Aus dem Institut für Experimentelle Pneumologie
Klinik der Universität München
Ehemaliger Direktor: Prof. Dr. Oliver Eickelberg

und der Abteilung Lung Repair and Regeneration (LRR) des
Helmholtz Zentrums München
Prof. Dr. Dr. Melanie Königshoff

Oncogenic Signaling in Idiopathic Pulmonary Fibrosis

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Henrik Marcel Ulke

aus Starnberg

2023

Mit Genehmigung der Medizinischen Fakultät der Universität München

Berichterstatterin:	Prof. Dr. Dr. Melanie Königshoff
Mitberichterstatter:	PD Dr. Sandra Frank Prof. Dr. Jürgen Behr Prof. Dr. Markus Ege
Mitbetreuung durch die promovierte Mitarbeiterin:	Dr. Kathrin Mutze
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	20.07.2023

Eidesstattliche Versicherung

Ulke, Henrik Marcel

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Titel

Oncogenic Signaling in Idiopathic Pulmonary Fibrosis

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Winterthur, 07.08.2023	Henrik Marcel Ulke
Ort, Datum	Unterschrift Doktorand

Für meine Familie und Freunde.

Contents

List of abbreviations

A

A	adenosine
AC	adenocarcinoma
ALAT	Latin American thoracic society
ANLN	anillin actin binding protein
APS	ammonium peroxodisulfate
ATII	alveolar epithelial type II
ATF4	activating transcription factor 4
ATS	American thoracic society
α-SMA	alpha-smooth muscle actin

B

BEBM	bronchial epithelial basal medium
Bleo	bleomycin sulfate
bp	base pairs
BSA	bovine serum albumin
BUB1	BUB1 mitotic checkpoint serine/threonine kinase
BW	body weight

C

°C	degree Celsius
C	cytosine
cDNA	complementary DNA
cm	centimeter
CNN1	calponin 1
CO_2	carbon dioxide
COL1A1	collagen type 1 alpha 1
COPD	chronic obstructive pulmonary disease

CT	computed tomography
C_t value	cycle threshold value
CTHRC1	collagen triple helix repeat containing 1
CTRL	control

D

d	day
Da	dalton
DEGs	differentially expressed genes
DLCO	diffusing capacity of the lung for carbon monoxide
DMEM	Dulbecco's modified Eagle medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DPBS	Dulbecco's phosphate-buffered saline
DPLD	diffuse parenchymal lung disease
DTT	dithiothreitol

E

ECM	extracellular matrix
ECT2	epithelial cell transforming sequence 2
EDTA	ethylenediaminetetraacetic acid
ENA	European nucleotide archive
EpCAM	epithelial cell adhesion molecule
ERS	European respiratory society
ES	enrichment score

F

FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
FDR	false discovery rate
FEV1	forced expiratory volume in the first second
FVC	forced vital capacity

FWER	familywise-error rate

G

G	gauge; guanosine
g	gram; gravity
GEO	gene expression omnibus
GO	gene ontology
GPR87	G protein-coupled receptor 87
GSEA	gene set enrichment analysis

H

h	hour
H&E	hematoxylin and eosin
H_2O	water
HBEC	human bronchial epithelial cell
HBSS	Hank's balanced salt solution
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGNC	HUGO gene nomenclature committee
HPRT	hypoxanthine-guanine phosphoribosyltransferase
HRCT	high-resolution computed tomography
HRP	horseradish peroxidase

I

i.p.	intraperitoneal
IgG	immunoglobulin G
IIP	idiopathic interstitial pneumonia
ILD	interstitial lung disease
IPF	idiopathic pulmonary fibrosis
IT	information technology

J

JRS	Japanese respiratory society

K

k	kilo
KCl	potassium chloride
kDa	kilodalton
KEGG	Kyoto encyclopedia of genes and genomes
kg	kilogram
KH_2PO_4	potassium dihydrogen phosphate
KRT14	cytokeratin 14
KRT5	cytokeratin 5
KRT6	cytokeratin 6

L

l	liter
LE	leading-edge
Log2	binary logarithm
LPA	lysophosphatidic acid

M

m	milli; meter
M	molar mass
mA	milliampere
MEM	minimum essential media
mg	milligram
$MgCl_2$	magnesium chloride
min	minute
ml	milliliter
mm	millimeter
mM	millimolar
MMF	medetomidine, midazolam, fentanyl
MRI	magnetic resonance imaging

mRNA	messenger ribonucleic acid
MUC5B	mucin 5B
μ	micro
μg	microgram
μl	microliter
μm	micrometer
μM	micromolar

N

n	nano
Na_2HPO_4	disodium hydrogen phosphate
NaCl	sodium chloride
NCBI	national center for biotechnology information
NES	normalized enrichment score
ng	nanogram
nl	nanoliter
nm	nanometer
nM	nanomolar
NRF2	nuclear factor erythroid 2-related factor 2
NSCLC	non-small cell lung cancer
NTC	no template control

O

P

p	passage
p.a.	*pro analysis*
Page	polyacrylamide gel electrophoresis
PAI1	plasminogen activator inhibitor 1
PBS	phosphate-buffered saline
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction

PET	positron-emission tomography
PFA	paraformaldehyde
pH	power of hydrogen
phBEC	primary human bronchial epithelial cell
PHGDH	phosphoglycerate dehydrogenase
pmATII	primary murine alveolar epithelial type II
PPI	protein-protein interaction
PSAT1	phosphoserine aminotransferase 1
PSPH	phosphoserine phosphatase

Q

qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative real-time polymerase chain reaction

R

%rH	percent relative humidity
RCF	relative centrifugal force
rel.	relative
RNA	ribonucleic acid
RNA-Seq	RNA sequencing
rpm	revolutions per minute

S

s.c.	subcutaneous
SCC	squamous cell carcinoma
SD	standard deviation
SDS	sodium dodecyl sulfate
sec	second
STAT3	signal transducer and activator of transcription 3

T

T	thymidine

T-PER	tissue protein extraction reagent
TBS	tris-buffered saline
TBST	tris-buffered saline with Tween 20
TEMED	N,N,N',N'-tetramethylethylenediamine
TGF-β	transforming growth factor beta 1
TNF-α	tumor necrosis factor-alpha
Tris	tris-(hydroxymethyl)-aminomethane
TTF-1	thyroid transcription factor 1

U

U	units; uridine
UIP	usual interstitial pneumonia
USA	United States of America

V

V	volt
v	volume

W

w	weight
WISP1	WNT1-inducible signaling pathway protein 1

X

Y

Z

Human gene symbols are *italicized* and spelled with all capital letters.

Mouse gene symbols are *italicized*, with only the first letter being capitalized.

Human and mouse protein symbols are spelled with all capital letters and not italicized.

Summary

Idiopathic pulmonary fibrosis (IPF) and non-small cell lung cancer (NSCLC) are two devastating pulmonary disorders, which are marked by excessive tissue production, irreversible damage to lung structure, and eventually loss of pulmonary function. The natural history of both diseases is characterized by chronic progression and represents a fatal prognosis for the affected patients. Treatments are able to slow down the diseases' progression, but to date no approved, curative treatment option exists. IPF is frequently associated with lung cancer and both diseases' pathogenesis share common hallmarks, such as altered cellular phenotypes, misregulated biological pathways and mediators, and similar genetic changes. We aimed to establish a link between the pathomechanisms of IPF and NSCLC by analyzing patterns of gene expression alterations and to further characterize the role of candidate genes in the pathogenesis of IPF.

NSCLC microarray datasets (*GSE44077*, *GSE43458*, *GSE18842*) from Gene Expression Omnibus (GEO) were analyzed and differentially expressed genes (DEGs) were extracted. Gene set enrichment analysis (GSEA) was used to determine the enrichment of DEGs from NSCLC in an IPF microarray dataset (*GSE47460*) and to describe the subsequent list of candidate genes. Further characterization of these genes of interest was achieved by annotation enrichment analysis, protein-protein interaction networks, BioGPS, and principal component analysis. The final candidate genes were verified by quantitative real-time polymerase chain reaction (qRT-PCR) as well as western blot analysis in human and mouse lung tissue samples, human bronchial epithelial cells, and primary murine alveolar epithelial type II (pmATII) cells. Finally, treatment of bronchial epithelial cells with pro-fibrotic transforming growth factor beta 1 (TGF-β) was performed and the expression of the candidate genes was analyzed.

IPF and NSCLC showed a significant pattern of shared gene expression alterations in the GSEA. Further analysis revealed a common set of 92 equally misregulated genes in IPF and NSCLC (log2 fold change > 1; adjusted p-value < 0.05), which demonstrated an IPF-specific signature in the principal component analysis. Annotation enrichment analysis of this gene set highlighted common themes, such as P53 regulation, extracellular matrix (ECM) organization, cell cycle, and proliferation. Western blot and qRT-PCR validated a significantly increased expression of the two candidate genes G protein-coupled receptor 87 (GPR87) and phosphoserine aminotransferase 1 (PSAT1) in NSCLC, IPF, and bleomycin-induced lung fibrosis in mice. TGF-β treatment of bronchial epithelial cells resulted in a significant upregulation of GPR87 *in vitro*.

In summary, we demonstrated a pathogenic link between IPF and NSCLC, which resulted in a subset of potential novel therapeutic targets. Further analysis of GPR87 and the other candidate genes might improve our understanding of IPF and enable novel therapeutic strategies.

Zusammenfassung

Die idiopathische Lungenfibrose (IPF) und das nichtkleinzellige Lungenkarzinom (NSCLC) sind verheerende Erkrankungen, die durch überschießende Gewebeproduktion, irreversible Lungenstrukturschäden und den Verlust der Lungenfunktion gekennzeichnet sind. Der natürliche Krankheitsverlauf ist chronisch progredient und stellt für betroffene Patienten eine tödliche Prognose dar. Eine Therapie kann das Fortschreiten dieser Erkrankungen verlangsamen, bis heute existiert jedoch keine zugelassene, kurative Behandlungsmöglichkeit. Die Pathogenese beider Erkrankungen weist gemeinsame Merkmale auf, wie den Wechsel der zellulären Phänotypen, dysregulierte Signalwege und Mediatoren sowie übereinstimmende genetische Veränderungen. Unser Ziel war, die Pathomechanismen von IPF und NSCLC zu verknüpfen, indem Veränderungen der Genexpression analysiert und die Rolle von spezifischen Genen in der Pathogenese von IPF charakterisiert wurden.

Mikroarray Datensätze zu NSCLC (*GSE44077*, *GSE43458*, *GSE18842*) von Gene Expression Omnibus (GEO) wurden analysiert und differentiell exprimierte Gene (DEGs) identifiziert. Mittels Gene Set Enrichment Analyse (GSEA) wurde die Akkumulation der DEGs aus NSCLC in einem Mikroarray Datensatz zu IPF (*GSE47460*) bestimmt und eine Liste spezifischer Gene erstellt, welche durch Annotation Enrichment Analyse, Protein-Protein-Interaktionsnetzwerke, BioGPS und Hauptkomponentenanalyse weiter charakterisiert wurde. Diese Gene wurden mittels quantitativer Echtzeit-Polymerase-Kettenreaktion (qRT-PCR) und Western Blot in Lungengewebeproben von Mensch und Maus, humanen Bronchialepithelzellen und primären murinen Alveolarepithelzellen Typ II (pmATII) verifiziert. Bronchialepithelzellen wurden mit pro-fibrotischem Transforming Growth Factor Beta 1 (TGF-β) behandelt und die Genexpression wurde analysiert.

IPF und NSCLC zeigten in der GSEA signifikant übereinstimmende Veränderungen der Genexpression. Die weiterführenden Analysen ergaben eine gemeinsame Gruppe von 92 gleichermaßen dysregulierten Genen in IPF und NSCLC (log2 fold change $>$ 1; p-Wert $<$ 0,05), welche in der Hauptkomponentenanalyse eine IPF-spezifische Signatur aufwiesen. Die Annotation Enrichment Analyse dieser Gruppe von Genen ergab übereinstimmende Aspekte, wie die Regulierung von P53, die Organisation der extrazellulären Matrix, den Zellzyklus und die Proliferation. Western Blot und qRT-PCR bestätigten eine signifikant erhöhte Expression des G Protein-Coupled Receptor 87 (GPR87) und der Phosphoserine Aminotransferase 1 (PSAT1) in NSCLC, IPF und Bleomycin-induzierter Lungenfibrose bei Mäusen. Die Behandlung von Bronchialepithelzellen mit TGF-β führte zu einer signifikanten Erhöhung von GPR87 *in vitro*.

Es konnte ein Zusammenhang der Pathogenesen von IPF und NSCLC gezeigt werden, was zur Generierung einer Gruppe potentieller therapeutischer Zielgene geführt hat. Weitere Analysen zu GPR87 und den restlichen Genen könnten unser Verständnis von IPF verbessern und neue therapeutische Strategien ermöglichen.

1 Introduction

1.1 Idiopathic pulmonary fibrosis

Chronic diseases of the respiratory system represent an enormous burden for the global health care, being one of the most common causes of mortality worldwide [1]. The expenses of personal health care, which were associated with chronic diseases of the respiratory system, were estimated at $132.1 billion in the United States of America (USA) in 2013 [2]. Moreover, these expenses have been subject to an average annual increase of 3.7% since 1996 [2], which also coincides with the significant increase of mortality related to chronic lung diseases in the USA from 1980 until 2014 [3].

Besides chronic obstructive pulmonary disease (COPD) and asthma, interstitial lung diseases (ILDs) are also part of the group of chronic respiratory diseases. ILD, also known as diffuse parenchymal lung disease (DPLD), is used as a generic term for at least 150 to 200 individual diseases [4, 5]. In 2013, the ILD related expenses of personal health care were estimated at $10.9 billion in the USA [2]. Apart from this tremendous economic impact, ILDs inflict an enormous burden on the affected patients due to their chronic and sometimes fatal progression combined with severe symptoms, which frequently results in comorbid depression [6] and a progressive reduction in quality of life [7].

Idiopathic pulmonary fibrosis (IPF) is one of the severe, irreversible ILDs and belongs to the subgroup of idiopathic interstitial pneumonias (IIPs), which are subdivided into major IIPs (acute interstitial pneumonia, cryptogenic organizing pneumonia, desquamative interstitial pneumonia, idiopathic nonspecific interstitial pneumonia, *idiopathic pulmonary fibrosis*, respiratory bronchiolitis-interstitial lung disease), rare IIPs (idiopathic lymphoid interstitial pneumonia, idiopathic pleuroparenchymal fibroelastosis), and unclassifiable IIPs, according to the latest American Thoracic Society (ATS) and European Respiratory Society (ERS) report [8]. In this context, IPF is considered the most frequent form of IIPs [9], with its annual incidence being assessed at 2.8 to 9.3 cases per 100,000 [10]. The prevalence of IPF has been reported for various regions and populations, which was summarized by Martinez *et al.* [11], indicating a prevalence between 10 to 60 cases per 100,000 [11].

1.1.1 Clinical presentation of IPF

Idiopathic pulmonary fibrosis represents a severe disorder of the human lung due to its irreversible and chronically progressing nature, which is marked by the advancing formation of abundant lung tissue and the resulting deterioration of pulmonary function.

While the pathogenesis of IPF still remains subject to further research, it is commonly accepted that age is related to the onset of IPF, which mostly affects the elderly popu-

lation aged 60 years and above [12]. In accordance with this, Raghu *et al.* [13] reported a prevalence of 494.5 cases per 100,000 among adults older than 65 years in the USA in 2011, which was significantly higher compared to the IPF prevalence among the general population. Besides the old age [12], multiple specific risk factors have been linked to the development of IPF. The most frequent include tobacco smoking [14, 15], male sex [9, 11, 12], work-related noxious substances [16], and several genetic alterations (e.g. mucin 5B (*MUC5B*)) [17, 18].

The diagnosis of IPF represents a fatal prognosis for affected patients, with a reported median survival time between 2.5 to 3.5 years [19]. Common symptoms described by IPF patients include increasing dyspnea, the appearance of clubbed fingers due to the chronic lack of oxygen, and dry cough, which are frequently accompanied by crackles on auscultation of the lung [9, 12].

The diagnostic concept in case of potential IPF has been recently revised and published as an international ATS/ERS/Japanese Respiratory Society (JRS)/Latin American Thoracic Society (ALAT) statement by Raghu *et al.* [12], which was further adopted by Behr *et al.* [20] for the recent German guidelines on IPF. The process of confirming the definite diagnosis of IPF still remains complex and sometimes lengthy, requiring the collaboration of different medical specialists and the performance of multiple examinations. Due to its non-invasive nature combined with a high diagnostic value, the high-resolution computed tomography (HRCT) is of significant importance in the context of diagnosing IPF [12, 20]. The representative HRCT findings of an IPF lung are summarized as so-called usual interstitial pneumonia (UIP) and are defined by specific hallmarks, such as honeycombing with reticulation, traction bronchiectasis/bronchiolectasis and ground-glass opacification, which are all typically observed in the basal parts of the lung and in close proximity to the pleura (Figure 1.1) [12, 20, 21].

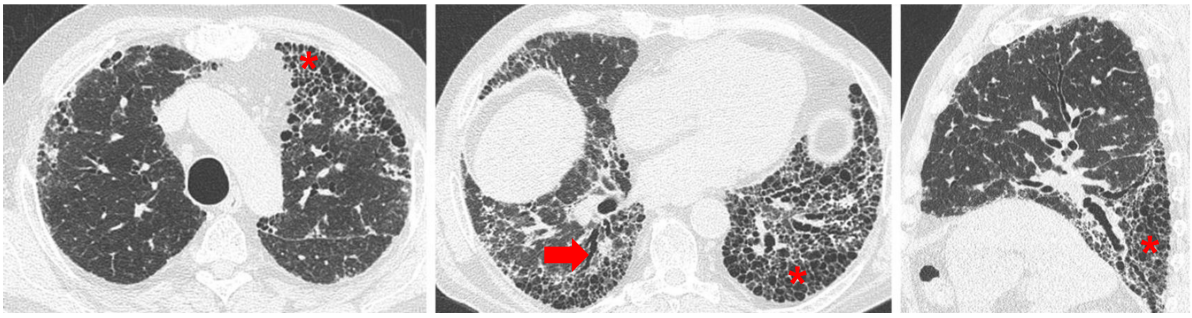

Figure 1.1: Hallmarks of usual interstitial pneumonia (UIP). High-resolution computed tomography (HRCT) showing subpleural and basal honeycombing with reticulation (asterisk) and traction bronchiectasis/bronchiolectasis (arrow) as the hallmarks of UIP [21] (Images were adapted and reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Der Radiologe, Hamer *et al.* [21], Copyright © 2020).

In order to confirm the definite diagnosis of IPF, it is essentially important to exclude other established fibrotic lung diseases in addition to the HRCT, and in case of indistinct HRCT findings, complementary examinations are recommended, such as serological testing, bronchoalveolar lavage, and lung biopsy for histopathology [12, 20].

The therapy of IPF was very limited for a long time, which has been recently changed by the introduction of the two substances nintedanib and pirfenidone. The latter has received approval for the therapy of IPF in 2011 (European Union) and 2014 (USA) [22]. Beforehand, the safety and benefit of pirfenidone intake had been evaluated by three large-scale, international, placebo-controlled and randomized phase 3 trials (ASCEND, CAPACITY 004 and 006), which proved a significant, positive impact of pirfenidone treatment on IPF disease progression, represented by progression-free survival and lesser alterations in forced vital capacity (FVC) [23, 24]. In addition, Nathan *et al.* [25] performed combined analyses of the ASCEND, CAPACITY 004 and 006 trials and were able to prove a significant decrease of mortality with pirfenidone treatment. The established route of administration is oral, enabling the anti-inflammatory and anti-fibrotic effects of pirfenidone [25], which still remain subject to further research and seem to be mediated through its broad impact on various pro-fibrotic and pro-inflammatory factors, such as transforming growth factor beta 1 (TGF-β) and tumor necrosis factor-alpha (TNF-α) [26].

Nintedanib was originally introduced as an anti-cancer drug [27], receiving approval for the therapy of IPF in 2015 (European Union) and 2014 (USA) [22]. The treatment with nintedanib was assessed by two international, randomized and placebo-controlled phase 3 trials (INPULSIS-1 and -2), which showed a significant deceleration of the IPF progression, measured by alterations in FVC [28]. The drug unfolds its anti-fibrotic effects as an inhibitor of multiple receptor tyrosine kinases, such as vascular endothelial growth factor receptor, platelet-derived growth factor receptor and fibroblast growth factor receptor [27].

The current international guidelines recommend the use of either nintedanib or pirfenidone, while lung transplantation remains the last resort [29]. Apart from that, only best supportive care (e.g. oxygen therapy) is available to IPF patients, which highlights the importance and urgent need of new therapeutic strategies.

1.1.2 Pathomechanisms of IPF

IPF demonstrates a heterogeneous course of disease with almost stable, slow, or rapid progression as well as acute exacerbations, eventually leading to early death of the patient [30]. Therefore, curative treatment options seem indispensable, but despite the important progress made, the pathomechanisms of IPF still remain subject to further research.

Idiopathic pulmonary fibrosis is characterized by the progressive formation of abundant lung tissue and the resulting deterioration of pulmonary function and structure. Even though (myo-)fibroblasts appear to be the effector cells of these changes in IPF lungs due to their redundant generation of extracellular matrix (ECM), the lung epithelium is assumed to be the actual initiator and driver of the IPF pathogenesis [31]. The current scientific consensus, based on recent findings, is that multiple, recurrent microinjuries to the lung epithelium lead to reprogramming of the affected epithelial cells [9, 11, 31]. In individuals with IPF, this process occurs in already prematurely aged lung epithelial cells, which seem to be particularly vulnerable [11]. Cellular senescence, a hallmark

of aging, has been demonstrated to be a prominent feature of alveolar epithelial type II (ATII) cells isolated from IPF lung tissue and bleomycin-induced lung fibrosis, and the experimental treatment of these cells with senolytic drugs had a significant anti-fibrotic effect [32]. In line with these findings, also telomere attrition is frequently detected in fibrotic ATII cells of IPF patients [33, 34] as well as damage-associated genomic instability [31]. Finally, genetic alterations have been proved to be a central feature of the IPF pathogenesis, especially alterations to mucin 5B (*MUC5B*) represent an important risk factor [35]. It is thought that these events lead to the release of numerous mediators (e.g. TGF-β, TNF-α, Wnt/β-catenin), (myo-)fibroblast activation and a distorted process of lung repair and regeneration with the subsequent deterioration of lung architecture with its defining histological changes (Figure 1.2) [9, 31].

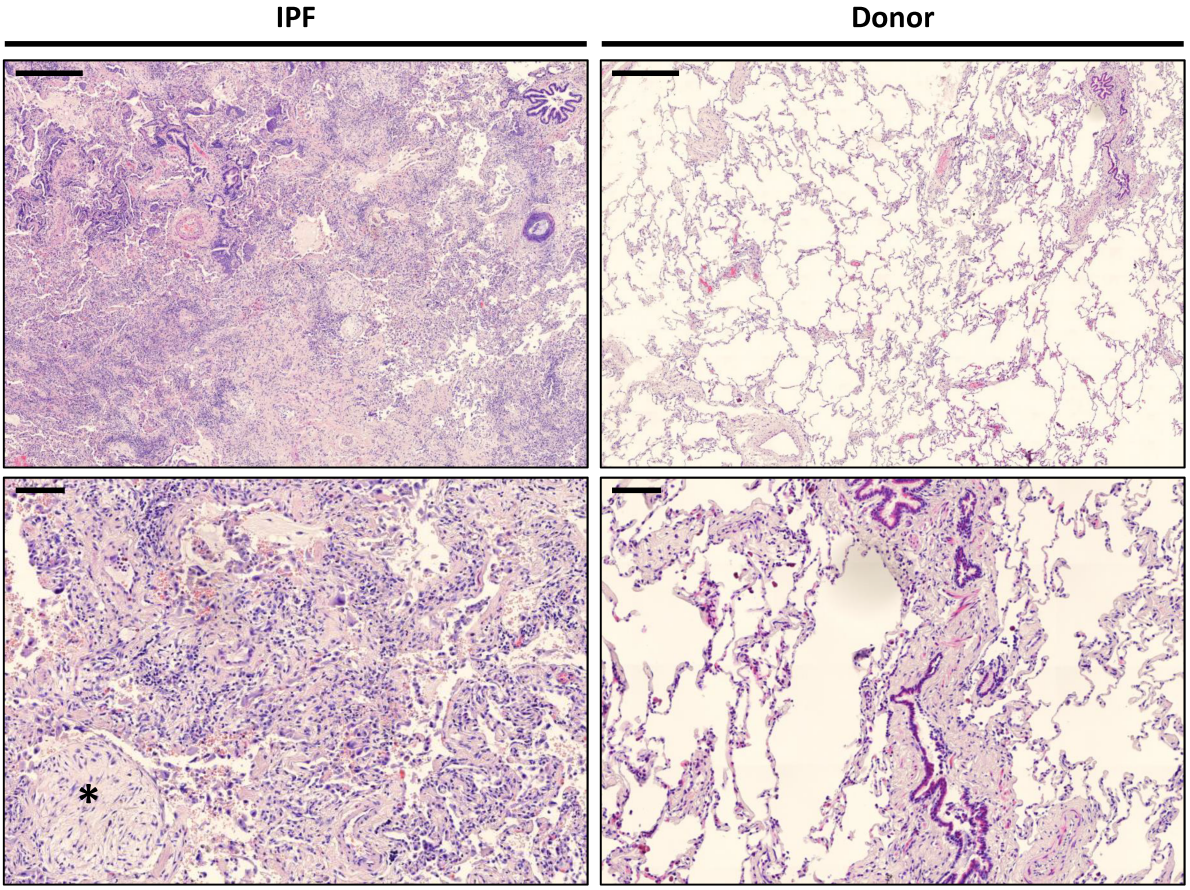

Figure 1.2: Deterioration of lung architecture in patients with IPF. Tissue sections of IPF and donor lungs were stained with hematoxylin and eosin. The fibrotic tissue specimens showed areas of dense fibrosis with destroyed lung architecture and fibroblast foci (asterisk). Scale bars: 500μm upper row, 100μm lower row.

1.1.3 Bleomycin-induced experimental model of IPF

The bleomycin-induced experimental model of pulmonary fibrosis in mice was used in this thesis to further study idiopathic pulmonary fibrosis in an animal model. Bleomycin was originally identified as anti-cancer drug and has been used for the therapy of multiple types of cancer, such as germ cell tumors [36] and Hodgkin lymphoma [37]. The downside to bleomycin treatment are its severe side effects on the lung in form of pneumonitis with a gradual progression to lung fibrosis [38]. Therefore, bleomycin is used to model idiopathic pulmonary fibrosis in various animal species.

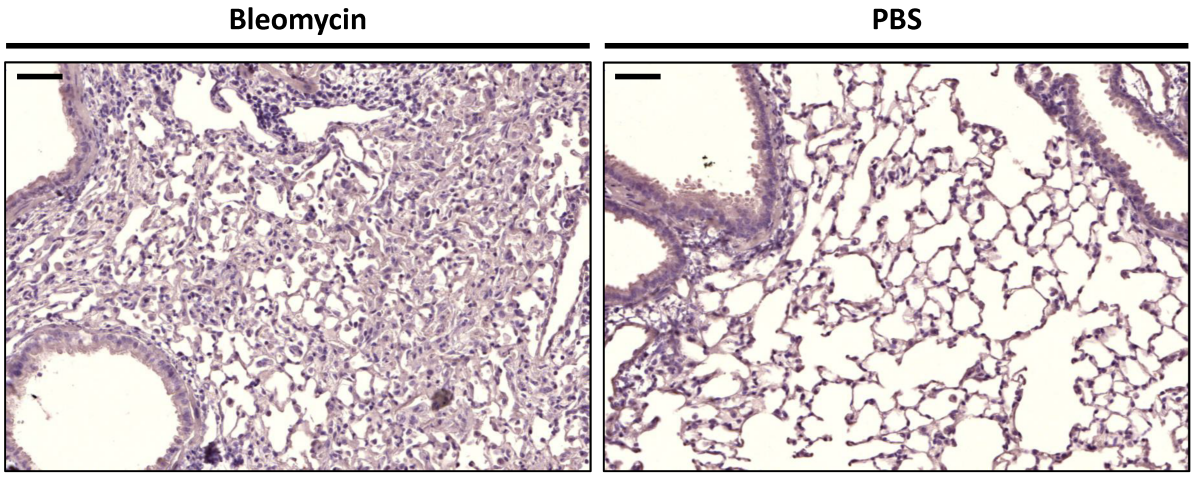

Figure 1.3: Bleomycin-induced pulmonary fibrosis in mice. Tissue sections of mice treated with bleomycin or phosphate-buffered saline (PBS) were harvested on day 14 after treatment and stained with hematoxylin and eosin. Lungs from bleomycin-treated mice showed dense pulmonary fibrosis. Scale bars: 50μm.

In the present study, bleomycin was intratracheally administered to C57BL/6N mice (Figure 1.3), which is to date the most favored and best described animal model of IPF [39–41]. The fibrotic destruction and remodeling of the alveolar structure upon bleomycin instillation are most evident on day 14 after the treatment and gradually develop after an early inflammatory phase, which is marked by infiltration of neutrophils and reaches its maximum around day 7 [42, 43]. Beyond day 21 after the administration of bleomycin, the course of the fibrotic phase has been reported to be highly variable with possible resolution of the lung fibrosis, which is one of the major criticism [40, 42]. Besides this discrepancy, it has been demonstrated that the bleomycin-induced experimental model of pulmonary fibrosis resembles several important features of human IPF, such as the appearance of fibroblast foci, initial injury of the alveolar epithelium, activation of (myo-)fibroblasts, and common mediators (e.g. TGF-β, TNF-α) [40, 43].

1.2 Lung cancer in IPF

1.2.1 Non-small cell lung cancer

Lung cancer is another severe disease of the human lung, which also shows an irreversible, chronic progression and is marked by exaggerated tissue production. Recent, international studies on cancer statistics reported that lung carcinomas represent globally the major cause of cancer-related deaths in males and females combined, with approximately 1.76 million deaths in 2018 [44–46]. Furthermore, lung cancer has been shown to have the highest incidence world-wide among all types of cancer [46]. In line with this, lung cancer accounts for the leading cause of cancer-related healthcare expenses in the European Union, with estimated 18.8 billion euros in 2009 [47].

Over the past decades research has made significant advancements in its understanding of lung cancer pathomechanisms and its risk factors. The exposure to (passive) tobacco smoking and its continuance remain the undisputed main etiologic factors for the development of lung cancer, while recent studies identified further risk factors, such as chronic or infectious lung diseases, genetic alterations, familial predisposition, and occupational or environmental carcinogens (air pollution, radiation, asbestos) [48].

Upon clinical examination, lung cancer patients present with rather unspecific symptoms of dyspnea, hemoptysis, chronic cough, pain of chest/shoulders, fatigue, weight loss, and fever, which usually appear not until advanced development of lung cancer [49]. Therefore, the diagnosis and staging of lung cancer requires diagnostic imaging with chest X-ray, computed tomography (CT), magnetic resonance imaging (MRI), or positron-emission tomography (PET) and eventually tissue biopsies (CT-guided, bronchoscopy, surgery) for the histological classification [50].

The World Health Organization (WHO) reworked the classification of lung cancer in 2015, dividing it into the major categories of small cell lung cancer (15%) and non-small cell lung cancer (85%, NSCLC), which is further subdivided into the main entities of squamous cell carcinoma (SCC), adenocarcinoma (AC), large cell carcinoma, and various minor subtypes [51]. The present thesis was specifically focused on NSCLC because of its epithelial origin.

Adenocarcinoma (40%) and squamous cell carcinoma (25-30%) represent the two major histological subtypes of NSCLC, but differ considerably from each other [52]. SCC mainly affects the male sex and patients with a smoking history, compared to AC, which has been observed more frequently in female and nonsmoking patients [53]. Furthermore, SCC predominantly originates from the bronchial epithelium and expresses cytokeratin 5 and 6 (KRT5/6), while AC predominantly develops from the alveolar epithelium, showing napsin A and thyroid transcription factor 1 (TTF-1) expression [52]. Overall, the adenocarcinoma shows a better prognosis with significantly higher survival rates [53].

The general therapy of lung cancer includes chemotherapy, radiotherapy, and surgery. Over the past decades cancer therapy has advanced considerably and targeted therapies have been developed, but there is still an urgent need for new, curative treatments.

1.2.2 Similarities of IPF and NSCLC

The majority of patients with IPF is suffering from various comorbidities, with lung cancer being frequently reported [7, 11]. Over the past years this association has received increasing attention and studies have elucidated this relationship, while the underlying pathomechanisms still remain unknown.

Recent, independent meta-analyses of epidemiological studies on lung cancer in IPF populations estimated a prevalence of 13.74% [54] or 13.54% [55]. In addition, it has been shown that lung cancer has a significantly increased incidence in IPF populations with an estimated rate of 2.07 per 100 person years, compared to 0.2 - 0.7 per 100 person years among non-IPF populations [54]. The risk for IPF patients to develop lung cancer increases notably with the time that has passed since the initial diagnosis of IPF, which is reflected by the rising cumulative incidence over the years (Figure 1.4) [56–60]. In line with these findings, IPF has been described as an independent risk factor for the development of lung cancer [54].

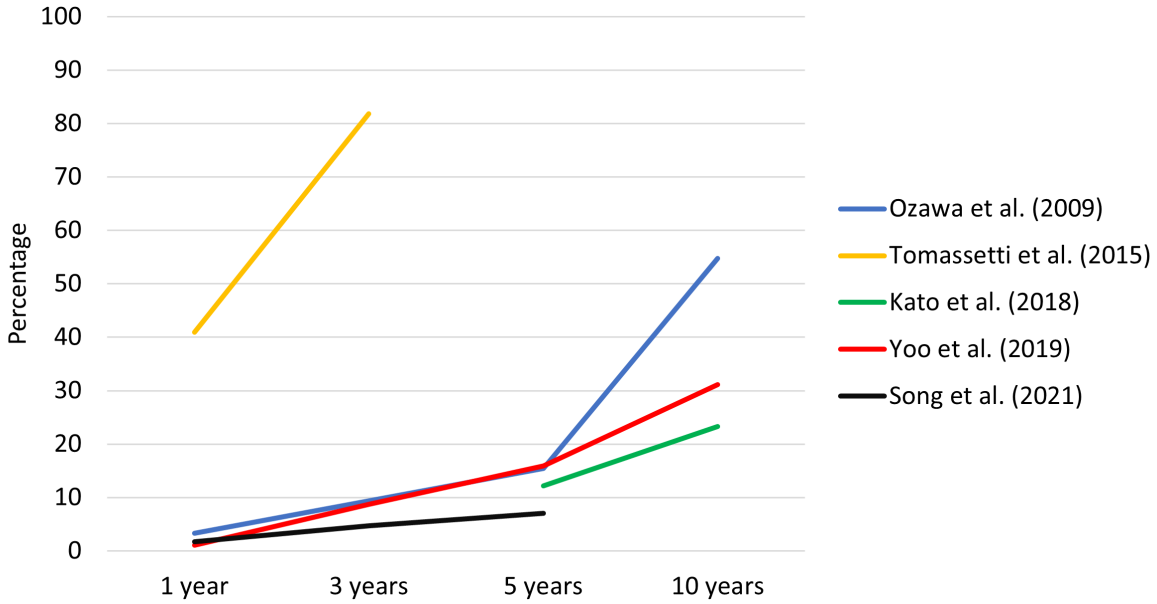

Figure 1.4: Cumulative incidence of lung cancer in IPF. Data has been extracted from recent studies by Ozawa *et al.* [56], Tomassetti *et al.* [57], Kato *et al.* [58], Yoo *et al.* [59], and Song *et al.* [60]. The cumulative incidence of lung cancer in IPF shows a notable increase over time. Variations of the cumulative incidence occur due to differences in patient selection criteria and study design.

The prognosis and outcome of IPF patients with diagnosed lung cancer have been demonstrated to be markedly worse compared to IPF patients without lung cancer, which is expressed by a significantly increased mortality and reduced survival rate [60]. Tomassetti *et al.* [57] have shown that lung cancer-associated complications, therapeutic procedures, and progression are mainly responsible for the worse outcome of IPF patients with lung cancer. In particular, the increased risk of acute exacerbations caused by lung cancer-related treatments is of great significance [57, 61].

The average IPF patient affected by lung cancer is characterized by male sex, old age, emphysema, and a long history of smoking [58–60]. In contrast to the general population, squamous cell carcinoma represents the predominant subtype of lung cancer in IPF populations and is frequently found in peripheral, lower areas of the lung, which are located in close proximity to the fibrotic lesions [55, 58, 59].

