136(12):1522–1527.*

These FF sandwiches were selected in FFPE sections obtained from both IPF/UIP cases and controls. Patients were selected only on the basis of the number of fibroblastic foci,

giving priority to those who were treated with steroids or mycophenolate and not with anti-fibrotic drugs.

3.2) Microdissection

In order to identify the best strategy for the isolation of FF sandwich from FFPE tissue, two microdissection methods were compared:

a) manual microdissection

b) laser microdissection

3.2.1) Manual microdissection

For the manual microdissection, a Leica M60 stereomicroscope (Leica Microsystems; Wetzlar, Germany), a 20 sterile syringe needle and a lysis solution tube were used to collect FF sandwiches.

3.2.1.1) Preparation of the manual microdissected slides

All the steps were carried out using RNase-free glass coplin jars that were prepared for staining by rinsing with 100% Ethanol (EtOH), followed by distilled water, RNaseZap™ RNase Decontamination Solution (ThermoFisher Scientific), followed by RNase-free water. The jars were allowed to dry before sealing with a glass lid and tape. Staining solutions were prepared: 50-ml falcon tubes were filled with RNase-free 50% EtOH, 70% EtOH, 95% EtOH, 100% EtOH, RNase-free water and cooled down to −20 °C (alcoholic solutions) or 4 °C (water). Coplin jars were also filled with hematoxylin

and eosin. RNase contamination was avoided by cleaning surfaces and tools with RNaseZap™. Blocks were cooling down to –20°C and cut with a microtome obtaining sections of about 4 μm thick (about 8 sections per case). Tissue slides were than incubated for 30 minutes in heat treatment drying oven.

Paraffin was removed by washing the slides with xylene 1 minute for three times (in three separate glass coplin jars) followed by a series of descending concentrations of ethanol as follows:

• 100% ethanol twice for 30 seconds

• 95% ethanol twice for 30 seconds

• 70% ethanol twice for 30 seconds

• 50% ethanol twice for 30 seconds

• distilled water 30 seconds

The sections were then stained with Hematoxylin and Eosin (H&E) following routine procedures. At the end, slides were air-dry at room temperature and a coverslip and mounting medium were not applied.

3.2.2) Laser microdissection (LMD)

To obtain high-quality DNA, mRNA, and proteins from (often small) tissue samples and even from single cells, laser microdissection (LMD) is one of the most useful techniques. Microdissected tissue material or single cells, free of contaminating, are extremely important for producing clean data. In most instances, RNA can be extracted from fresh frozen material in high quality. Unfortunately sometimes the morphology of this material is inadequate. Sections from FFPE tissue typically have well-preserved morphology at the expense of poor RNA integrity, due to the cross linking of proteins

and nucleic acids caused by formalin fixation. RNA also becoming increasingly fragmented with prolonged storage[129].

LMD is a technology for precision sample preparation that makes it possible to obtain homogenous, ultrapure samples from heterogenous starting material. A researcher can selectively and routinely analyze regions of interest down to single cells and chromosomes to obtain results that are reproducible, and specific. LMD experiments were performed using a Leica LMD6500 laser microscope (Leica Microsystems; Wetzlar, Germany) cutting into a 0.5-ml tube cap (Eppendorf, Hamburg, Germany). This Leica LMD system performs sample preparation for molecular biology analysis directly from the tissue section using a UV laser. The system is based on an inverted light microscope, fitted with a laser device to facilitate the visualization and procurement of cells. This platform consists of an inverted microscope, a solid state near infrared laser diode, a laser control unit, a joy stick controlled microscope stage with a vacuum chuck for slide immobilization, a CCD camera, and a monitor. The LMD microscope is connected to a computer.

LMD applications require specifically designed slides, to free the dissectate from the section. Tissue slices are placed on a membrane slides for LMD that are used as regular glass slides (glass slides covered with membrane on one side). This membrane is easily cut together with the tissue and acts as a stabilizing backbone during lifting: in this system most of the energy is absorbed by the membrane, the maximum temperatures reached by the tissue upon laser activation are in the range of 90°C for several milliseconds; thus leaving biological macromolecules of interest intact. This platforms use tissue that has been mounted on a 6 μm membrane and placed on a glass slide, onto which the operator directs an UV laser beam under direct visualization the desired cells were subsequently catapulted against gravity into an overhanging cap (figure 20).

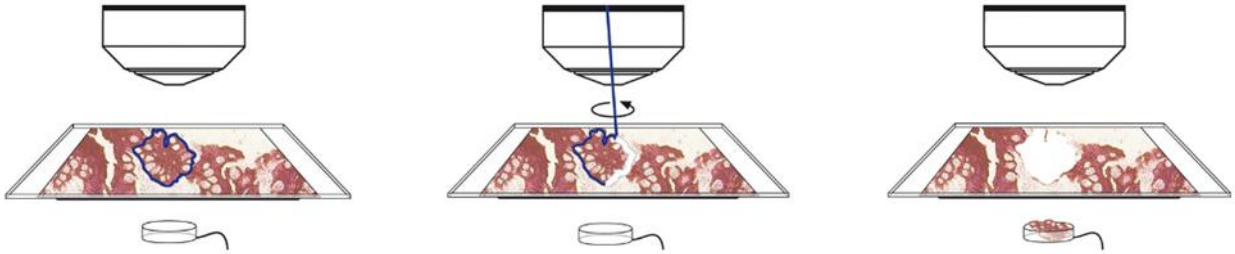

Figure 20: *the Leica LMD process (from the left). Step 1: Define the region of interest; Step 2: Laser beam precisely steered by prisms along your definition; Step 3: Dissectate is collected by gravity. From Leica LMD6500-7000-Brochure (Leica Microsystems).*

This specimen collection via gravity is simple, gentle, contact- and contamination-free, allows standard consumables for collection, no limitation of size or shape of dissectate and a pool unlimited amounts of dissectates.

3.2.2.1) Preparation of the LMD slides

All the steps were carried out using RNase-free glass coplin jars that were prepared for staining by rinsing with 100% Ethanol (EtOH), followed by distilled water, RNaseZap™ RNase Decontamination Solution (ThermoFisher Scientific), followed by RNase-free water. The jars were allowed to dry before sealing with a glass lid and tape. Staining solutions were prepared: 50-ml falcon tubes were filled with RNase-free 50% EtOH, 70% EtOH, 95% EtOH, 100% EtOH, RNase-free water and cooled down to −20 °C (alcoholic solutions) or 4 °C (water). Coplin jars were also filled with hematoxylin and eosin. RNase contamination was avoided by cleaning surfaces and tools with RNaseZap™.

Before starting with the experiment, PEN membrane glass slides were sterilized by UV-treatment: slides were incubated in a UV chamber at maximum power for 30 – 45 minutes.

Blocks were cooling down to –20°C and then the tissues were cut into 4-8 μm thick sections with a microtome processing a minimum of 8 to a maximum of 15 sections per patient. Paraffin ribbons were place on the surface of RNase-free water in a waterbath at 37°C; in this way, the floating ribbons are easier to stretch and subsequently can be placed without folds on the special membrane slides (PEN-Membrane Slides 2,0 μm; Leica Microsystems; Wetzlar, Germany). Tissue slides were then incubated for 30 minutes in heat treatment drying oven.

Paraffin must be removed prior to staining the paraffin-embedded sections. This is achieved by washing slides with xylene three times for 1 minute (in three separate glass coplin jars) followed by a series of descending concentrations of ethanol as follows:

- 100% ethanol twice for 30 seconds
- 95% ethanol twice for 30 seconds
- 70% ethanol twice for 30 seconds
- 50% ethanol twice for 30 seconds
- distilled water 30 seconds

The sections were then stained with Hematoxylin and Eosin (H&E) following routine procedures. At the end, slides were air-dry at room temperature. To allow LMD (cutting and lifting), a coverslip and mounting medium must not be applied.

3.3) RNA extraction

3.3.1) RNA extraction from the manual microdissected samples

RNA isolation was performed immediately after microdissection (during the same day). RNA extraction from the manual microdissected tissue was obtained by using RNeasy® FFPE kit (Qiagen, Hilden, Germany) according to the manufacture's protocol. Briefly, 150 µl of buffer PKD was added to the samples and then incubated at 56°C for 15 min with 10 µl proteinase K, then at 80°C for 15 min. Subsequently, after the addition of 320 µl Buffer RBC and 720 µl ethanol (100%) RNA was purified using RNeasy MinElute spin column. RNA was eluted in RNase-free water. DNase I was used to eliminate contaminating DNA. All the samples were then stored at -80 ° C.

3.3.2) RNA extraction from the laser microdissected samples

RNA isolation was performed immediately after microdissection (during the same day) using a modification of the original protocol. RNA extraction from the LMD tissue was obtained using RNeasy® FFPE kit (Qiagen, Hilden, Germany) according to the manufacture's protocol with the following two variations.

1) 150 µl of buffer PKD was added to the samples, then incubated overnight at 43°C with 10 µl proteinase K (*original protocol: 15 min at 56*°), followed by incubation at 80°C for 15 minutes. Subsequently, after the addition of 320 µl Buffer RBC and 720 µl ethanol (100%) RNA was purified using RNeasy MinElute spin column. RNA was eluted in RNase-free water. DNase I was used to eliminate DNA.

2) A volume of 0.75 µl SUPERase•IN™ RNase inhibitor was added to the extracted RNA (15 µl), then stored at -80 ° C (*original protocol: this step was not foreseen*). 1 µl was then aliquoted for Agilent testing.

3.4) RNA purification

To increase the amount of final RNA some samples, in particular 2 controls (sample 11, CR + CR), 2 VATS (sample 6, VG + VG) and 2 explanted lungs (sample 9, FN + FN), were microdissected several times and subsequently extracted individually. This RNA was then unified and processed with the RNeasy minElute Cleanup (Qiagen, Hilden, Germany) according to the manufacturer's protocol, that enabled cleanup and concentration of RNA (from 45 µg down to picogram amounts, less than 1 cell) from enzymatic reactions or other samples using specialized RNeasy MinElute spin columns based on silica-membrane technology. With the RNeasy MinElute procedure, all RNA fragments with a length of at last 200 nucleotides were isolated. The procedure provided enrichment for mRNA since most RNAs <200 nucleotides were selectively excluded. Briefly, 350 µl of buffer RLT and 250 µl of EtOH were added to the eluted RNA. RNeasy MinElute Spin Column was used for elution in RNase-free water after several washings with buffer RPE and 80% EtOH.

3.5) Quantity and quality controls

Nucleic acid isolation was followed by a quantity and quality (purity, integrity) evaluation before down-stream preparation steps. RNA quantification was important

for the standardization of the PCR input material and is crucial in NGS RNA library preparation.

3.5.1) Spectrophotometric method

Nucleic acids quantification was performed by using spectrophotometers that allowed ultraviolet light absorption measurement at the wavelength of 260 nm and applying the Lambert-Beer law that correlates absorbance, molar extinction co-efficient and nucleic acids concentration. Absorbance measurements were carried out on a Nanovue spectrophotometer (GE Healthcare Life Sciences, UK) by directly pipetting onto the pedestals 2 μl of RNA sample dissolved in RNase-free water, after an initial blank measurement. Nanovue automatically calculated the RNA concentration by applying specific extinction coefficients (40 for RNA).

The ratios $A_{260/280}$ and $A_{260/230}$ were used as indicators of sample purity. The nucleic acid sample ratio $A_{260/280}$ is generally used as indicator of protein contamination. Indeed the 280 nm is the absorbance wavelength of aromatic amino acid side chains and phenol groups. Pure RNA should present a $A_{260/280}$ ratio between 2 and 2.2. The $A_{260/230}$ ratio is generally used as indication of organic contaminants. 230 nm is the absorbance wavelength of many organic compounds (i.e. phenol, TRIzol, and chaotropic salts present in the most common lysis buffers). In pure DNA/RNA samples the ratio $A_{260/230}$ should be between 2 and 2.2.

3.5.2) Qubit Fluorometer

RNAs quantities were further assessed by using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Walthman, MA, USA). The Qubit 2.0 is a benchtop fluorometer that uses fluorescent dyes that specifically bind to RNA therefore being able to selective quantify it. According to the manufacturer's instructions, analyses were conducted with the RNA BR Assay kit (Thermo Fisher Scientific, Walthman, MA, USA) which presents a sensitivity range from 1 ng/µl to 1 µg/µl.

Briefly, 1 µl of the sample was mixed with 199 µl of "working solution" containing specific fluorescent dyes, composed by 199 µl RNA buffer and 1 µl of RNA reagent. After an incubation of 2 min the measurement was performed on the Qubit 2.0 by setting the RNA protocol.

3.5.3) Agilent Bioanalyzer

The Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, Carlsbad, USA), a microfluidics-based platform that can perform quantification and quality control of nucleic acid, leads to an electrophoretic separation of samples on micro channels containing fluorescent dyes. One µl of RNA samples was sufficient for analysis on 2100 Bioanalyzer with the RNA 6000 Pico Kit (Agilent Technologies, Santa Clara, Carlsbad, USA) and RNA 6000 Nano kit (Agilent Technologies, Santa Clara, Carlsbad, USA), according to the manufacture's guideline. Briefly, RNA chips were prepared by adding 9 µl gel-dye mix to the assigned well under pressure, and 9 µl of gel-dye mix to the next 2 walls. 9 µl of the RNA 6000 Pico conditioning solution was pipetted into the corresponding well. After adding 5 µl of marker to each well, 1 µl of

RNA ladder and 1 µl of sample were added to the separate wells. Then, chips were vortexed for 1 minute at 2400 rpm on an IKA vortex mixer and were run on the Agilent Bioanalyzer 2100. The run data were analyzed by Agilent 2100 expert software version B.02.08.SI648 (SR2). The Agilent RNA 6000 Pico Kit and 6000 Nano kit evaluate the ratio between the 18S and the 28S ribosomal subunits, and the presence of degraded small RNA fragments, in order to calculate the RNA Integrity Number (RIN) that expresses an estimation of the integrity of the RNA samples. RIN values range from 10 (intact) to 1 (totally degraded).

RNA 6000 Nano kit is able to assess the RNA of a concentration between 25 to 500 ng/µl, while RNA 6000 Pico kit between 50 to 5000 pg/µl.

The process of fixing the tissue sample and embedding it in paraffin can cause severe degradation of the RNA. The formalin fixation process causes crosslinkage between nucleic acids and proteins, and the covalent modification of RNA. This fixation process and storage can cause significant RNA degradation in some samples. Many researchers used RIN to determine RNA quality for gene expression analysis; however Illumina (Illumina, San Diego, CA, USA) have found that RIN values from degraded FFPE samples were not a sensitive measure of RNA quality nor were they a reliable predictor of successful library preparation. Instead, they had found that mean RNA fragment size is a more reliable determinant of RNA quality. They developed the DV_{200} metric—the percentage of RNA fragments > 200 nucleotides. Using DV_{200} to accurately assess FFPE RNA quality, they had fine-tuned the minimal RNA input required for successful library preparation. By adjusting RNA input amounts, high-quality libraries can be prepared from poor-quality FFPE samples (table 2).

Table 2: Recommended RNA Input Based on DV_{200}

Parameter Quality*†‡	DV_{200}	Recommended Input Quantity	
		TruSeq RNA Access	TruSight RNA Pan-Cancer or TruSight RNA Fusion
High	> 70%	20 ng	20 ng
Medium	50–70%	20-40 ng	20-50 ng
Low	30–50%	40–100 ng	50–100 ng
Too Degraded	< 30%	Not Recommended	

* to ensure successful library prep, we recommend using an RNA isolation method that includes a reverse-crosslinking step and DNase1 treatment (we used QIAGEN RNeasy FFPE Kit or QIAGEN AllPrep DNA/RNA FFPE Kit).
† FFPE RNA concentration was determined by NanoDrop.
‡ To ensure best performance on samples close to edge of a quality classification, err towards the higher end of the input recommendation.

Table 2: *recommended RNA input based on DV_{200}. From "Evaluating RNA quality from FFPE samples", Illumina*

3.6) Polymerase chain reaction (PCR)

PCR, a method developed by Mullis in the 80s, allows the exponential amplification of specific targeted DNA regions. A PCR reaction requires template DNA, deoxynucleotide triphosphates (dNTPs), oligonucleotide primers flanking the target DNA sequence, DNA polymerase enzyme, reaction buffer and magnesium.

3.6.1) Reverse transcription

RNAs isolated from FF+MECs with the manual microdissection technique were reverse transcribed. Briefly 500 ng of total RNA was mixed with 1 µl of 50 µM Random Primers hexamers (Invitrogen, ThermoFischer Scientific, Walthman, MA, USA), 0.4 µl of 10 mM dNTPs mix (Invitrogen, ThermoFischer Scientific, Walthman, MA, USA) and water to a total volume of 10 µl, then incubated on a thermal cycler at 70°C for 8 minutes.