Besides these epidemiological connections, the underlying pathomechanisms still remain poorly understood - but both diseases seem to share common hallmarks in their pathogenesis. The development and progression of lung cancer relies on the occurrence of multiple genetic alterations in oncogenes and tumor suppressor genes, such as the *P53* gene [62, 63]. Similar mutations of the *P53* gene and other lung cancer-related genes have also been observed in IPF patients [64]. Epigenetic alterations, such as the methylation of deoxyribonucleic acid (DNA), not only have been frequently identified in IPF and NSCLC but also show a highly similar pattern in both diseases [65]. Furthermore, a variety of biological pathways and mediators has been shown to be similarly misregulated in lung cancer as well as in IPF. The Wnt/β-catenin signaling pathway, for instance, is involved in the carcinogenesis of NSCLC, while in IPF it seems to be important for cell proliferation and activation of fibroblasts [66, 67]. Eventually, other common features, such as senescence, invasive behavior, and aberrant healing processes, have been frequently discussed for both diseases.

Based on these similarities, we hypothesized that analyzing the commonalities of the pathomechanisms of IPF and NSCLC will help to advance the knowledge about the pathogenesis of both diseases and to discover potential therapeutic targets in IPF.

1.3 PSAT1 - Phosphoserine aminotransferase 1

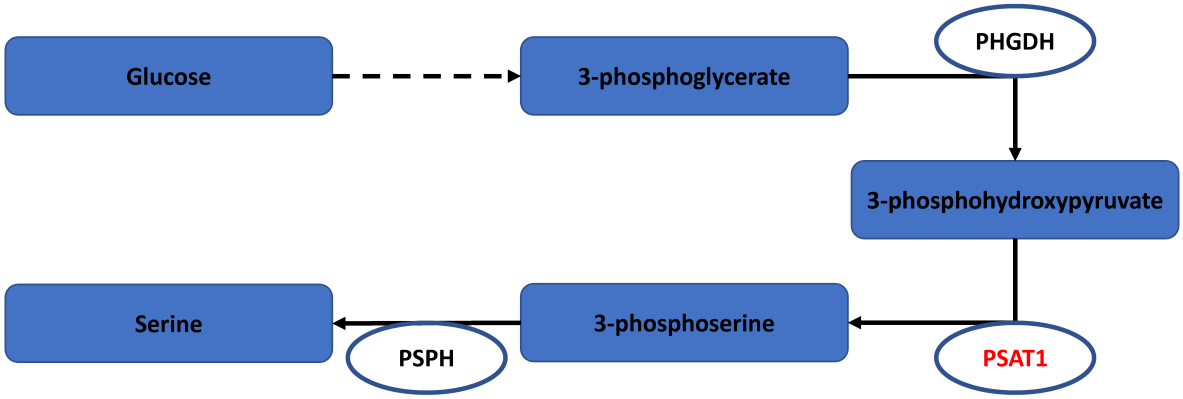

Figure 1.5: Serine biosynthesis pathway. Schematic presentation of the serine biosynthesis pathway with its main enzymes phosphoserine aminotransferase 1 (PSAT1), phosphoglycerate dehydrogenase (PHGDH), and phosphoserine phosphatase (PSPH).

The phosphoserine aminotransferase 1 (*PSAT1*) gene encodes one of the three major enzymes for the biosynthesis of serine (Figure 1.5). The amino acid serine has important direct or indirect roles in many biological processes and pathways of the healthy cell,

such as the folate cycle, methionine cycle, glutathione redox system, and the generation of various amino acids, proteins, lipids, and nucleosides [68]. The gene expression of PSAT1 is increased by the binding of activating transcription factor 4 (ATF4) to its promoter region, which has been shown to be positively controlled by the nuclear factor erythroid 2-related factor 2 (NRF2) [69, 70]. In addition, Guo *et al.* [71] demonstrated an indirect upregulation of PSAT1 via the long non-coding ribonucleic acid (RNA) MEG8 in NSCLC cells, which supported the development and growth of lung cancer.

Alterations of the PSAT1 expression have been described in various types of cancer, such as NSCLC [69, 72–74], esophageal squamous cell carcinoma [75], breast cancer [70, 76], and ovarian cancer [77]. Increased levels of PSAT1 were significantly connected to a worse prognosis in these cancer patients [69, 70, 72, 74, 75, 77]. Furthermore, PSAT1 overexpression facilitates cell proliferation in a cyclin D1-dependent manner [70, 74] and is related to metastasis, invasive behavior, and cell migration [72, 76].

Besides its contribution to oncogenesis, recent studies demonstrated that TGF-β promotes the expression of PSAT1 in human lung fibroblasts, which in turn enables the formation of collagen by myofibroblasts [78, 79]. Therefore, we hypothesized that PSAT1 represents a potential therapeutic target and its possible contribution to the pathogenesis of IPF requires further investigation.

1.4 GPR87 - G protein-coupled receptor 87

The G protein-coupled receptor 87 (*GPR87*) gene is located on the chromosome 3q24 in humans and encodes a seven-transmembrane receptor with 358 amino acids, which was originally identified and published in 2001 by Wittenberger *et al.* [80]. The specific biological function of GPR87 still remains unknown, while Tabata *et al.* [81] deorphanized the receptor and demonstrated lysophosphatidic acid (LPA) as a ligand for GPR87, which was subsequently confirmed by Ochiai *et al.* [82]. However, recent data suggests that LPA is a rather unspecific ligand with low binding affinity to GPR87, but to date no other ligands have been identified [83]. The GPR87 signaling cascade is mediated through activation of the G protein α_{q}, α_{i}, and $\alpha_{12/13}$ subunits (Figure 1.6) [83].

Given that GPR87 represents a member of the seven-transmembrane G protein-coupled receptor family, studies were able to verify its predominant location on the cell membrane [84]. Furthermore, the expression pattern of GPR87 in healthy, regular tissue seems to be very specific and limited to only certain types of cells or tissue [84]. In particular, the regular lung tissue and its epithelium showed a markedly low expression of GPR87 [84, 85]. Increased expression levels of GPR87 have been demonstrated in various cancer types, such as lung cancer [84–86], hepatocellular carcinoma [87], urothelial carcinoma [88], pancreatic cancer [89], and multiple squamous cell carcinomas (skin, cervix, pharynx, and larynx) [84]. Notably, this altered expression of GPR87 in lung cancer was primarily found in NSCLC and particularly in SCC of the lung [85, 86].

The regular biological function of GPR87 remains unexplained, whereas it has been shown to induce cell survival, proliferation, and anti-apoptotic effects in cancer, which is at least partially mediated through its interaction with P53 [84, 89–92]. In addition, the

significantly increased expression level of GPR87 in various types of cancer is correlated with a worse prognosis and supports metastasis, migration, and invasion [86, 87, 89].

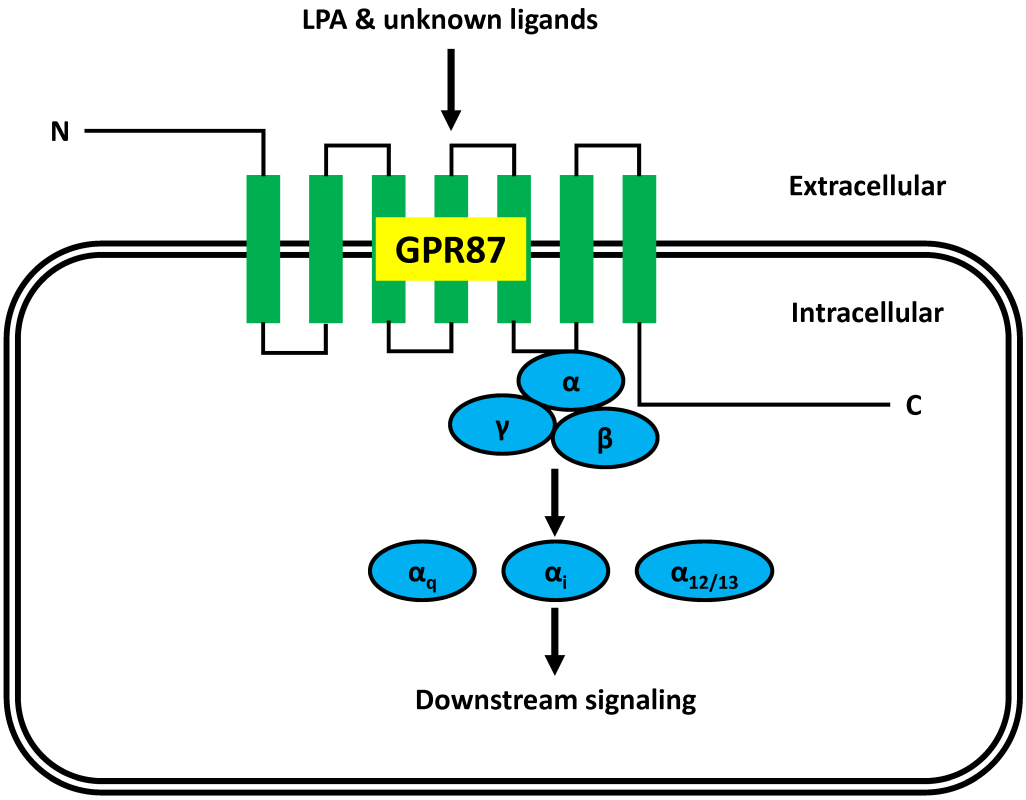

Figure 1.6: GPR87 signaling cascade. Schematic presentation of the G protein-coupled receptor 87 (GPR87) signaling cascade. Lysophosphatidic acid (LPA) or other unknown ligands bind to GPR87 and induce various subsequent signaling pathways through α_q, α_i, and $\alpha_{12/13}$ [83].

Recent studies demonstrated that GPR87 overexpression in cancer is promoted by activation of transcription through the binding of either histone H3.3 or the signal transducer and activator of transcription 3 (STAT3) [93, 94].

The contribution of GPR87 to the pathogenesis of IPF is still unknown. Thus, we hypothesized that GPR87 might contribute to cellular reprogramming in IPF and could represent a novel therapeutic target, which needed further research.

2 Aims of the study

IPF and NSCLC represent severe pulmonary disorders with a chronic progression and irreversible damage to the lung. It is commonly accepted that IPF and NSCLC share epidemiologic similarities and common pathomechanistic features. To date no approved, curative treatment exists for both diseases, which emphasizes the importance of further understanding the involved common pathomechanisms to develop novel therapeutic strategies and to discover possible targets for treatments. Hence, we hypothesized that oncogenic, cancer-related signaling acts as an important driver of the IPF pathogenesis and contributes to the lung epithelial cell alterations in IPF, such as cellular reprogramming.

The major aims of the presented thesis were to:

1. Establish a possible relationship between the pathomechanisms of IPF and NSCLC by elucidating common patterns of gene expression alterations.
2. Assemble and characterize a set of candidate genes commonly misregulated in IPF and NSCLC.
3. Determine expression levels of particular candidate genes in lung tissue from bleomycin-induced experimental lung fibrosis and IPF.
4. Analyze cell-specific expression of these candidate genes and identify mechanisms that drive their misregulation in IPF.

3 Materials and methods

3.1 Materials

3.1.1 Cell lines and primary cells

Table 3.1: Cell lines.

Cell line	Origin	Manufacturer/Distributor
16HBE14o$^-$	human bronchial epithelial cells, SV40 large T-antigen transformed	D. C. Gruenert, University of California; San Francisco, CA, USA
A549	human epithelial cells from lung carcinoma	ATCC, LGC Standards; Wesel, Germany

Table 3.2: Primary cells.

Cell type	Origin	Source
Primary human bronchial epithelial cells (phBEC)	human lung tissue	Asklepios Biobank for Lung Diseases; Gauting, Germany Lonza Group AG; Basel, Switzerland
Primary murine alveolar epithelial type II (pmATII) cells	C57BL/6N mice	Charles River Laboratories; Sulzfeld, Germany

3.1.2 Human tissue

Human lung tissue specimens were provided by different organizations. The Biobank of the Universities of Giessen and Marburg Lung Center (UGMLC), part of the DZL Biobanking Platform, provided human idiopathic pulmonary fibrosis (IPF) and healthy control specimens (Table 3.3). The human lung tissue samples were originally assembled within the European IPF Registry (eurIPFreg) and Biobank project. The ethics committee of the Justus Liebig University, Giessen, Germany was officially informed and the study protocol was authorized (Project 111/08 and 58/15). Informed consent was given in written form by all participants included in the study.

Lung resections, explanted lungs, and lung tissue biopsies from the university hospital of the Ludwig-Maximilians-University (LMU), Munich, Germany and the Asklepios

Table 3.3: Human IPF and control tissue, UGMLC Biobank.

Identification	Disease status	Age	Gender	Storage
IPF_1	IPF	24	female	liquid nitrogen
IPF_2	IPF	49	male	liquid nitrogen
IPF_3	IPF	55	male	liquid nitrogen
IPF_4	IPF	39	female	liquid nitrogen
IPF_5	IPF	62	female	liquid nitrogen
IPF_6	IPF	62	male	liquid nitrogen
IPF_7	IPF	36	male	liquid nitrogen
IPF_8	IPF	48	male	liquid nitrogen
IPF_9	IPF	31	female	liquid nitrogen
IPF_10	IPF	44	female	liquid nitrogen
IPF_11	IPF	unknown	unknown	liquid nitrogen
IPF_12	IPF	unknown	unknown	liquid nitrogen
IPF_13	IPF	unknown	unknown	liquid nitrogen
IPF_14	IPF	48	male	liquid nitrogen
Donor_1	healthy	unknown	unknown	liquid nitrogen
Donor_2	healthy	72	female	liquid nitrogen
Donor_3	healthy	61	male	liquid nitrogen
Donor_4	healthy	42	male	liquid nitrogen
Donor_5	healthy	44	female	liquid nitrogen
Donor_6	healthy	31	male	liquid nitrogen
Donor_7	healthy	41	female	liquid nitrogen
Donor_8	healthy	58	male	liquid nitrogen
Donor_9	healthy	53	male	liquid nitrogen
Donor_10	healthy	48	female	liquid nitrogen

Table 3.4: Lung tissue specimens for the isolation of phBECs and histology.

Identification	Airway diameter	Age	Gender	Storage
CPC_1	5-7mm	48	male	liquid nitrogen
CPC_2	9mm	65	female	liquid nitrogen
CPC_3	-	unknown	unknown	liquid nitrogen
CPC_4	-	54	unknown	liquid nitrogen

Biobank for Lung Diseases, Gauting, Germany were used for the isolation of phBECs and histology (Table 3.4). Informed consent was given in written form by all participants included in the study. The ethics committee of the Ludwig-Maximilians-University, Munich, Germany was officially informed and the study protocol was authorized (Project

333-10). Furthermore, isolated phBECs were purchased from Lonza Group AG, Basel, Switzerland (Table 3.5).

Table 3.5: phBECs from Lonza.

Identification	Age	Gender	Storage
Lonza_1	49	female	liquid nitrogen
Lonza_2	13	male	liquid nitrogen
Lonza_3	62	female	liquid nitrogen

Paired human lung specimens of lung cancer and corresponding control (CTRL) tissue were provided by the Asklepios Biobank for Lung Diseases, Gauting, Germany (Table 3.6). The project was approved by the ethics committee of the Ludwig-Maximilians-University, Munich, Germany (Project 623-15). Written informed consent was received from all participants.

Table 3.6: Human lung cancer and control tissue, Asklepios Biobank.

Identification	Disease status	Age	Gender	Storage
027-001 T/N	squamous cell carcinoma	64	female	RNAlater
027-002 T/N	squamous cell carcinoma	78	male	RNAlater
027-003 T/N	squamous cell carcinoma	64	male	RNAlater
027-004 T/N	squamous cell carcinoma	60	female	RNAlater
027-005 T/N	squamous cell carcinoma	50	female	RNAlater
027-006 T/N	squamous cell carcinoma	70	male	RNAlater
027-007 T/N	adenocarcinoma	45	female	RNAlater
027-008 T/N	adenocarcinoma	49	male	RNAlater
027-009 T/N	adenocarcinoma	42	female	RNAlater
027-010 T/N	adenocarcinoma	68	female	RNAlater
027-011 T/N	adenocarcinoma	58	female	RNAlater

3.1.3 Animals

In general, the wild-type C57BL/6N mouse strain was used for all animal experiments in this thesis. The pathogen-free mice were purchased from Charles River Laboratories, Sulzfeld, Germany and were able to settle down for at least one week in the animal facility of the Helmholtz Zentrum München.

3.1.4 Laboratory equipment and software

Table 3.7: Laboratory equipment.

Product	Manufacturer/Distributor
-20°C Freezer GNP 5255 Premium NoFrost	Liebherr-International AG; Bulle, Switzerland
-20°C Freezer LGex 3410 MediLine	Liebherr-International AG; Bulle, Switzerland
-80°C Freezer New Brunswick U570 HEF	Eppendorf AG; Hamburg, Germany
+4°C Fridge LKv 3910 MediLine	Liebherr-International AG; Bulle, Switzerland
A.S. Standard Fume Cupboard	Vinitex Laboratory Systems; Sint-Oedenrode, Netherlands
Autoclave DX-45	Systec GmbH; Linden, Germany
Autoclave VX-120	Systec GmbH; Linden, Germany
Axio Imager 2	Carl Zeiss AG; Oberkochen, Germany
Axiovert 40 C	Carl Zeiss AG; Oberkochen, Germany
BD LSR II Flow Cytometer	BD Biosciences; Franklin Lakes, NJ, USA
Biological Safety Cabinet, Class II, Herasafe KS 18	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Centrifuge 5424 R	Eppendorf AG; Hamburg, Germany
Centrifuge 5430	Eppendorf AG; Hamburg, Germany
Centrifuge Corning LSE Mini Microcentrifuge	Corning, Inc.; Corning, NY, USA
Centrifuge MIKRO 200 R	Andreas Hettich GmbH & Co. KG; Tuttlingen, Germany
Centrifuge MiniSpin Plus	Eppendorf AG; Hamburg, Germany
Centrifuge ROTINA 420 R	Andreas Hettich GmbH & Co. KG; Tuttlingen, Germany
ChemiDoc XRS+ System	Bio-Rad Laboratories, Inc.; Hercules, CA, USA
Cleaning Machine Miele G 7893	Miele & Cie. KG; Gütersloh, Germany
CO_2 Incubator BBD 6220	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Counting Chamber BLAUBRAND Neubauer Improved	BRAND GmbH & Co. KG; Wertheim, Germany
Easypet Electronic Pipet Filler	Eppendorf AG; Hamburg, Germany
Excellence Precision Balance XS4002S	Mettler-Toledo, LLC; Columbus, OH, USA
Forma 8600 Series -86°C Ultra-Low Temperature Chest Freezer	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Hyrax M55 Rotary Microtome	Carl Zeiss AG; Oberkochen, Germany
Ice Machine ZBE 110-35	ZIEGRA-Eismaschinen GmbH; Isernhagen, Germany
LightCycler 480 Instrument II	Roche Diagnostics; Rotkreuz, Switzerland
Liquid Nitrogen - APOLLO 200 Supply Vessel	Cryotherm GmbH & Co. KG; Kirchen/Sieg, Germany
Liquid Nitrogen - BIOSAFE 420 SCβ-Sample Storage System	Cryotherm GmbH & Co. KG; Kirchen/Sieg, Germany

Product	Manufacturer/Distributor
Magnetic Stirrer Big Squid White	IKA-Werke GmbH & Co. KG; Staufen, Germany
Magnetic Stirrer KMO 2 Basic	IKA-Werke GmbH & Co. KG; Staufen, Germany
Mastercycler Nexus	Eppendorf AG; Hamburg, Germany
Microbiological Incubator Heratherm IGS 60	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Microm EC 350 - Modular Tissue Embedding Center	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Microm STP420D Tissue Processor	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Mikro-Dismembrator S	Sartorius Stedim Biotech GmbH; Goettingen, Germany
Milli-Q Advantage A10 Water Purification System	Merck KGaA; Darmstadt, Germany
Mini-PROTEAN Tetra Vertical Electrophoresis Cell	Bio-Rad Laboratories, Inc.; Hercules, CA, USA
MIRAX SCAN, Slide Scanner	Carl Zeiss AG; Oberkochen, Germany
Mr. Frosty Freezing Container	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Multi-Axle-Rotating-Mixer, RM10W-30V	VWR International, LLC; Radnor, PA, USA
Multipette E3x	Eppendorf AG; Hamburg, Germany
NanoDrop 1000 Spectrophotometer	Thermo Fisher Scientific Inc.; Waltham, MA, USA
pH Meter inoLab pH 720	Xylem Analytics Germany Sales GmbH & Co. KG, WTW; Weilheim, Germany
Pipettes Research Plus	Eppendorf AG; Hamburg, Germany
PowerPac Basic Power Supply	Bio-Rad Laboratories, Inc.; Hercules, CA, USA
QuadroMACS Separator	Miltenyi Biotec; Bergisch Gladbach, Germany
Shaker Duomax 1030	Heidolph Instruments GmbH & Co. KG; Schwabach, Germany
Shaker MS 3 Basic	IKA-Werke GmbH & Co. KG; Staufen, Germany
Shaker Polymax 1040	Heidolph Instruments GmbH & Co. KG; Schwabach, Germany
Sunrise Absorbance Microplate Reader	Tecan Group Ltd.; Männedorf, Switzerland
Surgical Instruments	Fine Science Tools GmbH; Heidelberg, Germany
ThermoMixer Comfort	Eppendorf AG; Hamburg, Germany
Vacuum Pump EcoVac System 4	schuett-biotec GmbH; Goettingen, Germany
Vacuum Pump N022 AN.18	KNF Neuberger GmbH; Freiburg, Germany
Water Bath Aqualine AL 12	LAUDA Dr. R. Wobser GmbH & Co. KG; Lauda-Königshofen, Germany

Table 3.8: Software.

Software	Developer/Distributor
AxioVision, Release 4.8	Carl Zeiss AG; Oberkochen, Germany

Software	Developer/Distributor
Bioconductor, Release 3.6	Bioconductor [95, 96]; https://www.bioconductor.org/
BioGPS	The Scripps Research Institute; La Jolla, CA, USA
FlowJo Software, Version 9.6.4	FlowJo, LLC; Ashland, OR, USA
Gene Set Enrichment Analysis (GSEA), Version 3.0	Broad Institute, Inc.; Cambridge, MA, USA
GraphPad Prism, Version 5.00	GraphPad Software, Inc.; San Diego, CA, USA
Image Lab Software, Version 5.2.1	Bio-Rad Laboratories, Inc.; Hercules, CA, USA
JabRef, Version 5.0	JabRef Authors; MIT License; http://www.jabref.org/
LightCycler 480 Software, Version 1.5	Roche Diagnostics; Rotkreuz, Switzerland
Magellan - Data Analysis Software	Tecan Group Ltd.; Männedorf, Switzerland
Microsoft Office 2010 Professional Plus	Microsoft Corporation; Redmond, WA, USA
Microsoft Windows 7 Enterprise	Microsoft Corporation; Redmond, WA, USA
R Statistical Software, Version 3.4.3	R Core Team; R Foundation for Statistical Computing; Wien, Austria
RStudio, Version 1.1.383	RStudio; Boston, MA, USA
Salmon Software, Version 0.9.1	Rob Patro; GNU General Public License v3.0; https://combine-lab.github.io/salmon/
TeX Live 2018	TeX user groups; https://www.tug.org/texlive/
TeXstudio 2.12.16	Benito van der Zander; GNU General Public License v2.0; https://www.texstudio.org/
ZEN Imaging Software	Carl Zeiss AG; Oberkochen, Germany

3.1.5 Consumables

Table 3.9: Consumables.

Product	Manufacturer/Distributor
Cannulas Sterican (20G, 21G, 26G)	B. Braun Melsungen AG; Melsungen, Germany
CD31 MicroBeads	Miltenyi Biotec; Bergisch Gladbach, Germany
CD45 MicroBeads	Miltenyi Biotec; Bergisch Gladbach, Germany
Cell Culture Dishes	Corning, Inc.; Corning, NY, USA
Cell Culture Flasks, EasYFlask 75cm^2	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Cell Culture Multiwell Plates	Corning, Inc.; Corning, NY, USA TPP Techno Plastic Products AG; Trasadingen, Switzerland
Cell Scrapers and Lifters	Corning, Inc.; Corning, NY, USA

Product	**Manufacturer/Distributor**
Chromium Steel Grinding Balls (9mm)	Sartorius Stedim Biotech GmbH; Goettingen, Germany
Combitips Advanced	Eppendorf AG; Hamburg, Germany
Cover Glasses	Paul Marienfeld GmbH & Co. KG; Lauda-Königshofen, Germany
Cryogenic Tubes, Nalgene	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Falcon Conical Centrifuge Tubes (15ml, 50ml)	Corning, Inc.; Corning, NY, USA
Falcon Round Bottom Polystyrene Tubes, 14ml	Corning, Inc.; Corning, NY, USA
Filter Pipette Tips, SafeSeal-Tips Professional	Biozym Scientific GmbH; Hessisch Oldendorf, Germany
LS Columns	Miltenyi Biotec; Bergisch Gladbach, Germany
Microscope Slides, Glass	DWK Life Sciences GmbH; Wertheim/Main, Germany
MicroSprayer Aerosolizer - Model IA-1C and FMJ-250 High Pressure Syringe	Penn-Century, Inc.; Wyndmoor, PA, USA
Nitrocellulose Membrane	Bio-Rad Laboratories, Inc.; Hercules, CA, USA
Nylon Filters (100μm, 20μm, 10μm)	Sefar AG; Heiden, Switzerland
Parafilm	Bemis Company, Inc.; Neenah, WI, USA
Pasteur Pipettes, Glass	VWR International, LLC; Radnor, PA, USA
Pipette Tips	Eppendorf AG; Hamburg, Germany Kisker Biotech GmbH & Co. KG; Steinfurt, Germany
Polymerase Chain Reaction (PCR) Plates	Kisker Biotech GmbH & Co. KG; Steinfurt, Germany
Pre-Separation Filters	Miltenyi Biotec; Bergisch Gladbach, Germany
Reaction Tubes (0.5ml, 1.5ml, 2.0ml)	Eppendorf AG; Hamburg, Germany Greiner Bio-One International GmbH; Kremsmünster, Austria
Sealing Film PCR	Kisker Biotech GmbH & Co. KG; Steinfurt, Germany
Sterile Scalpels	B. Braun Melsungen AG; Melsungen, Germany
Sterile Serological Pipettes (2ml, 5ml, 10ml, 25ml, 50ml)	Greiner Bio-One International GmbH; Kremsmünster, Austria
Syringes (1ml, 10ml, 20ml)	B. Braun Melsungen AG; Melsungen, Germany
Whatman Grade 3MM Chr Blotting Paper	GE Healthcare; Chicago, IL, USA

3.1.6 Chemicals and recipes

Table 3.10: Chemicals.

Product	Manufacturer/Distributor
10X PCR Buffer II	Applied Biosystems, Thermo Fisher Scientific Inc.; Waltham, MA, USA
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) Buffer Solution (1M)	Sigma-Aldrich; Saint Louis, MO, USA
87% Glycerol	AppliChem GmbH; Darmstadt, Germany
Acetic Acid (glacial) 100%	Merck KGaA; Darmstadt, Germany
Agarose, Low Gelling Temperature	Sigma-Aldrich; Saint Louis, MO, USA
AKASOLV Aqua Care	Carl Roth GmbH & Co. KG; Karlsruhe, Germany
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare; Chicago, IL, USA
Ammonium Peroxodisulfate (APS)	AppliChem GmbH; Darmstadt, Germany
Atipamezole	ORION Pharma GmbH; Hamburg, Germany
autoMACS Rinsing Solution	Miltenyi Biotec; Bergisch Gladbach, Germany
Bleomycin Sulfate (Bleo)	Sigma-Aldrich; Saint Louis, MO, USA
Bovine Serum Albumin (BSA)	Sigma-Aldrich; Saint Louis, MO, USA
Bromophenol Blue	AppliChem GmbH; Darmstadt, Germany
Collagen, Type I Solution	Sigma-Aldrich; Saint Louis, MO, USA
cOmplete, Mini, Ethylenediaminetetraacetic Acid (EDTA)-Free Protease Inhibitor Cocktail	Roche Diagnostics; Rotkreuz, Switzerland
Deoxyribonucleoside Triphosphate (dNTP) Mix (10mM each)	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Dimethyl Sulfoxide (DMSO)	Carl Roth GmbH & Co. KG; Karlsruhe, Germany
Disodium Hydrogen Phosphate (Na_2HPO_4)	AppliChem GmbH; Darmstadt, Germany
Distilled Water	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Dithiothreitol (DTT)	AppliChem GmbH; Darmstadt, Germany
Dulbecco's Phosphate-Buffered Saline (DPBS)	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA
EDTA	AppliChem GmbH; Darmstadt, Germany
Entellan Mounting Medium	Sigma-Aldrich; Saint Louis, MO, USA
Eosin Y Solution 0.5%	Carl Roth GmbH & Co. KG; Karlsruhe, Germany
Ethanol Absolute, p.a.	AppliChem GmbH; Darmstadt, Germany
Fentanyl	Janssen-Cilag GmbH; Neuss, Germany
Fetal Bovine Serum (FBS), Heat Inactivated	PAN-Biotech GmbH; Aidenbach, Germany
Flumazenil	Hexal AG; Holzkirchen, Germany
Glucose	AppliChem GmbH; Darmstadt, Germany

Product	Manufacturer/Distributor
GlutaMAX Supplement	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Glycine, p.a.	AppliChem GmbH; Darmstadt, Germany
Hank's Balanced Salt Solution (HBSS), Calcium, Magnesium, No Phenol Red	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Hemalum Solution Acid, Mayer	Carl Roth GmbH & Co. KG; Karlsruhe, Germany
HEPES Buffered Saline Solution	Lonza Group AG; Basel, Switzerland
Hydrochloric Acid (HCl)	AppliChem GmbH; Darmstadt, Germany
Isopropanol, p.a.	AppliChem GmbH; Darmstadt, Germany
Ketamine	bela-pharm GmbH & Co. KG; Vechta, Germany
LightCycler 480 SYBR Green I Master	Roche Diagnostics; Rotkreuz, Switzerland
MACS BSA Stock Solution	Miltenyi Biotec; Bergisch Gladbach, Germany
Magnesium Chloride ($MgCl_2$) (25mM)	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Medetomidine	ORION Pharma GmbH; Hamburg, Germany
Methanol, p.a.	AppliChem GmbH; Darmstadt, Germany
Midazolam	Roche Pharma AG; Grenzach-Wyhlen, Germany
N,N,N',N'-Tetramethylethylenediamine (TEMED)	AppliChem GmbH; Darmstadt, Germany
Naloxone	PUREN Pharma GmbH & Co. KG; München, Germany
Nonfat Dried Milk Powder	AppliChem GmbH; Darmstadt, Germany
Paraffin	Richard-Allan Scientific, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Paraformaldehyde (PFA)	AppliChem GmbH; Darmstadt, Germany
Penicillin-Streptomycin (10,000U/ml)	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA
PhosSTOP Phosphatase Inhibitor Cocktail Tablets	Roche Diagnostics; Rotkreuz, Switzerland
Ponceau S Solution	Sigma-Aldrich; Saint Louis, MO, USA
Potassium Chloride (KCl)	AppliChem GmbH; Darmstadt, Germany
Potassium Dihydrogen Phosphate (KH_2PO_4)	AppliChem GmbH; Darmstadt, Germany
Quick Start Bradford 1x Dye Reagent	Bio-Rad Laboratories, Inc.; Hercules, CA, USA
Random Hexamers (50μM)	Invitrogen, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Recombinant Human Transforming Growth Factor Beta 1 (TGF-β) Protein	R&D Systems, Inc.; Minneapolis, MN, USA
Restore PLUS Western Blot Stripping Buffer	Thermo Fisher Scientific Inc.; Waltham, MA, USA
RNAlater Stabilization Solution	Invitrogen, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Roti-Block	Carl Roth GmbH & Co. KG; Karlsruhe, Germany
Rotiphorese Gel 30 (37.5:1)	Carl Roth GmbH & Co. KG; Karlsruhe, Germany
Saline 0.9%	B. Braun Melsungen AG; Melsungen, Germany
Sodium Chloride (NaCl)	AppliChem GmbH; Darmstadt, Germany

Product	Manufacturer/Distributor
Sodium Dodecyl Sulfate (SDS)	Carl Roth GmbH & Co. KG; Karlsruhe, Germany
SuperSignal West Dura Extended Duration Substrate	Thermo Fisher Scientific Inc.; Waltham, MA, USA
SuperSignal West Femto Maximum Sensitivity Substrate	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Tissue Protein Extraction Reagent (T-PER)	Thermo Fisher Scientific Inc.; Waltham, MA, USA
Tris-(Hydroxymethyl)-Aminomethane (Tris)	AppliChem GmbH; Darmstadt, Germany
Trypan Blue Solution	Sigma-Aldrich; Saint Louis, MO, USA
Tween 20	AppliChem GmbH; Darmstadt, Germany
UltraPure DNase/RNase-Free Distilled Water	Invitrogen, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Xylazine	bela-pharm GmbH & Co. KG; Vechta, Germany
Xylene	AppliChem GmbH; Darmstadt, Germany

Table 3.11: Recipes.

Solution	Compound	Concentration/ Amount
(-)medium	Dulbecco's modified Eagle medium (DMEM)	500ml
	Glucose	1.8g
	GlutaMAX	10ml
	Penicillin-Streptomycin	5ml
	HEPES	5ml
(+)medium	(-)medium	250ml
	DNase I	10mg
APS 10%	APS	10g
	H_2O	100ml
DTT solution	DTT	0.61g
	H_2O	1ml
FACS buffer	Phosphate-buffered saline (PBS) (1x)	500ml
	BSA	2% (w/v)
	EDTA	2mM
Laemmli sample buffer (4x)	87% Glycerol	4ml
	Tris/HCl 1.5M pH 8.8	1.3ml
	Bromophenol blue	0.002g
	SDS	0.8g
	H_2O *(add up to a total volume of 9ml)*	
	DTT solution *(freshly added)*	1ml
PBS pH 7.4 (10x)	KCl	4g
	KH_2PO_4	4g
	Na_2HPO_4	28.8g
	NaCl	160g
	H_2O *(add up to a total volume of 2,000ml)*	
	pH is adjusted to 7.4	

Solution	**Compound**	**Concentration/ Amount**
Running buffer (10x)	Glycine	288.2g
	SDS	20g
	Tris	60.6g
	H_2O *(add up to a total volume of 2,000ml)*	
SDS 10%	SDS	10g
	H_2O	100ml
Transfer buffer (10x)	Glycine	288g
	Tris	60.6g
	H_2O *(add up to a total volume of 2,000ml)*	
Transfer buffer (1x)	Transfer buffer (10x)	100ml
	Methanol	200ml
	H_2O	700ml
Tris/HCl 0.5M pH 6.8	Tris	30.285g
	H_2O *(add up to a total volume of 500ml)*	
	HCl *(pH is adjusted to 6.8)*	
Tris/HCl 1.5M pH 8.8	Tris	90.855g
	H_2O *(add up to a total volume of 500ml)*	
	HCl *(pH is adjusted to 8.8)*	
Tris-buffered saline (TBS) pH 7.6 (10x)	NaCl	80g
	Tris	24.2g
	H_2O *(add up to a total volume of 1,000ml)*	
	HCl *(pH is adjusted to 7.6)*	
Tris-buffered saline with Tween 20 (TBST) (1x)	TBS pH 7.6 (10x)	100ml
	H_2O	900ml
	Tween 20	0.5ml
Western blot resolving gel (10%)	H_2O	4.0ml
	Rotiphorese gel 30	3.3ml
	Tris/HCl 1.5M pH 8.8	2.5ml
	SDS 10%	0.1ml
	APS 10%	0.1ml
	TEMED	4μl
Western blot stacking gel (5%)	H_2O	1.4ml
	Rotiphorese gel 30	0.33ml
	Tris/HCl 0.5M pH 6.8	0.25ml
	SDS 10%	0.02ml
	APS 10%	0.02ml
	TEMED	2μl

3.1.7 Standards and kits

Table 3.12: Standards.

Product	Manufacturer/Distributor
peqGOLD Protein Marker V	VWR International, LLC; Radnor, PA, USA
Quick Start Bovine Serum Albumin Standard (2mg/ml)	Bio-Rad Laboratories, Inc.; Hercules, CA, USA

Table 3.13: Kits.

Product	Manufacturer/Distributor
peqGOLD DNase I Digest Kit	VWR International, LLC; Radnor, PA, USA
peqGOLD Total Ribonucleic Acid (RNA) Kit	VWR International, LLC; Radnor, PA, USA
Roti-Quick-Kit	Carl Roth GmbH & Co. KG; Karlsruhe, Germany

3.1.8 Enzymes

Table 3.14: Enzymes.