2 µl of 10X PCR buffer II, 4 µl of $MgCl_2$ solution at 25 mM (Applied Biosystems, ThermoFischer Scientific, Walthman, MA, USA), 1 µl of 40 U/µl RNase OUT, 1 µl of M-MLV reverse transcriptase 40,000 U (200 U/µl) (Invitrogen, ThermoFischer Scientific, Walthman, MA, USA) and 2 µl of water were added to the mix, mixed, heated at 38°C for 50 minutes and inactivated at 97°C for 3'.

3.6.2) Standard PCR

Amplification reactions were carried out in a final volume of 50 µl containing 10 µl of cDNA, 4 µl of 10X PCR Buffer II, 2 µl of 25mM $MgCl_2$ solution, 0.4 µl of 10 mM dNTPs mix, 1 µl of 10 µM forward and reverse primer, 0.25 µl of AmpliTaq Gold DNA polymerase 250U (Applied Biosystems, ThermoFischer Scientific, Walthman, MA, USA) and 31.35 µl of water. Amplification reactions were performed on 2720 thermal cycler (Applied Biosystems, ThermoFisher Scientific, Walthman, MA, USA) at specific annealing temperature, optimized based on the GC content of the sequence and melting temperatures (Tm) of primers. Primers specific for glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene were used to verify adequate nucleic acid extraction (see table 3).

Primer	**Sequence (5'⟶ 3')**	**Annealing temperature**
GAPDH Fw	AGGTGAAGGTCGGAGTCAACG	**50°C**
GAPDH Rv	GCTCCTGGAAGATGGTGATGG	

Table 3: *sequences and annealing temperature of GAPDH primers.*

Briefly, each sample was denatured at 95°C for 10’ to allow the activation of the hot start DNA polymerase, and then cDNAs were amplified by 40 three-step cycles (denaturation for 30’’ at 95°C, annealing for 30’’ at 50°C and extension at 72°C for 1’) followed by a final extension step of 7 min at 72°C to enhance the amplicon elongation. After the initial denaturation at 95°C for 10 min, the (30 sec at 95°C, 30 sec at annealing temperature, 1 min at 72°C).

All samples were processed with negative controls (reaction mixture without cDNA templates). Precautions were taken to avoid false positives as a result of contamination by PCR product carry over, by strictly following the guidelines for the general handling of the PCR procedure, such as separation of rooms, boards, and lab benches (i.e. extraction of nucleic acids, PCR amplification and gene sequencing performed in different rooms with separate equipment and pipettes). PCR products (234 bp) were then analyzed on a 2% agarose gel.

3.6.3) Agarose gel

For standard 3.5% agarose gel electrophoresis, 3.5 g of ultrapure Agarose gel (Thermo Fisher Scientific, Walthman, MA, USA) were added to 100 ml of 1X Tris-Acetate-EDTA (TAE) buffer (Tris-acetate 0.04 M and EDTA 0.001 M). The solution was heated in a microwave to dissolve agarose, gel was added with 5 µl of ethidium bromide (Sigma-Aldrich, Saint Louis, MO, USA), poured into a sealed tray and then a proper comb was inserted. Aliquots of 15 µl of PCR products and 3 µl of bromophenol blue loading dye were mixed and loaded into each gel well. In addition, 1.5 µl of DNA Marker VIII (Roche Applied Science, Manheim, Germany) were loaded in order to correctly recognize the fragment sizes. Electrophoretic run was performed at 100 V in

1X TAE buffer. The gels were visualized by UV transillumination and photographed with an Alliance 2.7 (UVITEC, Cambridge, UK).

3.7) Next Generation Sequencing and RNA-Sequencing

Microarray gene expression studies pioneered the use of genome wide techniques in the hunt of sets of gene implicated in complex phenotypes. Microarray technologies were limited by their dependence on the use of known probes, requiring specific chips. The advent of Next Generation Sequencing (NGS), also known as "deep sequencing" and "high-throughput sequencing" revolutionized gene expression analysis by obviating the need for pre-existing transcripts probes. NGS employ new sequencing able to produce millions of small DNA sequence (reads) in a single run and had the potential to focus on the analysis of entire genomes (Whole Genome Sequencing, WGS), on the sole coding part of the genome (Whole Exome Sequencing, WES) or on specific target genes (Target Resequencing). NGS technologies may also be employed for the detection of DNA methylation sites (Methylation Sequencing, Methyl-Seq), for DNA-protein interaction studies (Chromatin Immunoprecipitation Sequencing, ChIp-Seq), for transcription factor profiling and gene expression quantification study (RNA-Seq), ribosomal Sequencing (Ribo-Seq), small RNA profiling including microRNAs and promoter associated RNAs (sRNA-Seq), thus enabling epigenetic and transcriptomics analysis[130].

RNA sequencing (RNA-Seq) uses high-throughput reads produced by NGS to represent the entire transcriptome. RNA-Seq is used for a variety of application, most commonly to discover lists of genes that are differentially expressed between experimental groups;

i.e. samples from different tissues, samples from different treatment groups or samples from different populations. To identify gene networks associated with hereditary diseases or other genetic traits, individuals can be grouped into disease status (affected vs non-affected) or by different hyplotype at the mapped loci[131].

Although microarrays have revolutionized the study of transcriptomics and proved useful in determining gene expression profiles, RNA-seq by comparison is more sensitive, provides absolute quantity levels, is not affected by on-chip sequence biases, and gives additional information on gene expression levels and splice junction variants.

In RNA-Seq, RNA is commonly first converted to a more stable cDNA through a combination of reverse transcription and the selection process to isolate the RNA from the abundant rRNA. The input RNA quality is very important in RNA-Seq preparation because RNAse enzymes are ubiquitous and extremely stable and fragmentation can also occur simply when a divalent cation is present. Library preparation and sequencing of cDNA follow the same sequencing procedure as DNA-Seq. However, numerous variations of RNA-Seq library preparations have been developed, each with its benefits and limitations in terms of relative costs and input requirements. The main differences in these various library preparations are the methods of purifying and isolating RNA of interest (i.e. mRNA). RNA-Seq libraries can be made using polyadenylated tail selection, not-so-random primers (for reverse transcription), and ribosomal depletion[132,133].

RNA-Seq is composed by three main steps: library preparation, sequencing and bioinformatics data analysis of results including differential gene expression analysis.

3.7.1) Library preparation

The SMARTer Stranded Total RNA-Seq Kit - Pico Input Mammalian (Takara Bio USA, Inc., Mountain View, USA) includes all components needed to generate indexed cDNA libraries suitable for NGS on any Illumina platform, with recommended input ranging from 250 pg to 10 ng of total mammalian RNA. This kit incorporates SMART® (Switching Mechanism At 5' end of RNA Template) technology[134] and locked nucleic acid (LNA) technology, included as part of the template-switching oligo (TSO). The inclusion of LNA technology into the TSO (LNA-TSO) stabilizes the interaction between the TSO and non-templated nucleotides added by the reverse transcriptase. PCR amplification generates Illumina-compatible libraries without the need for adapter ligation. The directionality of the template-switching reaction preserves the strand orientation of the original RNA, making it possible to obtain strand-specific sequencing data from the synthesized cDNA.

The SMARTer Stranded Total RNA-Seq Kit - Pico Input Mammalian is compatible with picogram-inputs of total RNA from high-quality or partially degraded samples (250 pg–10 ng) and FFPE samples (5–10 ng). In order to generate library inserts of an appropriate size for subsequent Illumina sequencing, high-quality or partially degraded RNA samples are processed to the suitable fragment size prior to cDNA synthesis. For highly degraded, low-quality starting material, the RNA fragmentation step is skipped. Ribosomal RNA (rRNA) comprises a significant proportion (~90%) of total RNA samples. Depleting these abundant transcripts from total RNA samples prior to generating libraries provides benefits by lowering sequencing costs and improving mapping statistics. However, with very low input amounts, initial rRNA depletion from total RNA is not very effective and often leaves an insufficient amount of material for preparation of a good library. The workflow used in this kit takes advantage of a

technology allowing removal of ribosomal cDNA (cDNA fragments originating from rRNA molecules) after cDNA synthesis using probes specific to mammalian rRNA. These ribosomal cDNA are then cleaved by ZapR in the presence of the mammalian-specific R-Probes that hybridize to ribosomal RNA. The protocol for the library preparation is composed by different steps (figure 21).

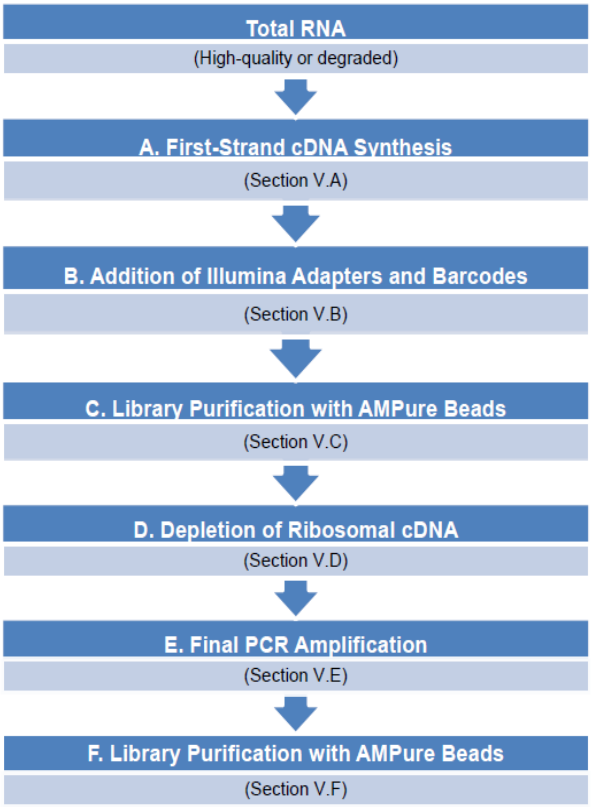

Figure 21: *SMARTer Stranded Total RNA-Seq Kit - Pico Input Mammalian protocol overview. This kit features an easy workflow that generates Illumina-compatible RNA-seq libraries. First, total RNA is converted to cDNA (Step A), and then adapters for Illumina sequencing (with specific barcodes) are added through PCR using only a limited number of cycles (Step B). The PCR products are purified (Step C) and then ribosomal cDNA is depleted (Step D). The cDNA fragments from Step D are further amplified (Step E) with primers universal to all libraries. Lastly, the PCR products are purified once more to yield the final cDNA library (Step F). From SMARTer® Stranded Total RNA-Seq Kit - Pico Input Mammalian User Manual Guide (Takara Bio).*

Fragmentation of the RNA. When the integrity of RNAs are not optimal (RIN ≤3), fragmentation of RNA is avoid. An adequate RNA quantity is used for libraries preparation, on the basis of table 2 indication. Subsequently, RNA for each sample is

diluted with nuclease-free ultrapure water to a final volume of 9 µl and SMART Pico Oligos Mix is adding.

First-Strand cDNA Synthesis. Random priming allowed the generation of cDNA from all RNA fragments in the sample, including rRNA. When the SMARTScribe™ Reverse Transcriptase (RT) reaches the 5' end of the RNA fragment, the enzyme's terminal transferase activity adds a few non-templated nucleotides to the 3' end of the cDNA. The carefully designed LNA-TSO (included in the Template Switching Oligo Mix) base-pairs with the non-templated nucleotide stretch, creating an extended template to enable the RT to continue replicating to the end of the oligonucleotide. The resulting cDNA contains sequences derived from the random primer and the LNA-TSO used in the reverse transcription reaction.

PCR1—Addition of Illumina Adapters and Indexes. The indexes (barcodes) that are used to distinguish pooled libraries from each other after sequencing are added at this step. PCR1 adds full-length Illumina adapters, including barcodes. The Forward PCR Primer binds to the LNA-TSO sequence, while the Reverse PCR Primer binds to sequence associated with the random primer.

Purification of the RNA-Seq Library Using AMPure Beads. The amplified RNA-seq library is purified by immobilization onto AMPure beads. The beads are then washed with 80% ethanol and the cDNA is eluted in Nuclease-Free Water.

Depletion of Ribosomal cDNA with ZapR and R-Probes. In this section, the library fragments originating from rRNA (18S and 28S) and mitochondrial rRNA (m12S and m16S) are cut by ZapR in the presence of R-Probes (mammalian-specific). These R-Probes hybridize to ribosomal RNA and mitochondrial rRNA sequences; however, the mitochondrial sequences are derived from the human mitochondrial genome and are therefore strictly human specific. This process leaves the library fragments originating

from non-rRNA molecules untouched, with priming sites available on both 5' and 3' ends for further PCR amplification.

PCR2—Final RNA-Seq Library Amplification. The library fragments not cleaved by the ZapR reaction will be further enriched in a second round of PCR using primers universal to all libraries. Since barcodes have already been added to the libraries, a single pair of primers can be used for all libraries.

Purification of Final RNA-Seq Library Using AMPure Beads. The amplified RNA-seq library is purified by immobilization onto AMPure beads. The beads are then washed with 80% ethanol and eluted in Stranded Elution Buffer.

Library validation. Quality control analysis of the sample library includes quantification of cDNA library templates and quality control:

1. **Quantification**: an accurate quantification of DNA library is essential to create optimal cluster densities across every lane of the flow cell and obtain high quality sequencing data. Libraries were quantify with Qubit dsDNA HS kit (Thermo Fisher Scientific, Walthman, MA, USA). A yield >3 ng/μl will provide enough material for further library validation and sequencing.

2. **Quality Control**: to evaluate library size distribution of the enriched fragments, 1 μl of library aliquot was running by running samples on the Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit (Agilent, Santa Clara, Carlsbad, USA). Libraries were dilute to about 1.5 ng/μl prior to loading the chip (for a consistent library-to-library profile).

Successful cDNA synthesis and amplification should produce a distinct curve spanning 200–1,000 bp, with a local maximum at ~300–400 bp, in the negative control no product or very minimal background over the corresponding range is necessary.

Normalize and Pool Libraries. Indexed DNA libraries are normalized to10 nM and then pooled together in equal volumes.

3.7.2) Sequencing

Cluster generation was achieved according to the Illumina Cluster Generation User Guide. In this procedure, templates were attached to the surface of an oligonucleotide coated flow cell and amplified to produce a cluster bound to the surface of the flow cell. Cluster generation workflow included cluster amplification, linearization, blocking, and primer hybridization. During cluster amplification the sample was hybridized on the flow cell and amplified, then the amplified sample was prepared for sequencing: one of the two adapters were cleaved off from the surface of the flow cell, 3' OH ends of the linearized dsDNA clusters was blocked, dsDNA was denatured and sequencing primers can hybridize. Briefly the template DNA was denatured with 0.1 N NaOH to a DNA concentration of 20 pM and added with Hybridization Buffer.

Amplification of the clusters is called bridge amplification: DNA polymerase is used to produce clusters of approximately one million copies of the original fragment. During sequencing the four labeled nucleotides were simultaneously added to the flow cell channels with the DNA polymerase, and were incorporated into the cluster fragments. The four nucleotides were labelled with base-specific fluorescents, the label contained a 3'-OH group that inhibits the fluorescence, the polymerase ligated the fluorescent labelled nucleotides in the clusters and, during this process, the 3'-OH group detached, and the fluorescence was detected. Illumina sequencer produces reads around 100 bases. In our case, the DNA library was sequenced on a flow cell with a paired-end sequencing 2x150 bp with a data yield of 1300-1500 Gb on a HiSeq 4000 System sequencer

(Illumina, San Diego, CA, USA) by sequencing by synthesis (SBS) technology using a fluorescently labelled method. The general workflow included single molecule amplification by bridge PCR and reversible terminator sequencing-by-synthesis. This kind of platforms produced reads of 50-150 bp in length, they employed flow cells with eight lanes, and multiple samples may be run on a single lane. To differentiate reads from different samples after sequencing on the same lane, a unique barcode may be attached to each sample during library preparation.

3.7.3) RNA-Seq bioinformatic workflow

Transcriptome profiling with RNA-Seq approach with the following goals:

1. Precise quantification of gene expression levels of individual samples of Homo sapiens;
2. Capture and compare all changes in gene expression levels among multiple samples;
3. Functional annotation of differentially expressed genes.

A typical bioinformatic workflow uses a reference genome and aim to identify the differentially expressed genes (DEGs) begins with raw reads, which are aligned to a reference genome (the human genome). Gene counts are then quantified from the alignment files and used in differential gene expression analysis. The second step involves the analysis of the read counts to each annotated locus in order to identify differential expression among samples. The third step includes the attachment of biological information to the differentially expressed genes, allowing a greater understanding about the relationship between an organism's genome and its phenotype.

Prior to further analysis, a quality check was performed on the raw sequencing data, removing low quality portions while preserving the longest high quality part of NGS reads. This trimming step was performed with the following parameters: the minimum

length was set to 35 bp and the quality score to 25. The software BBDuk was used for this scope.