Product	Manufacturer/Distributor
Dispase	Corning, Inc.; Corning, NY, USA
DNase I	AppliChem GmbH; Darmstadt, Germany
MuLV Reverse Transcriptase (50U/μl)	Applied Biosystems, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Pronase E	Sigma-Aldrich; Saint Louis, MO, USA
RNase Inhibitor (20U/μl)	Applied Biosystems, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Trypsin Neutralizing Solution	Lonza Group AG; Basel, Switzerland
Trypsin/EDTA Solution	Lonza Group AG; Basel, Switzerland
Trypsin-EDTA (0.25%)	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA

3.1.9 Cell culture media

Table 3.15: Cell culture media.

Product	Manufacturer/Distributor
Bronchial Epithelial Basal Medium (BEBM) ± BEGM SingleQuots	Lonza Group AG; Basel, Switzerland
DMEM - Low Glucose	Sigma-Aldrich; Saint Louis, MO, USA
DMEM/F-12	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA
Minimum Essential Media (MEM) - L-Glutamine	Gibco, Thermo Fisher Scientific Inc.; Waltham, MA, USA

3.1.10 Oligonucleotides

Primers for quantitative real-time polymerase chain reaction (qRT-PCR) were designed with the Primer-BLAST web application [97] (`https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi`; National Center for Biotechnology Information (NCBI); last accessed: 10 October 2018). Therefore, the required sequences were extracted from the NCBI nucleotide database (`https://www.ncbi.nlm.nih.gov/nuccore`; National Center for Biotechnology Information (NCBI); last accessed: 10 October 2018).

Primers were designed to display a melting temperature of 61-67°C (optimum 64°C) and to include all transcript variants of the respective gene. Furthermore, primers were required to be exon-exon junction spanning, to provide a PCR product of 80-150 base pairs (bp) length, and to display a low self complementarity with a GC content of 40-60%. The oligonucleotides were purchased from Eurofins Genomics Germany GmbH, Ebersberg, Germany. Reconstitution of the primers was performed with DNase/RNase-free distilled water. Subsequently, the primers were tested with serial dilutions of complementary deoxyribonucleic acid (cDNA) samples (undiluted, 1:8, 1:64, 1:512) for their efficiency, specificity, and the appearance of primer dimers.

Table 3.16: Human qRT-PCR primers.

Gene	Direction	Sequence (5' - 3')
ANLN	Forward	GTTCTGGACAAGGTCCCCTTT
	Reverse	CCAGGCACCAAAACCACTAAC
BUB1	Forward	CCTTCAAAACCAAAGGAGGAAGT
	Reverse	GTGAAGTCTCCTGGGCTCTT
CTHRC1	Forward	GTGGCTCACTTCGGCTAAAA
	Reverse	CACAAAGTCCTTCCACAGAAGA

Gene	Direction	Sequence (5' - 3')
GPR87	Forward	CCAGCCACCACAATGAAAGAAAT
	Reverse	CGCTCCTGTTGCCTGAATTG
HPRT	Forward	AAGGACCCCACGAAGTGTTG
	Reverse	GGCTTTGTATTTTGCTTTTCCA
PAI1	Forward	GACATCCTGGAACTGCCCTA
	Reverse	GGTCATGTTGCCTTTCCAGT
PSAT1	Forward	GTGATTGTCCGTGATGACCTG
	Reverse	TGACGTAGATGCTGAAACATGG

Table 3.17: Murine qRT-PCR primers.

Gene	Direction	Sequence (5' - 3')
Anln	Forward	GCCTCCAAGTCCAAGGCTAT
	Reverse	GGGTAAAGTTGCTGGTACGC
Bub1	Forward	GCCTGATGTTAGTCTAGTATGTGTT
	Reverse	CCTGTGGTTTCTCTCCTGAAC
Cthrc1	Forward	AGGGAGGTGGTAGACCTGTATAA
	Reverse	CCCCTTTGAATCCATCCCGA
Gpr87	Forward	CGCCACAATGAAAGAAATGAAACC
	Reverse	AGCCCCATTTTCTAGTTGTGATG
Hprt	Forward	CCTAAGATGAGCGCAAGTTGAA
	Reverse	CCACAGGACTAGAACACCTGCTAA
Psat1	Forward	TGGAAGGAGTGCTGACTACG
	Reverse	GCTTGGGTCTGGAATTTTTGTG

3.1.11 Antibodies

The primary antibodies for western blot were diluted in Roti-Block, which permitted the storage of the antibodies at 4°C for several weeks and multiple reuse of the antibodies.

Table 3.18: Primary antibodies, western blot.

Antibody	Manufacturer/Distributor	Species	Dilution
Anti-α-SMA (A5228)	Sigma-Aldrich; Saint Louis, MO, USA	mouse	1:2,000
Anti-β-ACTIN (A3854)	Sigma-Aldrich; Saint Louis, MO, USA	mouse	1:50,000
Anti-CNN1 (NB110-55650)	Novus Biologicals, LLC; Centennial, CO, USA	rabbit	1:1,000

Antibody	Manufacturer/Distributor	Species	Dilution
Anti-GPR87 (NBP2-16728)	Novus Biologicals, LLC; Centennial, CO, USA	rabbit	1:1,000
Anti-PSAT1 (PA5-22124)	Invitrogen, Thermo Fisher Scientific Inc.; Waltham, MA, USA	rabbit	1:1,000

Table 3.19: Secondary antibodies, western blot.

Antibody	Manufacturer/Distributor	Species	Dilution
Amersham ECL Mouse IgG, HRP-Linked (NA931)	GE Healthcare; Chicago, IL, USA	sheep	1:4,000
Amersham ECL Rabbit IgG, HRP-Linked (NA934)	GE Healthcare; Chicago, IL, USA	donkey	1:10,000

Table 3.20: Primary antibodies and isotype controls, FACS analysis.

Antibody	Manufacturer/Distributor	Species	Dilution
Anti-CD326 (EpCAM) (12-9326-42)	Invitrogen, Thermo Fisher Scientific Inc.; Waltham, MA, USA	mouse	1:50
Anti-GPR87 (LS-A1580)	LifeSpan Biosciences, Inc.; Seattle, WA, USA	rabbit	1:200
Mouse IgG1 Kappa Isotype Control, PE (12-4714-42)	Invitrogen, Thermo Fisher Scientific Inc.; Waltham, MA, USA	mouse	1:50

Table 3.21: Secondary antibodies, FACS analysis.

Antibody	Manufacturer/Distributor	Species	Dilution
Goat Anti-Rabbit IgG H&L (ab130805)	Abcam plc; Cambridge, United Kingdom	goat	1:50

3.2 Methods

3.2.1 In silico methods

3.2.1.1 Public transcriptome profiling datasets

For the subsequent study we used publicly available transcriptome profiling datasets from either microarray or RNA sequencing (RNA-Seq) experiments. Raw data was downloaded from Gene Expression Omnibus (GEO) [98, 99] (`https://www.ncbi.nlm.nih.gov/geo/`; National Center for Biotechnology Information (NCBI); last accessed: 10 October 2018) and the European Nucleotide Archive (ENA) [100] (`https://www.ebi.ac.uk/ena`; European Molecular Biology Laboratory, European Bioinformatics Institute; last accessed: 10 October 2018). Table 3.22 summarizes the analyzed datasets.

Table 3.22: Transcriptome profiling datasets.

Accession number	Samples	Description	Contributors
GSE47460 [101–104]	• Idiopathic pulmonary fibrosis (n = 121) • Controls (n = 91)	Transcriptome analysis of different interstitial lung diseases and chronic obstructive pulmonary disease (COPD), collected by the Lung Genomics Research Consortium.	Tedrow J, Kaminski N, Guardela BJ, Schwartz DA
GSE32537 [105]	• Idiopathic pulmonary fibrosis (n = 119) • Controls (n = 50)	Transcriptome analysis of different interstitial lung diseases.	Yang IV, Coldren CD, Leach SM, Murphy E, Lin J, Burton R, Groshong S, Cool C, Cosgrove GP, Lynch D, Brown KK, Schwarz MI, Fingerlin TE, Schwartz DA
GSE44077 [106]	• Non-small cell lung cancer (n = 37) • Controls (n = 45)	Transcriptome analysis of NSCLC and the proximate lung tissue.	Kadara H, Yoo S, Wistuba II
GSE43458 [107]	• Non-small cell lung cancer (n = 80) • Controls (n = 30)	Transcriptome analysis of lung adenocarcinoma in nonsmoker and smoker specimens.	Kadara H, Wistuba II

Accession number	Samples	Description	Contributors
GSE18842 [108]	• Non-small cell lung cancer (n = 46) • Controls (n = 45)	Transcriptome analysis of NSCLC.	Ramos AS, Morales MG, Gomez-Capilla JA, Muriel VP, Farez-Vidal ME
GSE10921 [109]	• IPF lung fibroblasts (n = 4) • Control lung fibroblasts (n = 3)	Transcriptome analysis of isolated human lung fibroblasts from IPF and control specimens.	Vuga LJ, Kaminski N
GSE94555 [110] *(SRP098915)*	• Idiopathic pulmonary fibrosis (n = 3) • Controls (n = 3)	Transcriptome analysis of human alveolar epithelial type II (ATII) cells in IPF.	Xu Y, Whitsett J, Stripp B, Mizuno T, Du Y
pmATII [111]	Each sample (n) consisted of six mice: • pmATII cells from bleomycin-induced experimental lung fibrosis (n = 3) • pmATII cells from control mice (n = 3)	Transcriptome analysis of isolated primary ATII cells from bleomycin-induced experimental lung fibrosis in mice.	Königshoff M, Kramer M, Balsara N, Wilhelm J, Amarie OV, Jahn A, Rose F, Fink L, Seeger W, Schaefer L, Günther A, Eickelberg O

3.2.1.2 Processing of RNA-Seq and microarray data

RNA-Seq data files were retrieved in the FASTQ data format and further handled with the Salmon v0.9.1 software [112]. The required transcriptome index was generated from the latest Ensembl GRCh38 cDNA release 91 transcriptome data [113] (`ftp://ftp.ensembl.org/pub/release-91/fasta/homo_sapiens/cdna`; European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Sanger Institute; last accessed: 10 October 2018) with the recommended k-mer size of 31. Subsequently, the raw RNA-Seq data files were processed with the quasi-mapping-based mode of the Salmon software. The results of the transcript-level abundance quantification were further analyzed with the R statistical software version 3.4.3 [114] and the RStudio version 1.1.383 [115], using the tximport package [116] to generate the definite gene-level count estimates. Eventually, the DESeq2 package [117] was applied to sort out those genes with less than 10 counts over all samples and to create a log2-based output of differential gene expression for subsequent experiments.

Microarray data files were handled with various software packages of the Bioconductor release 3.6 [95, 96]. The affy package [118] was used to achieve background correction, normalization, and log2 transformation of the expression values from Affymetrix microarray files. Subsequently, the Biobase [95] and limma [119] packages were applied to generate an output file of differential gene expression for further analyses.

3.2.1.3 Differentially expressed genes

Microsoft Excel was used to assign the HUGO gene nomenclature committee (HGNC) gene symbols to the respective microarray probe IDs or Ensembl gene IDs of each data output (3.2.1.2). Entries without an assigned HGNC gene symbol were excluded from further analyses. Subsequently, differentially expressed genes (DEGs) were extracted and the respective gene sets were created. DEGs were specified as those genes with a $|\log2 \text{ fold change}| > 1.0$ of expression between two conditions (e.g. disease and control) and an adjusted p-value < 0.05 [120].

3.2.1.4 Venn diagram

Venn diagrams represent a significant tool to compare multiple lists of items. Thus, they ease the creation and visualization of the overlap of predefined sets of genes. HGNC gene symbols of up to three gene sets were extracted and the creation of Venn diagrams was achieved with R statistical software version 3.4.3 [114], the RStudio version 1.1.383 [115], and the `venn.diagram` function of the VennDiagram package [121] with custom parameters. All genes included in a Venn diagram were set as 100 percent and the percentage of the respective areas was calculated [120]. Venn diagrams are not pictured area-proportional. Finally, gene lists of the overlapping areas were generated using the `calculate.overlap` function.

3.2.1.5 Annotation enrichment analysis

To analyze the enrichment of gene annotations, the public reference databases of the Gene Ontology (GO) Consortium [122, 123], the Kyoto Encyclopedia of Genes and Genomes (KEGG) [124, 125], and the REACTOME project [126] were utilized. The actual annotation enrichment analyses were conducted either with the stringApp v1.3.2 [127, 128] of the Cytoscape software platform v3.6.1 [129] or the GeneSCF v1.1 tool [130]. Results were further analyzed with Microsoft Excel. KEGG and REACTOME pathways were stated significantly overrepresented if the analysis encountered more than two genes of a particular pathway in a predefined set of genes, combined with a Benjamini and Hochberg false discovery rate (FDR) < 0.05 [120]. GO terms were stated significantly overrepresented if the analysis encountered at least two genes of a specific GO term in a predefined set of genes, combined with a FDR < 0.05 [120].

3.2.1.6 Protein-protein interaction networks

The stringApp v1.3.2 [127, 128] and the yFiles Layout Algorithms v1.0 of the Cytoscape software platform v3.6.1 [129] were used to generate protein-protein interaction (PPI) networks. The limit for the confidence score of protein-protein interactions was set to 0.9 for the subsequent compilation. Further analysis of the PPI networks was performed with the stringApp v1.3.2.

3.2.1.7 BioGPS

The gene annotation database BioGPS [131–133] (`http://biogps.org/#goto=welcome`; The Scripps Research Institute (TSRI); last accessed: 25 January 2019) was used to explore the tissue- and cell-specific expression of single genes. The "human gene atlas" microarray dataset (*GSE1133*) [134] was selected for the analysis. Results were saved as high resolution images. Expression values were presented as arbitrary fluorescence units.

3.2.1.8 Principal component analysis

Principal component analysis (PCA) was performed with the R statistical software v3.4.3 [114] and the RStudio v1.1.383 [115], using the R packages FactoMineR v1.41 [135], devtools v1.13.6, and factoextra v1.0.5. Gene expression values were extracted from the respective microarray dataset (Table 3.22). Subsequently, the `PCA` function was used for the actual principal component analysis. Scree plots were generated with the `fviz_screeplot` function and PCA biplots with the `fviz_pca_biplot` function.

3.2.1.9 Gene set enrichment analysis

Microarray and RNA-Seq data (Table 3.22), as well as the generated lists of DEGs (3.2.1.3) were prepared for gene set enrichment analysis (GSEA) [136, 137] according to the GSEA Wiki (`http://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats`; Broad Institute of the Massachusetts Institute of Technology and Harvard; last accessed: 25 January 2019) and the respective data files were created (Table 3.23).

Table 3.23: GSEA data files.

Data file	Extension	Content
Gene cluster text file format	*.gct	unranked datasets
Ranked list file format	*.rnk	pre-ranked datasets
Gene matrix file format	*.gmx	gene set files
Categorical class file format	*.cls	phenotype data

Subsequently, the required data files were imported and gene set enrichment analyses were executed with the GSEA software v3.0 [136, 137]. Basic and required settings were applied as listed below (Table 3.24). The *"phenotype"* permutation type was used for datasets with ≥ 14 samples, *"gene_set"* was used for datasets with less than 14 samples. Default options were selected for the advanced settings. Gene sets were considered to be significantly enriched with a nominal p-value < 0.05 and a FDR q-value < 0.05 [120]. The enrichment plots and heatmaps were created with the GSEA software.

Table 3.24: GSEA software settings.

Option	**Value**
Number of permutations	1,000
Permutation type	phenotype OR gene_set
Enrichment statistic	weighted
Metric for ranking genes	signal2noise
Gene list sorting mode	real
Gene list ordering mode	descending

3.2.1.10 Statistical analysis

In general, either GraphPad Prism version 5.00 for Windows or the R statistical software v3.4.3 with RStudio v1.1.383 was used to conduct the statistical analyses. Furthermore, all statistics on the enrichment of gene sets were computed automatically by the GSEA software v3.0. Experimental groups were analyzed with the paired t-test (two-tailed), the unpaired t-test (two-tailed), the repeated measures ANOVA with Bonferroni's multiple comparison test, or the one-way ANOVA with Bonferroni's multiple comparison test. Association of gene expression and pulmonary function was assessed with the Pearson correlation analysis and linear regression. Further details on the respective statistical method are described in the corresponding figure legend or paragraph of the results section. If applicable, all values and data are presented as mean ± standard deviation (SD). Results were stated statistically significant with a p-value < 0.05 (significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$).

3.2.2 Animal model of bleomycin-induced pulmonary fibrosis

The official instructions and guidelines on animal experiments by the Helmholtz Zentrum München were applied and all animal experiments were beforehand authorized by the district government of Upper Bavaria, Germany (AZ: 55.2-1-54-2532-88-2012).

To standardize the mouse model used for this thesis, the wild-type C57BL/6N mouse strain with an average age of six to eight weeks and body weight (BW) of 21-23g was used for all experiments. All mice were subject to standard housing conditions, including unrestricted access to drinking water and laboratory rodent diet.

Subsequently, mice were anesthetized with an intraperitoneal (i.p.) injection of medetomidine (500μg/kg BW), midazolam (5mg/kg BW), and fentanyl (50μg/kg BW) (MMF) and placed on a warming plate to keep their body temperature equal during the remaining process. The sedated mice were randomly divided into the *control* or *bleomycin group*, weighed, marked with ear punches, and intratracheally intubated with an intravenous 20G catheter. Mice of the *bleomycin group* were administered a single dose of 2U/kg BW bleomycin sulfate (Bleo) dissolved in sterile Dulbecco's phosphate-buffered saline (DPBS) [120]. Mice of the *control group* were administered 50μl sterile DPBS per animal [120]. Administration into the lung was achieved with the MicroSprayer Aerosolizer. The anesthesia was antagonized by a subcutaneous (s.c.) injection of atipamezole (2.5mg/kg BW), flumazenil (500μg/kg BW), and naloxone (1200μg/kg BW).

Mice were housed for another 14 days to develop bleomycin-induced pulmonary fibrosis while the BW was controlled on day (d) 1, d7, d10, and d14. Eventually, mice were sacrificed by i.p. injection of ketamine (100mg/kg BW) combined with xylazine (5mg/kg BW) and subsequent exsanguination. To remove all of the blood, lungs were washed out through the right ventricle of the heart with 20ml of 0.9% saline and eventually harvested.

3.2.3 Cell biological methods

3.2.3.1 Isolation of primary murine alveolar epithelial type II cells

Isolation of primary murine alveolar epithelial type II (pmATII) cells was performed by Julia Kipp, Anastasia van den Berg, and Kathrin Mutze. The applied method has been published [111, 138] and was further modified by Mutze *et al.* [139].

Harvested lungs (3.2.2) were intratracheally intubated to inject 1.5ml of dispase and 0.3ml of 1% low gelling temperature agarose per lung. Afterwards, lungs were rinsed from the outside with 0.9% saline and placed in 1.0ml dispase for 45min at room temperature. Isolation of pmATII cells was further accomplished by separating each lung into its lobes, which were placed in 5ml (+)medium and minced with forceps. The cell suspension was filtered consecutively with 100μm, 20μm, and 10μm nylon filters. Subsequently, the filtrate was centrifuged for 10min at 200g/RCF and 15°C. The remaining cell pellet was resuspended in 1.0ml (+)medium and supplemented with (-)medium to a total amount of 40ml, followed by incubation for 30min at 37°C in a humidified CO_2 incubator on uncoated petri dishes. Supernatant with floating cells was collected and centrifuged for 10min at 200g/RCF and 15°C, whereas adherent fibroblasts were discarded.

MACS cell separation buffer was added and cells were counted using the Neubauer chamber. For depletion of $CD31^+$ endothelial cells and $CD45^+$ leukocytes, 90μl of MACS cell separation buffer and 10μl of mouse CD31 and CD45 MicroBeads were added per 10^7 cells. The MACS separator was used, and the flow-through was collected and centrifuged for 10min at 200g/RCF and 15°C. Eventually, the remaining cell pellet was resuspended in (-)medium with 10% FBS and cells were counted using the Neubauer chamber. Cell viability was assessed and each sample was pooled of four individual mice. The successful isolation of pmATII cells was continuously monitored by verifying the cell purity with

fluorescence-activated cell sorting (FACS) or immunofluorescence. Samples were either snap-frozen or directly used for cell culture (3.2.3.3).

3.2.3.2 Isolation of primary human bronchial epithelial cells

Isolation of primary human bronchial epithelial cells (phBECs) was performed by Anastasia van den Berg according to the previously published protocol by Smirnova *et al.* [140] with modifications.

Only healthy lung tissue with regular histology was included for the isolation of phBECs. The human lung specimens were placed in cold minimum essential media (MEM) complemented with L-glutamine and 1% penicillin/streptomycin. Peripheral airways were carefully dissected, the surrounding tissue was removed, and the remaining bronchi were sliced lengthwise. The bronchial specimens were thoroughly rinsed with MEM complemented with L-glutamine and 1% penicillin/streptomycin and placed in 1.0mg/ml Pronase E for 20h at 4°C.

Subsequently, the epithelial cells were detached by scraping on the inner bronchial surface with a scalpel. Cell aggregates were further divided and a single-cell suspension was created by aspirating the cells multiple times with 20G and 26G cannulas. Finally, the cells were centrifuged for 5min at 300g/RCF and 4°C, and the cell pellet was resuspended in BEBM basal medium supplemented with BEGM SingleQuots and 1% penicillin/streptomycin. This cell suspension was incubated for 2-3h at 37°C in a humidified CO_2 incubator on uncoated petri dishes. Adherent fibroblasts were discarded, while the supernatant with isolated phBECs was used for cell culture (3.2.3.3) or frozen.

3.2.3.3 Cell culture

If applicable, all cells were stored frozen in liquid nitrogen with 10% dimethyl sulfoxide (DMSO) added. Before the initial seeding procedure, the frozen cells were quickly thawed in a warm water bath at 37°C and resuspended in 13ml of their respective medium (Table 3.25). This cell suspension was centrifuged for 10min at 300g/RCF and room temperature, the supernatant was aspirated, and the remaining cell pellet was resuspended in the appropriate cell culture medium with 1% penicillin/streptomycin (Table 3.25) and seeded in cell culture flasks or petri dishes.

All cell lines and isolated primary cells were cultured in their specific medium complemented with 1% penicillin/streptomycin and FBS or individual growth supplements (Table 3.25), in a humidified CO_2 incubator with 37°C, 5.0% CO_2, and 96.0% relative humidity (%rH).

Passaging of all cell lines was carried out with their respective split ratio (Table 3.25), as soon as the cell density reached approximately 80% of the cell culture flask or petri dish. In this case, cells were washed with prewarmed sterile DPBS and incubated with 2ml trypsin for 5min at 37°C. For enhanced detachment, the cell culture flask or petri dish was gently tapped against the hand. Cell culture medium was added to inactivate the trypsin digestion. By thoroughly pipetting, a single-cell suspension was created and

resuspended in its appropriate medium in a new cell culture flask or petri dish. Cell lines were used and subcultured up to passage (p) 25.

Primary human bronchial epithelial cells were cultured on 10cm petri dishes coated with collagen type 1 solution. The BEBM basal medium supplemented with BEGM SingleQuots and 1% penicillin/streptomycin was changed every second day until approximately 80% confluence was reached. Cells were washed with 5ml of HEPES buffered saline solution to prevent the inactivation of trypsin. Subsequently, the cells were incubated in 2ml trypsin/EDTA solution for 3-5min at 37°C and further detached by taping against the petri dish. Trypsin was deactivated with 4ml trypsin neutralizing solution. The cell suspension was centrifuged for 5min at 300g/RCF and room temperature and resuspended in BEBM basal medium supplemented with BEGM SingleQuots and 1% penicillin/streptomycin. Undifferentiated phBECs were cultured from p0 to p4.

Cell culture of primary murine alveolar epithelial type II cells was only possible directly after the isolation. In general, the pmATII cells were cultured for at most 5 days because of increasing alterations in the cell morphology [120].

Table 3.25: Cell culture media, additives, and split ratios.

Cell type	Medium and additives	Split ratio
16HBE14o$^-$	MEM + L-glutamine + 10% FBS	1:3
A549	DMEM/F-12 + 10% FBS	1:10
Primary human bronchial epithelial cells	BEBM + BEGM SingleQuots	1:3, 1:4
Primary murine ATII cells	(-)medium + 10% FBS	none

3.2.3.4 Cell treatments

For the treatment of cells, a single-cell suspension was created and cells were counted using the Neubauer chamber. The indicated number of cells was seeded in cell culture well plates or petri dishes (Table 3.26) and incubated in their appropriate medium for 24h at 37°C (Table 3.25).

Table 3.26: Number of cells and cell culture dish for cell treatments.

Cell type	Number of cells	Cell culture dish
16HBE14o$^-$	300,000	6-well cell culture plate
A549	300,000	6-well cell culture plate
Primary human bronchial epithelial cells	10,000 cells/cm^2	6-well cell culture plate, petri dishes
Primary murine ATII cells	1,000,000	12-well cell culture plate

Subsequently, the corresponding starvation medium with 0.1% FBS or without the BEGM SingleQuots was applied and the cells were starved for 24h at 37°C, to synchronize their cell cycle and to deprive the cells of the present growth factors. Finally, the specific treatment was added for the indicated period of time (Table 3.27). The treatment was stopped by washing the cells with cold sterile DPBS. Cell culture well plates or petri dishes were stored at -80°C for further processing.

Table 3.27: Cell treatment, concentrations, and time points.

Cell treatment	Concentrations	Time points
TGF-β	2ng/ml	24h
	5ng/ml	4h, 6h, 24h

3.2.4 RNA expression analysis

3.2.4.1 RNA isolation

The complete process of RNA isolation from cell and tissue specimens was achieved with the peqGOLD Total RNA Kit by adopting the manufacturer's protocol with some modifications.

3.2.4.1.1 Cells Cells were lysed by adding 350μl of RNA lysis buffer per well and thoroughly scraping the cells from the cell culture plate with a cell lifter. The lysate was pipetted onto deoxyribonucleic acid (DNA) removing columns, centrifuged for 1min at 12,000g/RCF and 25°C, and 70% ethanol was added (1:1 ratio). This mixture was transferred onto PerfectBind RNA columns. Hereafter, the same protocol was used for the RNA isolation of cells and tissue samples (3.2.4.1.2).

The PerfectBind RNA columns were centrifuged for 1min at 10,000g/RCF and 25°C, the flow-through was discarded, 400μl RNA wash buffer 1 were added, and the columns were again centrifuged for 30sec at 10,000g/RCF and 25°C. The remaining DNA was digested by using the peqGOLD DNase I Digest Kit with 15min incubation time at room temperature. Subsequently, the PerfectBind RNA columns were washed successively with 400μl RNA wash buffer 1 and twice with 600μl RNA wash buffer 2. The columns were dried by centrifugation for 2min at 10,000g/RCF and 25°C. Eventually, the RNA was eluted by adding 43μl of RNase-free water to each column and incubation for 5min at room temperature. The isolated RNA was stored at -80°C.

3.2.4.1.2 Tissue Tissue specimens were homogenized in cryogenic tubes with 9mm steel balls, using the Mikro-Dismembrator S with 3,000 revolutions per minute (rpm) for 30sec. The homogenized tissue powder was dissolved in 1ml Roti-Quick 1 of the Roti-Quick-Kit and incubated for 20min on ice. Subsequently, the Mikro-Dismembrator S was used again with 3,000rpm for 10sec and 1ml of Roti-Quick 2 was added. The tubes

were incubated for another 10min on ice and centrifuged for 15min at 13,000g/RCF and 4°C. The upper aqueous layer was recovered for further processing, and 1ml of 70% ethanol was added. This mixture was pipetted onto PerfectBind RNA columns. The samples were further processed as described in 3.2.4.1.1.

3.2.4.2 Determination of RNA concentration

The RNA concentration was determined using the NanoDrop 1000 with 1.0μl of isolated RNA. RNase-free water was used for the calibration and as the blank measurement. A 260/280 ratio of absorbance greater or equal to 2.0 for RNA was defined as pure RNA.

3.2.4.3 Synthesis of complementary DNA

The synthesis of complementary DNA (cDNA) was achieved with the Mastercycler nexus. Therefore, a total amount of either 2μg or 1μg RNA diluted in 20μl RNase-free water was denaturated for 10min at 70°C and subsequently cooled down to 4°C. Reverse transcription was started by adding 20μl of the mixed reagents (Table 3.28), followed by 10min incubation at 20°C, 75min at 43°C, 5min at 99°C, and cooling down to 4°C. In general, the cDNA was diluted with RNase-free water in a 1:5 ratio and stored at -20°C.

Table 3.28: Master mix for synthesis of cDNA by reverse transcription.

Reagent	**Stock concentration**	**Final concentration**	**Volume**
10X PCR Buffer II	100mM	20mM	4μl
$MgCl_2$ solution	25mM	10mM	8μl
dNTP mix	10mM	1mM	2μl
Random Hexamers	50μM	5μM	2μl
RNase Inhibitor	20U/μl	1U/μl	1μl
MuLV Reverse Transcriptase	50U/μl	5U/μl	2μl
H_2O			1μl
Total volume			20μl

3.2.4.4 Quantitative real-time polymerase chain reaction

Quantitative real-time polymerase chain reaction (qRT-PCR) was performed with a total volume of either 10μl or 20μl per sample, using the LightCycler 480 Instrument II. The corresponding compositions of the master mix for qRT-PCRs are shown in table 3.29. The specific qRT-PCR conditions are indicated in table 3.30. In general, all samples were pipetted at least as duplicates, and for every qRT-PCR no template controls (NTC) of RNase-free water were included. Normalization of the messenger ribonucleic acid

(mRNA) expression data was achieved by including the housekeeping gene human or mouse hypoxanthine-guanine phosphoribosyltransferase (*HPRT*, *Hprt*) for every sample.

The qRT-PCR results were analyzed using the LightCycler 480 software version 1.5. Melting curves were generated for every primer pair and analyzed for contamination of the samples or primer dimers. The mRNA expression data and therefore the amount of gene transcripts were calculated and presented as ΔCt values. This was achieved by the following calculation $\Delta Ct\ value = Ct^{\ housekeeping\ gene}$ - $Ct^{\ target}$.

Table 3.29: Master mix for qRT-PCRs.

Reagent	Stock concentration	Final concentration	Volume human	Volume mouse
LightCycler 480 SYBR Green I Master	2x	1x	10μl	5μl
Primer mix	10μM Forward primer 10μM Reverse primer	0.5μM Forward primer 0.5μM Reverse primer	1μl	0.5μl
H_2O			4μl	2μl
cDNA			5μl	2.5μl
Total volume			20μl	10μl

Table 3.30: Specific qRT-PCR conditions.

qRT-PCR stage	Number of cycles	Temperature	Time
Pre-incubation	1x	50°C	2min
Initial denaturation	1x	95°C	5min
Denaturation Annealing Extension	45x	95°C 59°C 72°C	5sec 5sec 10sec
Melting curve	1x	95°C 60°C 97°C	5sec 1min continuous
Cooling	1x	40°C	30sec

3.2.5 Protein analysis

3.2.5.1 Protein isolation

Frozen tissue specimens were homogenized in cryogenic tubes with 9mm steel balls, using the Mikro-Dismembrator S with 3,000rpm for 30sec. The homogenized tissue powder was dissolved in 600μl of the lysis buffer, which was made up of 10ml tissue protein extraction reagent (T-PER) supplemented with one tablet of cOmplete, Mini, EDTA-free

protease inhibitor cocktail and PhosSTOP phosphatase inhibitor, respectively. Lysates were incubated on ice for at least 30min, thoroughly mixed by pipetting, and eventually centrifuged for 30min at 14,000g/RCF and 4°C. The cell pellet was discarded, and the protein containing supernatant was collected and stored at -80°C for further analysis.

3.2.5.2 Preparation of protein samples

The protein concentrations of the isolated samples were quantified with the Bradford protein assay by adopting the manufacturer's protocol. The quick start BSA standard (2mg/ml) was diluted to prepare BSA samples with the concentrations of 2,000μg/ml, 1,500μg/ml, 1,000μg/ml, 750μg/ml, 500μg/ml, 250μg/ml, 125μg/ml, and 0μg/ml. Subsequently, 5μl of each protein sample and BSA standard were pipetted into separate wells of a 96-well plate and 250μl quick start Bradford 1x dye reagent were added per well. All protein samples and BSA standards were pipetted at least as duplicates. In a final step, the absorbance at 595nm was measured with the Sunrise absorbance microplate reader, the BSA standard curve was generated, and the final protein concentrations were calculated.

For western blot analysis, an equal amount of 10-15μg protein per sample was mixed with Laemmli sample buffer, dithiothreitol (DTT), and T-PER. Subsequently, the samples were incubated at 95°C for 5min, cooled down on ice, and stored at -20°C.

3.2.5.3 Western blot analysis

Equal amounts of the preprocessed protein samples were loaded onto 10% SDS polyacrylamide gels and SDS-polyacrylamide gel electrophoresis (Page) was performed with the Mini-PROTEAN tetra vertical electrophoresis cell and running buffer (1x). A voltage of 90V was applied for 15min, which allowed for even migration of all protein samples. Subsequently, the voltage was increased to 120V for 90min.

The transfer of protein samples from 10% SDS polyacrylamide gels to nitrocellulose membranes was achieved in transfer buffer (1x) by wet electroblotting with 300mA for 90min. Subsequently, the nitrocellulose membranes were blocked for 1h at room temperature in either Roti-Block or TBST (1x) with 5% nonfat dried milk powder added, depending on the respective primary antibody. Afterwards, the required primary antibodies were diluted with Roti-Block in the indicated ratio (Table 3.18), followed by incubation of the membranes in the respective primary antibody solution for 1h at room temperature or overnight at 4°C.

The blots were thoroughly washed three times with TBST (1x) and subsequently incubated for 1h at room temperature in the appropriate secondary antibody (Table 3.19), which was diluted in TBST (1x) with 5% nonfat dried milk powder added. Thereafter, the blots were washed again three times with TBST (1x). The detection of proteins was enabled with enhanced chemiluminescence, which was imaged with the ChemiDoc XRS+ system. To estimate the correct protein size, the peqGOLD protein-marker V was included for every membrane. Furthermore, β-ACTIN was used as loading control.

3.2.5.4 Densitometric analysis

The western blot results were saved as image files and analyzed with the Image Lab software version 5.2.1. The volume and intensity of each band was calculated and exported. Furthermore, each band intensity was normalized using the corresponding β-ACTIN band intensity.

3.2.6 Histology

3.2.6.1 Preparation of lung tissue specimens

Fixation of either lung tissue blocks (human) or whole lungs (mouse) was accomplished by incubation of the tissue sections in 4% PFA for 18-24h. Subsequently, the indicated process (Table 3.31) was executed with the Microm STP420D tissue processor and the processed lung specimens were embedded in paraffin blocks with the Microm EC 350 modular tissue embedding center. The paraffin-embedded tissue blocks were stored at 4°C and sliced into 3-4μm thick sections with the Hyrax M55 rotary microtome. Tissue slices were mounted onto microscope slides.

Table 3.31: Protocol for the Microm STP420D tissue processor.

Program cycle	Reagent	Time
1	4% PFA	60min
2	4% PFA	60min
3	50% ethanol	60min
4	70% ethanol	60min
5	96% ethanol	60min
6	96% ethanol	60min
7	100% ethanol	60min
8	100% ethanol	60min
9	Xylene	60min
10	Xylene	60min
11	Paraffin	30min
12	Paraffin	45min
13	Paraffin	45min
14	Paraffin	45min

3.2.6.2 Hematoxylin and eosin staining

To perform the hematoxylin and eosin (H&E) staining, the mounted tissue slices were deparaffinized and rehydrated (Table 3.32). Subsequently, tissue slices were incubated in hematoxylin for 5min, rinsed with distilled H_2O, placed in tap water for 15min, and finally dipped in 70% ethanol with 0.3% HCl for differentiation.

The mounted tissue slices were washed again with tap water for 5min and subsequently rinsed with distilled H_2O for 2min. This was followed by the incubation in Eosin Y solution for 3min and another washing step with distilled H_2O.

In a final step, the tissue was dehydrated with ethanol (70%, 80%, 90%, and 100%) and incubated twice in xylene for 5min. The slides were sealed for long-term storage by mounting a cover glass with Entellan mounting medium.

Table 3.32: Deparaffinization and rehydration of tissue slices.

Stage	Reagent	Time
1	Xylene	5min
2	Xylene	5min
3	100% ethanol	2min
4	100% ethanol	2min
5	90% ethanol	1min
6	80% ethanol	1min
7	70% ethanol	1min
8	Distilled H_2O	30sec

3.2.7 Fluorescence-activated cell sorting

Initially, a single-cell suspension was created and cells were counted using the Neubauer chamber (see 3.2.3.3). The required number of cells was transferred into the wells of a 96-well plate for the staining procedure. A total number of 250,000 cells per single staining and 500,000 cells as unstained control were used. Cells were washed twice with FACS buffer and incubated in the dark for 30min at 4°C in the primary antibody, which was diluted in FACS buffer (Table 3.20). Thereafter, cells were washed again twice with the FACS buffer.

If required, cells were incubated in the dark for 30min at 4°C in the appropriate secondary antibody (Table 3.21). Subsequently, the cells were washed twice with FACS buffer. Isotype controls were always included to check for non-specific background staining (Table 3.20). Finally, cells were incubated in 4% PFA for 15min at room temperature, washed twice with FACS buffer, and transferred into FACS tubes.

The fluorescence-activated cell sorting was performed with the BD LSR II flow cytometer, and the results were analyzed with the FlowJo software version 9.6.4.

4 Results

Indicated parts of the presented results have been previously published in Ulke *et al.* [120]. Adapted and reprinted with permission of the American Thoracic Society (p. III).

4.1 Common pattern of gene expression alterations in IPF and NSCLC

Given the various reports about an epidemiological correlation and similarities in the pathomechanisms of IPF and NSCLC, we hypothesized that further elucidating this possible relationship between the two diseases might lead to a better understanding of the IPF pathogenesis. Initially, we sought to identify and compare differentially expressed genes (DEGs) from IPF and NSCLC patients by analyzing microarray data.

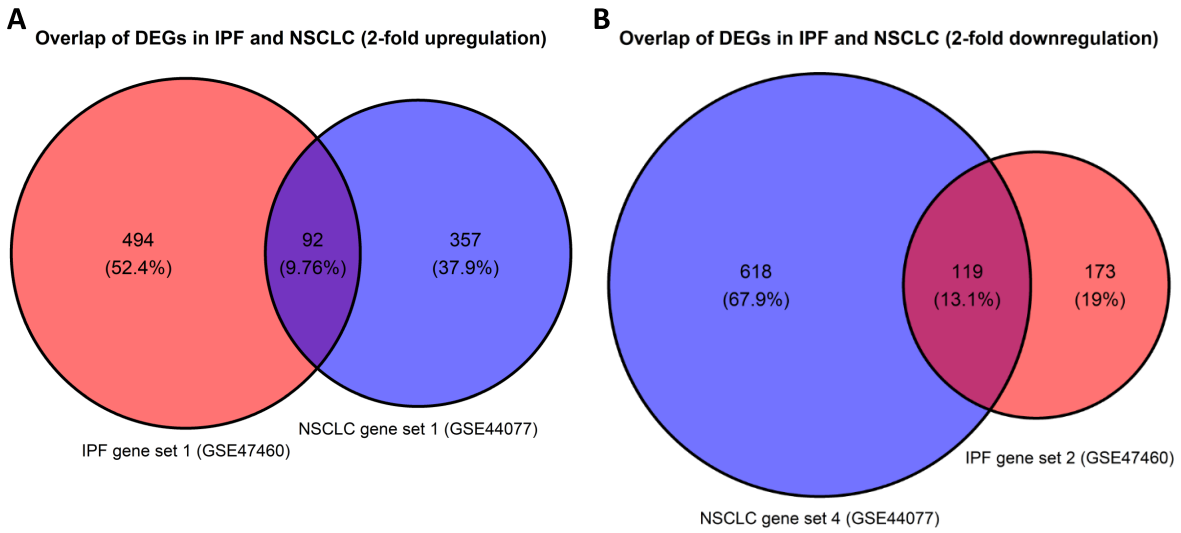

Figure 4.1: Overlap of DEGs from IPF and NSCLC. Venn diagrams were generated from gene sets that included the DEGs (|log2 fold change| > 1; adjusted p-value < 0.05) of the IPF (*GSE47460*) or NSCLC (*GSE44077*) microarray dataset. Overlap of **(A)** upregulated (9.76%; n = 92) and **(B)** downregulated DEGs (13.1%; n = 119) from IPF and NSCLC. Total number of genes in the respective Venn diagram was defined as 100 percent.

To start with, the two publicly available microarray datasets *GSE47460* (chronic lung diseases) [101, 103, 104] and *GSE44077* (NSCLC) [106] were downloaded from Gene Expression Omnibus (GEO) [98, 99] (`https://www.ncbi.nlm.nih.gov/geo/`; National Center for Biotechnology Information (NCBI); last accessed: 10 October 2018). The

GSE47460 microarray dataset consisted of samples from COPD, interstitial lung diseases, and control cases [101, 103]. Therefore, the dataset was reworked by retaining only the IPF (n = 121) and control samples (n = 91). Afterwards, differentially expressed genes (|log2 fold change| > 1 and adjusted p-value < 0.05) were extracted from both datasets. Gene sets were created by including either up- or downregulated genes from the respective dataset and were labeled as IPF gene set 1 (*GSE47460*; upregulated DEGs; n = 586), IPF gene set 2 (*GSE47460*; downregulated DEGs; n = 292), NSCLC gene set 1 (*GSE44077*; upregulated DEGs; n = 449), and NSCLC gene set 4 (*GSE44077*; downregulated DEGs; n = 737). The generated overlap (Figure 4.1) revealed a shared subset of equally misregulated genes in IPF and NSCLC. These results gave a first impression of possible similarities among DEGs in both diseases but required further validation.

4.1.1 Significant enrichment of DEGs from NSCLC in the GSE47460 microarray

Gene set enrichment analysis (GSEA) allows the parallel analysis of multiple gene lists for their enrichment in one of two individual phenotypes (e.g. IPF *vs.* control) from a single dataset [136, 137]. Therefore, GSEA was used to verify and statistically quantify the aforementioned overlap of DEGs from IPF and NSCLC (Figure 4.1).

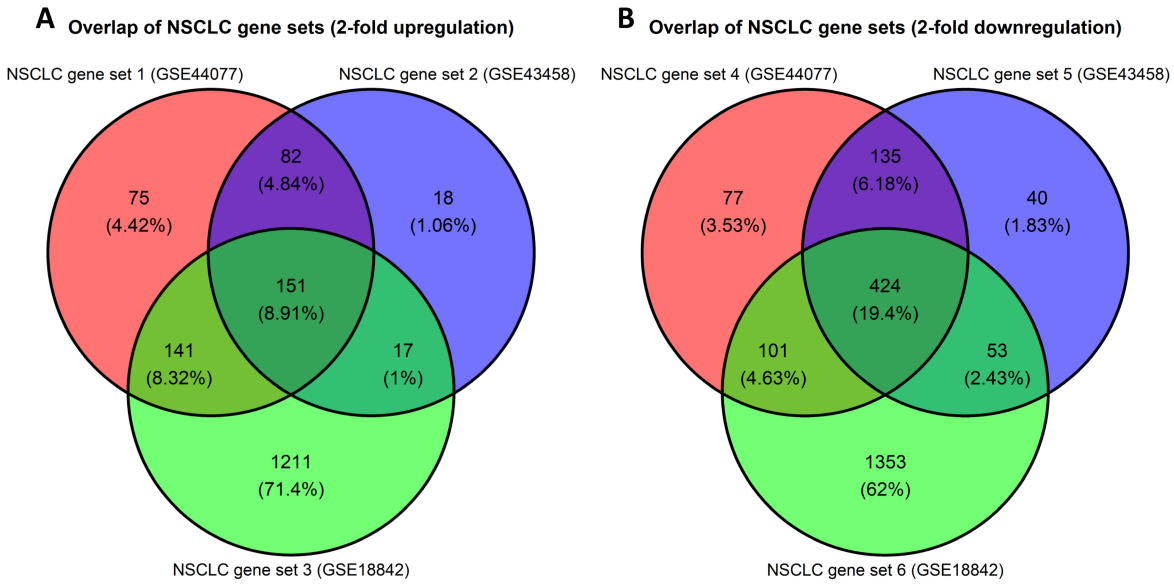

Figure 4.2: Overlap of DEGs from three individual NSCLC microarray datasets. The NSCLC gene sets 1-6 of DEGs (|log2 fold change| > 1; adjusted p-value < 0.05) from three independent NSCLC microarray datasets (*GSE44077*, *GSE43458*, *GSE18842*) were combined as Venn diagrams. Overlap of **(A)** upregulated (8.91%; n = 151) and **(B)** downregulated DEGs (19.4%; n = 424) in NSCLC. Total number of genes in the respective Venn diagram was defined as 100 percent.

To further increase the significance of the GSEA, the two additional and independent NSCLC microarray datasets *GSE43458* [107] and *GSE18842* [108] were downloaded from GEO. Subsequently, the DEGs were identified and added to the respective gene sets, which were labeled as NSCLC gene set 2 (*GSE43458*; upregulated DEGs; n =

268), NSCLC gene set 5 (*GSE43458*; downregulated DEGs; n = 652), NSCLC gene set 3 (*GSE18842*; upregulated DEGs; n = 1520), and NSCLC gene set 6 (*GSE18842*; downregulated DEGs; n = 1931). Comparison of the DEGs from all three NSCLC datasets showed a surprisingly small central overlap for upregulated (8.91%; n = 151) and downregulated genes (19.4%; n = 424) (Figure 4.2). Hence, these data demonstrated the individuality of the different NSCLC microarray datasets and confirmed the importance of including all three in the subsequent GSEA with their corresponding gene sets.

GSEA was conducted, using the reworked *GSE47460* microarray dataset with its associated phenotypes (IPF *vs.* control) as the rank-ordered gene list. The enrichment of the NSCLC gene sets 1-6 was analyzed and is demonstrated in Figure 4.3. The NSCLC gene sets 1-3 (Figure 4.3A-C) showed a significant enrichment (FDR q-value < 0.05; Nominal p-value < 0.05) in the IPF phenotype, while the NSCLC gene sets 4-6 (Figure 4.3D-F) showed a significant negative enrichment (FDR q-value < 0.05; Nominal p-value < 0.05) in the control phenotype [120]. Enrichment results are shown in Table 4.1.

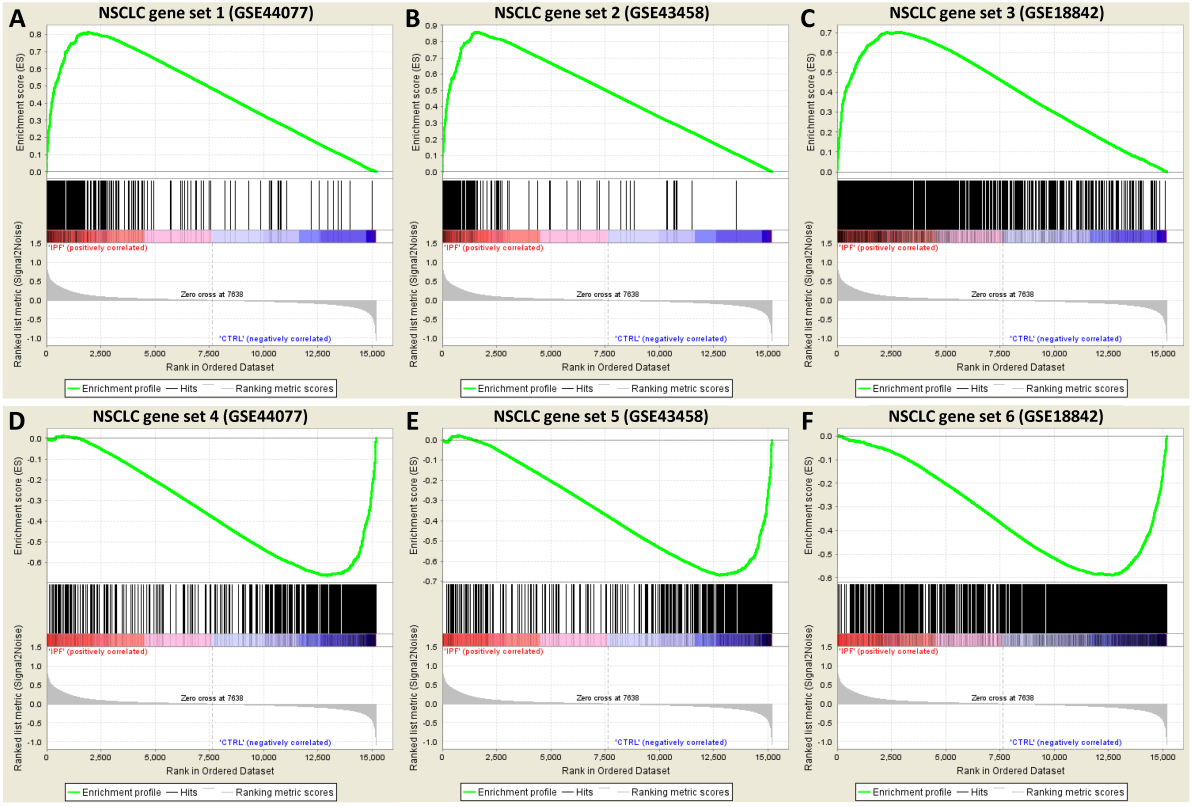

Figure 4.3: Representation of the NSCLC gene sets in IPF *vs.* control. Enrichment plots of the **(A)** NSCLC gene set 1 (*GSE44077*), **(B)** NSCLC gene set 2 (*GSE43458*), **(C)** NSCLC gene set 3 (*GSE18842*), **(D)** NSCLC gene set 4 (*GSE44077*), **(E)** NSCLC gene set 5 (*GSE43458*), and **(F)** NSCLC gene set 6 (*GSE18842*) in the IPF dataset (*GSE47460*). Significant enrichment (FDR q-value < 0.05; Nominal p-value < 0.05) was observed for **(A-C)** NSCLC gene sets 1-3 in the IPF phenotype and **(D-F)** NSCLC gene sets 4-6 in the control phenotype [120]. (Figure 4.3A-C were published in Ulke *et al.* [120] - *modified.*)

Table 4.1: GSEA results for Figure 4.3. (Data was published in Ulke *et al.* [120] - *modified.*)

Dataset & Phenotype	Enrichment plot	Enrichment Score (ES)	Normalized Enrichment Score (NES)	Nominal *p*-value	FDR q-value	FWER *p*-value
GSE47460 (IPF *vs.* control)	Figure 4.3A	0.8133	1.5512	< 0.001	0.0197	0.0340
	Figure 4.3B	0.8585	1.5517	< 0.001	0.0295	0.0340
	Figure 4.3C	0.7023	1.6995	< 0.001	0.0062	0.0040
	Figure 4.3D	-0.6623	-1.5307	0.0294	0.0207	0.0370
	Figure 4.3E	-0.6705	-1.5373	0.0277	0.0295	0.0350
	Figure 4.3F	-0.5902	-1.6607	0.0098	0.0124	0.0100

The *IPF phenotype* was defined as those genes with a higher mean expression in the IPF cases compared to the controls, and therefore included all the genes that were upregulated in IPF. In contrary, the *control phenotype* was defined as those genes with a higher mean expression in the control cases compared to IPF, and therefore included the genes that were downregulated in IPF. Hence, the enrichment results in Table 4.1 revealed that the genes upregulated in the three individual NSCLC microarray datasets were found to be significantly enriched in those genes upregulated over all IPF samples from the *GSE47460* microarray dataset [120]. In line with these results, the genes downregulated in the three individual NSCLC microarray datasets were found to be significantly enriched among the genes downregulated over all IPF samples [120].

In a final step, all DEGs included in the respective NSCLC gene set 1-6 were listed and their expression values were extracted from the *GSE47460* dataset and visualized as heatmaps (Figure 4.4 and 4.5). Interestingly, all six NSCLC gene sets were able to produce a noticeable clustering of the IPF and control patients, which implied that DEGs from NSCLC were also differentially expressed in IPF compared to control [120].

Overall, these results indicated the existence of a shared pattern of gene expression alterations, which could be jointly responsible for a pathomechanistical connection of IPF and NSCLC [120].

4.1.2 Generation of the leading-edge overlap

To narrow down the list of possible candidate genes for an in-depth analysis, we decided to focus on the genes that on the one hand were upregulated in NSCLC and on the other hand were enriched among those genes upregulated in IPF. GSEA generates the leading-edge list of genes for each run, which consists of those genes responsible for the enrichment score of a gene set in one of two individual phenotypes and constitutes the key element of a gene set [136]. Furthermore, it has been reported that the analysis of the leading-edge genes is most likely to uncover relevant mechanisms [136].

The leading-edge genes compiled by the GSEA runs for the NSCLC gene sets 1-3 of upregulated DEGs (Figure 4.3A-C) were withdrawn to their respective subset and

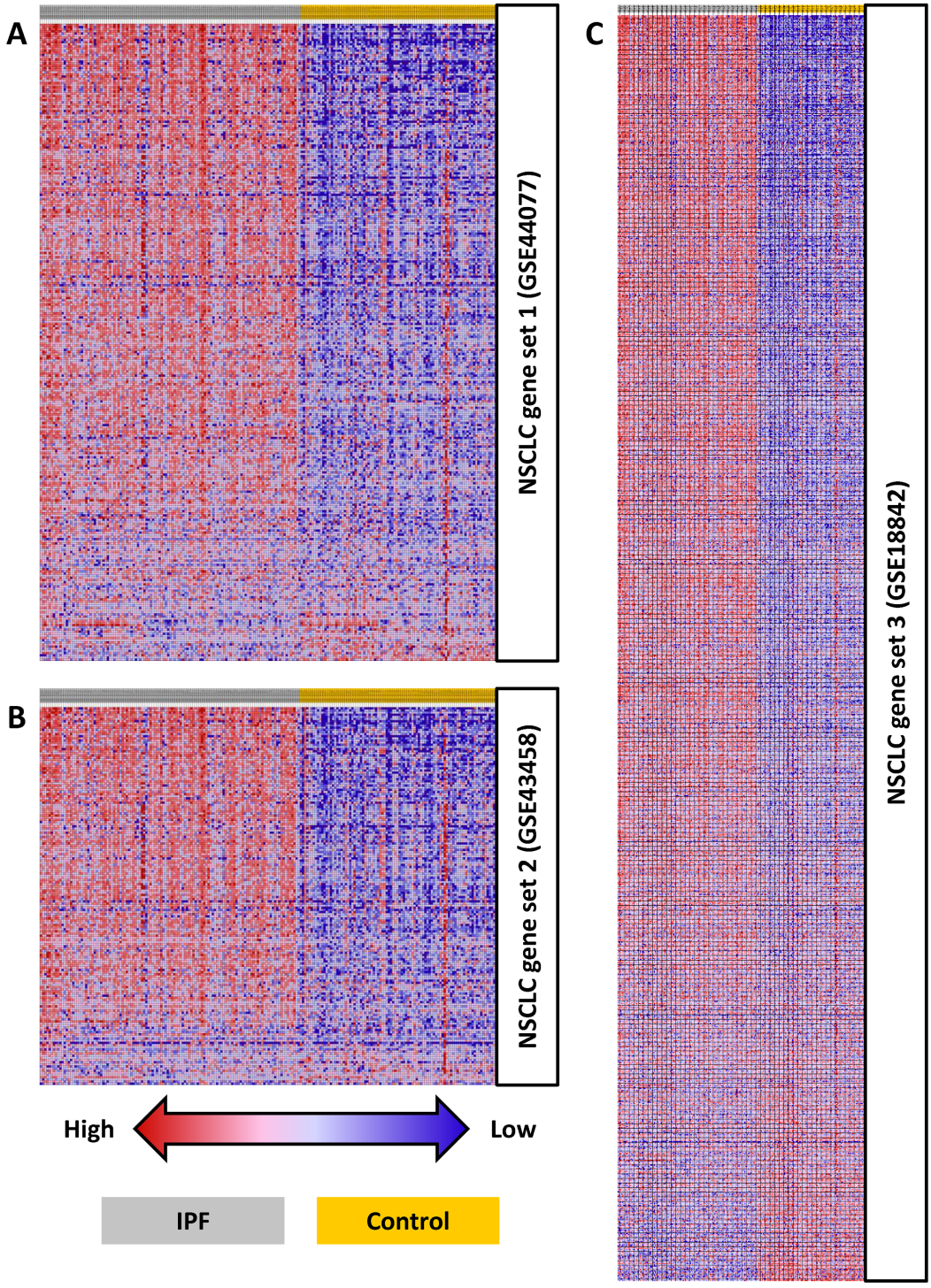

Figure 4.4: Heatmaps visualizing the distribution of IPF and control samples via genes upregulated in NSCLC. Expression levels for all genes included in the **(A)** NSCLC gene set 1 (*GSE44077*), **(B)** NSCLC gene set 2 (*GSE43458*), or **(C)** NSCLC gene set 3 (*GSE18842*) were extracted from the *GSE47460* dataset and visualized as heatmaps, revealing a noticeable distribution in either IPF or control cases [120]. (Figure 4.4A was published in Ulke *et al.* [120] - *modified.*)

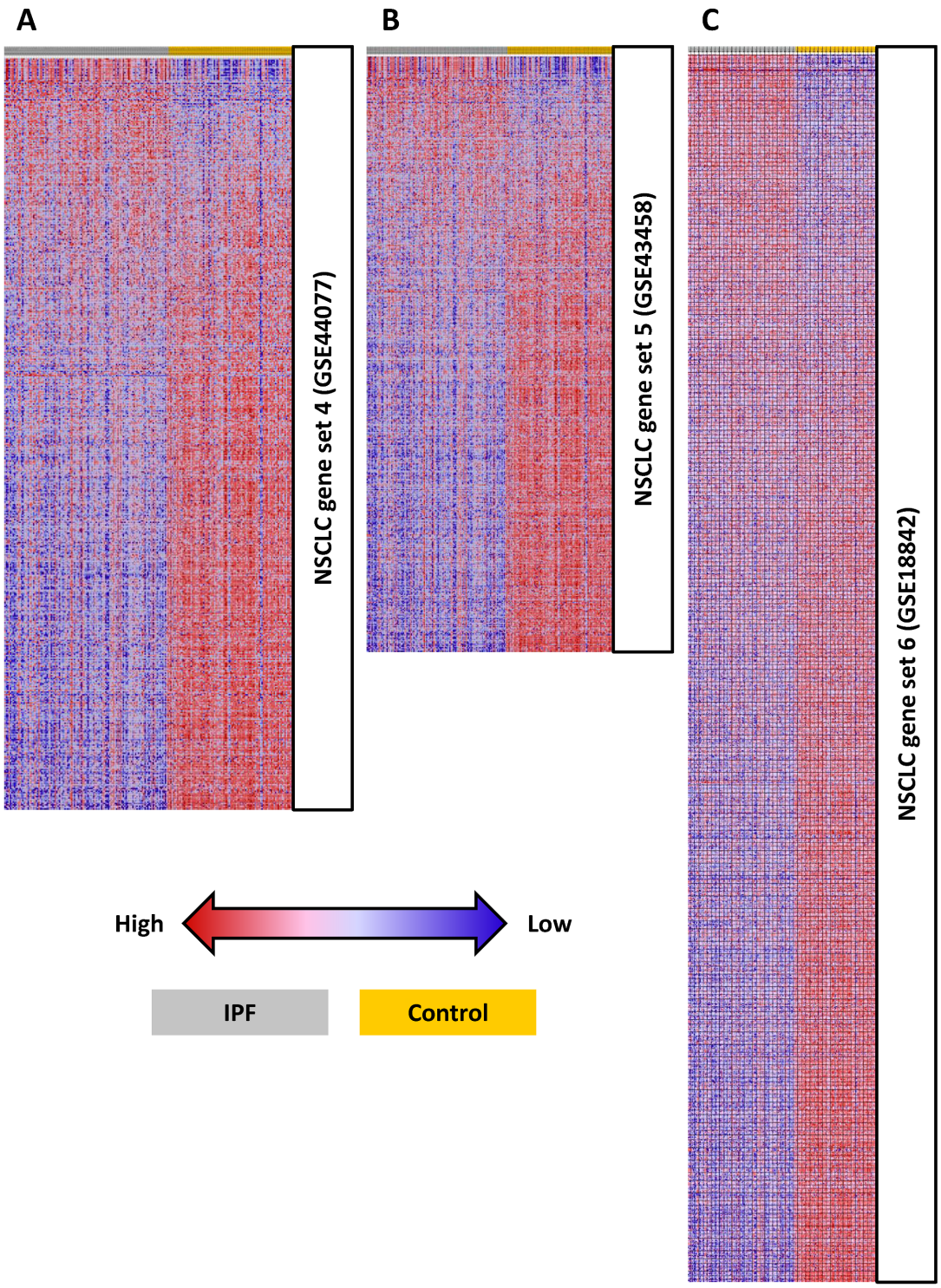

Figure 4.5: Heatmaps visualizing the distribution of IPF and control samples via genes downregulated in NSCLC. Expression levels for all genes included in the **(A)** NSCLC gene set 4 (*GSE44077*), **(B)** NSCLC gene set 5 (*GSE43458*), or **(C)** NSCLC gene set 6 (*GSE18842*) were extracted from the *GSE47460* dataset and visualized as heatmaps, revealing a noticeable distribution in either IPF or control cases.

labeled as leading-edge gene list 1 (n = 206; 46% of the NSCLC gene set 1; *GSE44077*), leading-edge gene list 2 (n = 136; 51% of the NSCLC gene set 2; *GSE43458*), and leading-edge gene list 3 (n = 545; 36% of the NSCLC gene set 3; *GSE18842*) [120].

Subsequently, the overlap of the three leading-edge lists was generated and visualized as Venn diagram (Figure 4.6), revealing the existence of a common core of 92 genes (15.4%), which were withdrawn and merged into a new list labeled as the *Overlap gene set* [120]. The complete process of assembling the *Overlap gene set* is shown in Figure 4.7, and an alphabetical list of the 92 genes is presented in Table 4.2.

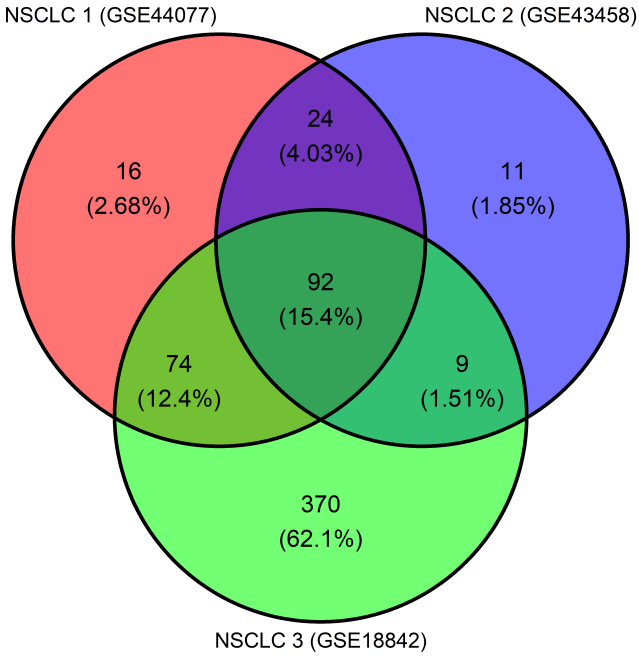

Figure 4.6: Overlap of the leading-edge lists. The leading-edge gene lists 1-3 were generated from their respective GSEA run (Figure 4.3A-C; upregulated DEGs) and visualized as Venn diagram, which demonstrated an overlap of 92 genes (15.4%) [120]. Total number of genes in the Venn diagram was defined as 100 percent. (Figure 4.6 was published in Ulke *et al.* [120] - *modified.*)

To conclude, the *Overlap gene set* represented a highly compressed list of 92 genes that were found significantly upregulated (log2 fold change > 1; adjusted p-value < 0.05) in all of the NSCLC microarray datasets (*GSE44077*, *GSE43458*, *GSE18842*) as well as the *GSE47460* (IPF) microarray dataset, and furthermore were part of all three leading-edge gene lists [120]. Thus, we hypothesized that analyzing the *Overlap gene set* could help to gain significant insight into the IPF pathogenesis.

4.1.3 Validation of the Overlap gene set

Next, we aimed to further validate the relevance of the *Overlap gene set* and to minimize the possibility of analyzing an incidental and unassociated group of genes. Therefore, an independent microarray dataset that characterized idiopathic interstitial pneumonias (*GSE32537*) [105] was downloaded from GEO and processed for GSEA by including

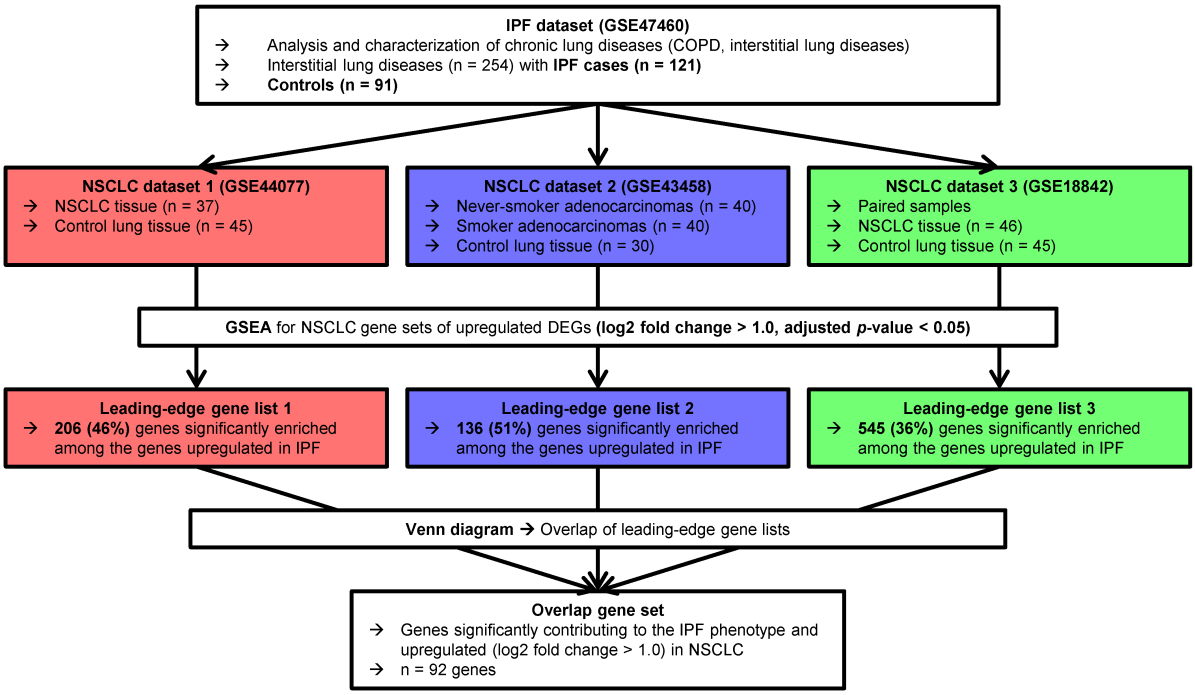

Figure 4.7: Process of generating the *Overlap gene set*. DEGs from three individual NSCLC microarray datasets (*GSE44077*, *GSE43458*, *GSE18842*) were extracted to the corresponding gene sets (NSCLC gene set 1-3). Subsequently, GSEA was conducted and the overlap of the leading-edge lists was created. (Figure 4.7 was published in Ulke *et al.* [120] - *modified*.)

Table 4.2: *Overlap gene set* in alphabetical order.

AIM2	*CDK1*	*CYP24A1*	*GPX2*	*MKI67*	*RAD51AP1*	*TNFRSF21*
AK4	*CDKN3*	*DEPDC1*	*GREM1*	*MMP1*	*RALGPS2*	*TNS4*
AKR1B10	*CEACAM5*	*DLGAP5*	*HELLS*	*MMP11*	*RPL39L*	*TOP2A*
ANLN	*CENPE*	*ECT2*	*HIST1H4K*	*MMP12*	*RRM2*	*TOX3*
ASPM	*CENPF*	*ETV4*	*HMMR*	*MMP7*	*SERINC2*	*TPX2*
ATP10B	*CEP55*	*EXO1*	*HS6ST2*	*MYBL2*	*SLC2A1*	*TTK*
BUB1	*CFB*	*FAM83A*	*ITGA11*	*NUF2*	*SLC44A5*	*TYMS*
BUB1B	*CKAP2L*	*FAP*	*KDELR3*	*PBK*	*SPINK1*	*UBE2T*
CCNA2	*COL10A1*	*FOXM1*	*KIAA0101*	*PDK1*	*SPP1*	
CCNB1	*COL3A1*	*FUT2*	*KIF11*	*PLK1*	*STEAP1*	
CCNB2	*CRABP2*	*GCNT3*	*KIF20A*	*PPAP2C*	*STIL*	
CDC45	*CST1*	*GJB2*	*KIF4A*	*PRSS2*	*SULF1*	
CDCA7	*CTHRC1*	*GOLM1*	*LCN2*	*PSAT1*	*TCN1*	
CDH3	*CXCL13*	*GPR87*	*MELK*	*PYCR1*	*TDO2*	

only IPF (n = 119) and control samples (n = 50). Hereupon, the *Overlap gene set* was analyzed for its enrichment in the *GSE32537* microarray dataset (Figure 4.8). GSEA revealed a significant enrichment (ES = 0.7680; NES = 1.4905; Nominal *p*-value = 0.0389; FDR q-value = 0.0389; FWER p-value = 0.0190) of the *Overlap gene set* among those genes upregulated in the IPF samples of the *GSE32537* microarray dataset [120].

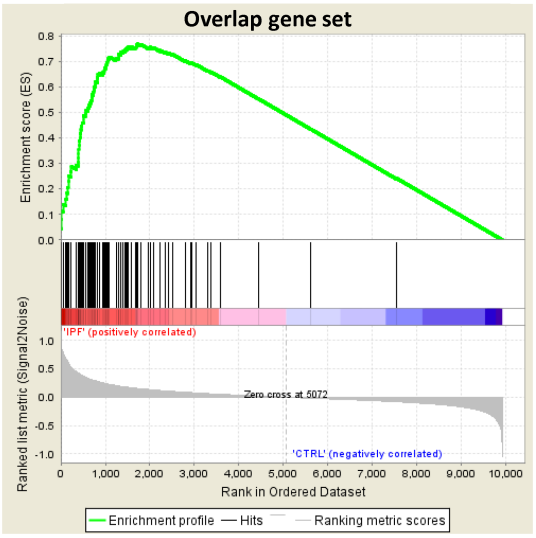

Figure 4.8: Representation of the *Overlap gene set* in the *GSE32537* dataset. GSEA was conducted to analyze the enrichment of the *Overlap gene set* in the *GSE32537* dataset. Significant enrichment (FDR q-value < 0.05; Nominal *p*-value < 0.05) was observed in the IPF phenotype [120]. (Figure 4.8 was published in Ulke *et al.* [120] - *modified*.)

4.2 Analysis of the Overlap gene set

4.2.1 Principal component analysis

Principal component analysis (PCA) enables the depiction of large amounts of data by reducing its complexity and thereby helps to detect distinctive features of the gene expression and distribution of samples [141, 142].

The expression values of all 92 genes included in the *Overlap gene set* were extracted for IPF (n = 122), COPD (n = 144), and control samples (n = 91) from the *GSE47460* microarray dataset. Subsequently, PCA was performed for this newly assembled expression dataset (357 x 92 values). A total of 92 components were required to preserve 100% of the variance of the original data, while the first 30 components already explained more than 95% of the original variance (Appendix Table 1). The scree plot in Figure 4.9 demonstrates the variance that was explained by the respective principal component 1-10. Results of the PCA were visualized as biplot for the first (PC1) and second component (PC2) (Figure 4.10), which preserved 61.2% of the original variance (PC1 = 50.7%; PC2 = 10.5%). Interestingly, there was a noticeable pattern of the sample distribution, especially along the PC1 (x-axis). IPF patients were clearly separated from

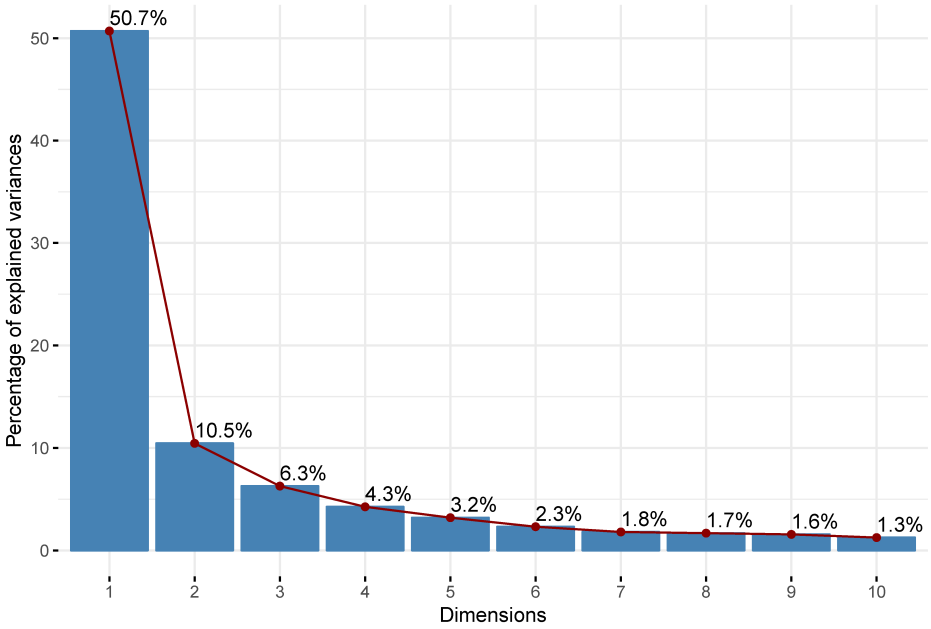

Figure 4.9: Preserved variance plotted for the first ten principal components. The scree plot was generated from the PCA results and displays the variance of the first ten PCs.