High quality reads were then aligned against the Homo sapiens reference genome sequence (GRCh38) [ENSEMBL] with STAR aligner (version 2.5.0c).

3.7.4) Differential gene expression analysis (DEGs analysis)

FeatureCounts (version 1.6.0) was used to calculate gene expression values as raw fragment counts (annotation GRCh38). Normalization was applied to the raw fragment counts by using the Trimmed Mean of M-values (TMM) normalization and Fragments Per Kilobase Million (FPKM) normalization. All the statistical analyses were performed with RStudio. The following step has been the removal of the not expressed genes and the ones showing too much variability for each comparison. The HTSFilter package was chosen for this scope which implements a filtering procedure for replicated transcriptome sequencing data based on a Jaccard similarity index.

The comparison 10 IPF cases versus 2 spontaneous pneumothorax (controls) were then performed: we have decided to use NOISeq package with the function “ARSyNseq” without specifying a specific metadata as batch effect in order to removal systematic noise from the data.

DEGs filtering was conducted following criteria: Log2(Fold Change, FC1)≥1; Log2(FC2)≤-1 and p-value adjust for False Discovery Rate (Benjamini-Hochberg correction) ≤0.05.

3.7.5) Gene Ontology enrichment analysis (GOEA)

The mission of the Gene Ontology (GO) Consortium is to develop a comprehensive, computational model of biological systems, ranging from the molecular to the organism level with a structured, precisely defined, common, controlled vocabulary for describing the roles of genes and gene products in any organism.

This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research. To this end, three independent ontologies accessible on the World-Wide Web (http://www.geneontology.org) are being constructed: biological process, molecular function and cellular component.

Biological process refers to a biological objective to which the gene or gene product contributes. Molecular function is defined as the biochemical activity (including specific binding to ligands or structures) of a gene product. This definition also applies to the capability that a gene product (or gene product complex) carries as a potential.

Cellular component refers to the place in the cell where a gene product is active. These terms reflect our understanding of eukaryotic cell structure[135].

For this study research, we used the online program NetworkAnalyst (http://www.networkanalyst.ca), a visual analytics platform for comprehensive gene expression profiling & meta-analysis. By coupling well-established statistical procedures with state-of-the-art data visualization techniques, NetworkAnalyst allows researchers to easily navigate large complex gene expression data sets to determine important features, patterns, functions and connections, thus leading to the generation of new biological hypotheses[136,137].

The KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway Database is a collection of manually drawn pathway maps representing our knowledge on the

molecular interaction, reaction and relation networks for: metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases and drug development[138].

3.8) Immunohistochemistry

MUC5B immunohistochemistry was performed in 44 interstitial lung disease (ILD) cases and 6 controls following the antibody manufacturer's protocol. Among ILD cases, 39 had a diagnosis of UIP/IPF and 5 of other ILDs (2 sarcoidosis, 1 nonspecific interstitial pneumonia, 2 smoking related interstitial lung disease and 1 emphysema).

Three μm-thick sections were processed for immunohistochemical analysis of MUC5B expression [monoclonal mouse anti-MUC5B antibody clone 4A10-H2, (1:200, Novus Biologicals, Centennial, Colorado, USA)]. For all immunohistochemical experiments, negative controls were obtained by incubation of the sections with the omission of primary antibody and using the antibody diluents alone or the appropriate non-immune IgG in each case. Immunohistochemistry (IHC) for both antibodies was performed in the Leica Bond-III Autostainer (Leica Microsystems Srl, Wetzlar, Germany) with the following protocol: after dewaxing and hydration, sections were incubated with Bond Epitope Retrieval 1 (Leica Microsystems Srl, Wetzlar, Germany) for antigen retrieval, then sections were treated for 15 minutes with the primary monoclonal antibody at room temperature and subsequently incubated with Bond Polymer Refine Detection system (Leica Microsystems Srl, Wetzlar, Germany) as detection system. The assay type was indeed

a commercial kit, according to the protocol of the manufacturer. Finally, the sections were counterstained with Mayer's haematoxylin.

MUC5B expression was evaluated in right lungs distinguishing upper lobe, median lobe, lower lobe: one sample for each lobe was randomly selected. If only the left lung was explanted, one sample taken from each of the two lobes was considered. In case of VATS, two different sections were considered.

Data were expressed by using the following scoring system:

- 0= no expression;
- 1=expression in less than 30% of epithelial cells;
- 2=expression in 30-60% of epithelial cells;
- 3=expression in more than 60% of epithelial cells.

We used bronchial positive staining as internal control.

4) RESULTS

4.1) Manual microdissection

The manual microdissection presented several difficulties from a practical point of view, mainly related to the small extension of the area of our interest, as the FF+MECs (FF sandwich, FF areas) have dimensions that vary greatly in terms of shape and volume (range, 1.3×10^4 to 9.9×10^7 μm^3)[139]. Microdissection was also difficult even when some laboratory strategies were adopted. In particular a stereomicroscope was used to have a greater freedom of movement to dissect the FF sandwich but no adequate vision of the section was possible (see figure 22). However, the main problem of this methodology was the contamination with adjacent tissue areas (e.g. fibrosis, inflammatory cells…) due to the use of a thin needle to remove the area of interest.

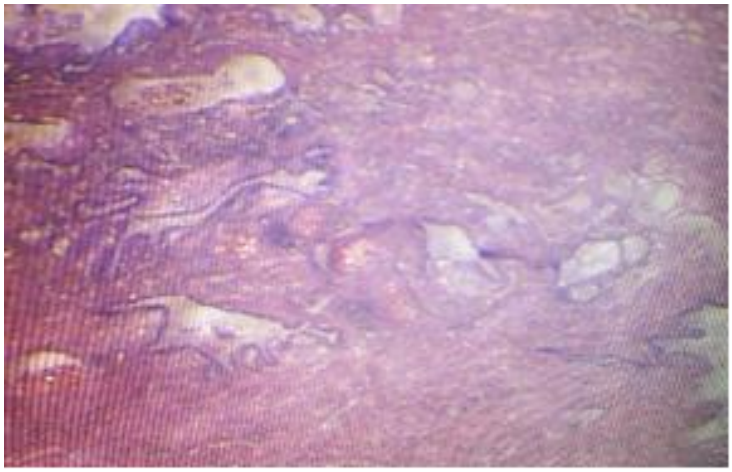

Figure 22: *view of the Leica M60 stereomicroscope preparation*

After the microdissection, tissue fragments were collected into tubes in PKD buffer and the RNA was extracted following the manufacturer's protocol.

Three attempts were performed trying to dissect only the FF sandwich from one IPF patient. The results are listed in the table below (table 4).

Attempt number	RNA concentration (ng/µl)	$A_{260/280}$	$A_{260/230}$
1	19.2	1.86	0.81
2	4.2	1.28	0.14
3	1.2	1.15	-6.20

Table 4: *absorbance measurements of RNA extracted from manual microdissected tissues on a Nanovue spectrophotometer (GE Healthcare Life Sciences, UK).*

RT-PCR for GAPDH (housekeeping gene) was performed in cases #1 and #2: amplicons were well visible in both (figure 23). This analysis was not performed in case #3 because of the poor quality and low quantity of extracted RNA.

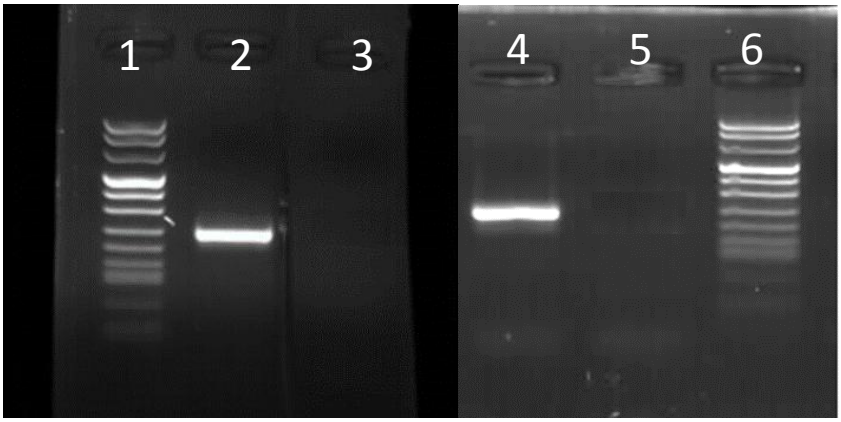

Figure 23: *agarose gel electrophoresis of amplicons obtained from RT-PCR for GAPDH (lane 2, 4). Molecular weight markers VIII (Roche Applied Science) are in lanes 1 and 6; negative controls in lane 3 and 5.*

However this methodology was abandoned because of the high probability of contamination of the area of interest (FF sandwich) with adjacent areas, shifting in a more selective procedure: laser microdissection (LMD).

4.2) Laser microdissection (LMD)

The LMD technique was more time consuming in comparison to manual microdissection (about 7 hours per patient) but more precise in the area selection (figure 24).

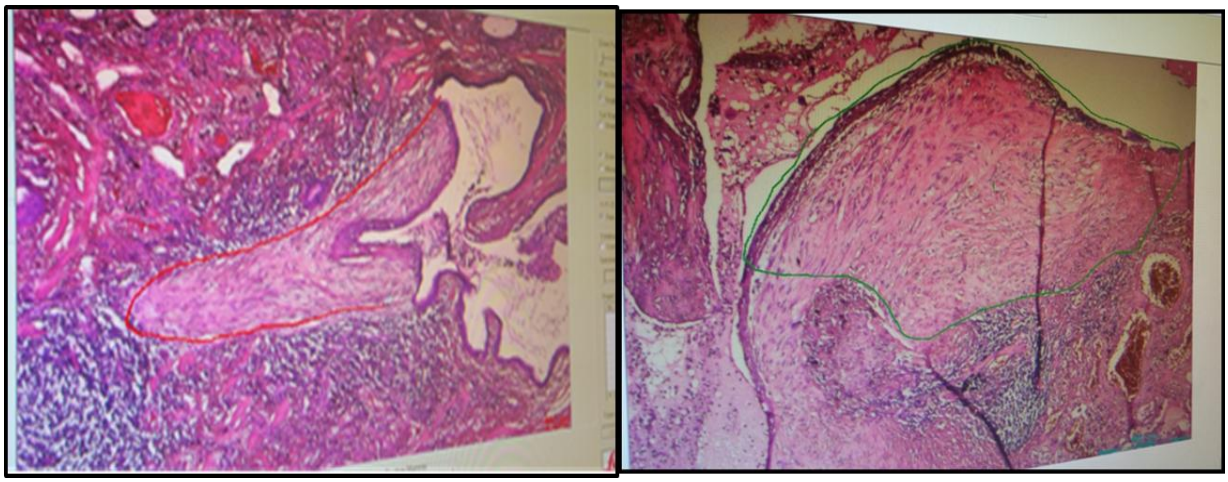

Figure 24: *laser microdissection of FF + MECs with Leica LMD6500 microscope in two emblematic IPF cases (#2 and #8 respectively)*

In IPF patients a mean (range) value of 8.3 (3-22) FF sandwich per section were microdissected in 8-15 sections/patient, thus the mean (± SD) total number of microdissected FF sandwich per patient was 87 (± 33). Considering the selected area, the mean (± SD) microdissected surface per patient was 396637 μm^2 (± 951132).

In control cases, the mean (range) value of microdissected FF sandwich per section was 5.2 (3-9) in 16 sections/patient thus the mean (± SD) total number of microdissected FF sandwich per patient was 69 (± 42). The mean (± SD) microdissected surface per patient was 190674 µm2 (± 154417).

These values were not statistically different between IPF cases and controls.

LMD was performed to select FF sandwich by three IPF patients (samples # 6, 7, 8). Even if no signal was detected using both RNA 6000 Nano kit and Qubit RNA BR Assay, RT-PCR for GAPDH was performed. As shown in figure 25, the amplicon was well visible in one sample (#8), while it was weak in the others (#6 and #7).

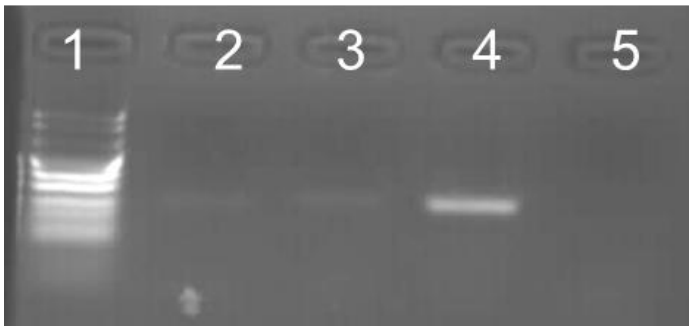

Figure 25: *agarose gel electrophoresis of amplicons obtained from RT-PCR for GAPDH. (1: marker VIII Roche Applied Science; 2: sample n.6; 3: sample n.7; 4: sample n.8; 5: negative control)*

The presence of a signal, even if weak, highlighted that RNA was present, but a more sensitive system was necessary for the detection. Thus, Agilent 6000 Pico Kit was used to evaluate quality and quantity of extracted RNA (Figure 26) and results are summarized in table 5. Four cases were excluded because of poor quality for the subsequent molecular analysis (highlighted in grey).

SAMPLE NUMBER	PATIENT	TRANSPLANT YEAR	STUDY GROUP	CONCENTRATION pg/ul	RIN value	DV200 %	QUALITY
12	B.G.	2003	IPF case	923	2.1	45	low
6	V.G.	2006	IPF case	978	2.1	46	low
/	F.V.	2008	IPF case	828	2.5	25	too degraded
8	G.G.	2011	IPF case	3660	2.2	39	low
5	P.A.	2014	IPF case	1133	2.4	43	low
7	P.S.	2014	IPF case	4724	2.1	44	low
1	B.R.	2015	IPF case	4203	2.4	41	low
2	P.S.	2015	IPF case	7530	2.1	59	medium
4	B.R.	2015	IPF case	840	2.2	52	medium
3	T.C.	2016	IPF case	7261	2.2	47	low
/	M.M.	2016	IPF case	160	2.3	25	too degraded
9	F.N.	2017	IPF case	2298	2.2	57	medium
10	S.M.	2017	control case	1403	2.3	58	medium
11	C.R.	2017	control case	960	2.2	70	high
/	D.A.A.	2017	control case	4235	2.5	21	too degraded
/	B.G.	2017	control case	154	2.3	28	too degraded

Table 5: *characteristics of RNAs evaluated using Agilent 6000 Pico Kit (Agilent Technologies, Santa Clara, Carlsbad, USA). In pink IPF samples that were used for libraries preparation, in yellow samples that were used as controls, in grey samples that were excluded due to their poor quality.*

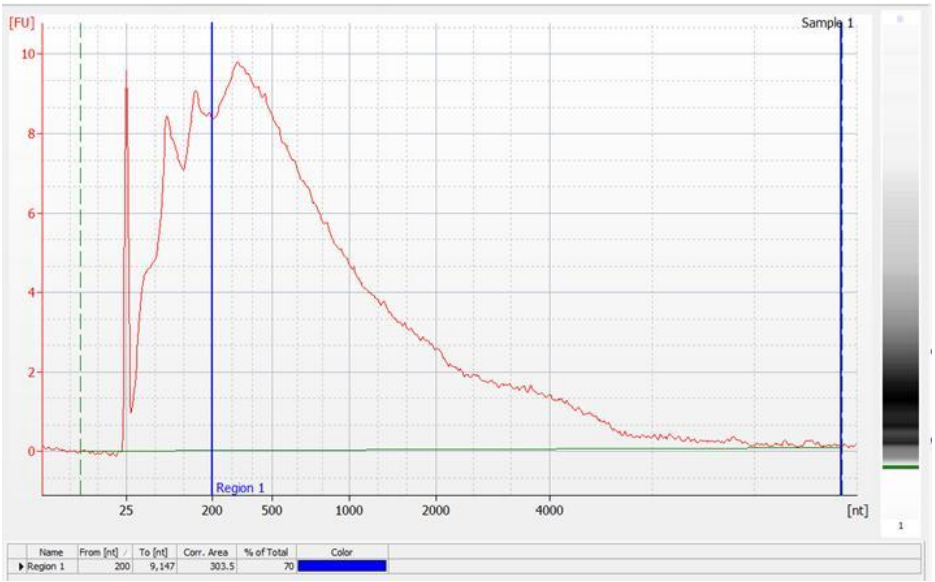

Figure 26: *emblematic electropherogram (sample #11) obtained with the Agilent 6000 Pico kit. Blue lines represent the* DV_{200} *area.*

Considering all microdissected cases, mean (±SD) RIN value was 2.2 (±0.1) indicating that 18S and 28S RNA subunits were not detected (as expected in RNA extracted from FFPE tissue samples). For what concerns DV_{200}, 58% of the samples showed a low quality (values ranging from 39% to 47%), 33% a medium quality (values ranging from 52% to 59%) and 9% a high quality (value of 70%). Mean (±SD) RNA concentration was 2993 (±2473) pg/ul.