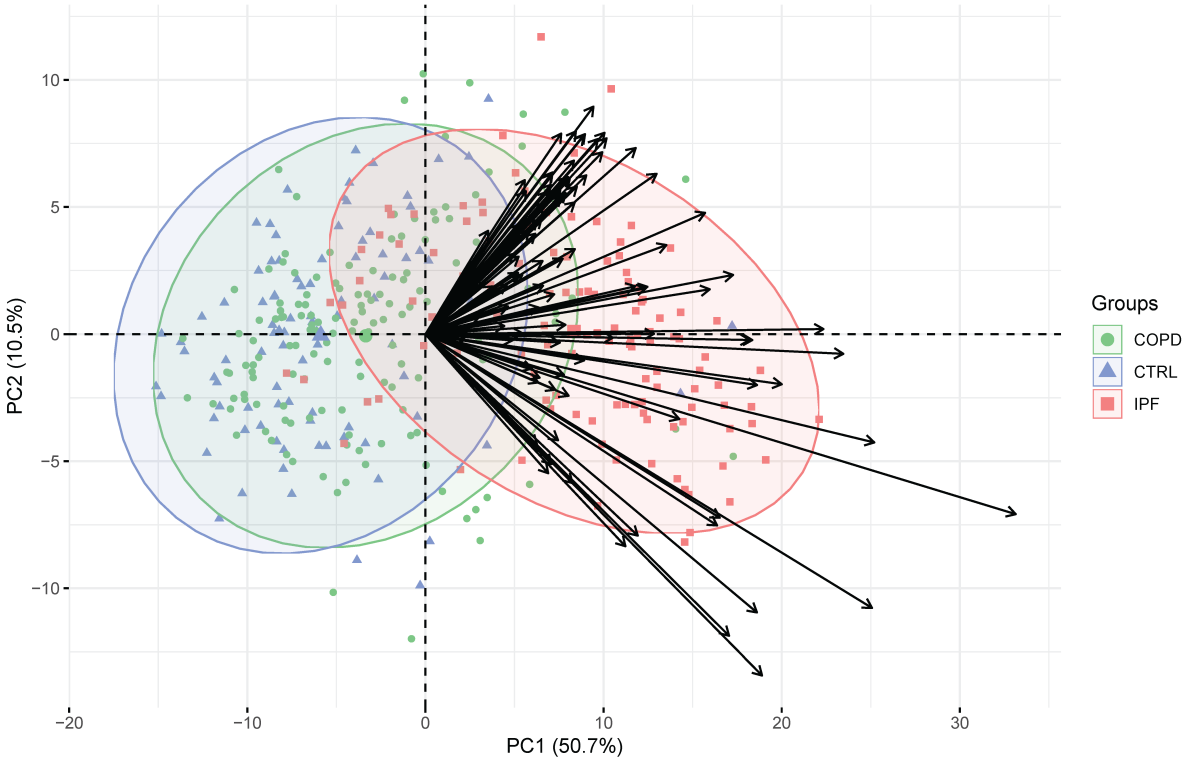

Figure 4.10: Principal component analysis presented as biplot. IPF *(red square)*, COPD *(green circle)*, and CTRL samples *(blue triangle)* were plotted onto the first two principal components (PC1; PC2). Weights of individual genes are presented as black arrows. Clear separation of IPF from COPD and CTRL samples was observed.

the other two groups, while the COPD patients and controls (CTRL) showed a striking overlap (confidence ellipses = 0.90). Furthermore, most of the 92 genes demonstrated a large weight for the displayed components (PC1/PC2) and pointed towards the cluster of IPF samples (Figure 4.10, black arrows). Overall, these PCA results indicated an IPF-specific pattern of the *Overlap gene set*.

4.2.2 Association of the Overlap gene set with specific cell types

The next question to be addressed, was if the *Overlap gene set* included a pattern of genes, which could be linked to a distinct cell type. Therefore, we selected three publicly available RNA-sequencing or microarray datasets that characterized the gene expression of individual cell types isolated from fibrotic and control lungs.

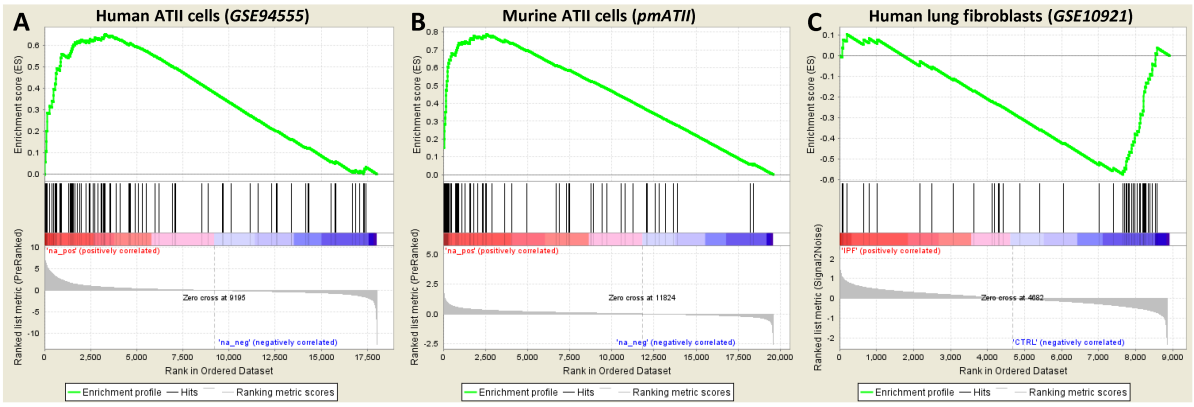

Figure 4.11: Representation of the *Overlap gene set* in different cell types associated with IPF. GSEA was executed to evaluate the enrichment of the *Overlap gene set* in three individual datasets (*GSE94555*, *pmATII* [111], *GSE10921*). Significant enrichment (FDR q-value < 0.05; Nominal p-value < 0.05) was discovered in **(A)** the IPF phenotype of isolated human ATII cells, **(B)** the fibrotic phenotype of isolated primary mouse ATII cells, and **(C)** the control phenotype of isolated human lung fibroblasts [120]. (Figure 4.11A-C were published in Ulke *et al.* [120] - *modified*.)

Table 4.3: GSEA results for Figure 4.11. (Data was published in Ulke *et al.* [120] - *modified*.)

Gene set	Enrichment plot	Enrichment Score (ES)	Normalized Enrichment Score (NES)	Nominal p-value	FDR q-value	FWER p-value
Overlap gene set	Figure 4.11A	0.6519	1.8353	< 0.001	< 0.001	< 0.001
	Figure 4.11B	0.7876	2.3798	< 0.001	< 0.001	< 0.001
	Figure 4.11C	-0.5760	-2.2948	< 0.001	< 0.001	< 0.001

The RNA-sequencing dataset *GSE94555* compared suspensions of primary human ATII cells from IPF and control patients [110], the *pmATII* microarray dataset described

primary murine ATII cells derived from the bleomycin model of lung fibrosis [111], and the *GSE10921* microarray dataset investigated primary human lung fibroblasts from IPF and control patients [109]. Subsequently, the datasets were preprocessed and GSEA was executed to evaluate a possible enrichment of the *Overlap gene set* in those three datasets (Figure 4.11). The corresponding enrichment results are summarized in Table 4.3. It was apparent from this table that there existed a significant positive enrichment (FDR q-value < 0.05; Nominal p-value < 0.05) of the *Overlap gene set* in both primary ATII cell datasets (Figure 4.11A and B), while a significant negative enrichment (FDR q-value < 0.05; Nominal p-value < 0.05) could be observed in the fibroblast dataset (Figure 4.11C) [120]. Hence, these 92 genes seemed to constitute an essential group of genes, which was significantly overrepresented within the upregulated genes of fibrotic human and murine ATII cells, as well as in the human lung fibroblasts from nondiseased control tissue [120].

4.2.3 Annotation enrichment analyses of the Overlap gene set

In a final step, we aimed to elucidate a common biological and mechanistic context of the 92 genes by performing functional annotation enrichment analyses. To begin with, the respective proteins for all genes of the *Overlap gene set* were identified and listed as the *Overlap protein list*.

Table 4.4: Enriched REACTOME pathways (TOP 5). The full results are shown in the Appendix Table 2. (Data was published in Ulke *et al.* [120] - *modified*.)

Stable identifier	**Pathway**	**FDR value**	***p*-value**	**Genes**
R-HSA-69278	Cell Cycle, Mitotic	4.06E-10	1.27E-12	*BUB1, BUB1B, CCNA2, CCNB1, CCNB2, CDC45, CDK1, CENPE, CENPF, FOXM1, HMMR, KIF20A, MYBL2, NUF2, PLK1, RRM2, TOP2A, TPX2, TYMS*
R-HSA-1640170	Cell Cycle	6.73E-10	4.19E-12	*BUB1, BUB1B, CCNA2, CCNB1, CCNB2, CDC45, CDK1, CENPE, CENPF, EXO1, FOXM1, HMMR, KIF20A, MYBL2, NUF2, PLK1, RRM2, TOP2A, TPX2, TYMS*
R-HSA-156711	Polo-like kinase mediated events	5.88E-08	5.49E-10	*CCNB1, CCNB2, CENPF, FOXM1, MYBL2, PLK1*
R-HSA-2500257	Resolution of Sister Chromatid Cohesion	2.43E-07	3.03E-09	*BUB1, BUB1B, CCNB1, CCNB2, CDK1, CENPE, CENPF, NUF2, PLK1*
R-HSA-69273	Cyclin A/B1/B2 associated events during G2/M transition	3.37E-07	5.25E-09	*CCNA2, CCNB1, CCNB2, CDK1, FOXM1, PLK1*

Table 4.5: Enriched KEGG pathway maps. (Data was published in Ulke *et al.* [120] - *modified.*)

KEGG identifier	Pathway map	FDR value	*p*-value	Genes
hsa04110	Cell cycle	1.81E-06	1.62E-08	*PLK1, BUB1, BUB1B, TTK, CDC45, CCNA2, CCNB1, CCNB2, CDK1*
hsa04914	Progesterone-mediated oocyte maturation	0.0007	1.18E-05	*PLK1, BUB1, CCNA2, CCNB1, CCNB2, CDK1*
hsa04218	Cellular senescence	0.0058	0.0002	*FOXM1, MYBL2, CCNA2, CCNB1, CCNB2, CDK1*
hsa04114	Oocyte meiosis	0.0091	0.0004	*PLK1, BUB1, CCNB1, CCNB2, CDK1*
hsa04115	P53 signaling pathway	0.0091	0.0004	*RRM2, CCNB1, CCNB2, CDK1*

Table 4.6: Enriched GO terms (TOP 5). The full results are shown in the Appendix Table 3. (Data was published in Ulke *et al.* [120] - *modified.*)

Accession number	GO term	FDR value	Genes
GO:0000278	Mitotic cell cycle	1.08E-13	*HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, MELK, PLK1, TPX2, PBK, BUB1, TYMS, CDKN3, FOXM1, HIST1H4K, RRM2, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDC45, TOP2A*
GO:1903047	Mitotic cell cycle process	1.08E-13	*HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, MELK, PLK1, TPX2, PBK, BUB1, TYMS, CDKN3, FOXM1, RRM2, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDC45, TOP2A*
GO:0007067	Mitotic nuclear division	1.34E-11	*MYBL2, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, PLK1, TPX2, PBK, BUB1, CENPF, ASPM, CEP55, KIF4A, CDK1*
GO:0022402	Cell cycle process	1.34E-11	*FAP, ECT2, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, MELK, TPX2, PBK, BUB1, TYMS, CDKN3, FOXM1, RRM2, CENPF, ASPM, MKI67, CEP55, STIL, KIF4A, KIF20A, CDC45, TOP2A*
GO:0051301	Cell division	1.51E-11	*ECT2, HELLS, CCNB1, KIF11, CENPE, ANLN, NUF2, CCNA2, BUB1B, CCNB2, PLK1, TPX2, BUB1, CENPF, ASPM, CEP55, KIF4A, KIF20A, CDK1, TOP2A*

Primarily, the *Overlap gene set* was analyzed for its internal, functional relationships and representation within various biological pathways. Therefore, the two reference databases of the Kyoto Encyclopedia of Genes and Genomes (KEGG) [124, 125] and the REACTOME [126] project were utilized. The comprehensive pathway enrichment analysis encountered a significant overrepresentation (FDR < 0.05 and > 2 pathway genes included in the *Overlap gene set*) of 67 individual REACTOME pathways (Table 4.4) and five KEGG pathway maps (Table 4.5) within the *Overlap gene set* [120]. Besides this, the Gene Ontology (GO) Consortium [122, 123] database was utilized to scan the *Overlap gene set* for commonly enriched GO terms. The analysis of all three major GO domains (molecular function, cellular component, biological process) revealed a total of 171 significantly enriched GO terms (FDR < 0.05 and ≥ 2 genes included in the *Overlap gene set*) (Table 4.6) [120]. Full results of the functional annotation enrichment analyses are shown in the Appendix Tables 2 and 3.

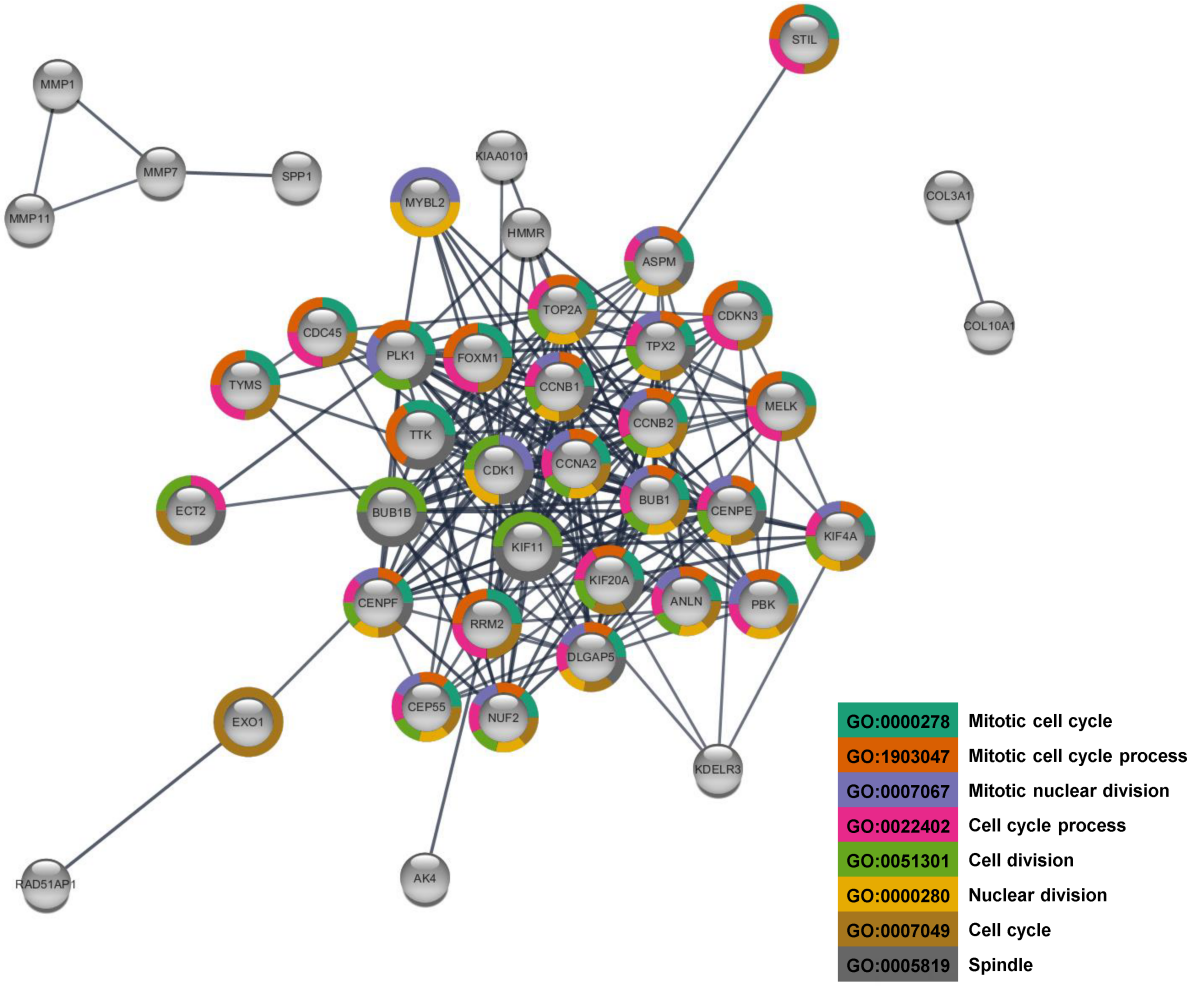

Figure 4.12: Protein-protein interaction network of the *Overlap protein list*. The eight most significant GO terms (Table 4.6) were selected. The corresponding proteins to those genes included in the eight GO terms were marked with the respective colors in the network. (Figure 4.12 was published in Ulke *et al.* [120] - reprinted with permission of the ATS.)

The *Overlap protein list* was used to create a protein-protein interaction (PPI) network with a preset limit of the confidence score of 0.9 (Figure 4.12). Thereby, only interactions with high relevance and likelihood were included. The final PPI network contained 210 interactions (edges) between 42 of the initial 92 proteins (nodes) and revealed that the *Overlap protein list* included a significant accumulation of interactions among its proteins (p-value < 1.0E-16; average node degree = 4.6667; average local clustering coefficient = 0.3557) [120].

To summarize this section, the *Overlap gene set* seemed to constitute a batch of functionally related genes, which shared a significant number of common interactions and were particularly connected to the principal themes of cell cycle and proliferation, extracellular matrix (ECM) organization, and P53 regulation [120].

4.3 Selection of candidate genes

After the in-depth analysis of the *Overlap gene set*, we aimed to select promising candidates from those 92 genes for further investigation. Therefore, an extensive literature research was performed for the *Overlap gene set*. The candidate genes were chosen by **(1)** their expression levels in the *GSE47460* microarray dataset (Figure 4.13), **(2)** the amount, quality, and types of reported scientific findings, **(3)** their individual weight and direction towards the IPF cluster in the PCA (Figure 4.10), and **(4)** their appearance in the functional annotation enrichment analyses (Section 4.2.3).

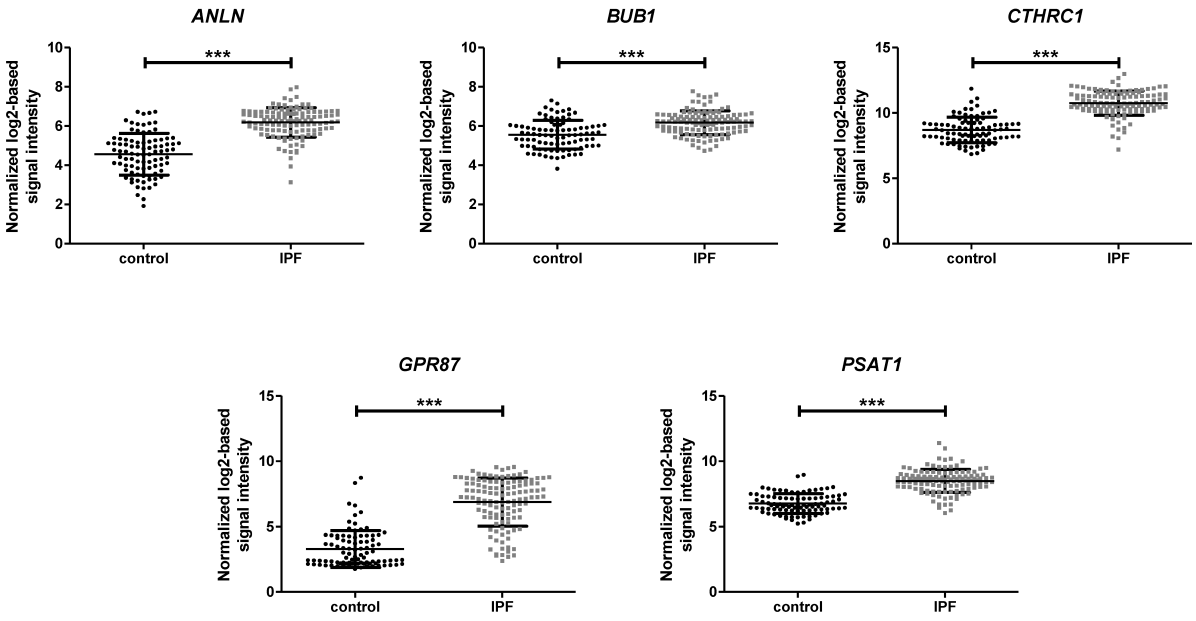

Figure 4.13: Expression of the candidate genes in the *GSE47460* microarray dataset. The expression values of *ANLN*, *BUB1*, *CTHRC1*, *GPR87*, and *PSAT1* were extracted as normalized log2-based signal intensities from the *GSE47460* microarray dataset. The mean expression was compared with the unpaired t-test for IPF (n = 121) and control (n = 91). Data is shown as mean±SD. Significance: * p<0.05; ** p<0.01; *** p<0.001.

Finally, the following six candidate genes were selected: *ANLN*, *BUB1*, *CTHRC1*, *ECT2*, *GPR87*, and *PSAT1*. The subsequent analysis of *ECT2* was published in Ulke *et al.* [120] and is not further described in this thesis. The five remaining candidate genes showed significantly increased expression levels in the IPF (n = 121) compared to control (n = 91) samples of the *GSE47460* microarray dataset (Figure 4.13; gene expression mean±SD IPF *vs.* control: *ANLN* 6.18±0.76 *vs.* 4.56±1.06; *BUB1* 6.17±0.61 *vs.* 5.56±0.73; *CTHRC1* 10.75±0.94 *vs.* 8.70±0.98; *GPR87* 6.88±1.85 *vs.* 3.28±1.42; *PSAT1* 8.50±0.88 *vs.* 6.77±0.75).

4.3.1 Expression in human tissue samples

To further decrease the number of possible candidate genes, we sought to verify the previously observed upregulation. Therefore, the mRNA expression levels of *ANLN*, *BUB1*, *CTHRC1*, *GPR87*, and *PSAT1* were analyzed by qRT-PCR in human whole tissue samples of IPF and healthy donor lungs. Significantly increased mRNA expression levels were observed for all five candidate genes in IPF compared to control specimens (Figure 4.14; mRNA expression mean±SD IPF *vs.* control: *ANLN* -5.56±0.22 (n = 3) *vs.* -7.87±0.49 (n = 5); *BUB1* -3.36±0.51 (n = 6) *vs.* -4.26±0.65 (n = 5); *CTHRC1* 1.45±1.05 (n = 6) *vs.* -2.53±1.23 (n = 6); *GPR87* -2.03±1.48 (n = 8) *vs.* -7.34±2.10 (n = 4); *PSAT1* -0.77±0.59 (n = 12) *vs.* -1.92±0.96 (n = 9)).

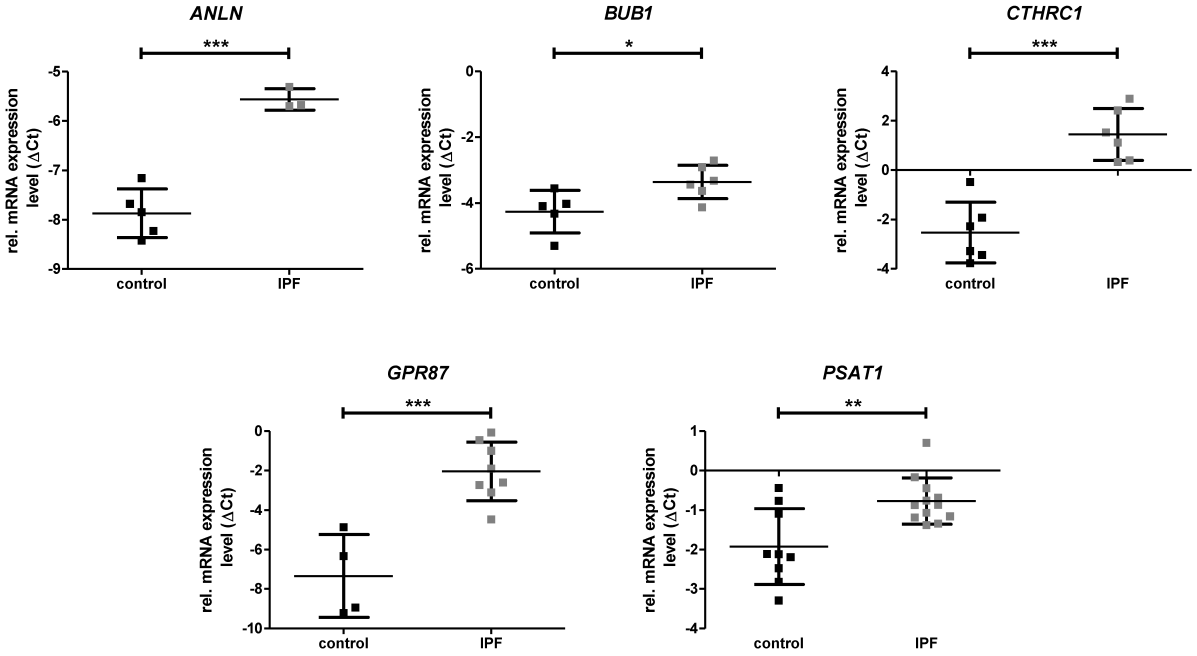

Figure 4.14: mRNA expression levels of the candidate genes in IPF and control lung tissue. The mRNA expression levels of *ANLN*, *BUB1*, *CTHRC1*, *GPR87*, and *PSAT1* were assessed by qRT-PCR in total RNA isolated from whole lung tissue specimens of IPF (n = 3-12) and donor (n = 4-9). Significantly increased gene expression was observed for all five genes. Expression levels were normalized to *HPRT* and are presented as ΔCt. Data is shown as mean±SD. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; unpaired t-test.

Subsequently, human whole lung tissue samples of non-small cell lung cancer (NSCLC) patients with either adenocarcinoma (AC) or squamous cell carcinoma (SCC) were examined by qRT-PCR for their gene expression levels of the candidate genes. For this purpose, paired specimens from single lung cancer patients were obtained (n = 4-6; tumor and nondiseased control tissue). The results demonstrated a significant upregulation of *ANLN* and *GPR87* in the SCC tissue compared to control (Figure 4.15; mRNA expression mean±SD SCC *vs.* control: *ANLN* -5.61±1.14 *vs.* -7.80±1.08; *GPR87* -2.41±2.49 *vs.* -5.89±1.29). Furthermore, *CTHRC1* and *PSAT1* were significantly increased in the AC tissue samples compared to control (Figure 4.15; mRNA expression mean±SD AC *vs.* control: *CTHRC1* 0.11±0.79 *vs.* -3.54±1.94; *PSAT1* 0.29±1.00 *vs.* -2.27±0.31).

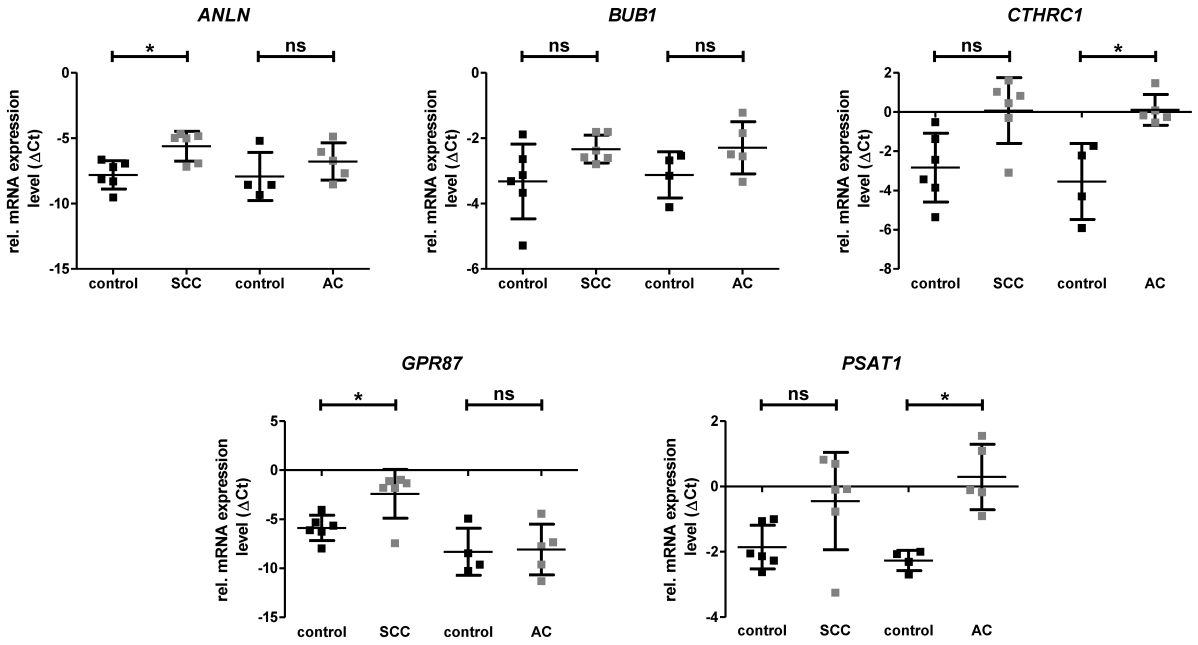

Figure 4.15: mRNA expression levels of the candidate genes in paired NSCLC and control lung tissue. Paired specimens from patients with either SCC (n = 6) or AC (n = 4-5) were analyzed with qRT-PCR for their mRNA expression levels of the candidate genes. Significantly increased gene expression was detected for *ANLN* and *GPR87* in the SCC samples as well as for *CTHRC1* and *PSAT1* in the AC samples. Expression levels were normalized to *HPRT* and are presented as ΔCt. Data is shown as mean±SD. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; paired t-test.

4.3.2 Expression in the bleomycin-induced fibrotic mouse model

The model of bleomycin-induced experimental lung fibrosis represents a commonly used and well characterized animal model of IPF. Therefore, we also aimed to verify the gene expression levels of the candidate genes in whole lung tissue specimens from bleomycin-treated mice and untreated controls. The mice were instilled with either 2U/kg body weight of bleomycin sulphate (Bleo) or 50μl pure DPBS, and lungs were harvested on day 7 or 14 for further analysis. Thereafter, the mRNA expression levels of *Anln*, *Bub1*,

Cthrc1, *Gpr87*, and *Psat1* were analyzed by qRT-PCR in the Bleo (n = 6-7) and DPBS (n = 3-6) samples. Interestingly, there was no difference for *Anln* and *Bub1*, whereas *Cthrc1* (d14), *Gpr87* (d7 and d14), and *Psat1* (d7 and d14) showed a significant upregulation in bleomycin-induced lung fibrosis (Figure 4.16; mRNA expression mean±SD Bleo *vs.* DPBS: *Cthrc1* (d14) -2.67±0.70 *vs.* -5.21±0.71; *Gpr87* (d7) -11.13±1.45 *vs.* -13.32± ±1.38; *Gpr87* (d14) -11.25±0.98 *vs.* -13.76±0.48; *Psat1* (d7) -1.70±0.58 *vs.* -2.65±0.43; *Psat1* (d14) -2.15±0.49 *vs.* -3.08±0.61).

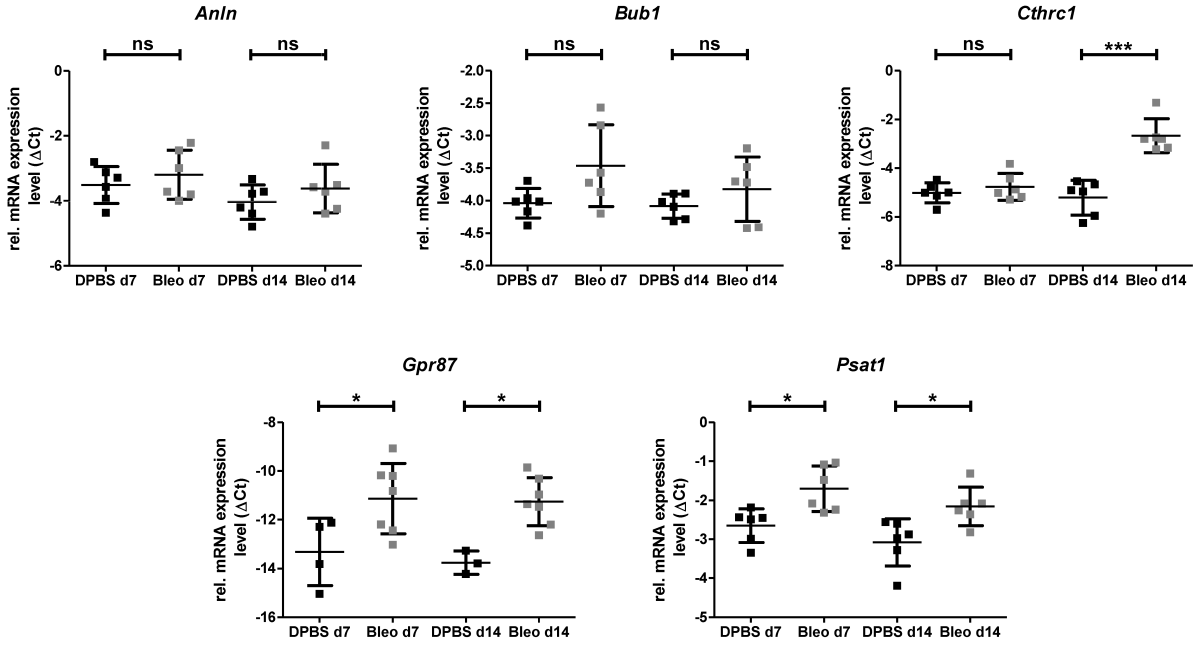

Figure 4.16: mRNA expression levels of the candidate genes in bleomycin-induced lung fibrosis. Mice were treated with either 2U/kg BW of Bleo or 50μl DPBS. Lungs were removed on day 7 or 14 after the instillation. mRNA levels of the candidate genes were evaluated with qRT-PCR for Bleo (n = 6-7) and DPBS (n = 3-6) specimens. Significant upregulation in the bleomycin-treated mice was detected for *Cthrc1* (d14), *Gpr87* (d7, d14), and *Psat1* (d7, d14). Expression levels were normalized to *Hprt* and are presented as ΔCt. Data is shown as mean±SD. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; one-way ANOVA with Bonferroni's multiple comparison test.

In conclusion, we observed the most reliable confirmation of altered gene expression for GPR87 and PSAT1 in our specimens of IPF, NSCLC, and experimental lung fibrosis. In combination with **(1)** the nonexistent scientific reports about their possible role in the pathogenesis of IPF and **(2)** the possibility to study them in our animal model of lung fibrosis, we decided to narrow down the candidate genes to GPR87 and PSAT1.

4.4 PSAT1 - Phosphoserine aminotransferase 1

The phosphoserine aminotransferase 1 (PSAT1) showed a significant upregulation on mRNA level in the lungs of IPF patients as well as in the bleomycin model of lung fibrosis. To confirm that this altered gene expression also led to increased protein levels of PSAT1,

western blot analysis was performed on whole lung tissue specimens of IPF/control (n = 6, each) and Bleo/DPBS control (n = 3, each). The PSAT1 protein expression was significantly increased in IPF (Figure 4.17; rel. PSAT1 protein level mean±SD IPF *vs.* control: 0.56±0.43 *vs.* 0.08±0.05) and bleomycin-induced lung fibrosis on day 7 and 14 (Figure 4.17; rel. PSAT1 protein level mean±SD Bleo *vs.* DPBS: *d7* 0.56±0.09 *vs.* 0.18±0.02; *d14* 0.35±0.05 *vs.* 0.18±0.01), which further supported the PSAT1 mRNA expression data (Figure 4.14 and 4.16).

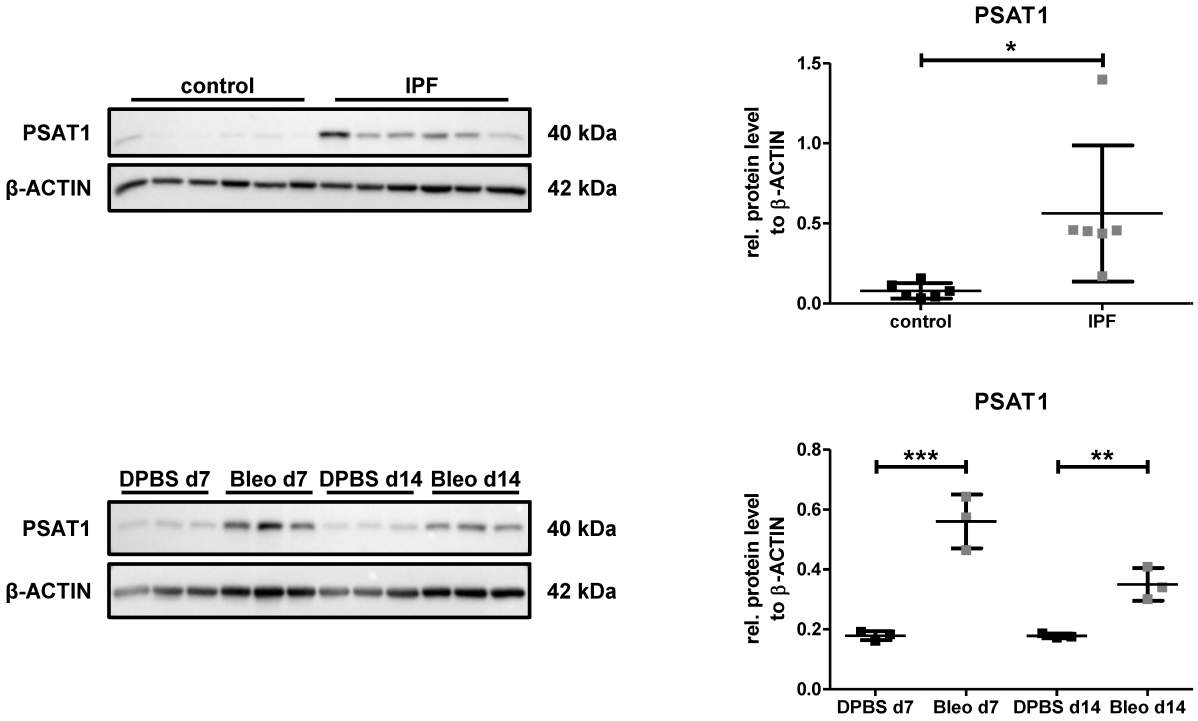

Figure 4.17: PSAT1 protein levels in IPF and experimental lung fibrosis. Western blot analysis was conducted to analyze PSAT1 protein expression in IPF *vs.* control (n = 6, each) and bleomycin-induced experimental lung fibrosis *vs.* DPBS control (n = 3, each). Significantly increased PSAT1 levels were detected in IPF as well as in bleomycin-treated lungs. Representative western blot (left) with its corresponding densitometric analysis (right). β-ACTIN was used as loading control. Data is shown as mean±SD. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; unpaired t-test, one-way ANOVA with Bonferroni's multiple comparison test.

Subsequently, we aimed to assess the correlation between disease severity of IPF and the expression level of *PSAT1*. The diffusing capacity of the lung for carbon monoxide (DLCO) and the forced vital capacity (FVC) represent two aspects measured by pulmonary function testing, which are regularly examined to estimate the disease stage of IPF [143]. Therefore, the available DLCO and FVC measurements of all IPF subjects were extracted from the *GSE47460* microarray dataset and correlated to their respective expression levels of *PSAT1*. The Pearson correlation analysis revealed a significant negative correlation between the *PSAT1* expression levels and DLCO (n = 109; R^2 = 0.14; $p < 0.0001$; Pearson r = -0.3741) as well as FVC (n = 118; R^2 = 0.1226; p =

0.0001; Pearson r = -0.3501) in IPF patients (Figure 4.18).

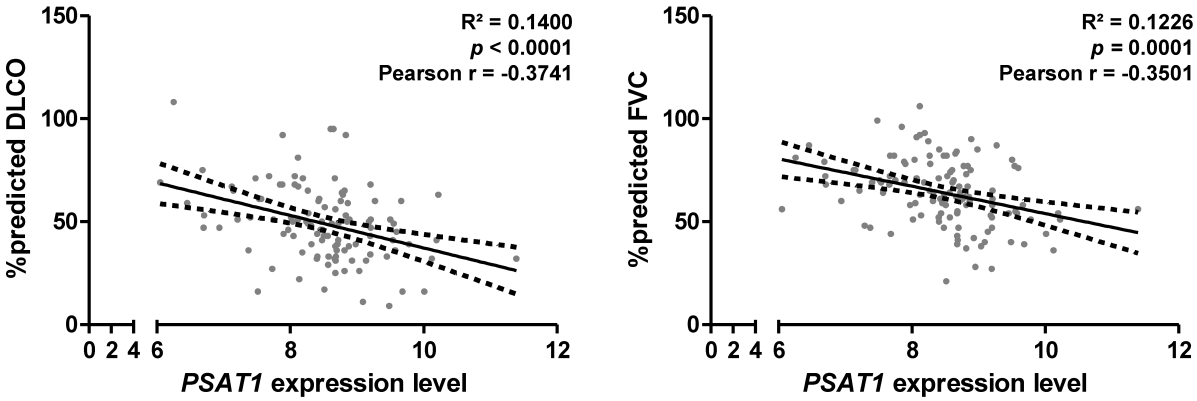

Figure 4.18: Correlation between *PSAT1* expression and pulmonary function in IPF. The DLCO, FVC, and *PSAT1* expression values were extracted from the *GSE47460* microarray dataset. Pearson correlation analysis revealed a significant negative correlation. Linear regression analysis is shown as continuous line with the 95% confidence interval (dashed lines).

Finally, we wanted to confirm the specific upregulation of PSAT1 in fibrotic alveolar epithelial cells. Therefore, isolated primary mouse ATII cells were analyzed for their mRNA expression levels of *Psat1*. qRT-PCR analysis revealed significantly increased mRNA levels of *Psat1* in the pmATII cells isolated from bleomycin-treated mice compared to DPBS-treated control on day 14 after instillation (Figure 4.19; *Psat1* mRNA expression mean±SD Bleo *vs.* DPBS: 1.11±0.62 *vs.* -0.59±0.36; ** $p<0.01$).

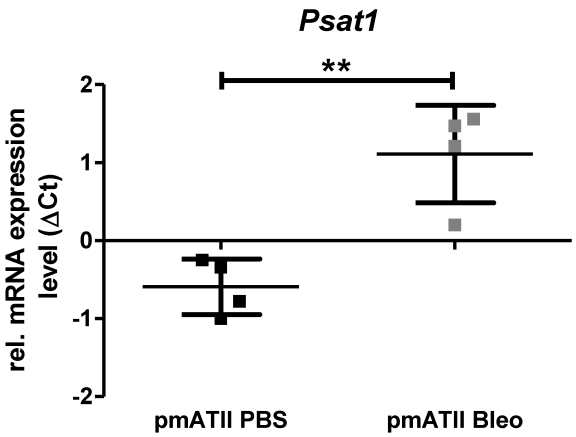

Figure 4.19: *Psat1* mRNA expression in primary mouse ATII cells. The mRNA levels of *Psat1* in isolated pmATII cells from bleomycin- and DPBS-treated mice on d14 after instillation (n = 4, each) were assessed by qRT-PCR. Significantly increased Psat1 was detected in the bleomycin-treated mice. Expression levels were normalized to *Hprt* and are presented as ΔCt. Data is shown as mean±SD. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; unpaired *t*-test.

4.5 GPR87 - G protein-coupled receptor 87

4.5.1 Expression analysis of GPR87 on protein level

The G protein-coupled receptor 87 (GPR87) was our second gene of interest that was significantly upregulated on mRNA level in lung specimens of IPF, SCC, and bleomycin-induced experimental lung fibrosis. This was furthermore confirmed on protein level by western blot analysis. GPR87 showed a significant upregulation in whole lung homogenate of IPF compared to donor specimens (Figure 4.20; rel. GPR87 protein level mean±SD IPF *vs.* control: 0.35±0.27 *vs.* 0.08±0.10). The fibrotic phenotype of the IPF samples was additionally supported by the significantly increased expression of the mesenchymal markers α-smooth muscle actin (α-SMA) and Calponin 1 (CNN1) (Figure 4.20; rel. protein level mean±SD IPF *vs.* control: α-SMA 0.42±0.32 *vs.* 0.01±0.02, CNN1 0.59±0.35 *vs.* 0.06±0.12).

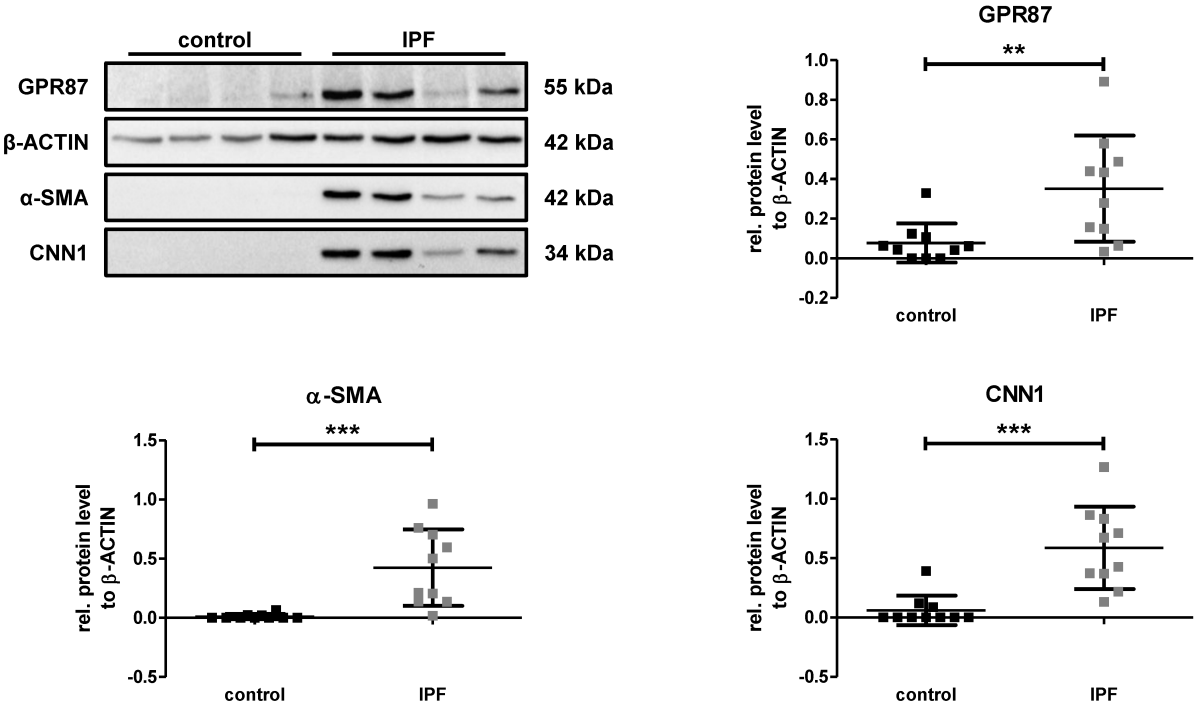

Figure 4.20: GPR87 protein expression levels in IPF. Western blot was performed to analyze the protein levels of GPR87 in whole lung homogenate of IPF and donors (n = 10, each). Significantly increased GPR87, as well as α-SMA and CNN1 expression was found in the IPF specimens. Representative western blot with the densitometric analyses. β-ACTIN was used as loading control. Data is shown as mean±SD. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; unpaired t-test.

In addition, GPR87 was significantly increased in whole lung homogenate of bleomycin-treated mice compared to control on d14 after the instillation (Figure 4.21; rel. GPR87 protein level mean±SD Bleo *vs.* DPBS: *d7* 0.35±0.08 *vs.* 0.26±0.10; *d14* 0.54±0.10 *vs.* 0.25±0.11).

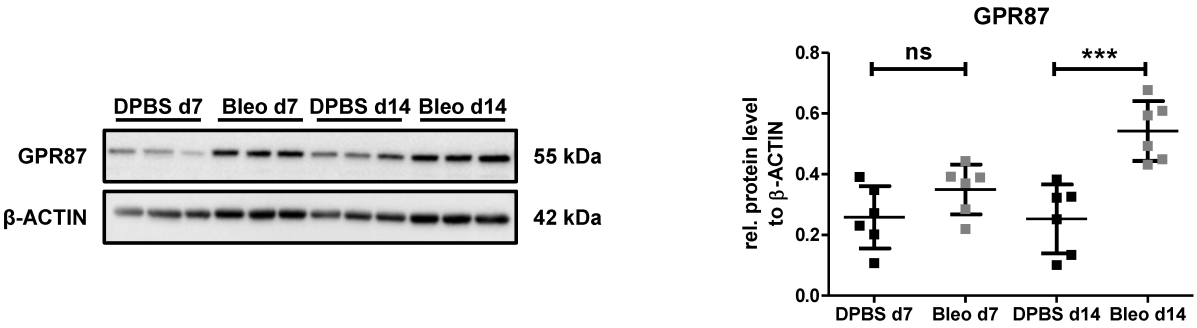

Figure 4.21: GPR87 protein levels in experimental lung fibrosis. Expression of GPR87 was analyzed by western blot in whole lung homogenate of bleomycin-treated mice compared to DPBS control (n = 6, each). Significant upregulation of GPR87 was found on d14 after the instillation of bleomycin. Representative western blot (left) with the densitometric analysis (right). β-ACTIN was used as loading control. Data is shown as mean±SD. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; one-way ANOVA with Bonferroni's multiple comparison test.

4.5.2 Correlation of GPR87 with IPF disease severity

Subsequently, also the *GPR87* expression levels from the *GSE47460* microarray dataset were correlated to their respective DLCO and FVC values. Results of the Pearson correlation analysis revealed a slight but significant negative correlation with the DLCO (n = 109; R^2 = 0.0737; p = 0.0043; Pearson r = -0.2714) and FVC (n = 118; R^2 = 0.0778; p = 0.0022; Pearson r = -0.2789) (Figure 4.22).

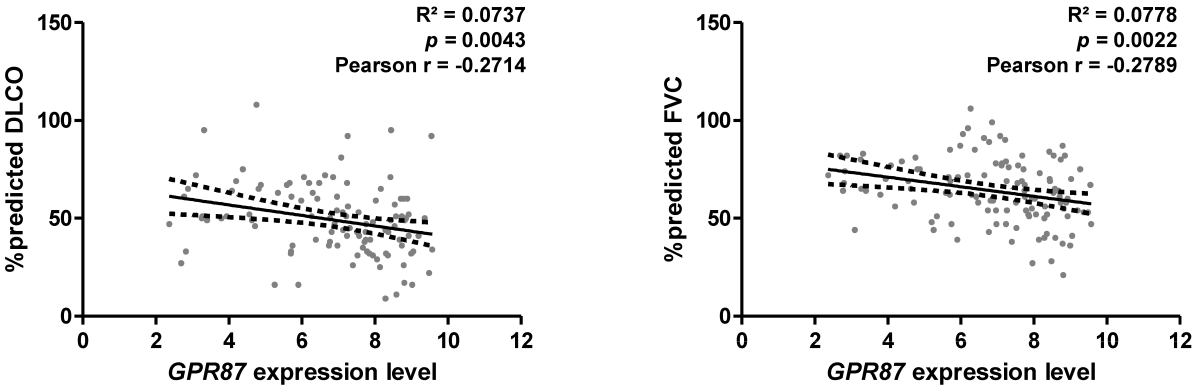

Figure 4.22: Correlation between *GPR87* expression and pulmonary function in IPF. The DLCO, FVC, and *GPR87* expression values were extracted from the *GSE47460* microarray dataset. Pearson correlation analysis revealed a significant negative correlation. Linear regression analysis is shown as continuous line with the 95% confidence interval (dashed lines).

4.5.3 Lung cell-specific expression of GPR87

In the beginning we also focused on alveolar epithelial type II cells, but besides a very low baseline expression of *Gpr87*, we were not able to observe any difference

Figure 4.23: Analysis of tissue- and cell-specific localization of GPR87 with BioGPS. The tissue- and cell-specific expression of GPR87 in the *GSE1133* microarray dataset (GeneAtlas U133A, gcrma) was assessed with the help of BioGPS [131–133]. Image was created with BioGPS and exported. Expression values are shown as arbitrary fluorescence units.

between isolated pmATII cells from experimental lung fibrosis compared to control (Figure 4.24A). To get an impression about a possible tissue- and cell-specific expression of GPR87, we used a gene annotation database. BioGPS [131–133] (`http://biogps.org/#goto=genereport&id=53836`; The Scripps Research Institute (TSRI); last accessed: 25 January 2019) was used to analyze the "human gene atlas" microarray dataset (*GSE1133*) [134]. Results revealed a prominent expression in bronchial epithelial cells, especially compared to the low expression in whole lung tissue (Figure 4.23). These findings were furthermore confirmed by qRT-PCR, measuring the expression levels of GPR87 in cultured A549 (human epithelial lung carcinoma cell line), pmATII, 16HBE14o$^-$ (human bronchial epithelial cell line), and primary human bronchial epithelial cells (phBEC) (Figure 4.24B; *GPR87, Gpr87* mRNA expression mean±SD: A549 -11.80±0.87, pmATII -6.31±0.35, 16HBE14o$^-$ 0.07±0.38, phBEC 1.22±0.13; *** p<0.001).

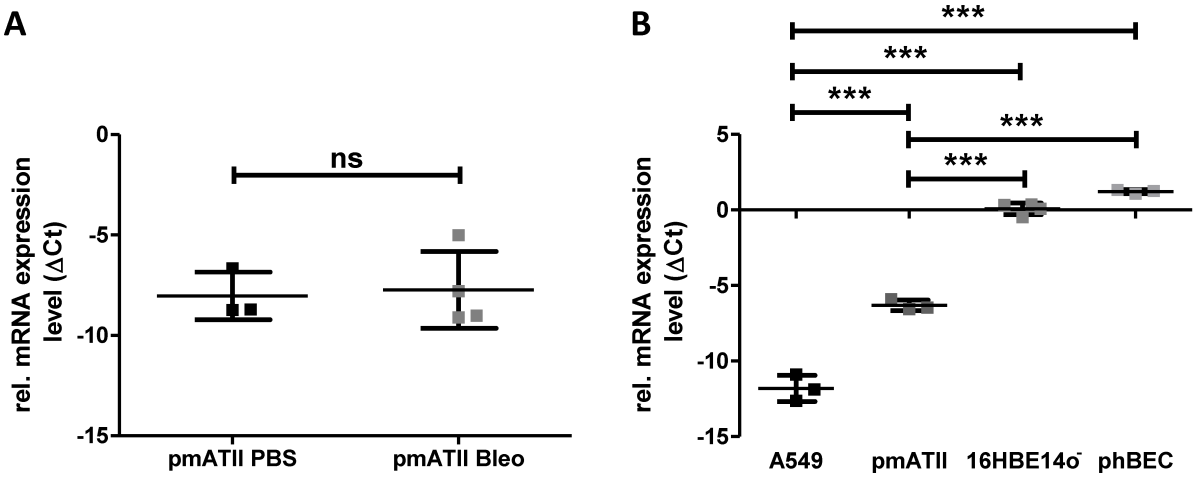

Figure 4.24: mRNA expression of *GPR87* in specific lung cells. (A) Freshly isolated pmATII cells from bleomycin- (n = 4) and PBS-treated (n = 3) mice were analyzed by qRT-PCR for their mRNA levels of *Gpr87*. No significant difference was observed. **(B)** Cultured A549 (n = 3), pmATII (n = 3), 16HBE14o$^-$ (n = 4), and phBEC (n = 3) cells were analyzed by qRT-PCR for their expression levels of *GPR87, Gpr87*. Significantly higher mRNA levels could be observed in the bronchial epithelial cells. Expression levels were normalized to *Hprt, HPRT* and are presented as ΔCt. Data is shown as mean±SD. Significance: * p<0.05; ** p<0.01; *** p<0.001; unpaired t-test, one-way ANOVA with Bonferroni's multiple comparison test.

Moreover, these results were complemented by a Pearson correlation analysis. Expression values of *GPR87* and bronchial epithelial basal cell markers, such as *KRT5*, *TP63*, and *KRT14* were extracted from the microarray datasets *GSE47460* or *GSE32537* for the IPF cases. Results showed a strong, significant, and positive correlation of *GPR87* with the corresponding expression levels of *KRT5* (Figure 4.25A, *GSE47460*, n = 121, R^2 = 0.8721, p < 0.0001, Pearson r = 0.9339; Figure 4.25B, *GSE32537*, n = 119, R^2 = 0.7501, p < 0.0001, Pearson r = 0.8661), *TP63* (Figure 4.25C, *GSE47460*, n = 121, R^2 = 0.1726, p < 0.0001, Pearson r = 0.4155; Figure 4.25D, *GSE32537*, n = 119, R^2 = 0.7550, p < 0.0001, Pearson r = 0.8689), and *KRT14* (Figure 4.25E, *GSE32537*, n =

119, $R^2 = 0.5088$, $p < 0.0001$, Pearson r = 0.7133).

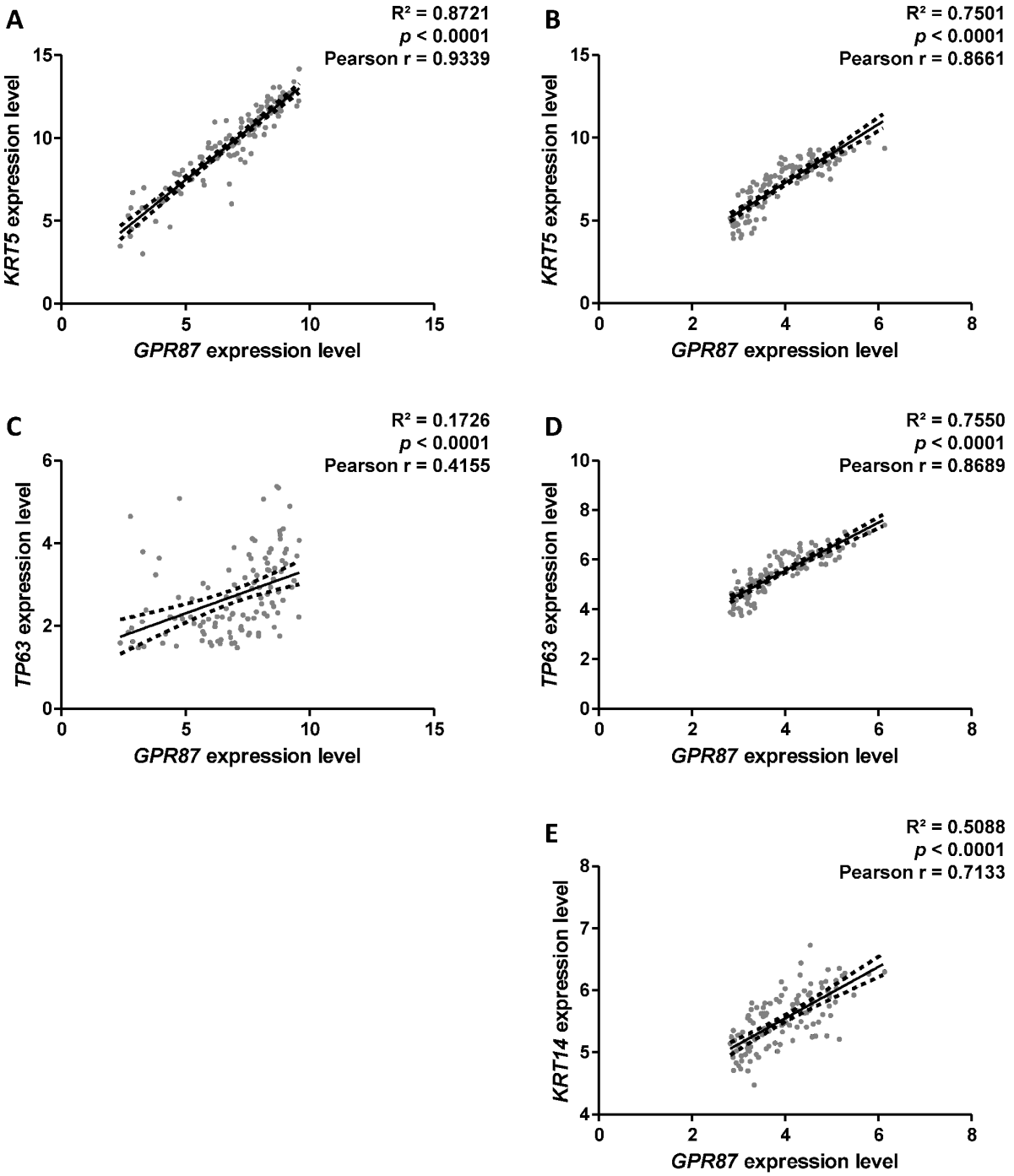

Figure 4.25: Correlation of *GPR87* with bronchial epithelial basal cell markers. Expression values were extracted from the respective microarray dataset for *GPR87* (*GSE47460*, *GSE32537*) as well as **(A)** *KRT5* (*GSE47460*), **(B)** *KRT5* (*GSE32537*), **(C)** *TP63* (*GSE47460*), **(D)** *TP63* (*GSE32537*), and **(E)** *KRT14* (*GSE32537*). Pearson correlation analysis revealed a significant positive correlation (A-E). Linear regression analysis is shown as continuous line with the 95% confidence interval (dashed lines).

4.5.4 TGF-β induced expression of GPR87 *in vitro*

Finally, we sought to investigate if the expression of GPR87 is influenced by the pro-fibrotic mediator transforming growth factor beta 1 (TGF-β). Therefore, the cultured human bronchial epithelial cell line 16HBE14o$^-$ was treated with 5ng/ml TGF-β for 4h, 6h, and 24h. The successful treatment was confirmed by significantly increased expression levels of the plasminogen activator inhibitor 1 (*PAI1*) (Figure 4.26B; mRNA expression mean±SD TGF-β *vs.* CTRL: *4h* 6.91±0.25 *vs.* 4.55±0.28; *6h* 6.61±0.31 *vs.* 4.44±0.59; *24h* 6.53±0.38 *vs.* 4.61±0.35), a well-known downstream target of TGF-β signaling [144]. Furthermore, the *GPR87* mRNA levels were significantly increased upon 4h, 6h, and 24h of treatment (Figure 4.26A; mRNA expression mean±SD TGF-β *vs.* CTRL: *4h* 0.96±0.47 *vs.* 0.33±0.41; *6h* 0.79±0.28 *vs.* 0.03±0.60; *24h* 0.50±0.25 *vs.* 0.07±0.38).

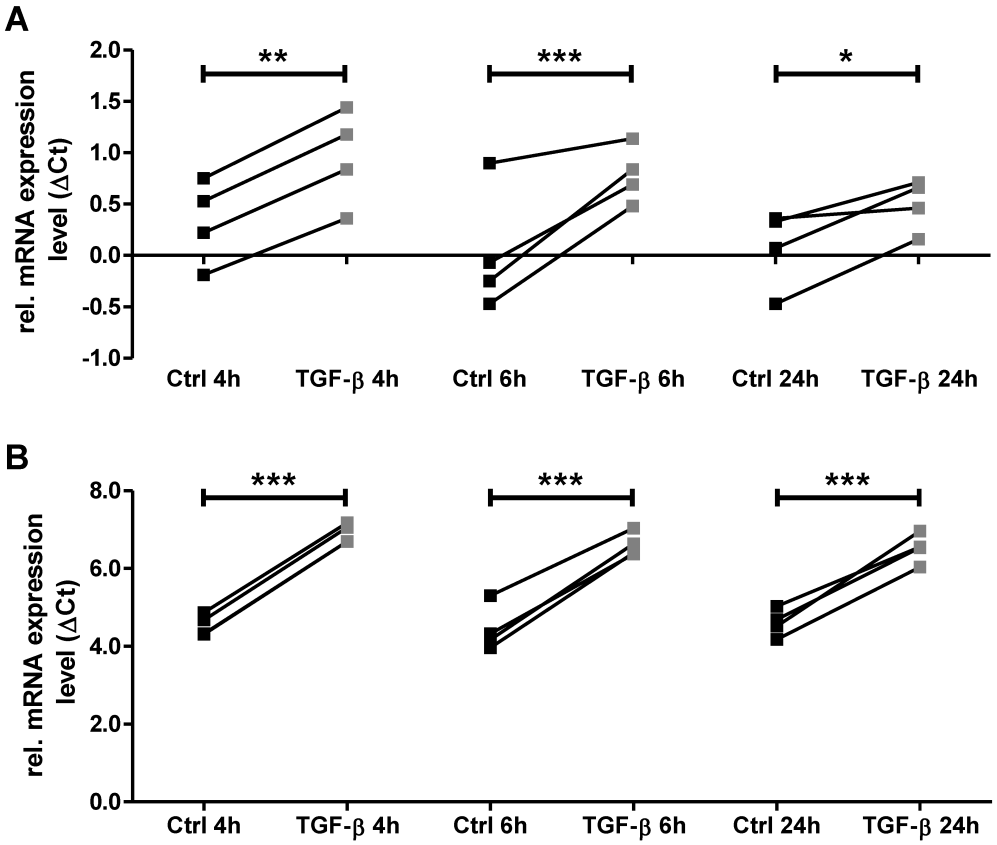

Figure 4.26: Increased *GPR87* expression after treatment with TGF-β. Cultured 16HBE14o$^-$ cells were treated with TGF-β for 4h, 6h, and 24h. The mRNA expression levels of **(A)** *GPR87* and **(B)** *PAI1* were analyzed by qRT-PCR. Results showed a significant increase of *GPR87* and *PAI1* for all time points after TGF-β treatment. Expression levels were normalized to *HPRT* and are presented as ΔCt. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; repeated measures ANOVA with Bonferroni's multiple comparison test.

Subsequently, isolated primary human bronchial epithelial cells (phBECs) were cul-

tured and treated with 2ng/ml TGF-β for 24h. qRT-PCR analysis revealed significantly elevated levels of *GPR87* (Figure 4.27A; mRNA expression mean±SD TGF-β *vs.* CTRL: *24h* 0.98±0.20 *vs.* 0.06±0.23) and *PAI1* (Figure 4.27B; mRNA expression mean±SD TGF-β *vs.* CTRL: *24h* 8.36±0.38 *vs.* 5.31±0.42) upon TGF-β treatment.

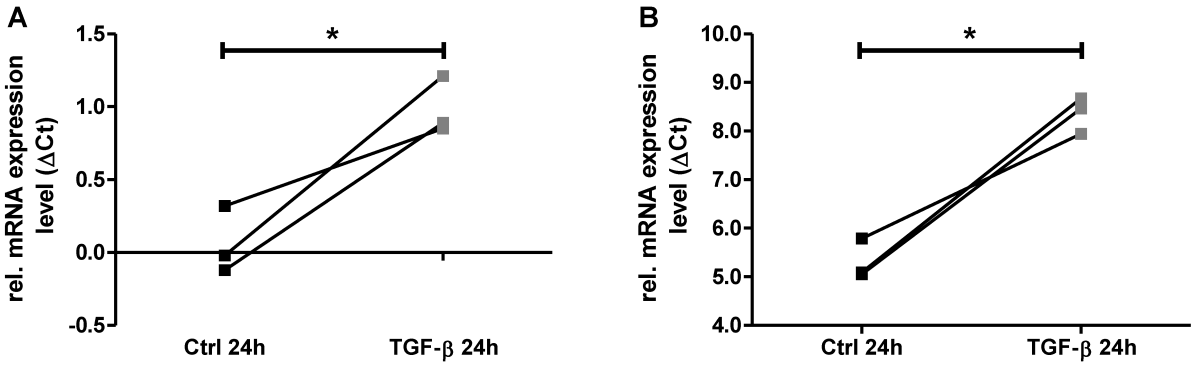

Figure 4.27: TGF-β induced upregulation of *GPR87* in phBECs. Treatment of phBECs with 2ng/ml TGF-β for 24h resulted in significantly increased mRNA expression levels of **(A)** *GPR87* and **(B)** *PAI1*. Expression levels were normalized to *HPRT* and are presented as ΔCt. Significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; paired t-test.

To summarize the results of the preceding sections, we were able to show that PSAT1 and GPR87 were also increased on protein level in IPF and bleomycin-induced lung fibrosis in mice. Furthermore, while ATII cells seemed to be one of the primary sources for the increased PSAT1 expression, GPR87 showed to be mainly expressed by bronchial epithelial cells. In addition, we observed a significant upregulation of GPR87 upon TGF-β treatment *in vitro*.

5 Discussion

Chronic respiratory diseases significantly impact the lives of affected patients due to their lasting nature and incurability. Two of these devastating and eventually lethal pulmonary disorders, the idiopathic pulmonary fibrosis and non-small cell lung cancer are marked by excessive tissue production and irreversible damage to lung structure as well as function. Recent scientific advancement showed common pathomechanistic and epidemiological characteristics of these diseases, which led to the assumption that there exists a connection of both disorders.

In the present thesis, we aimed to elucidate this connection by analyzing patterns of gene expression alterations in IPF and NSCLC. Furthermore, we focused on generating a set of misregulated candidate genes and subsequently analyzing them in the context of pulmonary fibrosis. We hypothesized that this approach leads to a better understanding of the IPF pathogenesis and to the detection of novel therapeutic targets.

The results of the current study showed for the first time a significant link between the pathogenesis of IPF and NSCLC, based on a systematic analysis of differentially expressed genes in both diseases. Up- and downregulated genes in NSCLC were extracted from three microarray datasets (*GSE44077*, *GSE43458*, *GSE18842*) and analyzed for their expression in IPF versus control samples (*GSE47460* dataset), which allowed for a strikingly clear separation of IPF and control. In accordance with these findings, GSEA demonstrated a significant enrichment of DEGs from the three NSCLC microarray datasets among those genes up- or downregulated in the IPF dataset (*GSE47460*). These results eventually led to the generation of a set of 92 candidate genes, which were significantly upregulated (log2 fold change > 1; adjusted p-value < 0.05) in NSCLC and contributing to the IPF phenotype. The gene set showed an IPF-specific signature and was especially enriched among fibrotic alveolar epithelial type II cells. Most importantly, this gene set was not only a random selection but showed a significant number of common interactions as well as an enrichment of functional relationships. Subsequently, those 92 candidate genes were evaluated and the G protein-coupled receptor 87 (GPR87) as well as the phosphoserine aminotransferase 1 (PSAT1) were chosen for further analysis. Both genes were found to be significantly upregulated in whole tissue samples of human IPF and bleomycin-induced lung fibrosis in mice. Moreover, PSAT1 upregulation was mainly detected in ATII cells, whereas GPR87 was primarily expressed by bronchial epithelial cells. Interestingly, the *in vitro* treatment of bronchial epithelial cells with the pro-fibrotic mediator TGF-β resulted in a significant upregulation of GPR87.

Taken together, this study demonstrated a pathogenic link between IPF and NSCLC by applying a top-down systems biology approach, which also resulted in a condensed set of 92 candidate genes. Initial experiments revealed two promising targets (GPR87, PSAT1) for novel therapeutic strategies, which require further analysis.

5.1 Systems biology

Recent advances in science as well as rapidly developing technologies have brought the field of systems biology more and more into the focus of general scientific interest. Systems biology is an approach to capture and understand the complexity of a system, from single cells to the whole organism, by analyzing each individual element, their interactions, and their functions in the respective system [145]. The omics techniques have enabled the generation of massive and all-encompassing amounts of data on various levels of an organism through the use of high-throughput methods, such as DNA microarrays or RNA sequencing. This progress is fostered by the development of novel bioinformatic analyses, which are possible due to the emergence of supercomputers and drastically increased performance of the general information technology (IT) infrastructure.

The present study was conducted as a top-down systems biology approach, using public transcriptome profiling datasets from RNA-Seq as well as microarrays. The top-down approach is designed to collect the most complete data possible about largely unknown systems in order to obtain a comprehensive overview and subsequently aims to analyze the bottom level of single elements (e.g. individual genes, molecules), their relationships, and interactions [145, 146]. This approach holds tremendous advantages, since it does not require any knowledge at all about the involved mechanisms, instead an overall picture is taken, which allows drawing conclusions and forming hypotheses based on the observed correlations [146, 147]. IPF and NSCLC share various similar pathogenetic mechanisms as well as epidemiological connections, such as IPF being an independent risk factor for the development of lung cancer and a significantly increased incidence rate of lung cancer in IPF populations [54]. This possible connection has been extensively discussed in recent years, but the underlying pathomechanisms have not been investigated beyond mere hypothesis generation. Due to this initial situation, we decided to use the aforementioned top-down approach in order to reduce the risk of biased decision-making based on speculations.

Another major strength of this approach lies within the possibility to simultaneously examine more than just individual or a handful of elements through the intensive use of omics techniques. The analysis of single elements at a time has usually required systematic trial and error, which has not only been very time consuming, but has also resulted in context being ignored and mechanisms being missed [146]. Moreover, the individual analysis of each possibly involved element seems to be an impossible task due to the large amount of unknown molecules [146]. In contrast, systems biology allowed us to characterize and compare genome-wide changes at a given point in time in different systems (e.g. healthy controls versus disease). Subsequently, we were able to use these data to generate groups and networks of significantly altered elements and analyze them for common patterns and biological functions. Finally, we have advanced to the molecular level to extract individual elements of interest and examine them for their function in the diseased system. Using this approach, we have significantly increased the likelihood of detecting biologically relevant processes, since a simultaneous and uniform change in the expression of a functionally related group of genes is significantly more meaningful than a change in the expression of an individual gene [148].