4.3) Evaluation of libraries

According the SMARTer Stranded Total RNA-Seq kit- Pico Input Mammalian (Takara Bio USA, Inc., Mountain View, USA) procedures, libraries were prepared and then quantified by Qubit dsDNA HS kit (Thermo Fisher Scientific, Walthman, MA, USA). Obtained values are summarized in table 6.

Sample number	Qubit concentration (ng/µl)	cDNA fragments (base pair)
1	4.86	278
2	16.9	303
3	18.5	295
4	16.9	295
5	18.9	292
6	5.74	257
7	16.2	285
8	11.7	278
9	22.6	299
10	21.4	348
11	11.5	287
12	3.4	252

Table 6: *concentration of cDNA libraries evaluated with Qubit dsDNA HS kit (Thermo Fisher Scientific, Walthman, MA, USA)*

Quantity and quality of all libraries were adequate to perform the subsequent RNA-Seq analysis (library electrophoresis run summary in the figure 27 and an example of electropherogram in the figure 28). The mean number of obtained cDNA fragments was 289 nucleotides.

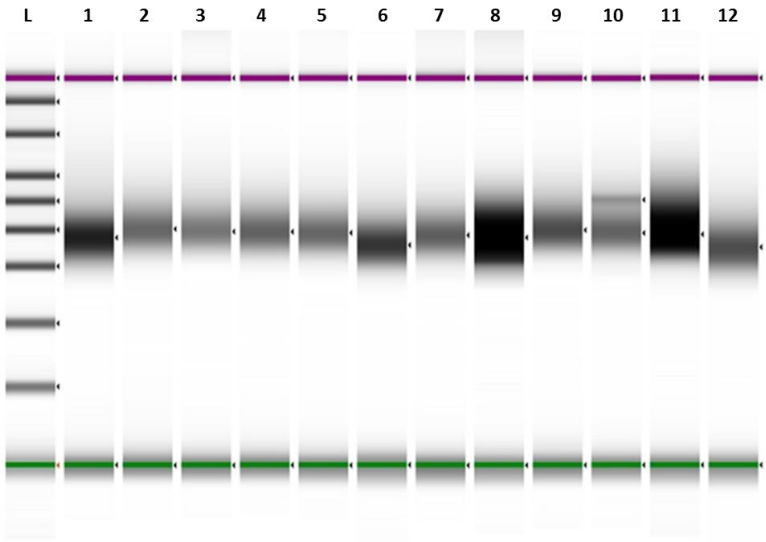

Figure 27: *electrophoresis run summary of libraries obtained using the SMARTer Stranded Total RNA-Seq kit- Pico Input Mammalian (Takara Bio USA, Inc., Mountain View, USA). Lane L= Electronic Ladder; lane 1 to 12 correspond to sample 1 to 12 respectively).*

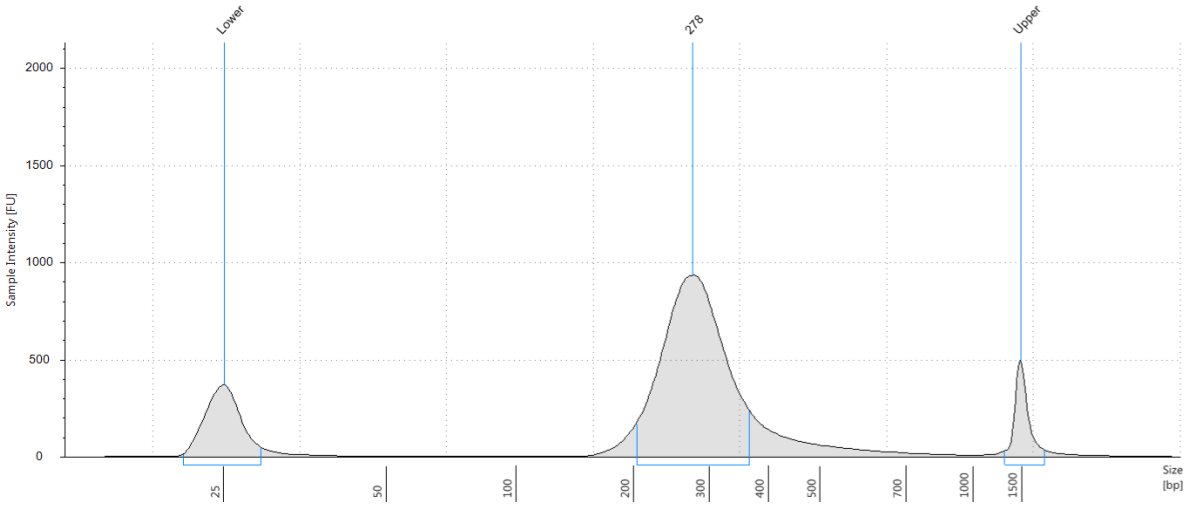

Figure 28: *example of an electropherogram obtained using Agilent High Sensitivity D5000 Screeen Type Assay (sample 7)*

4.4) RNA-sequencing and DEGs analysis

RNA libraries were sequenced and the results are reported in table 7 (quality threshold: ≥Q30).

sample ID	Barcode sequence	PF Clusters	# Reads	Yield (Mbases)	% >= Q30
1	ATTACTCG	77.940.408	155.880.816	23.538	88,08
2	TCCGGAGA	69.304.070	138.608.140	20.930	88,31
3	CGCTCATT	71.011.715	142.023.430	21.446	87,43
4	GAGATTCC	57.967.516	115.935.032	17.506	88,56
5	ATTCAGAA	62.679.508	125.359.016	18.929	88,6
6	GAATTCGT	65.305.723	130.611.446	19.722	89
7	CTGAAGCT	68.332.654	136.665.308	20.636	87,79
8	TAATGCGC	58.854.383	117.708.766	17.774	89,7
9	CGGCTATG	56.889.185	113.778.370	17.181	91,29
10	TCCGCGAA	58.762.460	117.524.920	17.746	90,63
11	TCTCGCGC	70.054.732	140.109.464	21.157	90,96
12	AGCGATAG	66.643.828	133.287.656	20.126	89,74

Table 7: *NGS reads obtained after the sequencing process*

Before performing the DEGs analysis, a quality check was done on the raw sequencing data removing low quality portions and preserving the longest high quality part of NGS reads. The trimming step was performed with the following parameters: the minimum length was set to 35 bp and the quality score to 25. The software BBDuk was used for this scope. The mean trimmed pair reads obtained with NGS was of 100 milions reads/sample (table 8).

Used Name	Condition	Raw pair reads	Trimmed pair reads
Sample_01	IPF	155,880,816	125,692,986
Sample_02	IPF	138,608,140	109,566,296
Sample_03	IPF	142,023,430	111,823,508
Sample_04	IPF	115,935,032	92,571,704
Sample_05	IPF	125,359,016	100,899,286
Sample_06	IPF	130,611,446	105,382,252
Sample_07	IPF	136,665,308	108,605,884
Sample_08	IPF	117,708,766	99,322,980
Sample_09	IPF	113,778,370	98,228,200
Sample_10	Control	117,524,920	100,133,242
Sample_11	Control	140,109,464	119,879,612
Sample_12	IPF	133,287,656	113,810,592

Table 8: *number of reads before and after quality check.*

The high quality reads were aligned to the Homo sapiens reference genome sequence (GRCh38) [ENSEMBL] using STAR aligner (version 2.5.0c). Alignment stats are shown in table 9.

Used Name	Condition	Percentage of Mapping (Uniquely mapped pairs) (%)	Percentage of Mapping (Multi-mapped pairs) (%)
Sample_01	IPF	71.54	9.19
Sample_02	IPF	77.98	14.14
Sample_03	IPF	78.34	13.28
Sample_04	IPF	80.52	13.23
Sample_05	IPF	81.35	11.09
Sample_06	IPF	49.31	8.30
Sample_07	IPF	77.12	15.06
Sample_08	IPF	63.24	29.95
Sample_09	IPF	82.77	9.08
Sample_10	Control	64.22	10.07
Sample_11	Control	73.41	12.36
Sample_12	IPF	51.69	16.78

Table 9: *alignment stats of the high quality reads against the Homo sapiens reference genome sequence (GRCh38)*

Table 9 shows that for some samples (i.e. Sample_06, Sample_10 and Sample_12) a low percentage of reads was mapped (considering both Unique and Multiple Mapping reads). In these samples, results were the same even using Local mode (STAR parameter).

4.4.1) Statistical analysis

FeatureCounts (version 1.6.0) was used to calculate gene expression values as raw fragment counts (annotation GRCh38). Normalization was applied to the raw fragment counts by using the Trimmed Mean of M-values (TMM) normalization and Fragments Per Kilobase Million (FPKM) normalization.

All the statistical analyses were performed with R and RStudio (version 1.1.463).

Genes that had no expression and/or showing too much too much variability for each comparison were excluded. The HTSFilter package was chosen for this scope, which implements a filtering procedure for replicated transcriptome sequencing data based on a Jaccard similarity index. The Trimmed Means of M-values (TMM) normalization strategy was used. First of all, the edgeR package was used in order to do a differential expression gene analysis due to this package is one of the most used methods to perform this kind of analysis. Using a generalized linear model it is possible to take into account confounding factors to increase the sensibility of the analysis but we did not find any significant results with this package.

Thus, to cover this issue, we used the following statistical analysis:

- first a " standard" analysis with NOISeq was performed[140]. NOISeq package was chosen since it was specifically designed to deal with "noisy" RNA-seq data. We find only 87 differentially expressed genes (all of them down-regulated).

- A new NOISeq analysis was performed, in this case with the function "ARSyNseq" which applies an ANOVA analysis to try to subtract the "noise" created by the batch effect. The batch effect was applied using Metadata (sex, age at transplantation, morphologic diagnosis, etc.) but we didn't find any differentially expressed genes.
- Finally, we decided to use NOISeq package with the function "ARSyNseq" without specifying a specific metadata as batch effect in order to removal systematic noise from the data. By this, 323 differentially expressed genes were found (14 of them up-regulated and 309 down-regulated).

Figure 29 shows a summary of the differentially expressed genes that were detected.

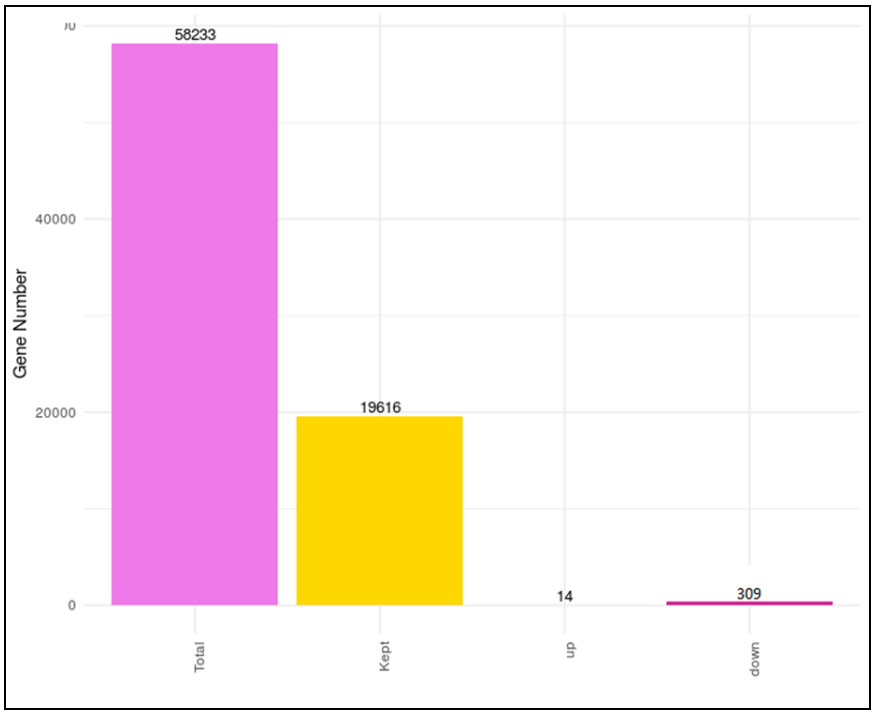

Figure 29: *the bars shown in this plot report the total number of annotated in genes in the reference genome (Total), the number of genes that passed the low expression filter (Kept), the total number of differentially expressed genes (DE) across the comparisons.*

An MA plot and a Volcano plot were also generated (figure 30). MA plot shows the relationship between the average expression value (on the X-axis) and the fold change (Y-axis) for each gene in the genome (figure 30 A). The distribution of the dots in the MA-plot can be useful to check if the differentially expressed genes are uniformly distributed across the different ranges of expression values and the relationship with the fold-change.

Volcano plot shows the relationship between the fold-change (on the X-axis) and the significance of the differential expression test (Y-axis) for each gene in the genome (figure 30 B). The distribution of the dots in the Volcano plot can be useful to check if a range of fold-changes is associated with a stronger or a weaker significance of differential expression.

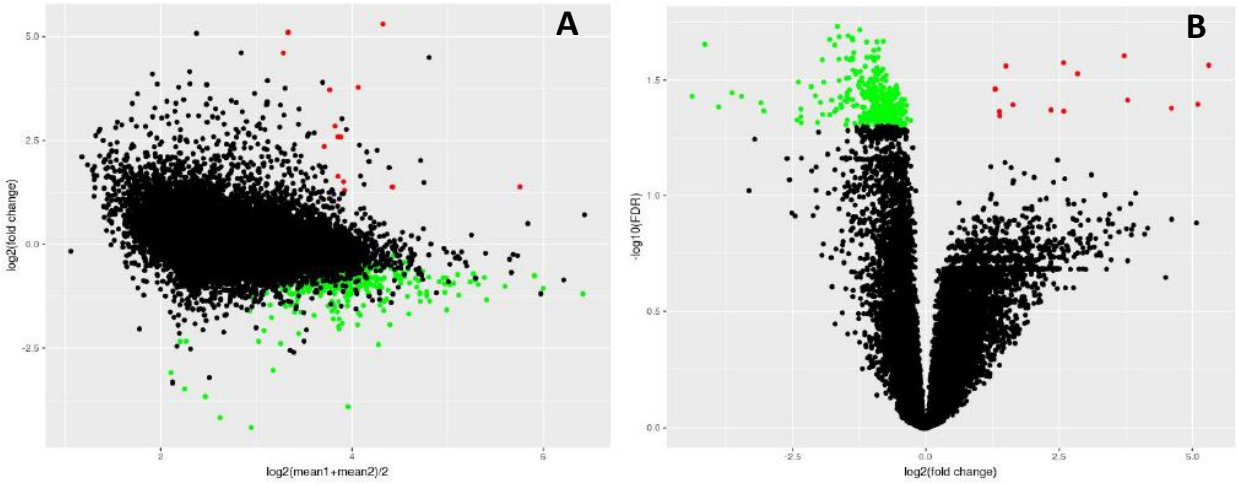

Figure 30: *MA plot (A) and Volcano plot (B). Black dots represent the genes that are not significantly different expressed, while red and green dots are the genes that are significantly up- and down-regulated, respectively.*

4.4.2) Up-regulated genes

The group of over-expressed genes included 14 genes that are summarize in table 10:

Gene ID	log2FC	FDR	Description
MUC5B	5,31	2,73E-02	mucin 5B
SCGB3A1	5,10	4,03E-02	secretoglobin family 3A member 1
LCN2	4,61	4,19E-02	lipocalin 2
PIGR	3,78	3,86E-02	polymeric immunoglobulin receptor
KRT5	3,72	2,48E-02	keratin 5
KRT17	2,85	2,96E-02	keratin 17
CEACAM6	2,59	4,31E-02	carcinoembryonic antigen related cell adhesion molecule 6
SERPINE1	2,58	2,66E-02	serpin family E member 1
NAMPT	2,35	4,26E-02	nicotinamide phosphoribosyltransferase
LAMC2	1,63	4,05E-02	laminin subunit gamma 2
GPRC5A	1,50	2,75E-02	G protein-coupled receptor class C group 5 member A
NEAT1	1,38	4,51E-02	nuclear paraspeckle assembly transcript 1 (non-protein coding)
VMP1	1,38	4,33E-02	vacuole membrane protein 1
BHLHE40	1,30	3,46E-02	basic helix-loop-helix family member e40

Table 10: *DEGs up-regulated genes. Log2 fold change (log2FC), false discovery rate (FDR)*

The most significant up-regulated gene was MUC5B. This gene has long been studied in IPF as described in the chapter 1.8.1.1.2.

In our research, other up-regulated genes were detected. We take into account only the genes with a Log2FC greater than 1.60. These genes are mainly involved in epithelial-to-mesenchymal transition (EMT) and cell proliferation process. Here below briefly described their principal function in the lung:

- **Secretoglobin family 3A member 1 (*SCGB3A1*)** is a positive regulator of myoblast fusion and is involved in the regulation of cell population proliferation (Uniprot, GO - Biological process; https://www.uniprot.org/help/gene_ontology). This gene is also up-regulated in lung adenocarcinoma[141].
- **Lipocalin 2 (*LCN2*)** is a secreted glycoprotein that transports small lipophilic ligands.

Its up-regulated expression lays a critical role in all growth and differentiation, EMT process, angiogenesis, and cell migration and with malignancy in various cancers[142,143]. Notably, LCN2 functions as an initiator of carcinogenesis and metastasis by involving multiple signaling pathways[144]. In a mouse xenograft model of human lung adenocarcinoma, down-regulation of *LCN2* was demonstrated to reduce tumor growth and promote extensive cell death *in* vivo[145].

- **Polymeric immunoglobulin receptor (*PIGR*)** plays a major role in reacting with antigens distributed over the huge area of mucosal surfaces that comprise the digestive, respiratory, and urogenital tracts. The main function of PIGR is to transport dimeric immunoglobulin A (IgA) and polymeric immunoglobulin M (IgM) from the lamina propria across the epithelial barrier to mucosal surface. IgA is the most abundant immunoglobulin found in mucosal secretions, and participates in frontline defense mechanisms found in the respiratory tract, along with mucociliary clearance. This gene is also largely regulated by airway epithelial cell differentiation, similar to mucin production by goblet cells. Frequent exacerbators COPD patients showed up-regulation of this gene[146].
- **Keratin 5 (*KRT5*)** and **Keratin 17 (*KRT17*)** are also found over-expressed in IPF patients in other research studies. These genes cluster with other genes related to bronchiolar epithelium including mucins, proline-rich secreted factors, serine protease inhibitors, ion channels and cilium components. This is consistent with reports of abnormal "bronchiolization" of alveolar spaces in IPF, which may represent epithelialization of honeycombed cystic spaces in regions of dense scarring[147]. In particular, KRT5 can be found in close association with MUC5B in honeycomb cysts[148]. *KRT17* was also find to regulate protein synthesis and epithelial cell growth in a mouse model[149].

• **Carcinoembryonic antigen related cell adhesion molecule 6 (*CEACAM6*)** overexpression modulates cancer progression through aberrant cell differentiation, anti-apoptosis, cell growth and resistance to therapeutic agents. In addition, CEACAM6 overexpression in multiple malignancies promotes cell invasion and metastasis, thereby representing an acquired advantage of tumor cells directly responsible for an invasive phenotype[150]. Recently it has been successfully proposed a therapeutic effect silencing CEACAM6 in a xenograft model of lung adenocarcinoma[151].

• **Serpin family E member 1 (*SERPINE1*)** also known as Plasminogen activator inhibitor-1 (PAI-1) is a primary inhibitor of tissue type and urokinase type plasminogen activator which convert plasminogen in plasmin, a serum proteinase, playing a major role in fibrinolysis. Besides suppression of fibrinolysis, *SERPINE1*, has many other functions, including modulation of cell adhesion, migration and proliferation, dependent or independent of its protease inhibitor activity. Several studies have shown that *SERPINE1* plays a critical role in the development of lung fibrosis, although the mechanism whereby *SERPINE1* promotes lung fibrosis remains elusive[152]. More recently several works have reported an emerging role of *SERPINE1* as mediator of pneumocytes senescence in IPF, even if the mechanisms how *SERPINE1* promotes cell senescence remain unclear[153].

• **Nicotinamide phosphoribosyltransferase (*NAMPT*)** has been implicated in many important biological processes including metabolism, stress response, apoptosis and aging.

NAMPT is a “cytozyme” dually functioning as a cytokine in the extracellular milieu (eNAMPT, which activates the TLR4 receptor to transduce pro-survival signals) and as an intracellular enzyme (iNAMPT, which plays a well described role as an inhibitor

of spoptosis). eNAMPT mediates TLR4/NOX4 dependent induction of iNAMPT and pro-fibrotic myofibroblast phenotypes (NOX4-dependent oxidative stress, differentiation, senescence and apoptosis-resistance). iNAMPT is persistently expressed in senescent and IPF myofibroblast that fail to undergo apoptosis and in lung of aged mice with non-resolving fibrosis. Reduction in NAMPT expression facilitated IPF myofibroblast apoptosis and led to protection from fibrosis in vivo[154].

- **Laminin subunit gamma 2 (*LAMC2*)** is involved in cell migration, extracellular matrix organization, positive regulator of cell population proliferation (Uniprot, GO - Biological process; https://www.uniprot.org/help/gene_ontology) and was already found overexpressed in the "Fibroblastic Foci sandwich" of UIP biopsies[155].

4.4.3) Down-regulated genes

The 309 down-regulated DEGs were studied by a Gene Ontology Enrichment Analysis (GOEA) to identify the most enriched Gene Ontology (GO) categories. This type of analysis was possible only for the down-regulated genes, since we could not find enough significantly enriched GOs for the up-regulated genes (their number were too small for a GOEA analysis).

In the followed graphical representation, statistical significance is represented by colored dots (red to yellow gradient) which describe a different level of significance, higher (red) to lower (pale yellow) (figures 31, 32, 33, 34). Under-expressed genes were annotated mainly as extracellular matrix structure and organization pathway.

GO cellular component analysis (figure 31) indicated that the enrichment of down-regulated DEGs was predominantly in extracellular region part, extracellular matrix and

proteinaceus extracellular matrix (red dots: i.e. *ACTB*, genes of the collagen superfamily, *DYNC1H1*, *FBLN1*, *FBLN2*, *EFEMP1*, *FLNA*, *FMOD*, *GPC1*, *HSPA8*, *HSPG2*, *IGF2*, Insulin-like growth factor binding proteins, *LAMA4*, *LAMB2*, *LGALS1*, *LTBP2*, *LUM*, *MFAP2*, *MFAP4*, *MGP*, *MMP2*, *MMP11*, *MMP14*, *MYH9*, *SERPINF1*, *SERPINA1*, *TIMP2*, *VEGFB*); extracellular region, extracellular matrix part and extracellular space (orange dots).

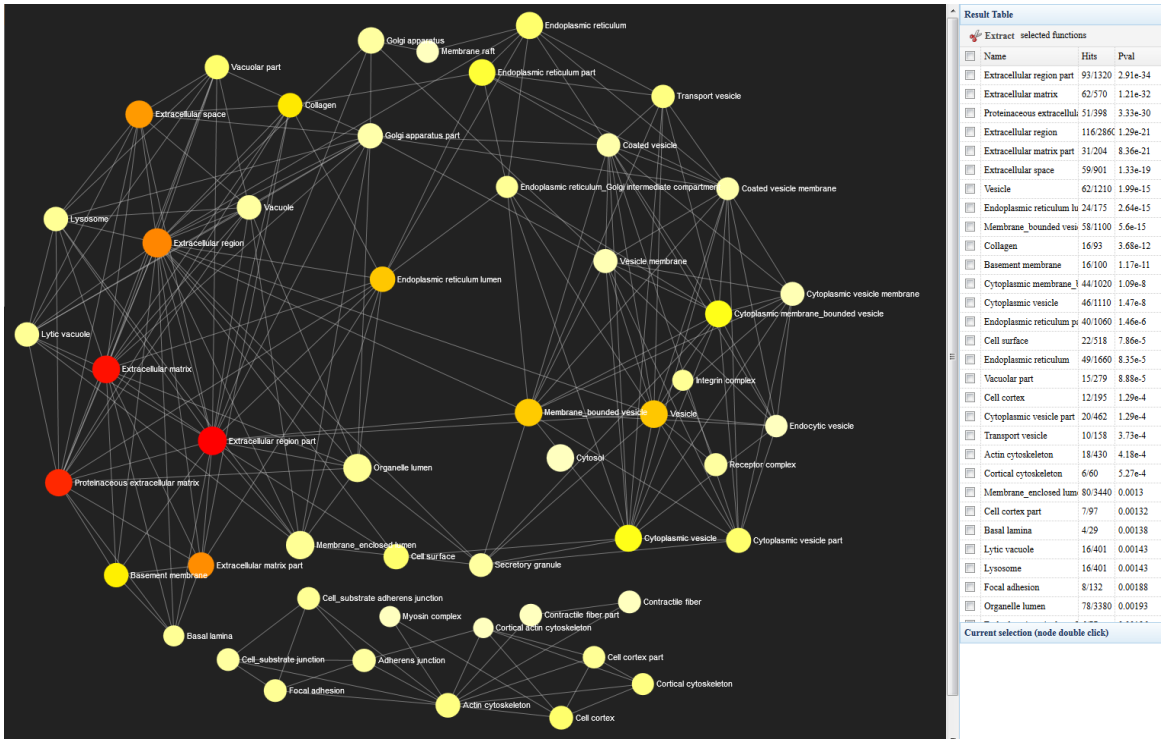

Figure 31: *GOEA cellular component (CC) analysis*

In molecular function (figure 32), GO analysis showed that down-regulated DEGs were mainly enriched in extracellular matrix structure (red dot: i.e. genes of the collagen superfamily, *FBLN1*, *FBLN2*, *LAMA4*, *EMILIN1*), growth factor binding (orange dots: i.e. genes of the collagen superfamily, Insulin-like growth factor binding proteins,

LTBP2, PDGFRA, PDGFRB, NRP1) and collagen binding (orange dots: *SERPINH1, DCN, LUM, NID1, PCOLCE, SPARC, COL14A1, NID2, ITGA11, ANTXR1, PODN*).

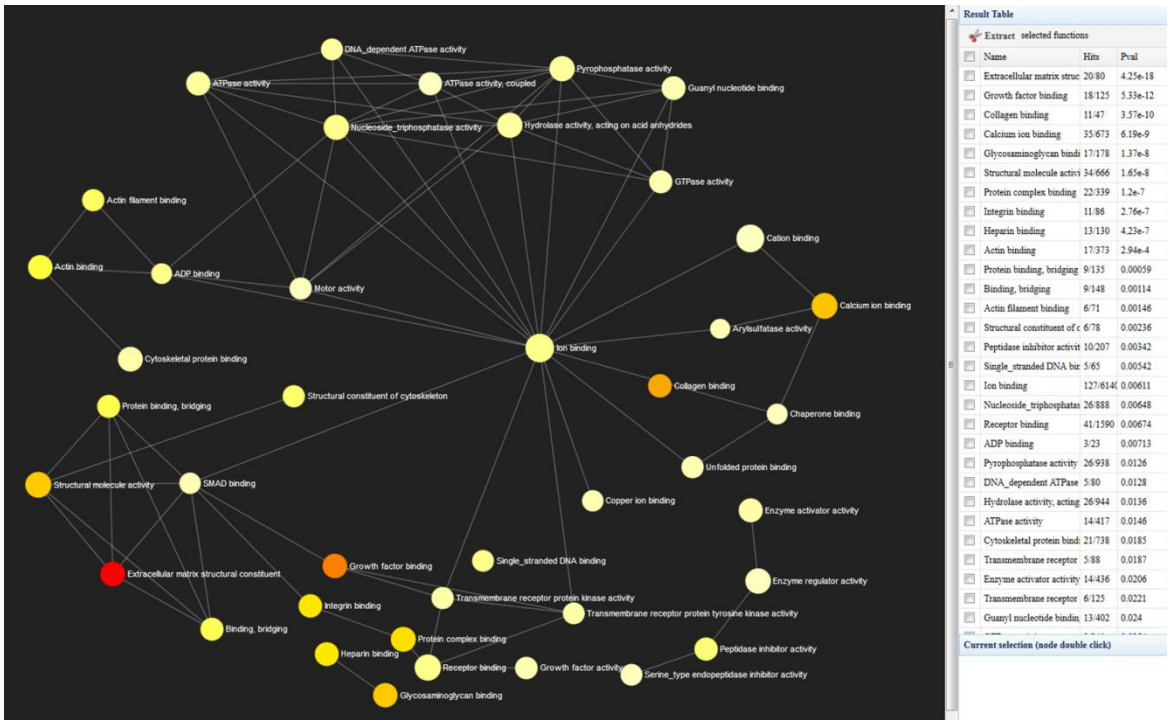

Figure 32: *GOEA molecular function (MF) analysis*

In the biological process group (figure 33), GO analysis showed that down-regulated DEGs were mainly enriched in extracellular structure organization (red dot): genes listed under this GO term were, for example, *APLP2, SERPINH1*, genes of the collagen superfamily,*MMP2, MMP11, MMP14, PDGFRA, TIMP2*.

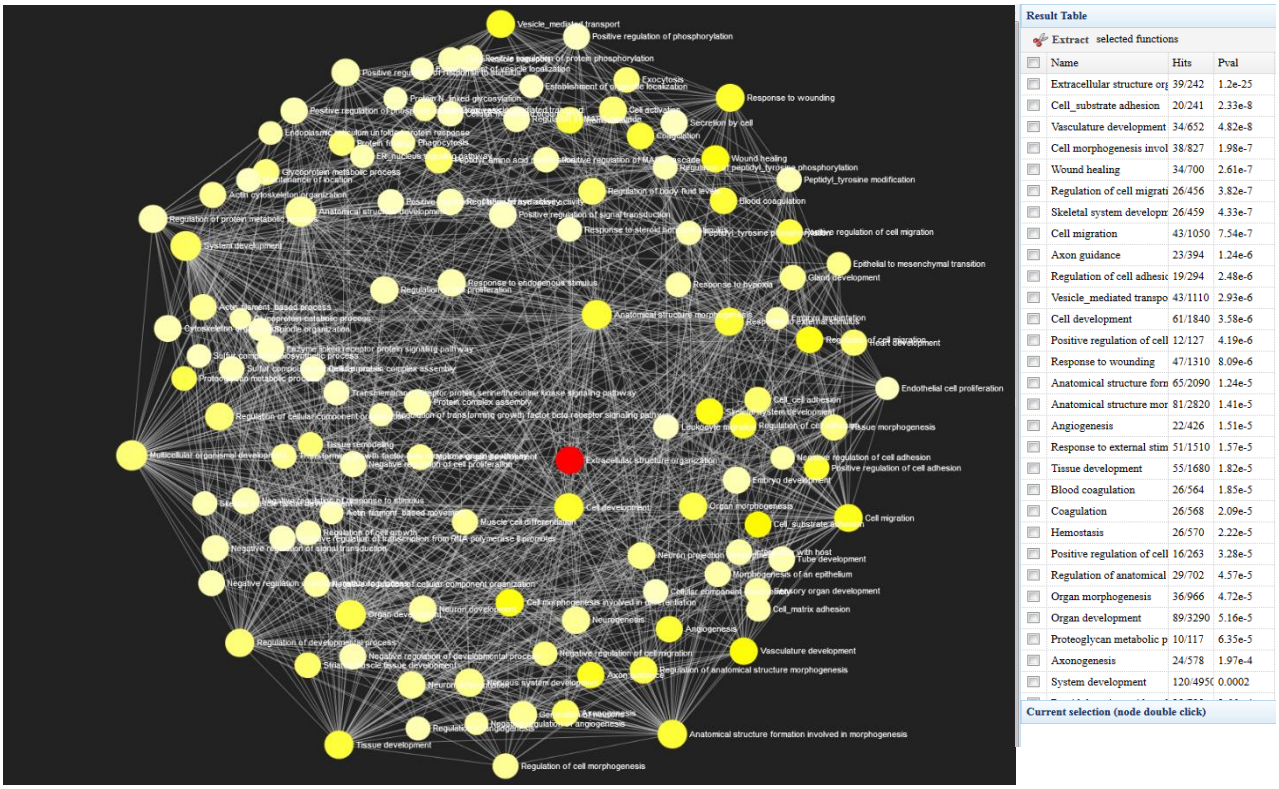

Figure 33: *GOEA biological process (BP) analysis*

Furthermore, the enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway of down-regulated DEGs mainly induced focal adhesion (i.e. *PARVA*, *MYLK*, *COL4A2*, *PDGFRA*, *ACTB*, *COL6A1*, *TLN1*, *ITGB5*, *LAMB2*, *THBS2*, *MYL9*, *FLNA*, *HGF*, *LAMA4*, *PDGFRB*, *VEGFB*, *COL1A2*, *COL6A2*, *COL1A1*, *ITGA11*), protein digestion and absorption, ECM-receptor interaction and phagosome pathway (figure 34).