Systems biology heavily relies on the generation of large amounts of data, which on the one hand provides the aforementioned advantages, but on the other hand can also be seen as a weakness of this approach. The main challenge in this context is caused by the exceedingly high dimensionality of the generated data, which requires bioinformatic methods to simplify the data by reducing their dimensionality and organizing them in such a way that we can evaluate and understand the results [147, 148]. Omics experiments, such as microarrays, include several thousand variables for each sample, which a human being could not grasp nor compare. However, simplifying data by reducing their dimensionality is challenging and bears the risk of overfitting the model by selecting too many variables as well as the risk of losing data, which can cause important aspects to be missed. To realize the reduction of dimensionality, in this study we relied on well established methods, such as gene set enrichment analysis and principal component analysis. Thus, we tried to avoid the loss of significant data.

Furthermore, the top-down systems biology approach is based solely on the analysis of distinct phenotypes, such as healthy controls, IPF, and NSCLC, which usually leads to the adoption of mechanisms that are based on correlations only, but do not necessarily reflect the reality [146, 147]. Therefore, to address this issue, we did not only focus on idiopathic pulmonary fibrosis in the approach of this study, but also included a second disease of the lung (NSCLC), since common pathomechanistic and epidemiological characteristics of these two diseases are frequently discussed by recent literature. Both disorders were initially compared to the respective controls, and subsequently an overlap of the two analyses was generated to further condense the results. By focusing on this overlap of two pathomechanistically similar diseases, we aimed to uncover significant mechanisms that are relevant in reality. Finally, the proposed model was experimentally validated by generating a protein-protein interaction network and performing annotation enrichment analysis. Additional data on protein level in terms of proteomics were not available at the time this study was performed.

To date, only few comparable studies have been published that performed a systematic comparison of IPF and NSCLC. Both Spek *et al.* [149] and Leng *et al.* [150] have adopted a top-down systems biology approach for their respective study and also started by analyzing public transcriptome profiling datasets. In addition, both studies partially used the same datasets, such as *GSE32537*, *GSE47460*, and *GSE18842*, which were already used for the present thesis. However, the total number of samples included in this thesis was higher than in the two comparable studies, especially the number of lung cancer samples. Furthermore, there exists a major difference in the definition of the cutoff values for differentially expressed genes, which we defined as those genes with a |log2 fold change| > 1.0 and an adjusted p-value < 0.05. Spek *et al.* [149] as well as Leng *et al.* [150] have chosen different approaches for the definition of their cutoff values with a fold change > 1.5 and p-value < 0.01 [149] or solely a p-value < 0.05 [150], respectively. The recommendations in current literature are to use a combined cutoff consisting of fold change and p-value, with a high fold change increasing reproducibility and a balanced p-value improving specificity and sensitivity [151]. Thus, we think that we meet these recommendations with our cutoff values, while in particular the use of p-values alone as in Leng *et al.* [150] is not advised [151]. Spek *et al.* [149] have not performed further

experiments of dimensionality reduction, but have analyzed the enrichment of gene ontology terms. In contrast, Leng *et al.* [150] performed various methods of dimensionality reduction as well as annotation enrichment analysis and finally supported their findings by immunohistochemistry of human lung tissue.

In summary, the top-down systems biology approach has been shown to be an appropriate strategy to elucidate the aforementioned research questions and to address the aims of this study. Furthermore, this has significantly increased the likelihood of discovering biologically relevant processes and making unbiased decisions.

5.1.1 Omics techniques

As already mentioned, almost all of the transcriptome profiling datasets used in this thesis were from microarray studies, while only one dataset was from a RNA-Seq experiment. In the past years, the popularity of the RNA-Seq technique and especially the single-cell analysis has significantly increased, whereas the use of microarrays has steadily declined [152]. DNA microarrays have played an important role in the area of omics and have profoundly changed our ability to study the transcriptome [153]. Due to their intensive use, DNA microarrays are very cost-effective and the available analytical tools are well developed, but there are some major restrictions to this technique [153, 154]. First of all, only the expression of genes whose sequence is known and for which a probe is present on the microarray can be analyzed, furthermore, it is not possible to accurately measure very low or high gene expression, the binding affinity of different probes is varying, and there exists cross-hybridization as well as background signaling [153, 154]. The RNA sequencing technique allows us to overcome many of the restrictions mentioned above. There is no knowledge required about the gene sequences, the technique works without the use of probes, background signaling is not an issue, and gene expression can be measured accurately, even for those genes with very low or high expression [153, 154]. Most importantly, the RNA sequencing technique can be done with small amounts of RNA, which also enables single-cell analysis [154]. Therefore, from today's perspective, we would have performed the experiments with RNA-Seq datasets. However, these were hardly available at the time of the experimental phase of this thesis, and thus we primarily used microarray datasets. In particular, the analysis of cell-specific expression (Figure 4.11) of the *Overlap gene set* would have been much easier to perform by using single-cell analysis. This is of great interest because it has been shown that the lung contains a wide variety of different cell types [155].

5.1.2 Gene set enrichment analysis

The use of omics techniques results in the generation of massive amounts of data, which raise the importance of applying methods for dimensionality reduction. Furthermore, the interpretation of omics data requires the exploration of common patterns and biological functions among groups of significantly altered elements, instead of analyzing individual targets. This is based on the assumption that diseases are associated with changes in specific groups of genes rather than single genes [137]. Therefore, we relied on gene set

enrichment analysis in the present thesis. GSEA has been first published and described by Subramanian *et al.* [136] and Mootha *et al.* [137], who used it to analyze various omics data of human cancer and diabetes mellitus.

GSEA is one of the various applications of gene set analysis, with the aim of studying whole groups of genes instead of analyzing individual genes. The original idea behind this method is to examine predefined sets of genes for their enrichment and thus their relationship with one of two phenotypes (e.g. healthy controls versus disease) [137]. This is achieved through a multi-stage process and the use of specialized mathematical algorithms, which analyze whether a predefined gene set shows a random, uniform distribution among a ranked dataset or a significant concentration within the differentially expressed genes either at the top or the end of the ranked dataset [136]. GSEA is computing an enrichment score for every gene set by using the so-called Kolmogorov-Smirnov statistic, which has been modified to be correlation-weighted [136, 156]. Because this statistical approach considers all genes of the ranked dataset, GSEA was classified as a competitive method by Goeman *et al.* [157], which is based on the examination of the null hypothesis that the gene set of interest contains at most as many DEGs as the rest of the ranked dataset. Subsequently, the statistical significance is calculated with a phenotype/sample permutation approach, which considers biologically relevant gene dependencies [136, 156]. Finally, multiple hypothesis testing is conducted and the leading-edge is assembled of those genes contributing most to the enrichment score [136].

To date, various approaches for gene set analysis have been published and multiple comparative studies have been conducted to identify the most accurate method. The scientific community has not reached agreement on a common standard for gene set analysis, which is mainly due to missing standard datasets for testing and benchmarking purposes [148]. All of the published methods have in common that they aim to provide a statistical statement about the significance of a gene set of interest [158]. This is achieved either by a phenotype/sample permutation approach (e.g. GSEA) or through gene sampling/permutation, which are based on the assumption of independence of either phenotypes/samples or genes, respectively [157]. The use of gene set analysis methods relying on gene sampling/permutation is strongly discouraged by current literature because of the biologically irrational assumption of gene independence and the proven impact of gene correlations/dependence on the significance of a gene set [148, 156–159]. On the other hand, the phenotype/sample permutation approach requires larger datasets with a certain amount of samples, which is not always feasible [159]. To address this issue, GSEA also offers the option of gene sampling/permutation for small datasets [136].

In addition to the competitive methods (e.g. GSEA), Goeman *et al.* [157] also described the self-contained methods, which consider only those genes included in the gene set of interest and which are favored by these authors. This view is based on the fact that the self-contained methods are more powerful and restrictive, but this is actually a major weakness of the approach and may result in a gene set being considered significant because of a single DEG [157, 159]. The null hypothesis being examined by the self-contained approach states that there exist no DEGs in the gene set of interest [157].

Currently, there are no commonly accepted recommendations, standards, or guidelines

for the specific use of the various gene set analysis methods. Therefore, we determined the most suitable method for our study purposes. In contrast to the self-contained methods, the competitive method GSEA seemed more reliable, balanced, and biologically relevant due to being less powerful and considering the whole dataset. In addition to these advantages, GSEA also respects gene dependencies, is frequently updated by its authors, and offers an easy-to-use software solution. In the present study, we relied on large transcriptome profiling datasets to be able to use the phenotype/sample permutation approach of GSEA, which is generally recommended by the current literature. However, we had to use the gene sampling/permutation approach in a few cases because the available datasets were too small. Gene sampling/permutation is not recommended by recent publications and the results obtained by this approach have to be interpreted with caution. This represents a limitation of the present study and requires the assembling of larger datasets with a higher number of individual samples for future research. But, although this approach is not recommended, it at least enabled us to perform gene set analysis of small datasets and to formulate a working hypothesis for subsequent analyses.

5.2 Animal models of IPF

The bleomycin-induced experimental model of pulmonary fibrosis in mice has already been introduced in section 1.1.3. In short, the administration of bleomycin has been shown to initiate pulmonary fibrosis as a severe side effect, which has been used for decades to mimic IPF in animal models. Various results of the present thesis were generated based on the murine bleomycin-induced experimental model of pulmonary fibrosis, namely on the *pmATII* [111] microarray dataset and the use of isolated ATII cells from bleomycin-treated mice. Furthermore, we planned on using this model for future experiments, such as a *Gpr87* knockout mouse.

To date, multiple animal models of IPF have been described, which use either different species (e.g. mice, hamsters, rats, dogs, primates) or different compounds (e.g. bleomycin, asbestos, silica, cytokines, radiation) [160, 161]. The bleomycin-induced experimental model of pulmonary fibrosis in mice has been recommended by an official ATS workshop report as standard for preclinical evaluation of IPF treatments [162]. This model has proved its clinical significance by contributing to the assessment and approval of the IPF treatments with nintedanib and pirfenidone [161]. Furthermore, the experimental pulmonary fibrosis caused by bleomycin has been thoroughly investigated and described because this model is frequently used in IPF research. This comprehensive knowledge represents a major strength of the bleomycin model. In addition, Peng *et al.* [163] demonstrated that bleomycin-induced pulmonary fibrosis in mice largely resembles human IPF at the molecular level.

In contrast to human IPF, the bleomycin model shows severe inflammation in the first days after administration and rapid progression to pulmonary fibrosis, which is self-limiting and eventually vanishes [42]. However, we relied on the murine bleomycin model for the *in vivo* experiments because of its aforementioned strengths and missing

alternatives. The dependence on this model is a limitation of the present study because the selection of potential candidate genes was tied to their expression in the murine bleomycin model. The two genes of interest *Anln* and *Bub1* did not show a significant and reproducible upregulation in the bleomycin model, unlike their significantly increased expression in human IPF samples. Therefore, we had to exclude both genes from further analyses. Future studies might reevaluate the original 92 candidate genes for further investigation with the help of novel experimental *in vivo* settings or *in vitro* lung organoids.

5.3 Relationship of IPF and NSCLC

In recent years, increasing evidence has been suggesting a pathomechanistic connection between lung cancer and IPF. The risk of developing lung cancer increases significantly over time for IPF patients, which is represented by an increase of the cumulative incidence (Figure 1.4) [56–60]. However, the approved treatments of IPF with pirfenidone and nintedanib have demonstrated a significant effect on decelerating the IPF disease progression [24, 164]. This in turn could further increase the time-dependent risk of developing lung cancer and might lead to a rising incidence in IPF patients in the future. The affected patients show a significantly worse prognosis and an increased risk of acute exacerbations as well as other complications caused by treatments [57, 60]. Thus, there exists an urgent need for curative therapeutic options, which stresses the importance of understanding the involved pathomechanisms and revealing possible targets for novel therapies. Initially, we aimed to further investigate and confirm the link between IPF and lung cancer in the present study by quantifying shared patterns of gene expression alterations and establishing a common foundation of both diseases.

We performed preliminary comparisons of differentially expressed genes in IPF and NSCLC by using publicly available microarray datasets and generating Venn diagrams (Figure 4.1). The analysis revealed an interesting overlap of DEGs and indicated a possible connection of both diseases but required further quantification and statistical evaluation. We decided not to include lung cancer in general for the subsequent analyses but to solely focus on the subgroup of NSCLC because of the following reasons. First, NSCLC has been demonstrated to mainly originate from the bronchial and alveolar epithelium, which is assumed to play a key role in the pathogenesis of IPF [52]. Second, the subgroup of NSCLC represents the most common form of lung cancer found in IPF patients [58]. Though, comparing the two diseases, it is also important to note that essential hallmarks of lung cancer have not been shown in IPF. Idiopathic pulmonary fibrosis does not feature metastatic potential or the invasion of other organs besides the lung, there exists no clonal expansion of affected cells, and somatic mutations have no part in the IPF pathogenesis [165].

At the time the experiments were performed, only one extensive IPF microarray dataset (*GSE47460*) with a sufficient number of individual samples existed. In contrast, multiple larger NSCLC microarray datasets (*GSE44077*, *GSE43458*, *GSE18842*) were available for the present study. Therefore, we sought to elucidate possible discrep-

ancies of DEGs in those three NSCLC microarray datasets and to combine them for the subsequent analysis. The generated overlap of the NSCLC datasets demonstrated a surprisingly small number of common DEGs (Figure 4.2) and gave an impression of their individuality. These results confirmed the importance of including all three NSCLC datasets in the GSEA to further increase its significance. Moreover, comparing the extent of the IPF-NSCLC overlap (upregulated 9.76%; downregulated 13.1%) with the NSCLC-NSCLC overlap (upregulated 8.91%; downregulated 19.4%), the percentage of upregulated DEGs was almost the same. This further strengthened a possible link between the two diseases. Recently, the diagnostic guidelines for IPF have been revised and published [12, 20]. The IPF microarray dataset has been assembled under the former diagnostic guidelines and might have contained few samples that are no longer classified as IPF from today's perspective. This should not have affected the actual results because of the large number of individual samples included in the dataset.

GSEA (Section 5.1.2) has been used to obtain an assessment and statistical statement about the enrichment of DEGs from NSCLC in the IPF dataset. The analysis uncovered a shared pattern of significant gene expression alterations in IPF and NSCLC, which proved a common foundation of both diseases on the level of gene expression and regulation. These results were further supported by heatmaps (Figure 4.4 and 4.5), showing a distinct separation of IPF and control cases based on their expression levels of those genes included in the generated NSCLC gene sets. This common subset of similarly misregulated genes may be involved in the common existence and the pathomechanistic connection of IPF and NSCLC. Leng *et al.* [150] confirmed these results by analyzing different publicly available microarray datasets of IPF and NSCLC. The authors reported a subset of 79 DEGs, which they identified as potential common mediators for the pathogenesis of both diseases [150]. However, no detailed list of the respective genes was published, and a comparison to our subset of 92 genes was not possible.

In contrast, based on their analysis of DEGs in IPF and NSCLC, Spek *et al.* [149] argued against a common foundation of similar gene expression alterations, although they reported a shared signature of DEGs and an enrichment of functional annotations. The authors asserted that the number of common DEGs is rather small, especially compared to the amount of directly opposed DEGs [149]. Relying solely on sheer numbers to draw this kind of conclusion, one could also argue that the three NSCLC datasets (Figure 4.2) do not share significant similarities of gene expression alterations, but this is the same disease in all three datasets. The total number might appear small because of the individuality of all samples, but there exists a common core of DEGs for NSCLC as well as for the comparison of IPF and NSCLC, which requires further analyses to verify its significance. In this context, the present study would have benefited from the analysis of datasets with paired IPF, NSCLC, and control samples. This kind of data did not exist but should be addressed by future research.

In summary, we were able to demonstrate a common core of gene expression alterations in IPF and NSCLC, which we could verify and statistically assess by using GSEA. These results emphasize the connection of IPF and NSCLC. Furthermore, this core of DEGs could be an important driver for the pathogenesis of both diseases, and its contribution requires further analysis.

5.4 Significance of the Overlap gene set

Subsequently, we aimed to use the results of the previous GSEA to assemble a gene set of common upregulated DEGs, which we planned to further characterize in the context of IPF. The leading-edge gene list is generated for each GSEA run and contains those genes, which form the maximum enrichment score, constitute a key element of the original gene set, and can represent a relevant functional subset [136]. Subramanian *et al.* [136] recommend to rely on the leading-edge gene list, especially for gene sets with high enrichment scores and when using manually assembled gene sets for GSEA. The original NSCLC gene sets 1-3 of upregulated DEGs had high enrichment scores (Table 4.1) and were manually generated by extracting DEGs from NSCLC microarray datasets. Therefore, we used the three leading-edge gene lists of the corresponding GSEA runs and generated an overlap, which was labeled *Overlap gene set*. This condensed subset included 92 genes that were significantly upregulated in all NSCLC microarray datasets, enriched among the genes upregulated in IPF, and part of the leading-edge gene lists.

Manually assembled gene sets are based on artificial cutoff values, which could lead to the inclusion of false positive or the exclusion of false negative genes. To avoid the inclusion of false positive genes, we selected rather restrictive cutoff values (log2 fold change > 1; adjusted p-value < 0.05) and combined this with the use of the leading-edge gene lists. This approach might have led to the exclusion of false negative genes, which had to be accepted. Therefore, we aimed to verify the expression of five candidate genes of the *Overlap gene set* in lung tissue samples of NSCLC, human IPF, and murine bleomycin-induced lung fibrosis. We were able to demonstrate a significantly increased expression of *ANLN*, *BUB1*, *CTHRC1*, *GPR87*, and *PSAT1* in human IPF. *Cthrc1*, *Gpr87*, and *Psat1* showed a significant upregulation on d14 in bleomycin-induced lung fibrosis. Finally, *ANLN* and *GPR87* were significantly upregulated in human lung tissue of squamous cell carcinoma, whereas *CTHRC1* and *PSAT1* were increased in human adenocarcinoma samples. These results indicated that individual candidate genes for subsequent mechanistic studies had to be selected with caution because we were not able to confirm all candidate genes in our tissue samples. Furthermore, it demonstrated the importance of differentiating between the NSCLC subtypes of squamous cell carcinoma and adenocarcinoma. Hence, *GPR87* and *PSAT1* were chosen for further analyses.

5.4.1 Overlap gene set reveals an IPF-specific signature

To analyze the *Overlap gene set* in the context of IPF, we sought to determine if this subset showed an IPF-specific signature. Therefore, GSEA was performed to evaluate the enrichment of the *Overlap gene set* in another independent IPF microarray dataset. We observed a significant enrichment of the *Overlap gene set* in the IPF phenotype of the dataset (Figure 4.8), which corroborated an IPF-specific signature of this gene set.

Subsequently, to further support these results, we performed principal component analysis of the *Overlap gene set*. PCA represents a common technique for dimensionality reduction, which is required to analyze, summarize, and display large amounts of data [148]. The individual principal components are computed with the help of mathe-

matical algorithms, represent the maximum possible variance of the original data, and are uncorrelated and perpendicular to each other [142]. In general, the first two or three principal components are used to graphically display the individual samples of a dataset in two or three dimensions, respectively. Instead of using all the gene expression values of each sample, these two or three principal components are sufficient to visualize the distribution of all samples, which helps to detect specific features, patterns, or groups of these samples [142]. The individual genes can be visualized as vectors/arrows, with the length representing the weight of the respective gene in regard to the principal components [142].

Performing PCA of the *Overlap gene set* and its corresponding gene expression values in the IPF, COPD, and control samples of the *GSE47460* microarray dataset, we were able to significantly reduce the dimensionality and to preserve a high amount of the original variance (Figure 4.9 and 4.10). We observed an interesting overlap of the COPD and control samples, whereas the IPF group showed a distinct, relevant separation from the other two groups. In addition, the 92 genes had a large weight for the principal components and directed at the IPF cluster. These results were supported by Leng *et al.* [150], who also performed PCA and demonstrated a similar clustering and separation of IPF and control samples, which was based on their subset of common DEGs. The use of PCA could also be seen as a limitation because of the inevitable data loss connected to the methods of dimensionality reduction. However, there exists no common alternative to plot and visualize such high dimensional data. In conclusion, these results verified an IPF-specific signature of the *Overlap gene set*, which strengthened its possible contribution to the pathogenesis of IPF.

5.4.2 Cell-specific association of the Overlap gene set

To further characterize the *Overlap gene set*, we analyzed the possibility of a cell-specific association. Using GSEA, we were able to confirm a significant enrichment of the *Overlap gene set* among those genes upregulated in fibrotic ATII cells from human IPF and murine bleomycin-induced pulmonary fibrosis (Figure 4.11 and Table 4.3). In contrast, there was no significant enrichment in the IPF phenotype of isolated human lung fibroblasts, which represent another major cell type involved in the development of IPF. This cell-specific enrichment in fibrotic lung epithelial cells might have been influenced by the use of NSCLC gene sets for the generation of the *Overlap gene set* because NSCLC mostly originates from the lung epithelium [52, 120]. However, these GSEA runs confirmed an epithelial cell-specific signature of the *Overlap gene set*, which highlighted the contribution of epithelial cells to the pathogenesis of IPF as well as NSCLC and established a possible pathomechanistic link between both diseases. In addition, these results supported the significance of the *Overlap gene set* in the context of IPF. Future studies should further evaluate and specify the cell-specific association of the *Overlap gene set* by using single-cell analysis.

Alterations of the phenotype, morphology, and function of epithelial cells in IPF were defined as *reprogramming* by a recent workshop report on IPF [166]. The commonly accepted model of the IPF pathogenesis includes repetitive injuries to the lung epithelium,

which eventually results in a distorted process of lung repair and regeneration and in *reprogramming* of the lung epithelium [9]. Both alveolar and bronchial/airway epithelial cells in IPF are significantly affected by this *reprogramming* [31]. The reprogrammed epithelium exhibits multiple alterations, such as increased proliferation, bronchiolar/basal and ATII cell hyperplasia, squamous metaplasia, areas of bronchiolization, and honeycombing [12, 120, 167, 168]. The honeycombed lesions represent fibrotic areas with cystic structures that contain bronchial epithelium [12]. This infiltration of the alveolar region by bronchial epithelial cells is also called bronchiolization [168]. Wistuba *et al.* [169] reported various of these changes as precancerous lesions for the development of NSCLC, especially for the formation of squamous cell carcinoma. This is of great interest because SCC is the most common subtype of NSCLC in IPF patients [58].

In conclusion, the epithelial cell-specific association of the *Overlap gene set* verified the hypothesis of lung epithelial cells as key drivers for the pathogenesis of IPF and NSCLC. Hence, this reprogrammed epithelium might represent the connecting element of both diseases. Common gene expression alterations (*Overlap gene set*) might induce or support epithelial *reprogramming*, foster the development of IPF, and eventually lead to the transition from precancerous lesions to the formation of NSCLC.

5.4.3 Overlap gene set includes enrichment of annotations

Performing annotation enrichment analyses of the *Overlap gene set*, we aimed to identify common pathways, processes, molecular activities, and cellular localization. This was required to ensure that the *Overlap gene set* was not just a random compilation but included functionally related genes with a common biological and mechanistic background. We used the public reference databases GO [122, 123], KEGG [124, 125], and REACTOME [126] to conduct the analyses, which demonstrated a significant enrichment of annotations in the *Overlap gene set* (Tables 4.4, 4.5, and 4.6; Appendix Tables 2 and 3). In particular, the results revealed the principal themes of cell cycle, proliferation, ECM organization, cellular senescence, and P53 signaling. Furthermore, the generation of a protein-protein interaction network showed a significant amount of interactions between the proteins, which corresponded to the 92 genes of the *Overlap gene set*. Overall, we were able to prove that the *Overlap gene set* consists of functionally related genes that exhibit multiple common interactions.

It is interesting to note that the annotation enrichment analyses uncovered principal themes that are commonly associated with the development of IPF and NSCLC, which ultimately confirmed the importance of the *Overlap gene set* and indicated an essential role in the pathogenesis of both diseases. The altered and distorted regulation of the cell cycle represents one of the key features of cancer development and causes unhampered proliferation of the affected cells [170]. Hence, Eymin *et al.* [170] proposed the cell cycle regulation as a possible target for novel cancer therapies. Besides NSCLC, the pathomechanistic feature of increased proliferation of the lung epithelium has also been reported in IPF, which was associated with an increased expression of WNT1-inducible signaling pathway protein 1 (WISP1) and epithelial cell transforming sequence 2 (ECT2), among others [11, 111, 120]. Moreover, this hyperproliferation and even fibrosis were reduced

by targeting and decreasing the activity of WISP1 or ECT2 [111, 120]. These results were complemented by the principal theme of P53 signaling, which represents a tumor suppressor gene that is commonly affected in lung cancer and leads to impaired cell cycle regulation [170]. Alterations of P53 signaling have also been observed in IPF [168]. Furthermore, recent evidence suggests that the extracellular matrix is actively supporting and adding to the development of IPF and NSCLC, which has been correlated with a distorted wound repair process in both diseases [171]. Finally, cellular senescence is another important hallmark of the fibrotic lung epithelium, and the treatment with senolytic drugs seemed to reduce the fibrotic potential [32]. Besides, it has been reported that senescence might foster the development of human cancer by inducing extracellular signaling [172]. These enriched principal themes seemed to represent important processes for the development of IPF and NSCLC. Therefore, future studies could elucidate individual themes with their corresponding set of genes and their contribution to the pathogenesis of both diseases.

5.5 PSAT1 is upregulated in fibrotic ATII cells

The final aim of the present thesis was to translate the results of the systems biology approach into mechanistic studies. Therefore, we addressed individual genes of the *Overlap gene set* and investigated their cell-specific expression as well as possible mechanisms that could have been responsible for their misregulation in IPF. The phosphoserine aminotransferase 1 (*PSAT1*) was the first candidate gene and represents a key enzyme of the serine biosynthesis pathway (Figure 1.5). Initially, we analyzed the PSAT1 protein levels in human IPF and bleomycin-induced lung fibrosis. Western blot analysis revealed significantly increased protein levels of PSAT1 in IPF and bleomycin-treated lung tissue (Figure 4.17). Most interestingly, we were able to confirm a cell-specific upregulation of the *Psat1* mRNA expression in isolated primary ATII cells from bleomycin-treated mice (Figure 4.19), which was consistent with our analyses of the *Overlap gene set*. Furthermore, there existed a significant negative correlation between the increased *PSAT1* expression levels in IPF and the decreased DLCO as well as FVC measurements in these patients (Figure 4.18). Overall, these results indicated a reliable upregulation of PSAT1 in pulmonary fibrosis, which was specifically expressed by the fibrotic ATII cells and might contribute to the disease severity of IPF.

Besides its contribution to the pathogenesis of various cancer types, increased PSAT1 levels have also been reported in NSCLC, where it supported cell proliferation and tumor progression [69, 72, 74]. PSAT1 expression is directly regulated by the transcription factor ATF4, which in turn is controlled by NRF2 [69, 70]. However, the role of PSAT1 in the IPF pathogenesis remains subject of current research. An increasing number of recent publications have reported that lung fibroblasts in human IPF are affected by the process of *metabolic reprogramming*, which stimulates the cellular metabolism, contributes to the transition from fibroblasts to myofibroblasts, supports ECM generation, and activates various processes [78, 79, 173, 174]. Furthermore, in this context of *metabolic reprogramming*, it has been shown that TGF-β and the fibrotic lung envi-

ronment induce an increased expression of the transcription factor ATF4, which leads to a significantly elevated expression of all enzymes of the *de novo* glycine and serine synthesis in lung fibroblasts and results in excessive collagen production [78, 79, 173, 174]. The activation of the serine biosynthesis pathway is particularly important because glycine represents by far the most frequent amino acid in the collagen structure and the production of glycine requires the amino acid serine as precursor [78, 174]. PSAT1 is one of the essential key enzymes of this serine biosynthesis pathway and showed significantly increased protein and mRNA levels in response to TGF-β stimulation in lung fibroblasts [79]. In addition, PSAT1 knockdown in lung fibroblasts disrupted the aforementioned mechanism of TGF-β/ATF4-related signaling and resulted in an inhibition of TGF-β-mediated collagen synthesis [79]. Zhu *et al.* [175] demonstrated that vitamin D3 treatment of human lung fibroblasts decreased the PSAT1 expression levels, which might affect *metabolic reprogramming* and diminish collagen production.

However, the contribution of PSAT1 upregulation in fibrotic ATII cells to the IPF development requires further research. Human ATII cells are able to acquire a TGF-β-mediated pro-fibrotic phenotype, express mesenchymal markers, and experience at least partial epithelial-to-mesenchymal transition [176]. Therefore, we hypothesize that pro-fibrotic ATII cells express increased TGF-β/ATF4-mediated PSAT1 in a fibrotic environment. Thus, pro-fibrotic ATII cells might undergo *metabolic reprogramming* and contribute to collagen production as well as ECM organization. This hypothesis is supported by the fact that Selvarajah *et al.* [173] and O'Leary *et al.* [174] demonstrated the highest ATF4 levels in hyperplastic ATII cells adjacent to fibroblast foci and only to a lesser extent in lung fibroblasts. In summary, PSAT1 represents an interesting target for the development of novel IPF and lung cancer therapies. Future studies should confirm the mechanism of PSAT1 upregulation in fibrotic ATII cells and address its contribution to the IPF pathogenesis.

5.6 GPR87 is expressed in HBECs and regulated by TGF-β

The G protein-coupled receptor 87 (*GPR87*) was the second candidate gene of the *Overlap gene set* that was further characterized by in-depth analyses. GPR87 is part of the G protein-coupled receptor family, but its biological function still remains unknown. Although LPA represents a possible ligand for GPR87, it seems to be rather unspecific with a low binding affinity [83]. To date, no other ligands for GPR87 have been described. In addition, its contribution to the pathogenesis of IPF requires further research.

Western blot analysis confirmed significantly elevated protein levels of GPR87 in bleomycin-induced murine lung fibrosis and human IPF, which was accompanied by increased expression of the mesenchymal markers α-SMA and CNN1 (Figure 4.20 and 4.21). Furthermore, *GPR87* expression levels showed a significant negative correlation with DLCO and FVC measurements in IPF, which indicated a correlation of GPR87 and disease severity (Figure 4.22). The analysis of the "human gene atlas" microarray dataset (*GSE1133*) [134] revealed an extraordinarily tissue- and cell-specific expression of *GPR87* in bronchial epithelial cells (Figure 4.23). We were able to confirm this

cell-specific pattern by demonstrating significantly higher *GPR87* expression in human bronchial epithelial cells, compared to the alveolar epithelium (Figure 4.24). These results were further supported by a remarkable, significant correlation of *GPR87* with the expression of the basal cell markers *KRT5*, *KRT14*, and *TP63* in human IPF (Figure 4.25). Moreover, a preliminary FACS analysis (n = 1) indicated the existence of a specific subset of GPR87 positive phBECs (Figure 5.1). Finally, we observed a significant upregulation of *GPR87* expression *in vitro* by treatment of human bronchial epithelial cells with the pro-fibrotic mediator TGF-β (Figure 4.26 and 4.27).

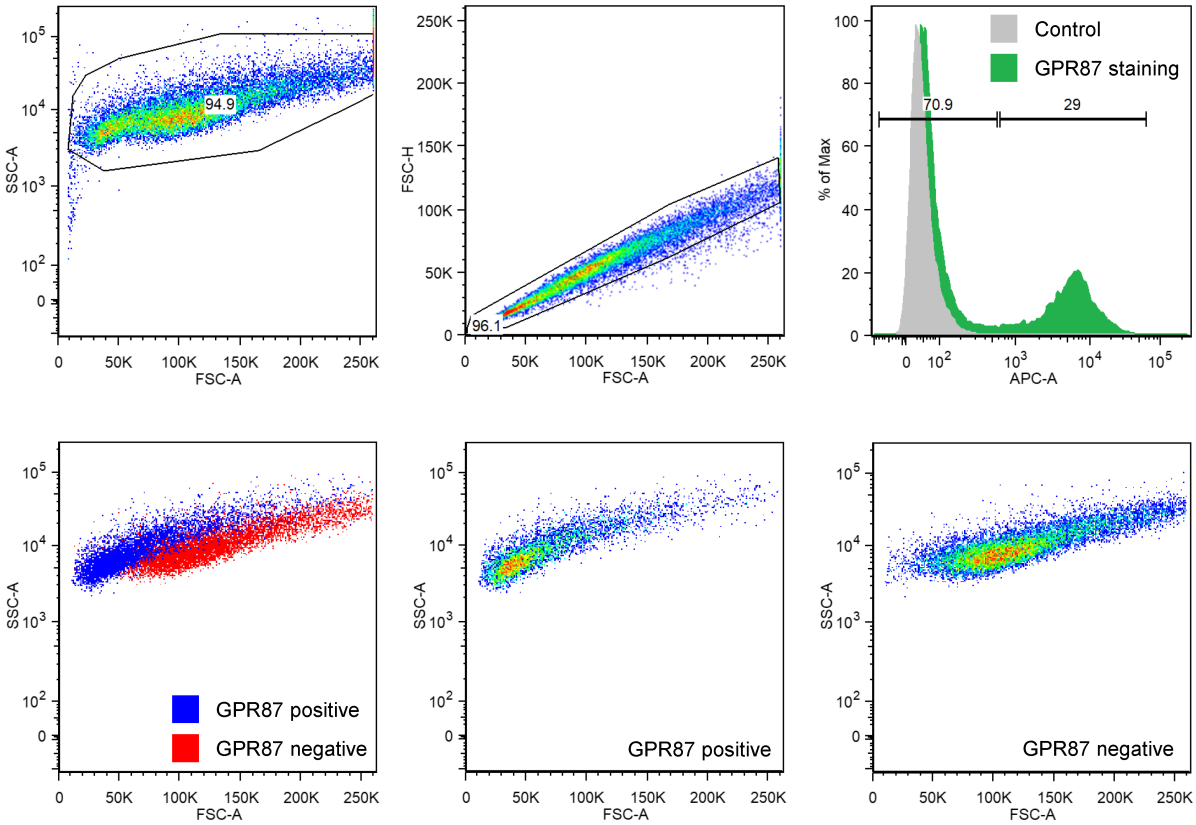

Figure 5.1: FACS analysis of GPR87 staining in isolated phBECs. Preliminary FACS analysis (n = 1) demonstrated successful GPR87 staining of a specific subset of isolated phBECs, compared to control.

The family of G protein-coupled receptors has been described as one of the largest groups of related receptors and represents a major target for the development of novel drugs [80, 177]. These features in combination with the highly cell-specific expression of GPR87 emphasize its possible role as target for novel IPF therapies. The restriction of GPR87 expression to bronchial epithelial cells might help to limit the impact of possible drugs on other cell types and to reduce the occurrence of side effects. Basal cells serve as progenitor cells for the repair and regeneration of the bronchial epithelium and express the basal cell markers KRT5, KRT14, and TP63 [168]. Taking into account the preliminary FACS results of GPR87 expression in a specific subset of bronchial

epithelial cells and the significant correlation of GPR87 expression with the basal cell markers in IPF, we hypothesized that basal cells could be the major source of increased GPR87 levels in IPF. We were able to confirm this hypothesis in our subsequent study by Heinzelmann *et al.* [178], which demonstrated GPR87 expression in KRT5 positive basal cells of IPF lung tissue and human cell culture experiments. Furthermore, we verified the regulatory mechanism of TGF-β-induced upregulation of GPR87 expression in bronchial epithelial cells, which has already been described in the present thesis [178]. This is an intriguing finding because TGF-β represents one of the major pro-fibrotic mediators in IPF. Besides TGF-β, the GPR87 expression has also been reported to be regulated by P53, which contributed to cell survival [90]. Although its precise role in IPF remains subject to further research, Heinzelmann *et al.* [178] proposed an association of GPR87 with the processes of airway remodeling and bronchiolization. The highlighted role of GPR87 within the bronchial epithelium of the IPF lung corresponds with its specific expression in SCC lung tissue, compared to AC samples (Figure 4.15).

Taken together, we propose that GPR87 is upregulated by TGF-β in basal cells of the IPF lung and might contribute to IPF progression, bronchiolization, squamous metaplasia, and distorted basal cells, which subsequently supports the initiation of lung cancer development in IPF. Therefore, GPR87 could represent a specific target for novel therapeutic options, and its contribution requires further research.

5.7 Conclusion and outlook

IPF and lung cancer are devastating disorders of the human lung, which impose a poor prognosis on affected patients because of the lack of curative treatment options. Hence, there exists an urgent need for the development of novel therapies. This may require the implementation of unusual strategies, which we have demonstrated in the present thesis. We were able to confirm the existence of oncogenic, cancer-related signaling in IPF, which may contribute to cellular reprogramming, support the IPF progression, and promote alterations of the epithelium that facilitate the development of lung cancer.