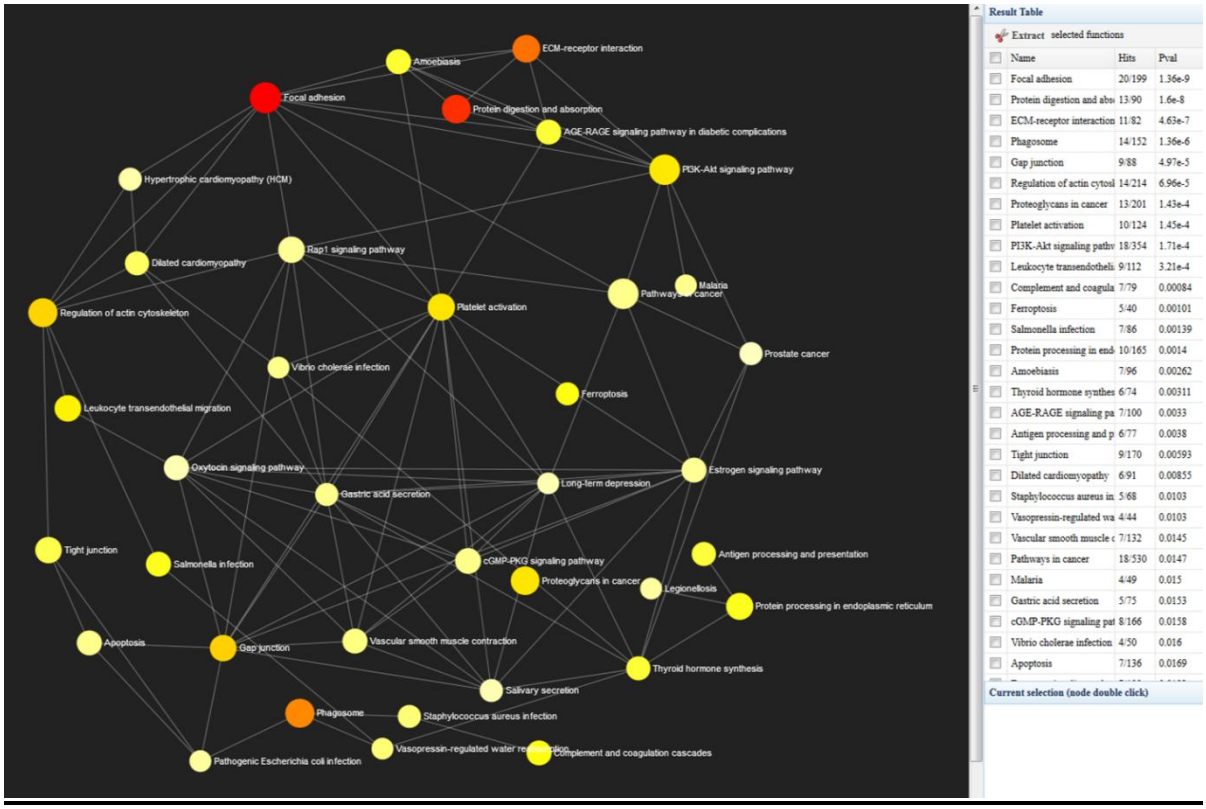

Figure 34: *KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway*

4.5) Validation of MUC5B by immunohistochemistry

MUC5B immunohistochemistry was performed in 44 ILD cases and 6 controls (including all the cases in whose RNA-seq were done).

MUC5B expression was mainly detected in metaplastic epithelial cells lining distal bronchial tract, honeycomb lesions, some hyperplastic type II pneumocytes lining alveolar spaces, metaplastic epithelial cells lining FF and some hybrid cells (cells with transitional features of both epithelial cells and fibroblasts) that were detected inside the FF areas[89] (figure 35). MUC5B was never detected in the control group.

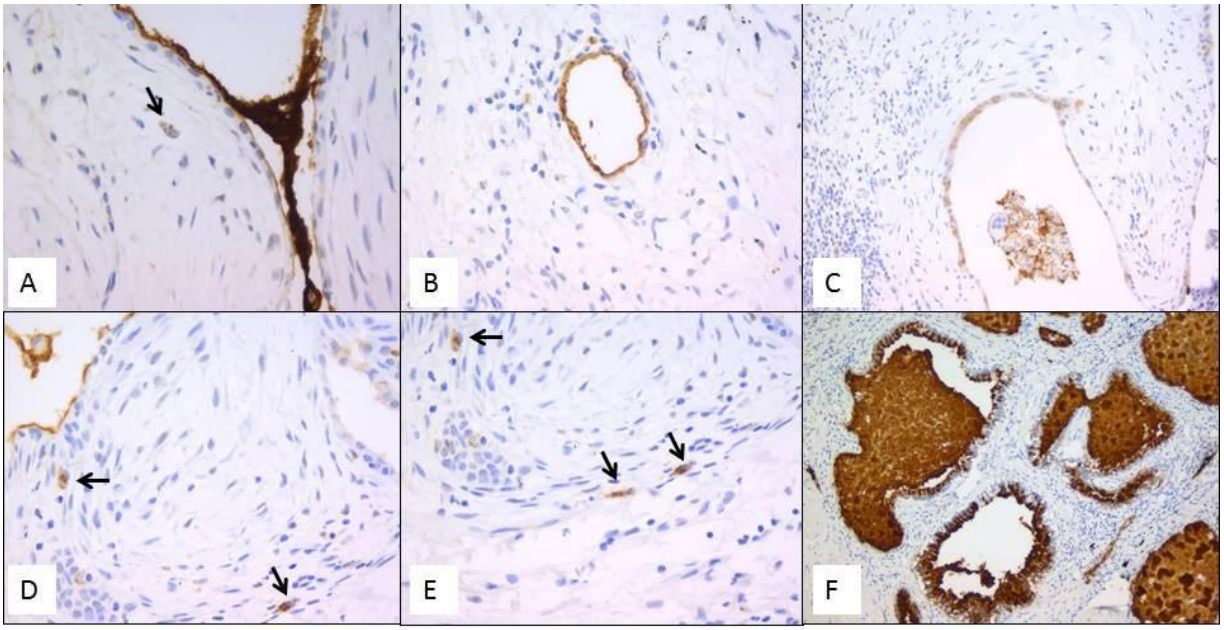

Figure 35: *representative futures of immunohistochemistry for MUC5B (clone 4A10-H2; 1:200, Novus Biologicals, Centennial, Colorado, USA) in emblematic UIP/IPF cases. Staining is visible in mucus between two FF sandwiches and in the metaplastic epithelial cells lining FF (A); hyperplastic type II pneumocytes lining alveolar spaces (B); hyperplastic type II pneumocytes and macrophages (C); hybrid cells in FF areas (D and E); honeycomb cysts filled with mucus (F). Arrows indicate the hybrid cells in figure A, D and E.*

MUC5B expression was significantly higher in ILD cases in comparison to controls, both considering separately each lobe and evaluating the median values (all $p \leq 0.001$) (table 11 and figure 36).

	ILDs (UIP/IPF=39; other ILDs=5)	**Controls (n=6)**	**p-value**
Upper lobe	2 (1-3)	0 (0-0)	0.001
Middle lobe	2 (1-3)	0 (0-0)	0.002
Lower lobe	3 (2-3)	0 (0-0)	0.0003
Median value of the lobes	2 (1-3)	0 (0-0)	0.0003

Table 11: *MUC5B expression score [median (Q1-Q3)] in ILDs vs controls.*

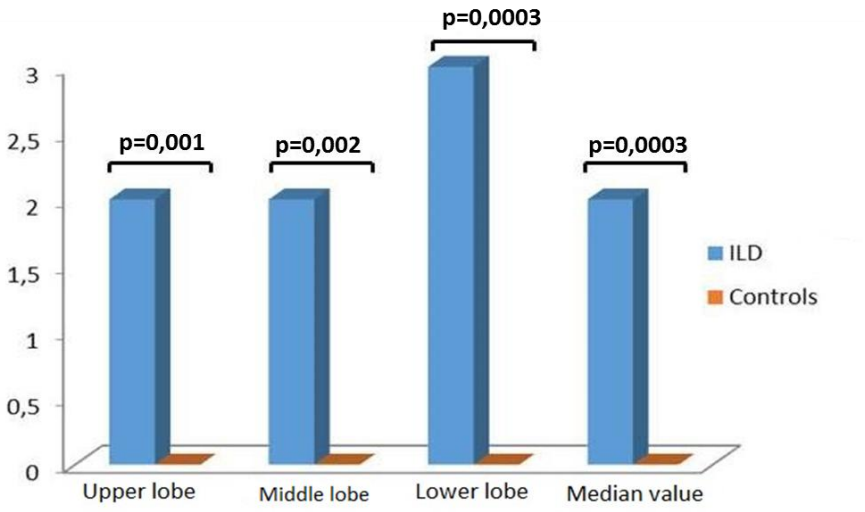

Figure 36: *Histogram showing MUC5B clone 4A10-H2 immunohistochemistry in ILDs vs controls.*

In particular, an increasing gradient of MUC5B protein expression was detected from upper to lower lobe in each patient.

MUC5B expression was not different between end-stage (lung transplanted) and early stage (VATS) cases.

MUC5B expression was equally expressed in the lower lobes of UIP/IPF and other ILDs cases (median values 3:3) while higher MUC5B expression was observed in upper and middle lobes in UIP/IPF in comparison to other ILDs, even if these values didn't reach a statistical significance (may be due to the small number of other ILDs cases) (table 12 and figure 37).

	UIP/IPF (n=39)	Other ILDs (n=5)	p-value
Upper lobe	2 (1-3)	1 (0-1)	0.1
Middle lobe	2 (1-3)	1 (0-2)	0.3
Lower lobe	3 (2-3)	3 (2-3)	0.7
Median value of the lobes	2 (2-3)	1 (0-2)	0.09

Table 12: *MUC5B expression score [median (Q1-Q3)] in UIP/IPF cases vs other ILDs.*

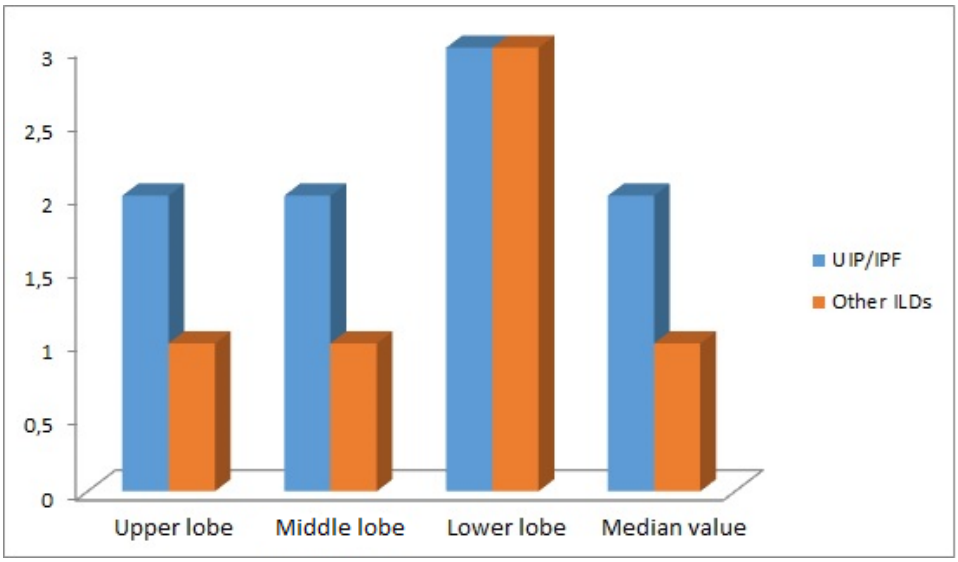

Figure 37: *Histogram showing MUC5B clone 4A10-H2 immunohistochemistry in UIP/IPF vs other ILDs.*

5) DISCUSSION

Idiopathic pulmonary fibrosis (IPF), the most common form of idiopathic interstitial pneumonias, is a chronic, progressive, irreversible, and lethal lung disease with a median survival of 2 to 3 years after diagnosis[128]. The etiology of IPF remains unclear and despite recent advances in the therapy, IPF persists as an incurable disease[100,156] with different and dynamic clinical phenotypes (from accelerated disease with early mortality to slowly progressive disease)[42]. To date, the current view about the pathogenesis of IPF is that it occurs in genetically susceptible individuals following repeated or persistent epithelial injuries that lead to the activation of fibroblasts and the excessive collagen deposition[157].

UIP, the histopathological pattern associated with IPF, is characterized by a heterogeneous appearance with areas of subpleural fibrosis and honeycombing alternating with areas of less affected or normal parenchyma (spatial heterogeneity). Small areas of active fibrosis (fibroblast foci, FF) are present in the background of collagen deposition, and these reflect the temporal heterogeneity of the process and indicate current ongoing disease[35]. Several papers[93,124-127] demonstrated a bad prognostic impact of number/extension of FF in IPF patients thus reinforcing the concept that this morphological change is the crucial lesion of UIP.

In the last few years great efforts have been made in investigating morphological and pathogenetic substrates of IPF in order to identify some candidate biomarkers: to date, lung tissue evaluation is still considered the bedrock to extrapolate the most clinically fitting ones especially in fibrotic disease that are unlikely to release specific biomarkers in other non-accessible compartment (e.g. blood)[36].

As in the oncological field, it would be desirable to apply high throughput OMIC technologies for a wide tissue screening of molecules rather than using a cherrypicking approach. Thus, the main goal of this PhD study was the identification of IPF crucial biomarkers in FF sandwich area using RNA-Seq approach, the most updated technology for the whole transcriptome analysis.

To the best of our knowledge, in literature only one study used RNA-Seq approach in FFPE tissue samples from IPF patients[113]. However, in this study the whole section was used, without selecting specific injured areas of the disease: this may represent a critical point when high sensitive molecular analyses are used for specific biomarker detection, especially in highly heterogeneous morphological lesions as UIP.

In the present study we used laser microdissection technology to select only the FF sandwich areas, avoiding contamination from other close components (e.g. inflammatory and endothelial cells, collagen). These very small amounts of lung tissue were then used for RNA extraction. The obtained RNA quality and quantity were in line with the previous studies, indeed in literature the quality (RIN number) of RNA extracted from laser microdissected FFPE tissues ranged between 2.1-2.6 and the length of the fragments were less than 300 ribonucleotides, thus optimal for OMIC approaches[113,158,159]. During my PhD study I optimized the RNA extraction protocol to obtain adequate RNA from these very small areas of FFPE tissues. The modifications I have introduced, the prolonged digestion step (an overnight proteinase K digestion rather than 15 minutes) and the addition of a RNAase inhibitor, allowed me to obtain a greater quantity and a better quality (avoiding fragmentation) of RNA. This represents an important methodological goal of my study and could provide crucial recommendations to other researchers interested in molecular approach of this disease, being archival tissues (FFPE) our major source of IPF lung tissue. As in lung

oncological field, the possibility to study gene expression of specific cells/areas of IPF lung tissue represents an important step forward to the improving of our knowledge in the pathogenesis and above all to the discovery of new druggable targets.

In this research study, RNA-seq analysis allowed the detection of 58,233 total numbers of annotated genes in the reference genome, with 19,616 numbers of genes that passed the low expression filter. In particular, 323 of them were differentially expressed between IPF and control: 14 of them were up-regulated and 309 down-regulated (FDR adjusted $p \leq 0.05$). The 14 up-regulated genes were all involved in epithelial proliferation/cancer progression and EMT process. These data corroborate the crucial role of metaplastic epithelial cells in the pathogenesis of IPF, in accordance with the theory that epithelial cell dysfunction represents the crucial actor not only for the development, but also for the progression of the disease[32]. Interestingly, the most significant up-regulated gene was *MUC5B*. This gene belongs to a 20 gene family but MUC5AC and MUC5B are the major mucins that are expressed in lungs. In normal lung, these mucins are mainly present in large airways and are responsible for trapping inhaled particles including bacteria, and transporting them out of the airways by ciliary and cough-driven forces[148,160]. Mucin dysregulation is commonly associated with a number of pulmonary diseases (chronic obstructive pulmonary disease, asthma, cystic fibrosis, etc.) however, the location of mucus production may have a deep impact effect on the consequent pathophysiological phenotype. For instance in contrast with IPF, where the excessive mucus production is localized to the distal airways/honeycomb[161,162], cystic fibrosis and chronic bronchitis are associated with enhanced mucus secretion in more proximal airways and submucosal glands (consequently, these diseases are airway centric).

Several works, including experimental models[61], have demonstrated an excess MUC5B protein in epithelial cells of the respiratory bronchiole and honeycomb cyst regions in lung fibrosis. However, it still remains unclear how MUC5B leads to the development of IPF. The prevailing theories are that overproduction of MUC5B impairs mucociliary clearance and contributes to mucociliary dysfunction, thus resulting in a chronic retention of several inhaled substances (air pollutants, cigarette smoke, microorganisms, etc.) and endogenous inflammatory debris that, over time, leads in temporally and spatially distinct areas of microscopic scaring and progressive lung fibrosis. Alternatively, MUC5B could interfere with alveolar repair, either by interfering with the interaction between the type II alveolar epithelial cells and the underlying matrix, or by interfering with the surface-tension properties of surfactant. This could enhance alveolar collapse and fibrosis of adjacent bronchoalveolar units, and eventually lead to the development of IPF.

The overexpressed MUC5B detected in the present research study, highlights the key role of injured epithelial cells even in a lesion (FF areas) where fibroblasts appear to be the predominant cells. To validate what we observed in RNA-Seq data, we performed immunohistochemistry for MUC5B. We found a strong MUC5B positive staining not only in distal bronchial tract and honeycomb lesions, but also in hyperplastic type II pneumocytes lining several alveolar spaces[163], in metaplastic epithelial cells lining FF and in some hybrid cells inside FF. In FF areas of control group, MUC5B was never detected.

Dysregulated expression of mucin, particularly as overexpression, has been reported to be associated with proliferation, altered cellular adhesion, epithelial mesenchymal transition, invasion and neoplastic transformation. The mucin mediated EMT process is a very complex process. A recent histological study showed that in the bronchiolar

epithelium (an important source of MUC5B) of UIP patients there is an increased expression of *TGFβ*, *MMP-7* and *MMP-9*, which are genes known to be involved in the EMT process[164]. Another work conduct by Kurche et al., highlighted the association of *MUC5B* risk allele rs35705950 and genes related to EMT process (*collagens 1A2, 3, 4; integrins α1, α3, αv, β3, β5, β6; SPARC;* and *MMP7*) in a whole transcriptome profiling analysis[165]. The expression of MUC5B is also regulated by Sp1, a transcription factor that may regulate the expression of several genes involved in EMT including vascular endothelial growth factor (*VEGF*), type I collagen (*COL1A2*), *TGFβ* and *MMPs*[62].

Besides MUC5B, we found other up-regulated genes in FF areas of IPF compared to controls: the majorities are involved in epithelial proliferation/cancer progression and EMT transition. We can speculate that key molecules triggering fibrosis and promoting neoplastic transformation (an important complication in IPF) are secreted in FF sandwich, thus confirming that this is the most important morphological lesion of the disease.

Consequently, we can assert that FF sandwich represent the most intriguing area for a more in depth investigation concentrated on dissecting out the pathogenesis and discovering druggable molecules.

A growing literature is now developing to advance the MUC5B anti-fibrotic effects using mucolytic treatment[61] or some transcriptional factors (such as forkhead box A2 -FOXA2-, activator protein 1 transcription factor -AP-1-, Sp1, NF-κB, TTF-1, TGT3, CREB, STAT, c-Myc) and inflammatory mediators/cytokines (ROS, IL-1β, IL-6, IL-13, and IL-17A)[166-169] that modulate the expression of MUC5B.

The use of target molecules/factors may facilitate in the future the development of novel treatments for this fatal disease with limited treatment options.

Limitations and future perspectives

The first limitation of the present research is the small number of patients; however, it must be taken into consideration that OMIC approaches are usually applied in limited case series. It is well known that this high throughput technology permits to obtain a greater quantity of data but it is time consuming and very expensive. This was designed as a pilot study, useful to plan a prospective multicenter project with a higher number of patients, supported by specific grants. A wider study population will lead to a higher statistical power (including moreover an adequate number of control cases) and will allow comparisons between early and end stage disease.

Another limitation is that transcript validation focuses only on MUC5B (the most significant overexpressed). We are now planning to investigate also some significant down regulated genes.

At the end, in order to confirm MUC5B role and dissect out pathogenetic mechanisms, we are planning to use a transgenic experimental model of pulmonary fibrosis (bleomycin-induced? γ-herpesvirus-induced?).

6) REFERENCES

1. Cordier JF, Cottin V. Neglected evidence in idiopathic pulmonary fibrosis: from history to earlier diagnosis. *Eur Respir J.* 2013;42(4):916-923.

2. Corrigan DJ. On cirrhosis of the lung. *Lancet*. 1838;31(790):149-151.

3. Osler W. The principles and Practice of Medicine. *D Appleton and Company, New York.* 1892.

4. ANDER L. IDIOPATHIC INTERSTITIAL FIBROSIS OF THE LUNGS. I. PROGNOSIS AS INDICATED BY RADIOLOGICAL FINDINGS. *Acta Med Scand.* 1965;178:47-58.

5. Reid JM, Cuthbert J, Craik JE. Chronic diffuse idiopathic fibrosing alveolitis. *Br J Dis Chest.* 1965;59(4):194-201.

6. Stack BH, Grant IW, Irvine WJ, Moffat MA. Idiopathic diffuse interstitial lung disease. A review of 42 cases. *Am Rev Respir Dis.* 1965;92(6):939-948.

7. Hamman L, Rich AR. Fulminating Diffuse Interstitial Fibrosis of the Lungs. *Trans Am Clin Climatol Assoc.* 1935;51:154-163.

8. L. Hamman ARR. Acute diffuse inhterstitial fibrosis of the lungs. *Bull Johns Hopkins Hosp.* 1944(74):177-212.

9. Robbins LL. Idiophatic Pulmonary Fibrosis: roentgenologic findings. *Radiology*. 1948;51(4):459-467.

10. A.A. Liebow. New concept and entities in pulmonary disease. *The Lung. Baltimore, The Williams and Wilkins Company.* 1968:332-365.

11. Crystal RG, Fulmer JD, Roberts WC, Moss ML, Line BR, Reynolds HY. Idiopathic pulmonary fibrosis. Clinical, histologic, radiographic, physiologic, scintigraphic, cytologic, and biochemical aspects. *Ann Intern Med.* 1976;85(6):769-788.

12. Carrington CB, Gaensler EA, Coutu RE, FitzGerald MX, Gupta RG. Natural history and treated course of usual and desquamative interstitial pneumonia. *N Engl J Med.* 1978;298(15):801-809.

13. Crystal RG, Bitterman PB, Rennard SI, Hance AJ, Keogh BA. Interstitial lung diseases of unknown cause. Disorders characterized by chronic inflammation of the lower respiratory tract (first of two parts). *N Engl J Med.* 1984;310(3):154-166.

14. Katzenstein AL, Fiorelli RF. Nonspecific interstitial pneumonia/fibrosis. Histologic features and clinical significance. *Am J Surg Pathol.* 1994;18(2):136-147.

15. Nagai S, Kitaichi M, Itoh H, Nishimura K, Izumi T, Colby TV. Idiopathic nonspecific interstitial pneumonia/fibrosis: comparison with idiopathic pulmonary fibrosis and BOOP. *Eur Respir J.* 1998;12(5):1010-1019.

16. Bjoraker JA, Ryu JH, Edwin MK, et al. Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 1998;157(1):199-203.

17. Fujita J, Yamadori I, Suemitsu I, et al. Clinical features of non-specific interstitial pneumonia. *Respir Med.* 1999;93(2):113-118.

18. Cottin V, Donsbeck AV, Revel D, Loire R, Cordier JF. Nonspecific interstitial pneumonia. Individualization of a clinicopathologic entity in a series of 12 patients. *Am J Respir Crit Care Med.* 1998;158(4):1286-1293.

19. Travis WD, Hunninghake G, King TE, et al. Idiopathic nonspecific interstitial pneumonia: report of an American Thoracic Society project. *Am J Respir Crit Care Med.* 2008;177(12):1338-1347.

20. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. *N Engl J Med.* 2018;379(8):797-798.

21. Park SW, Baek AR, Lee HL, et al. Korean Guidelines for Diagnosis and Management of Interstitial Lung Diseases: Part 1. Introduction. *Tuberc Respir Dis (Seoul).* 2019.

22. Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med.* 2013;188(6):733-748.

23. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med.* 2000;161(2 Pt 1):646-664.

24. Navaratnam V, Fleming KM, West J, et al. The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax.* 2011;66(6):462-467.

25. Hutchinson J, Fogarty A, Hubbard R, McKeever T. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J.* 2015;46(3):795-806.

26. Duchemann B, Annesi-Maesano I, Jacobe de Naurois C, et al. Prevalence and incidence of interstitial lung diseases in a multi-ethnic county of Greater Paris. *Eur Respir J.* 2017;50(2).

27. Richeldi L, Rubin AS, Avdeev S, Udwadia ZF, Xu ZJ. Idiopathic pulmonary fibrosis in BRIC countries: the cases of Brazil, Russia, India, and China. *BMC Med.* 2015;13:237.

28. Raghu G, Chen SY, Yeh WS, et al. Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001-11. *Lancet Respir Med.* 2014;2(7):566-572.

29. Tsakiri KD, Cronkhite JT, Kuan PJ, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A.* 2007;104(18):7552-7557.

30. Talbert JL, Schwartz DA, Steele MP. Familial Interstitial Pneumonia (FIP). *Clin Pulm Med.* 2014;21(3):120-127.

31. Lee HL, Ryu JH, Wittmer MH, et al. Familial idiopathic pulmonary fibrosis: clinical features and outcome. *Chest.* 2005;127(6):2034-2041.

32. Wolters PJ, Blackwell TS, Eickelberg O, et al. Time for a change: is idiopathic pulmonary fibrosis still idiopathic and only fibrotic? *Lancet Respir Med.* 2018;6(2):154-160.

33. Sauleda J, Núñez B, Sala E, Soriano JB. Idiopathic Pulmonary Fibrosis: Epidemiology, Natural History, Phenotypes. *Med Sci (Basel).* 2018;6(4).

34. Raghu G, Amatto VC, Behr J, Stowasser S. Comorbidities in idiopathic pulmonary fibrosis patients: a systematic literature review. *Eur Respir J.* 2015;46(4):1113-1130.

35. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med.* 2018;198(5):e44-e68.

36. Lunardi F, Balestro E, Nannini N, Vuljan SE, Rea F, Calabrese F. Idiopathic pulmonary fibrosis: Are any of the morphological-molecular markers useful in clinical management? *Histol Histopathol.* 2017;32(7):661-672.

37. Masai K, Tsuta K, Motoi N, et al. Clinicopathological, Immunohistochemical, and Genetic Features of Primary Lung Adenocarcinoma Occurring in the Setting of Usual Interstitial Pneumonia Pattern. *J Thorac Oncol.* 2016;11(12):2141-2149.

38. Chung JH, Lynch DA. The Value of a Multidisciplinary Approach to the Diagnosis of Usual Interstitial Pneumonitis and Idiopathic Pulmonary Fibrosis: Radiology, Pathology, and Clinical Correlation. *AJR Am J Roentgenol.* 2016;206(3):463-471.

39. Lynch DA, Sverzellati N, Travis WD, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med.* 2018;6(2):138-153.

40. Kolb M, Richeldi L, Behr J, et al. Nintedanib in patients with idiopathic pulmonary fibrosis and preserved lung volume. *Thorax.* 2017;72(4):340-346.

41. Albera C, Costabel U, Fagan EA, et al. Efficacy of pirfenidone in patients with idiopathic pulmonary fibrosis with more preserved lung function. *Eur Respir J.* 2016;48(3):843-851.

42. Martinez FJ, Safrin S, Weycker D, et al. The clinical course of patients with idiopathic pulmonary fibrosis. *Ann Intern Med.* 2005;142(12 Pt 1):963-967.

43. Selman M, Carrillo G, Estrada A, et al. Accelerated variant of idiopathic pulmonary fibrosis: clinical behavior and gene expression pattern. *PLoS One.* 2007;2(5):e482.

44. Balestro E, Calabrese F, Turato G, et al. Immune Inflammation and Disease Progression in Idiopathic Pulmonary Fibrosis. *PLoS One.* 2016;11(5):e0154516.

45. Shaw TJ, Martin P. Wound repair at a glance. *J Cell Sci.* 2009;122(Pt 18):3209-3213.

46. West JB, Mathieu-Costello O. Structure, strength, failure, and remodeling of the pulmonary blood-gas barrier. *Annu Rev Physiol.* 1999;61:543-572.

47. Selman M, King TE, Pardo A, Society AT, Society ER, Physicians ACoC. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Ann Intern Med.* 2001;134(2):136-151.

48. Martinez FJ, Collard HR, Pardo A, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers.* 2017;3:17074.

49. Arish N, Petukhov D, Wallach-Dayan SB. The Role of Telomerase and Telomeres in Interstitial Lung Diseases: From Molecules to Clinical Implications. *Int J Mol Sci.* 2019;20(12).

50. Kropski JA, Mitchell DB, Markin C, et al. A novel dyskerin (DKC1) mutation is associated with familial interstitial pneumonia. *Chest.* 2014;146(1):e1-e7.

51. Alder JK, Stanley SE, Wagner CL, Hamilton M, Hanumanthu VS, Armanios M. Exome sequencing identifies mutant TINF2 in a family with pulmonary fibrosis. *Chest.* 2015;147(5):1361-1368.

52. Stanley SE, Gable DL, Wagner CL, et al. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. *Sci Transl Med.* 2016;8(351):351ra107.

53. Newton CA, Batra K, Torrealba J, et al. Telomere-related lung fibrosis is diagnostically heterogeneous but uniformly progressive. *Eur Respir J.* 2016;48(6):1710-1720.

54. Borie R, Bouvry D, Cottin V, et al. Regulator of telomere length 1 (*Eur Respir J.* 2019;53(2).

55. Snetselaar R, van Batenburg AA, van Oosterhout MFM, et al. Short telomere length in IPF lung associates with fibrotic lesions and predicts survival. *PLoS One.* 2017;12(12):e0189467.

56. Povedano JM, Martinez P, Serrano R, et al. Therapeutic effects of telomerase in mice with pulmonary fibrosis induced by damage to the lungs and short telomeres. *Elife.* 2018;7.

57. McDonough JE, Martens DS, Tanabe N, et al. A role for telomere length and chromosomal damage in idiopathic pulmonary fibrosis. *Respir Res.* 2018;19(1):132.

58. Mulugeta S, Nureki S, Beers MF. Lost after translation: insights from pulmonary surfactant for understanding the role of alveolar epithelial dysfunction and cellular quality control in fibrotic lung disease. *Am J Physiol Lung Cell Mol Physiol.* 2015;309(6):L507-525.

59. Lawson WE, Grant SW, Ambrosini V, et al. Genetic mutations in surfactant protein C are a rare cause of sporadic cases of IPF. *Thorax.* 2004;59(11):977-980.

60. Markart P, Ruppert C, Wygrecka M, et al. Surfactant protein C mutations in sporadic forms of idiopathic interstitial pneumonias. *Eur Respir J.* 2007;29(1):134-137.

61. Hancock LA, Hennessy CE, Solomon GM, et al. Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. *Nat Commun.* 2018;9(1):5363.

62. Zhang Q, Wang Y, Qu D, Yu J, Yang J. The Possible Pathogenesis of Idiopathic Pulmonary Fibrosis considering MUC5B. *Biomed Res Int.* 2019;2019:9712464.

63. Seibold MA, Wise AL, Speer MC, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med.* 2011;364(16):1503-1512.

64. Yang IV, Fingerlin TE, Evans CM, Schwarz MI, Schwartz DA. MUC5B and Idiopathic Pulmonary Fibrosis. *Ann Am Thorac Soc.* 2015;12 Suppl 2:S193-199.

65. Ponnusamy MP, Seshacharyulu P, Lakshmanan I, Vaz AP, Chugh S, Batra SK. Emerging role of mucins in epithelial to mesenchymal transition. *Curr Cancer Drug Targets.* 2013;13(9):945-956.

66. Taskar VS, Coultas DB. Is idiopathic pulmonary fibrosis an environmental disease? *Proc Am Thorac Soc.* 2006;3(4):293-298.

67. Tsukamoto K, Hayakawa H, Sato A, Chida K, Nakamura H, Miura K. Involvement of Epstein-Barr virus latent membrane protein 1 in disease progression in patients with idiopathic pulmonary fibrosis. *Thorax.* 2000;55(11):958-961.

68. Tang YW, Johnson JE, Browning PJ, et al. Herpesvirus DNA is consistently detected in lungs of patients with idiopathic pulmonary fibrosis. *J Clin Microbiol.* 2003;41(6):2633-2640.

69. Han MK, Zhou Y, Murray S, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med.* 2014;2(7):548-556.

70. Molyneaux PL, Cox MJ, Willis-Owen SA, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2014;190(8):906-913.

71. Wilson EC, Shulgina L, Cahn AP, et al. Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole: an economic evaluation alongside a randomised controlled trial. *Pharmacoeconomics.* 2014;32(1):87-99.

72. Raghu G, Freudenberger TD, Yang S, et al. High prevalence of abnormal acid gastro-esophageal reflux in idiopathic pulmonary fibrosis. *Eur Respir J.* 2006;27(1):136-142.

73. Budinger GRS, Kohanski RA, Gan W, et al. The Intersection of Aging Biology and the Pathobiology of Lung Diseases: A Joint NHLBI/NIA Workshop. *J Gerontol A Biol Sci Med Sci.* 2017;72(11):1492-1500.

74. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell.* 2013;153(6):1194-1217.

75. Buendía-Roldán I, Mejía M, Navarro C, Selman M. Idiopathic pulmonary fibrosis: Clinical behavior and aging associated comorbidities. *Respir Med.* 2017;129:46-52.

76. Bueno M, Lai YC, Romero Y, et al. PINK1 deficiency impairs mitochondrial homeostasis and promotes lung fibrosis. *J Clin Invest.* 2015;125(2):521-538.

77. Lehmann M, Korfei M, Mutze K, et al. Senolytic drugs target alveolar epithelial cell function and attenuate experimental lung fibrosis. *Eur Respir J.* 2017;50(2).

78. Hecker L, Logsdon NJ, Kurundkar D, et al. Reversal of persistent fibrosis in aging by targeting Nox4-Nrf2 redox imbalance. *Sci Transl Med.* 2014;6(231):231ra247.

79. Schafer MJ, White TA, Iijima K, et al. Cellular senescence mediates fibrotic pulmonary disease. *Nat Commun.* 2017;8:14532.

80. Huang WT, Akhter H, Jiang C, et al. Plasminogen activator inhibitor 1, fibroblast apoptosis resistance, and aging-related susceptibility to lung fibrosis. *Exp Gerontol.* 2015;61:62-75.

81. Yang IV, Schwartz DA. Epigenetics of idiopathic pulmonary fibrosis. *Transl Res.* 2015;165(1):48-60.

82. Dakhlallah D, Batte K, Wang Y, et al. Epigenetic regulation of miR-17~92 contributes to the pathogenesis of pulmonary fibrosis. *Am J Respir Crit Care Med.* 2013;187(4):397-405.

83. Issa JP. Aging and epigenetic drift: a vicious cycle. *J Clin Invest.* 2014;124(1):24-29.

84. Yang IV, Pedersen BS, Rabinovich E, et al. Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2014;190(11):1263-1272.

85. Sgalla G, Iovene B, Calvello M, Ori M, Varone F, Richeldi L. Idiopathic pulmonary fibrosis: pathogenesis and management. *Respir Res.* 2018;19(1):32.

86. Korfei M, Skwarna S, Henneke I, et al. Aberrant expression and activity of histone deacetylases in sporadic idiopathic pulmonary fibrosis. *Thorax.* 2015;70(11):1022-1032.

87. Liu F, Killian JK, Yang M, et al. Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene.* 2010;29(25):3650-3664.

88. Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. *Annu Rev Pathol.* 2014;9:157-179.

89. Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *J Cell Biol.* 2006;172(7):973-981.

90. Morbini P, Inghilleri S, Campo I, Oggionni T, Zorzetto M, Luisetti M. Incomplete expression of epithelial-mesenchymal transition markers in idiopathic pulmonary fibrosis. *Pathol Res Pract.* 2011;207(9):559-567.

91. Heukels P, van Hulst JAC, van Nimwegen M, et al. Fibrocytes are increased in lung and peripheral blood of patients with idiopathic pulmonary fibrosis. *Respir Res.* 2018;19(1):90.

92. Moeller A, Gilpin SE, Ask K, et al. Circulating fibrocytes are an indicator of poor prognosis in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2009;179(7):588-594.

93. King TE, Schwarz MI, Brown K, et al. Idiopathic pulmonary fibrosis: relationship between histopathologic features and mortality. *Am J Respir Crit Care Med.* 2001;164(6):1025-1032.

94. Dowman L, McDonald CF, Hill C, et al. The benefits of exercise training in interstitial lung disease: protocol for a multicentre randomised controlled trial. *BMC Pulm Med.* 2013;13:8.

95. Laporta Hernandez R, Aguilar Perez M, Lázaro Carrasco MT, Ussetti Gil P. Lung Transplantation in Idiopathic Pulmonary Fibrosis. *Med Sci (Basel).* 2018;6(3).

96. Valapour M, Lehr CJ, Skeans MA, et al. OPTN/SRTR 2017 Annual Data Report: Lung. *Am J Transplant.* 2019;19 Suppl 2:404-484.

97. Raghu G, Rochwerg B, Zhang Y, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *Am J Respir Crit Care Med.* 2015;192(2):e3-19.

98. Lunardi F, Pezzuto F, Vuljan SE, Calabrese F. Idiopathic Pulmonary Fibrosis and Antifibrotic Treatments: Focus on Experimental Studies. *Arch Pathol Lab Med.* 2018;142(9):1090-1097.

99. Azuma A, Nukiwa T, Tsuboi E, et al. Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2005;171(9):1040-1047.

100. King TE, Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med.* 2014;370(22):2083-2092.

101. Taniguchi H, Ebina M, Kondoh Y, et al. Pirfenidone in idiopathic pulmonary fibrosis. *Eur Respir J.* 2010;35(4):821-829.

102. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet.* 2011;377(9779):1760-1769.

103. Canestaro WJ, Forrester SH, Raghu G, Ho L, Devine BE. Drug Treatment of Idiopathic Pulmonary Fibrosis: Systematic Review and Network Meta-Analysis. *Chest.* 2016;149(3):756-766.

104. Rochwerg B, Neupane B, Zhang Y, et al. Treatment of idiopathic pulmonary fibrosis: a network meta-analysis. *BMC Med.* 2016;14:18.

105. Vancheri C, Kreuter M, Richeldi L, et al. Nintedanib with Add-on Pirfenidone in Idiopathic Pulmonary Fibrosis. Results of the INJOURNEY Trial. *Am J Respir Crit Care Med.* 2018;197(3):356-363.

106. Kaminski N, Allard JD, Pittet JF, et al. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc Natl Acad Sci U S A.* 2000;97(4):1778-1783.

107. Zuo F, Kaminski N, Eugui E, et al. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A.* 2002;99(9):6292-6297.

108. Selman M, Pardo A, Barrera L, et al. Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *Am J Respir Crit Care Med.* 2006;173(2):188-198.

109. Wang XM, Zhang Y, Kim HP, et al. Caveolin-1: a critical regulator of lung fibrosis in idiopathic pulmonary fibrosis. *J Exp Med.* 2006;203(13):2895-2906.

110. Yang IV, Burch LH, Steele MP, et al. Gene expression profiling of familial and sporadic interstitial pneumonia. *Am J Respir Crit Care Med.* 2007;175(1):45-54.

111. Boon K, Bailey NW, Yang J, et al. Molecular phenotypes distinguish patients with relatively stable from progressive idiopathic pulmonary fibrosis (IPF). *PLoS One.* 2009;4(4):e5134.

112. Konishi K, Gibson KF, Lindell KO, et al. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2009;180(2):167-175.

113. Vukmirovic M, Herazo-Maya JD, Blackmon J, et al. Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis. *BMC Pulm Med.* 2017;17(1):15.

114. Gangwar I, Kumar Sharma N, Panzade G, Awasthi S, Agrawal A, Shankar R. Detecting the Molecular System Signatures of Idiopathic Pulmonary Fibrosis through Integrated Genomic Analysis. *Sci Rep.* 2017;7(1):1554.

115. Sivakumar P, Thompson JR, Ammar R, et al. RNA sequencing of transplant-stage idiopathic pulmonary fibrosis lung reveals unique pathway regulation. *ERJ Open Res.* 2019;5(3).

116. Spek CA, Duitman J. Is idiopathic pulmonary fibrosis a cancer-like disease? Transcriptome analysis to fuel the debate. *ERJ Open Res.* 2019;5(1).

117. Lee JU, Cheong HS, Shim EY, et al. Gene profile of fibroblasts identify relation of CCL8 with idiopathic pulmonary fibrosis. *Respir Res.* 2017;18(1):3.

118. Sheu CC, Chang WA, Tsai MJ, Liao SH, Chong IW, Kuo PL. Bioinformatic analysis of next‐generation sequencing data to identify dysregulated genes in fibroblasts of idiopathic pulmonary fibrosis. *Int J Mol Med.* 2019;43(4):1643-1656.

119. Emblom-Callahan MC, Chhina MK, Shlobin OA, et al. Genomic phenotype of non-cultured pulmonary fibroblasts in idiopathic pulmonary fibrosis. *Genomics.* 2010;96(3):134-145.

120. Mullenbrock S, Liu F, Szak S, et al. Systems Analysis of Transcriptomic and Proteomic Profiles Identifies Novel Regulation of Fibrotic Programs by miRNAs in Pulmonary Fibrosis Fibroblasts. *Genes (Basel).* 2018;9(12).

121. Reyfman PA, Walter JM, Joshi N, et al. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med.* 2019;199(12):1517-1536.

122. Xu Y, Mizuno T, Sridharan A, et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight.* 2016;1(20):e90558.

123. E. Valenzi TT, C. Morse, M. Rojas, A.L. Mora, R. Lafyatis. Single cell RNA-Seq analysis of idiopathic pulmonary fibrosis reveals alterations in lung cell populations and in selective cell expression of biomarkers MMP7, CCL18 and MUC1 *American Journal of Respiratory and Critical Care Medicine.* 2018;197(A4356).

124. Nicholson AG, Fulford LG, Colby TV, du Bois RM, Hansell DM, Wells AU. The relationship between individual histologic features and disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med.* 2002;166(2):173-177.

125. Harada T, Watanabe K, Nabeshima K, Hamasaki M, Iwasaki H. Prognostic significance of fibroblastic foci in usual interstitial pneumonia and non-specific interstitial pneumonia. *Respirology.* 2013;18(2):278-283.

126. Walsh SL, Wells AU, Sverzellati N, et al. Relationship between fibroblastic foci profusion and high resolution CT morphology in fibrotic lung disease. *BMC Med.* 2015;13:241.

127. Calabrese F, Lunardi F, Giacometti C, et al. Overexpression of squamous cell carcinoma antigen in idiopathic pulmonary fibrosis: clinicopathological correlations. *Thorax.* 2008;63(9):795-802.

128. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med.* 2011;183(6):788-824.

129. Burgemeister R. Nucleic acids extraction from laser microdissected FFPE tissue sections. *Methods Mol Biol.* 2011;724:117-129.

130. McCombie WR, McPherson JD, Mardis ER. Next-Generation Sequencing Technologies. *Cold Spring Harb Perspect Med.* 2018.

131. Wood DL, Nones K, Steptoe A, et al. Recommendations for Accurate Resolution of Gene and Isoform Allele-Specific Expression in RNA-Seq Data. *PLoS One.* 2015;10(5):e0126911.

132. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet.* 2019.

133. Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One.* 2017;12(12):e0190152.

134. Lukyanov K, Diatchenko L, Chenchik A, et al. Construction of cDNA libraries from small amounts of total RNA using the suppression PCR effect. *Biochem Biophys Res Commun.* 1997;230(2):285-288.

135. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25-29.

136. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc.* 2015;10(6):823-844.

137. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* 2019;47(W1):W234-W241.

138. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457-462.

139. Jones MG, Fabre A, Schneider P, et al. Three-dimensional characterization of fibroblast foci in idiopathic pulmonary fibrosis. *JCI Insight.* 2016;1(5).

140. Tarazona S, Furió-Tarí P, Turrà D, et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* 2015;43(21):e140.

141. Li W, Zheng H, Qin H, et al. Exploration of differentially expressed plasma proteins in patients with lung adenocarcinoma using iTRAQ-coupled 2D LC-MS/MS. *Clin Respir J.* 2018;12(6):2036-2045.

142. Shi H, Gu Y, Yang J, Xu L, Mi W, Yu W. Lipocalin 2 promotes lung metastasis of murine breast cancer cells. *J Exp Clin Cancer Res.* 2008;27:83.

143. Leng X, Lin H, Ding T, et al. Lipocalin 2 is required for BCR-ABL-induced tumorigenesis. *Oncogene.* 2008;27(47):6110-6119.

144. Hu C, Yang K, Li M, Huang W, Zhang F, Wang H. Lipocalin 2: a potential therapeutic target for breast cancer metastasis. *Onco Targets Ther.* 2018;11:8099-8106.

145. Song B, Zhang H, Jiang L, et al. Down-regulation of lipocalin 2 suppresses the growth of human lung adenocarcinoma through oxidative stress involving Nrf2/HO-1 signaling. *Acta Biochim Biophys Sin (Shanghai).* 2015;47(10):805-814.

146. Sun P, Ye R, Wang C, Bai S, Zhao L. Identification of proteomic signatures associated with COPD frequent exacerbators. *Life Sci.* 2019;230:1-9.

147. DePianto DJ, Chandriani S, Abbas AR, et al. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. *Thorax.* 2015;70(1):48-56.

148. Seibold MA, Smith RW, Urbanek C, et al. The idiopathic pulmonary fibrosis honeycomb cyst contains a mucocilary pseudostratified epithelium. *PLoS One.* 2013;8(3):e58658.

149. Kim S, Wong P, Coulombe PA. A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. *Nature.* 2006;441(7091):362-365.

150. Rizeq B, Zakaria Z, Ouhtit A. Towards understanding the mechanisms of actions of carcinoembryonic antigen-related cell adhesion molecule 6 in cancer progression. *Cancer Sci.* 2018;109(1):33-42.

151. Son SM, Yun J, Lee SH, et al. Therapeutic Effect of pHLIP-mediated CEACAM6 Gene Silencing in Lung Adenocarcinoma. *Sci Rep.* 2019;9(1):11607.

152. Citrin DE, Prasanna PGS, Walker AJ, et al. Radiation-Induced Fibrosis: Mechanisms and Opportunities to Mitigate. Report of an NCI Workshop, September 19, 2016. *Radiat Res.* 2017;188(1):1-20.

153. Jiang C, Liu G, Luckhardt T, et al. Serpine 1 induces alveolar type II cell senescence through activating p53-p21-Rb pathway in fibrotic lung disease. *Aging Cell.* 2017;16(5):1114-1124.

154. Hecker L, Wang T, Garcia J. Nampt Mediates Age-Associated Pro-Fibrotic Myofibroblast Phenotypes and Persistent Fibrosis. In. Vol 195:A2656. Am J Respir Crit Care Med ed2017.

155. Chilosi M, Zamò A, Doglioni C, et al. Migratory marker expression in fibroblast foci of idiopathic pulmonary fibrosis. *Respir Res.* 2006;7:95.

156. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med.* 2014;370(22):2071-2082.

157. King TE, Pardo A, Selman M. Idiopathic pulmonary fibrosis. *Lancet.* 2011;378(9807):1949-1961.

158. Datta S, Malhotra L, Dickerson R, Chaffee S, Sen CK, Roy S. Laser capture microdissection: Big data from small samples. *Histol Histopathol.* 2015;30(11):1255-1269.

159. Butler AE, Matveyenko AV, Kirakossian D, Park J, Gurlo T, Butler PC. Recovery of high-quality RNA from laser capture microdissected human and rodent pancreas. *J Histotechnol.* 2016;39(2):59-65.

160. Fahy JV, Dickey BF. Airway mucus function and dysfunction. *N Engl J Med.* 2010;363(23):2233-2247.

161. Nakano Y, Yang IV, Walts AD, et al. MUC5B Promoter Variant rs35705950 Affects MUC5B Expression in the Distal Airways in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med.* 2016;193(4):464-466.

162. Chen G, Ribeiro CMP, Sun L, et al. XBP1S Regulates MUC5B in a Promoter Variant-Dependent Pathway in Idiopathic Pulmonary Fibrosis Airway Epithelia. *Am J Respir Crit Care Med.* 2019;200(2):220-234.

163. Liptzin DR, Watson AM, Murphy E, et al. MUC5B expression and location in surfactant protein C mutations in children. *Pediatr Pulmonol.* 2015;50(12):1270-1276.

164. Figueira de Mello GC, Ribeiro Carvalho CR, Adib Kairalla R, et al. Small airway remodeling in idiopathic interstitial pneumonias: a pathological study. *Respiration*. 2010;79(4):322-32.

165. Kurche JS, Huber J, Dobrinskikh E, et al. Pulmonary Fibrosis MUC5B Risk Allele rs35705950 Is Associated with Transcriptional Profiles of Senescence and Epithelial-to-Mesenchymal Transition in Airway Cells Cultured at Air-Liquid Interface. *Am J Respir Crit Care Med.* 2018;197:A3836.

166. Song S, Jung E, Bae c, et al. Visfatin induces MUC8 and MUC5B expression via p38 MAPK/ROS/NF-κB in human airway epithelial cells. *J Biomed Sci*. 2014;20,21:49.

167. Wynn TA. IL-13 effector functions. *Annu Rev Immunol*. 2003;21:425-56.

168. Hao Y, Kuang Z, Jing J et al. Mycoplasma pneumoniae modulates STAT3-STAT6/EGFR-FOXA2 signaling to induce overexpression of airway mucins. *Infect Immun.* 2014;82(12):5246-55.

169. Zhert G, Sung WP, Nguyenvu LT et al. IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production. *Am J Respir Cell Mol Biol.* 2007;36(2):244-53.