In summary, this thesis established a relationship between the pathogenesis of IPF and NSCLC by conducting a top-down systems biology approach and demonstrating a common pattern of differentially expressed genes in both diseases. The subset of common, upregulated DEGs exhibited an IPF-specific signature, was associated with fibrotic lung epithelial cells, and showed a significant enrichment of shared interactions and annotations. These results verified the pathomechanistic relationship by confirming the existence of a common subset of functionally related genes in IPF and NSCLC, which included a shared biological and mechanistic background. Further analysis of these common features may improve our understanding of the IPF pathogenesis and lead to novel therapeutic targets. We were also able to prove the expression of selected candidate genes in our tissue samples of IPF, NSCLC, and bleomycin-treated mice. Finally, we provided evidence that GPR87 and PSAT1 represent promising targets for the development of IPF treatments. Overall, the present thesis implemented an uncommon, novel strategy to broaden our knowledge about IPF and proved the practical relevance

of an exploratory systems biology approach for the identification of specific targets and subsequent mechanistic studies.

In conclusion, we propose that common gene expression alterations in the lung epithelium of IPF patients lead to morphological and cellular changes, which support IPF progression and represent a basis for malignant transformations and the development of lung cancer. Therefore, the oncogenic signaling in IPF provides an exciting opportunity to target disease progression and cancer formation.

The present thesis provides various ideas for subsequent studies and future research. In order to further elucidate the connection of IPF and NSCLC, future studies should address the downregulated DEGs as well as those genes with an opposed misregulation in both diseases. Furthermore, single-cell analysis of IPF lung tissue could help to determine and verify the cell-specific enrichment of the *Overlap gene set* in IPF. Besides this, the contribution of our results should also be evaluated in terms of cancer research because we mainly focused our analyses on the context of IPF. Eventually, GPR87 represents a promising target for therapeutic strategies, and its contribution requires further investigation of *in vivo* models, such as GPR87 knockout mice.

References

1. Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, H. T., Abebe, M., Abebe, Z., Abejie, A. N., Abera, S. F., Abil, O. Z., Abraha, H. N., Abrham, A. R., Abu-Raddad, L. J., Accrombessi, M. M. K., Acharya, D., Adamu, A. A., Adebayo, O. M., Adedoyin, R. A., Adekanmbi, V., Adetokunboh, O. O., Adhena, B. M., Adib, M. G., Admasie, A., Afshin, A., Agarwal, G., Agesa, K. M., Agrawal, A., Agrawal, S., Ahmadi, A., Ahmadi, M., Ahmed, M. B., Ahmed, S., Aichour, A. N., Aichour, I., Aichour, M. T. E., Akbari, M. E., Akinyemi, R. O., Akseer, N., Al-Aly, Z., Al-Eyadhy, A., Al-Raddadi, R. M., Alahdab, F., Alam, K., Alam, T., Alebel, A., Alene, K. A., Alijanzadeh, M., Alizadeh-Navaei, R., Aljunid, S. M., Alkerwi, A., Alla, F., Allebeck, P., Alonso, J., Altirkawi, K., Alvis-Guzman, N., Amare, A. T., Aminde, L. N., Amini, E., Ammar, W., Amoako, Y. A., Anber, N. H., Andrei, C. L., Androudi, S., Animut, M. D., Anjomshoa, M., Ansari, H., Ansha, M. G., Antonio, C. A. T., Anwari, P., Aremu, O., Ärnlöv, J., Arora, A., Arora, M., Artaman, A., Aryal, K. K., Asayesh, H., Asfaw, E. T., Ataro, Z., Atique, S., Atre, S. R., Ausloos, M., Avokpaho, E. F. G. A., Awasthi, A., Quintanilla, B. P. A., Ayele, Y., Ayer, R., Azzopardi, P. S., Babazadeh, A., Bacha, U., Badali, H., Badawi, A., Bali, A. G., Ballesteros, K. E., Banach, M., Banerjee, K., Bannick, M. S., Banoub, J. A. M., Barboza, M. A., Barker-Collo, S. L., Bärnighausen, T. W., Barquera, S., Barrero, L. H., Bassat, Q., Basu, S., Baune, B. T., Baynes, H. W., Bazargan-Hejazi, S., Bedi, N., Beghi, E., Behzadifar, M., Behzadifar, M., Béjot, Y., Bekele, B. B., Belachew, A. B., Belay, E., Belay, Y. A., Bell, M. L., Bello, A. K., Bennett, D. A., Bensenor, I. M., Berman, A. E., Bernabe, E., Bernstein, R. S., Bertolacci, G. J., Beuran, M., Beyranvand, T., Bhalla, A., Bhattarai, S., Bhaumik, S., Bhutta, Z. A., Biadgo, B., Biehl, M. H., Bijani, A., Bikbov, B., Bilano, V., Bililign, N., Bin Sayeed, M. S., Bisanzio, D., Biswas, T., Blacker, B. F., Basara, B. B., Borschmann, R., *et al.* Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392,** 1736–1788 (2018).

2. Dieleman, J. L., Baral, R., Birger, M., Bui, A. L., Bulchis, A., Chapin, A., Hamavid, H., Horst, C., Johnson, E. K., Joseph, J., Lavado, R., Lomsadze, L., Reynolds, A., Squires, E., Campbell, M., DeCenso, B., Dicker, D., Flaxman, A. D., Gabert, R., Highfill, T., Naghavi, M., Nightingale, N., Templin, T., Tobias, M. I., Vos, T. & Murray, C. J. L. US Spending on Personal Health Care and Public Health, 1996-2013. *JAMA* **316,** 2627–2646 (2016).

3. Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Shirude, S., Naghavi, M., Mokdad, A. H. & Murray, C. J. L. Trends and Patterns of Differences in Chronic Respiratory Disease Mortality Among US Counties, 1980-2014. *JAMA* **318,** 1136–1149 (2017).

4. Eickelberg, O. & Selman, M. Update in Diffuse Parenchymal Lung Disease 2009. *American Journal of Respiratory and Critical Care Medicine* **181,** 883–888 (2010).

5. Demedts, M., Wells, A., Antó, J., Costabel, U., Hubbard, R., Cullinan, P., Slabbynck, H., Rizzato, G., Poletti, V., Verbeken, E., Thomeer, M., Kokkarinen, J., Dalphin, J. & Newman Taylor, A. Interstitial lung diseases: an epidemiological overview. *European Respiratory Journal* **18,** 2s–16s (2001).

6. Ryerson, C. J., Arean, P. A., Berkeley, J., Carrieri-Kohlman, V. L., Pantilat, S. Z., Landefeld, C. S. & Collard, H. R. Depression is a common and chronic comorbidity in patients with interstitial lung disease. *Respirology* **17,** 525–532 (2012).

7. Kreuter, M., Swigris, J., Pittrow, D., Geier, S., Klotsche, J., Prasse, A., Wirtz, H., Koschel, D., Andreas, S., Claussen, M., Grohé, C., Wilkens, H., Hagmeyer, L., Skowasch, D., Meyer, J. F., Kirschner, J., Gläser, S., Kahn, N., Welte, T., Neurohr, C., Schwaiblmair, M., Held, M., Bahmer, T., Oqueka, T., Frankenberger, M. & Behr, J. The clinical course of idiopathic pulmonary fibrosis and its association to quality of life over time: longitudinal data from the INSIGHTS-IPF registry. *Respiratory Research* **20,** 59 (2019).

8. Travis, W. D., Costabel, U., Hansell, D. M., King, T. E., Lynch, D. A., Nicholson, A. G., Ryerson, C. J., Ryu, J. H., Selman, M., Wells, A. U., Behr, J., Bouros, D., Brown, K. K., Colby, T. V., Collard, H. R., Cordeiro, C. R., Cottin, V., Crestani, B., Drent, M., Dudden, R. F., Egan, J., Flaherty, K., Hogaboam, C., Inoue, Y., Johkoh, T., Kim, D. S., Kitaichi, M., Loyd, J., Martinez, F. J., Myers, J., Protzko, S., Raghu, G., Richeldi, L., Sverzellati, N., Swigris, J. & Valeyre, D. An Official American Thoracic Society/European Respiratory Society Statement: Update of the International Multidisciplinary Classification of the Idiopathic Interstitial Pneumonias. *American Journal of Respiratory and Critical Care Medicine* **188,** 733–748 (2013).

9. Lederer, D. J. & Martinez, F. J. Idiopathic Pulmonary Fibrosis. *New England Journal of Medicine* **378,** 1811–1823 (2018).

10. Hutchinson, J., Fogarty, A., Hubbard, R. & McKeever, T. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *European Respiratory Journal* **46,** 795–806 (2015).

11. Martinez, F. J., Collard, H. R., Pardo, A., Raghu, G., Richeldi, L., Selman, M., Swigris, J. J., Taniguchi, H. & Wells, A. U. Idiopathic pulmonary fibrosis. *Nature Reviews Disease Primers* **3,** 17074 (2017).

12. Raghu, G., Remy-Jardin, M., Myers, J. L., Richeldi, L., Ryerson, C. J., Lederer, D. J., Behr, J., Cottin, V., Danoff, S. K., Morell, F., Flaherty, K. R., Wells, A., Martinez, F. J., Azuma, A., Bice, T. J., Bouros, D., Brown, K. K., Collard, H. R., Duggal, A., Galvin, L., Inoue, Y., Jenkins, R. G., Johkoh, T., Kazerooni, E. A., Kitaichi, M., Knight, S. L., Mansour, G., Nicholson, A. G., Pipavath, S. N. J., Buendía-Roldán, I., Selman, M., Travis, W. D., Walsh, S. L. F. & Wilson, K. C. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *American Journal of Respiratory and Critical Care Medicine* **198,** e44–e68 (2018).

13. Raghu, G., Chen, S.-Y., Yeh, W.-S., Maroni, B., Li, Q., Lee, Y.-C. & Collard, H. R. Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001-11. *The Lancet Respiratory Medicine* **2,** 566–572 (2014).

14. Harari, S., Davì, M., Biffi, A., Caminati, A., Ghirardini, A., Lovato, V., Cricelli, C. & Lapi, F. Epidemiology of idiopathic pulmonary fibrosis: a population-based study in primary care. *Internal and Emergency Medicine.* doi:10.1007/s11739-019-02195-0 (2019).

15. Baumgartner, K. B., Samet, J. M., Stidley, C. A., Colby, T. V. & Waldron, J. A. Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine* **155,** 242–248 (1997).

16. Baumgartner, K. B., Samet, J. M., Coultas, D. B., Stidley, C. A., Hunt, W. C., Colby, T. V. & Waldron, J. A. Occupational and Environmental Risk Factors for Idiopathic Pulmonary Fibrosis: A Multicenter Case-Control Study. *American Journal of Epidemiology* **152,** 307–315 (2000).

17. Allen, R. J., Porte, J., Braybrooke, R., Flores, C., Fingerlin, T. E., Oldham, J. M., Guillen-Guio, B., Ma, S.-F., Okamoto, T., John, A. E., Obeidat, M., Yang, I. V., Henry, A., Hubbard, R. B., Navaratnam, V., Saini, G., Thompson, N., Booth, H. L., Hart, S. P., Hill, M. R., Hirani, N., Maher, T. M., McAnulty, R. J., Millar, A. B., Molyneaux, P. L., Parfrey, H., Rassl, D. M., Whyte, M. K. B., Fahy, W. A., Marshall, R. P., Oballa, E., Bossé, Y., Nickle, D. C., Sin, D. D., Timens, W., Shrine, N., Sayers, I., Hall, I. P., Noth, I., Schwartz, D. A., Tobin, M. D., Wain, L. V. & Jenkins, R. G. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *The Lancet Respiratory Medicine* **5,** 869–880 (2017).

18. Kaur, A., Mathai, S. K. & Schwartz, D. A. Genetics in Idiopathic Pulmonary Fibrosis Pathogenesis, Prognosis, and Treatment. *Frontiers in Medicine* **4,** 154 (2017).

19. King Jr, T. E., Pardo, A. & Selman, M. Idiopathic pulmonary fibrosis. *The Lancet* **378,** 1949–1961 (2011).

20. Behr, J., Günther, A., Bonella, F., Dinkel, J., Fink, L., Geiser, T., Geißler, K., Gläser, S., Handzhhiev, S., Jonigk, D., Koschel, D., Kreuter, M., Leuschner, G., Markart, P., Prasse, A., Schönfeld, N., Schupp, J. C., Sitter, H., Müller-Quernheim, J. & Costabel, U. S2K-Leitlinie zur Diagnostik der idiopathischen Lungenfibrose. *Pneumologie.* doi:10.1055/a-1120-3531 (2020).

21. Hamer, O. W., Rehbock, B. & Schaefer-Prokop, C. Idiopathische pulmonale Fibrose. *Radiologe.* doi:10.1007/s00117-020-00675-5 (2020).

22. Martino, E. D., Provenzani, A., Vitulo, P. & Polidori, P. Systematic Review and Meta-analysis of Pirfenidone, Nintedanib, and Pamrevlumab for the Treatment of Idiopathic Pulmonary Fibrosis. *Annals of Pharmacotherapy.* doi:10.1177/1060028-020964451 (2020).

23. Noble, P. W., Albera, C., Bradford, W. Z., Costabel, U., Glassberg, M. K., Kardatzke, D., King Talmadge E, J., Lancaster, L., Sahn, S. A., Szwarcberg, J., Valeyre, D. & du Bois, R. M. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *The Lancet* **377,** 1760–1769 (2011).

24. King, T. E., Bradford, W. Z., Castro-Bernardini, S., Fagan, E. A., Glaspole, I., Glassberg, M. K., Gorina, E., Hopkins, P. M., Kardatzke, D., Lancaster, L., Lederer, D. J., Nathan, S. D., Pereira, C. A., Sahn, S. A., Sussman, R., Swigris, J. J. & Noble, P. W. A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis. *The New England Journal of Medicine* **370,** 2083–2092 (2014).

25. Nathan, S. D., Albera, C., Bradford, W. Z., Costabel, U., Glaspole, I., Glassberg, M. K., Kardatzke, D. R., Daigl, M., Kirchgaessler, K.-U., Lancaster, L. H., Lederer, D. J., Pereira, C. A., Swigris, J. J., Valeyre, D. & Noble, P. W. Effect of pirfenidone on mortality: pooled analyses and meta-analyses of clinical trials in idiopathic pulmonary fibrosis. *The Lancet Respiratory Medicine* **5,** 33–41 (2017).

26. Oku, H., Shimizu, T., Kawabata, T., Nagira, M., Hikita, I., Ueyama, A., Matsushima, S., Torii, M. & Arimura, A. Antifibrotic action of pirfenidone and prednisolone: Different effects on pulmonary cytokines and growth factors in bleomycin-induced murine pulmonary fibrosis. *European Journal of Pharmacology* **590,** 400–408 (2008).

27. Wollin, L., Wex, E., Pautsch, A., Schnapp, G., Hostettler, K. E., Stowasser, S. & Kolb, M. Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. *European Respiratory Journal* **45,** 1434–1445 (2015).

28. Richeldi, L., du Bois, R. M., Raghu, G., Azuma, A., Brown, K. K., Costabel, U., Cottin, V., Flaherty, K. R., Hansell, D. M., Inoue, Y., Kim, D. S., Kolb, M., Nicholson, A. G., Noble, P. W., Selman, M., Taniguchi, H., Brun, M., Le Maulf, F., Girard, M., Stowasser, S., Schlenker-Herceg, R., Disse, B. & Collard, H. R. Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis. *New England Journal of Medicine* **370,** 2071–2082 (2014).

29. Raghu, G., Rochwerg, B., Zhang, Y., Garcia, C. A. C., Azuma, A., Behr, J., Brozek, J. L., Collard, H. R., Cunningham, W., Homma, S., Johkoh, T., Martinez, F. J., Myers, J., Protzko, S. L., Richeldi, L., Rind, D., Selman, M., Theodore, A., Wells, A. U., Hoogsteden, H. & Schünemann, H. J. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *American Journal of Respiratory and Critical Care Medicine* **192,** e3–e19 (2015).

30. Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Colby, T. V., Cordier, J.-F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., Hansell, D. M., Johkoh, T., Kim, D. S., King, T. E., Kondoh, Y., Myers, J., Müller, N. L., Nicholson, A. G., Richeldi, L., Selman, M., Dudden, R. F., Griss, B. S., Protzko, S. L. & Schünemann, H. J. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *American Journal of Respiratory and Critical Care Medicine* **183,** 788–824 (2011).

31. Selman, M. & Pardo, A. The leading role of epithelial cells in the pathogenesis of idiopathic pulmonary fibrosis. *Cellular Signalling* **66,** 109482 (2020).

32. Lehmann, M., Korfei, M., Mutze, K., Klee, S., Skronska-Wasek, W., Alsafadi, H. N., Ota, C., Costa, R., Schiller, H. B., Lindner, M., Wagner, D. E., Günther, A. & Königshoff, M. Senolytic drugs target alveolar epithelial cell function and attenuate experimental lung fibrosis ex vivo. *European Respiratory Journal* **50,** 1602367 (2017).

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

34. Snetselaar, R., van Batenburg, A. A., van Oosterhout, M. F. M., Kazemier, K. M., Roothaan, S. M., Peeters, T., van der Vis, J. J., Goldschmeding, R., Grutters, J. C. & van Moorsel, C. H. M. Short telomere length in IPF lung associates with fibrotic lesions and predicts survival. *PLOS ONE* **12,** e0189467 (2017).

35. Moore, C., Blumhagen, R. Z., Yang, I. V., Walts, A., Powers, J., Walker, T., Bishop, M., Russell, P., Vestal, B., Cardwell, J., Markin, C. R., Mathai, S. K., Schwarz, M. I., Steele, M. P., Lee, J., Brown, K. K., Loyd, J. E., Crapo, J. D., Silverman, E. K., Cho, M. H., James, J. A., Guthridge, J. M., Cogan, J. D., Kropski, J. A., Swigris, J. J., Bair, C., Kim, D. S., Ji, W., Kim, H., Song, J. W., Maier, L. A., Pacheco, K. A., Hirani, N., Poon, A. S., Li, F., Jenkins, R. G., Braybrooke, R., Saini, G., Maher, T. M., Molyneaux, P. L., Saunders, P., Zhang, Y., Gibson, K. F., Kass, D. J., Rojas, M., Sembrat, J., Wolters, P. J., Collard, H. R., Sundy, J. S., O'Riordan, T., Strek, M. E., Noth, I., Ma, S.-F., Porteous, M. K., Kreider, M. E., Patel, N. B., Inoue, Y., Hirose, M., Arai, T., Akagawa, S.,

Eickelberg, O., Fernandez, I. E., Behr, J., Mogulkoc, N., Corte, T. J., Glaspole, I., Tomassetti, S., Ravaglia, C., Poletti, V., Crestani, B., Borie, R., Kannengiesser, C., Parfrey, H., Fiddler, C., Rassl, D., Molina-Molina, M., Machahua, C., Worboys, A. M., Gudmundsson, G., Isaksson, H. J., Lederer, D. J., Podolanczuk, A. J., Montesi, S. B., Bendstrup, E., Danchel, V., Selman, M., Pardo, A., Henry, M. T., Keane, M. P., Doran, P., Vašáková, M., Sterclova, M., Ryerson, C. J., Wilcox, P. G., Okamoto, T., Furusawa, H., Miyazaki, Y., Laurent, G., Baltic, S., Prele, C., Moodley, Y., Shea, B. S., Ohta, K., Suzukawa, M., Narumoto, O., Nathan, S. D., Venuto, D. C., Woldehanna, M. L., Kokturk, N., de Andrade, J. A., Luckhardt, T., Kulkarni, T., Bonella, F., Donnelly, S. C., McElroy, A., Armstong, M. E., Aranda, A., Carbone, R. G., Puppo, F., Beckman, K. B., Nickerson, D. A., Fingerlin, T. E. & Schwartz, D. A. Resequencing Study Confirms That Host Defense and Cell Senescence Gene Variants Contribute to the Risk of Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine* **200,** 199–208 (2019).

36. Alsdorf, W., Seidel, C., Bokemeyer, C. & Oing, C. Current pharmacotherapy for testicular germ cell cancer. *Expert Opinion on Pharmacotherapy* **20,** 837–850 (2019).

37. Shanbhag, S. & Ambinder, R. F. Hodgkin lymphoma: A review and update on recent progress. *CA: A Cancer Journal for Clinicians* **68,** 116–132 (2018).

38. Sleijfer, S. Bleomycin-Induced Pneumonitis. *CHEST* **120,** 617–624 (2001).

39. B. Moore, B., Lawson, W. E., Oury, T. D., Sisson, T. H., Raghavendran, K. & Hogaboam, C. M. Animal Models of Fibrotic Lung Disease. *American Journal of Respiratory Cell and Molecular Biology* **49,** 167–179 (2013).

40. Liu, T., De Los Santos, F. G. & Phan, S. H. in *Fibrosis: Methods and Protocols* (ed Rittié, L.) 27–42 (Springer New York, 2017). doi:10.1007/978-1-4939-7113-8_2.

41. Carrington, R., Jordan, S., Pitchford, S. C. & Page, C. P. Use of animal models in IPF research. *Pulmonary Pharmacology & Therapeutics* **51,** 73–78 (2018).

42. Izbicki, G., Segel, M. J., Christensen, T. G., Conner, M. W. & Breuer, R. Time course of bleomycin-induced lung fibrosis. *International Journal of Experimental Pathology* **83,** 111–119 (2002).

43. Williamson, J. D., Sadofsky, L. R. & Hart, S. P. The pathogenesis of bleomycin-induced lung injury in animals and its applicability to human idiopathic pulmonary fibrosis. *Experimental Lung Research* **41,** 57–73 (2015).

44. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A. & Bray, F. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer* **144,** 1941–1953 (2019).

45. Global Burden of Disease Cancer Collaboration. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncology* **5,** 1749–1769 (2019).

46. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. & Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68,** 394–424 (2018).

47. Luengo-Fernandez, R., Leal, J., Gray, A. & Sullivan, R. Economic burden of cancer across the European Union: a population-based cost analysis. *The Lancet Oncology* **14,** 1165–1174 (2013).

48. Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C. & Boffetta, P. Risk factors for lung cancer worldwide. *European Respiratory Journal* **48,** 889–902 (2016).

49. Xing, P.-Y., Zhu, Y.-X., Wang, L., Hui, Z.-G., Liu, S.-M., Ren, J.-S., Zhang, Y., Song, Y., Liu, C.-C., Huang, Y.-C., Liao, X.-Z., Xing, X.-J., Wang, D.-B., Yang, L., Du, L.-B., Liu, Y.-Q., Zhang, Y.-Z., Liu, Y.-Y., Wei, D.-H., Zhang, K., Shi, J.-F., Qiao, Y.-L., Chen, W.-Q., Li, J.-L., Dai, M. & the LuCCRES Group. What are the clinical symptoms and physical signs for non-small cell lung cancer before diagnosis is made? A nation-wide multicenter 10-year retrospective study in China. *Cancer Medicine* **8,** 4055–4069 (2019).

50. Postmus, P. E., Kerr, K. M., Oudkerk, M., Senan, S., Waller, D. A., Vansteenkiste, J., Escriu, C. & Peters, S. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* **28,** iv1–iv21 (2017).

51. Inamura, K. Lung Cancer: Understanding Its Molecular Pathology and the 2015 WHO Classification. *Frontiers in Oncology* **7.** doi:10.3389/fonc.2017.00193 (2017).

52. Duma, N., Santana-Davila, R. & Molina, J. R. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clinic Proceedings* **94,** 1623–1640 (2019).

53. Wang, B.-Y., Huang, J.-Y., Chen, H.-C., Lin, C.-H., Lin, S.-H., Hung, W.-H. & Cheng, Y.-F. The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients. *Journal of Cancer Research and Clinical Oncology* **146,** 43–52 (2020).

54. Brown, S.-A. W., Dobelle, M., Padilla, M., Agovino, M., Wisnivesky, J. P., Hashim, D. & Boffetta, P. Idiopathic Pulmonary Fibrosis and Lung Cancer. A Systematic Review and Meta-analysis. *Annals of the American Thoracic Society* **16,** 1041–1051 (2019).

55. JafariNezhad, A. & YektaKooshali, M. H. Lung cancer in idiopathic pulmonary fibrosis: A systematic review and meta-analysis. *PLOS ONE* **13,** e0202360 (2018).

56. Ozawa, Y., Suda, T., Naito, T., Enomoto, N., Hashimoto, D., Fujisawa, T., Nakamura, Y., Inui, N., Nakamura, H. & Chida, K. Cumulative incidence of and predictive factors for lung cancer in IPF. *Respirology* **14,** 723–728 (2009).

57. Tomassetti, S., Gurioli, C., Ryu, J. H., Decker, P. A., Ravaglia, C., Tantalocco, P., Buccioli, M., Piciucchi, S., Sverzellati, N., Dubini, A., Gavelli, G., Chilosi, M. & Poletti, V. The Impact of Lung Cancer on Survival of Idiopathic Pulmonary Fibrosis. *CHEST* **147,** 157–164 (2015).

58. Kato, E., Takayanagi, N., Takaku, Y., Kagiyama, N., Kanauchi, T., Ishiguro, T. & Sugita, Y. Incidence and predictive factors of lung cancer in patients with idiopathic pulmonary fibrosis. *ERJ Open Research* **4,** 00111–2016 (2018).

59. Yoo, H., Jeong, B.-H., Chung, M. J., Lee, K. S., Kwon, O. J. & Chung, M. P. Risk factors and clinical characteristics of lung cancer in idiopathic pulmonary fibrosis: a retrospective cohort study. *BMC Pulmonary Medicine* **19,** 149 (2019).

60. Song, M. J., Kim, S. Y., Park, M. S., Kang, M. J., Lee, S. H. & Park, S. C. A nationwide population-based study of incidence and mortality of lung cancer in idiopathic pulmonary fibrosis. *Scientific Reports* **11,** 2596 (2021).

61. Ichihara, E., Miyahara, N., Maeda, Y. & Kiura, K. Managing Lung Cancer with Comorbid Interstitial Pneumonia. *Internal Medicine* **59,** 163–167 (2020).

62. The Cancer Genome Atlas Research Network, Genome sequencing centres: Broad Institute. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).

63. The Cancer Genome Atlas Research Network, Disease analysis working group. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).

64. Demopoulos, K., Arvanitis, D. A., Vassilakis, D. A., Siafakas, N. M. & Spandidos, D. A. MYCL1, FHIT, SPARC, p16INK4 and TP53 genes associated to lung cancer in idiopathic pulmonary fibrosis. *Journal of Cellular and Molecular Medicine* **6,** 215–222 (2002).

65. Rabinovich, E. I., Kapetanaki, M. G., Steinfeld, I., Gibson, K. F., Pandit, K. V., Yu, G., Yakhini, Z. & Kaminski, N. Global Methylation Patterns in Idiopathic Pulmonary Fibrosis. *PLOS ONE* **7,** e33770 (2012).

66. Stewart, D. J. Wnt Signaling Pathway in Non-Small Cell Lung Cancer. *JNCI: Journal of the National Cancer Institute* **106,** djt356 (2013).

67. Königshoff, M., Balsara, N., Pfaff, E.-M., Kramer, M., Chrobak, I., Seeger, W. & Eickelberg, O. Functional Wnt Signaling Is Increased in Idiopathic Pulmonary Fibrosis. *PLOS ONE* **3,** e2142 (2008).

68. Yang, M. & Vousden, K. H. Serine and one-carbon metabolism in cancer. *Nature Reviews Cancer* **16,** 650–662 (2016).

69. DeNicola, G. M., Chen, P.-H., Mullarky, E., Sudderth, J. A., Hu, Z., Wu, D., Tang, H., Xie, Y., Asara, J. M., Huffman, K. E., Wistuba, I. I., Minna, J. D., DeBerardinis, R. J. & Cantley, L. C. NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nature Genetics* **47,** 1475–1481 (2015).

70. Gao, S., Ge, A., Xu, S., You, Z., Ning, S., Zhao, Y. & Pang, D. PSAT1 is regulated by ATF4 and enhances cell proliferation via the GSK3β/β-catenin/cyclin D1 signaling pathway in ER-negative breast cancer. *Journal of Experimental & Clinical Cancer Research* **36,** 179 (2017).

71. Guo, K., Qi, D. & Huang, B. LncRNA MEG8 promotes NSCLC progression by modulating the miR-15a-5p-miR-15b-5p/PSAT1 axis. *Cancer Cell International* **21,** 84 (2021).

72. Chan, Y.-C., Chang, Y.-C., Chuang, H.-H., Yang, Y.-C., Lin, Y.-F., Huang, M.-S., Hsiao, M., Yang, C.-J. & Hua, K.-T. Overexpression of PSAT1 promotes metastasis of lung adenocarcinoma by suppressing the IRF1-IFNγ axis. *Oncogene* **39,** 2509–2522 (2020).

73. Stewart, P. A., Welsh, E. A., Slebos, R. J. C., Fang, B., Izumi, V., Chambers, M., Zhang, G., Cen, L., Pettersson, F., Zhang, Y., Chen, Z., Cheng, C.-H., Thapa, R., Thompson, Z., Fellows, K. M., Francis, J. M., Saller, J. J., Mesa, T., Zhang, C., Yoder, S., DeNicola, G. M., Beg, A. A., Boyle, T. A., Teer, J. K., Ann Chen, Y., Koomen, J. M., Eschrich, S. A. & Haura, E. B. Proteogenomic landscape of squamous cell lung cancer. *Nature Communications* **10,** 3578 (2019).

74. Yang, Y., Wu, J., Cai, J., He, Z., Yuan, J., Zhu, X., Li, Y., Li, M. & Guan, H. PSAT1 regulates cyclin D1 degradation and sustains proliferation of non-small cell lung cancer cells. *International Journal of Cancer* **136,** E39–E50 (2015).

75. Liu, B., Jia, Y., Cao, Y., Wu, S., Jiang, H., Sun, X., Ma, J., Yin, X., Mao, A. & Shang, M. Overexpression of Phosphoserine Aminotransferase 1 (PSAT1) Predicts Poor Prognosis and Associates with Tumor Progression in Human Esophageal Squamous Cell Carcinoma. *Cellular Physiology and Biochemistry* **39,** 395–406 (2016).

76. Metcalf, S., Dougherty, S., Kruer, T., Hasan, N., Biyik-Sit, R., Reynolds, L. & Clem, B. F. Selective loss of phosphoserine aminotransferase 1 (PSAT1) suppresses migration, invasion, and experimental metastasis in triple negative breast cancer. *Clinical & Experimental Metastasis* **37,** 187–197 (2020).

77. Zheng, M.-J., Li, X., Hu, Y.-X., Dong, H., Gou, R., Nie, X., Liu, Q., Ying-Ying, H., Liu, J.-J. & Lin, B. Identification of molecular marker associated with ovarian cancer prognosis using bioinformatics analysis and experiments. *Journal of Cellular Physiology* **234,** 11023–11036 (2019).

78. Nigdelioglu, R., Hamanaka, R. B., Meliton, A. Y., O'Leary, E., Witt, L. J., Cho, T., Sun, K., Bonham, C., Wu, D., Woods, P. S., Husain, A. N., Wolfgeher, D., Dulin, N. O., Chandel, N. S. & Mutlu, G. M. Transforming Growth Factor (TGF)-β Promotes de Novo Serine Synthesis for Collagen Production. *Journal of Biological Chemistry* **291,** 27239–27251 (2016).

79. Hamanaka, R. B., O'Leary, E. M., Witt, L. J., Tian, Y., Gökalp, G. A., Meliton, A. Y., Dulin, N. O. & Mutlu, G. M. Glutamine Metabolism Is Required for Collagen Protein Synthesis in Lung Fibroblasts. *American Journal of Respiratory Cell and Molecular Biology* **61,** 597–606 (2019).

80. Wittenberger, T., Schaller, H. C. & Hellebrand, S. An expressed sequence tag (EST) data mining strategy succeeding in the discovery of new G-protein coupled receptors. *Journal of Molecular Biology* **307,** 799–813 (2001).

81. Tabata, K.-i., Baba, K., Shiraishi, A., Ito, M. & Fujita, N. The orphan GPCR GPR87 was deorphanized and shown to be a lysophosphatidic acid receptor. *Biochemical and Biophysical Research Communications* **363,** 861–866 (2007).

82. Ochiai, S., Furuta, D., Sugita, K., Taniura, H. & Fujita, N. GPR87 mediates lysophosphatidic acid-induced colony dispersal in A431 cells. *European Journal of Pharmacology* **715,** 15–20 (2013).

83. Niss Arfelt, K., Fares, S., Sparre-Ulrich, A. H., Hjortø, G. M., Gasbjerg, L. S., Mølleskov-Jensen, A.-S., Benned-Jensen, T. & Rosenkilde, M. M. Signaling via G proteins mediates tumorigenic effects of GPR87. *Cellular Signalling* **30,** 9–18 (2017).

84. Glatt, S., Halbauer, D., Heindl, S., Wernitznig, A., Kozina, D., Su, K.-C., Puri, C., Garin-Chesa, P. & Sommergruber, W. hGPR87 contributes to viability of human tumor cells. *International Journal of Cancer* **122,** 2008–2016 (2008).

85. Gugger, M., White, R., Song, S., Waser, B., Cescato, R., Rivière, P. & Reubi, J. C. GPR87 Is an Overexpressed G-Protein Coupled Receptor in Squamous Cell Carcinoma of the Lung. *Disease Markers* **24,** 857474 (2008).

86. Nii, K., Tokunaga, Y., Liu, D., Zhang, X., Nakano, J., Ishikawa, S., Kakehi, Y., Haba, R. & Yokomise, H. Overexpression of G protein-coupled receptor 87 correlates with poorer tumor differentiation and higher tumor proliferation in non-small-cell lung cancer. *Molecular and Clinical Oncology* **2,** 539–544 (2014).

87. Yan, M., Li, H., Zhu, M., Zhao, F., Zhang, L., Chen, T., Jiang, G., Xie, H., Cui, Y., Yao, M. & Li, J. G Protein-Coupled Receptor 87 (GPR87) Promotes the Growth and Metastasis of CD133+ Cancer Stem-Like Cells in Hepatocellular Carcinoma. *PLOS ONE* **8,** e61056 (2013).

88. Okazoe, H., Zhang, X., Liu, D., Shibuya, S., Ueda, N., Sugimoto, M. & Kakehi, Y. Expression and Role of GPR87 in Urothelial Carcinoma of the Bladder. *International Journal of Molecular Sciences* **14,** 12367–12379 (2013).

89. Wang, L., Zhou, W., Zhong, Y., Huo, Y., Fan, P., Zhan, S., Xiao, J., Jin, X., Gou, S., Yin, T., Wu, H. & Liu, T. Overexpression of G protein-coupled receptor GPR87 promotes pancreatic cancer aggressiveness and activates NF-κB signaling pathway. *Molecular Cancer* **16,** 61 (2017).

90. Zhang, Y., Qian, Y., Lu, W. & Chen, X. The G Protein-Coupled Receptor 87 Is Necessary for p53-Dependent Cell Survival in Response to Genotoxic Stress. *Cancer Research* **69,** 6049–6056 (2009).

91. Zhang, X., Liu, D., Hayashida, Y., Okazoe, H., Hashimoto, T., Ueda, N., Sugimoto, M. & Kakehi, Y. G Protein-Coupled Receptor 87 (GPR87) Promotes Cell Proliferation in Human Bladder Cancer Cells. *International Journal of Molecular Sciences* **16,** 24319–24331 (2015).

92. Kita, Y., Go, T., Nakashima, N., Liu, D., Tokunaga, Y., Zhang, X., Nakano, T., Nii, K., Chang, S. S. & Yokomise, H. Inhibition of Cell-surface Molecular GPR87 With GPR87-suppressing Adenoviral Vector Disturb Tumor Proliferation in Lung Cancer Cells. *Anticancer Research* **40,** 733–741 (2020).

93. Park, S.-M., Choi, E.-Y., Bae, M., Kim, S., Park, J. B., Yoo, H., Choi, J. K., Kim, Y.-J., Lee, S.-H. & Kim, I.-H. Histone variant H3F3A promotes lung cancer cell migration through intronic regulation. *Nature Communications* **7,** 12914 (2016).

94. Jiang, J., Yu, C., Guo, X., Zhang, H., Tian, S., Cai, K., He, Z. & Sun, C. G Protein-Coupled Receptor GPR87 Promotes the Expansion of PDA Stem Cells through Activating JAK2/STAT3. *Molecular Therapy - Oncolytics* **17,** 384–393 (2020).

95. Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., Pagès, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L. & Morgan, M. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* **12,** 115–121 (2015).

96. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. H. & Zhang, J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5,** R80 (2004).

97. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13,** 134 (2012).

98. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30,** 207–210 (2002).

99. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S. & Soboleva, A. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* **41,** D991–D995 (2013).

100. Harrison, P. W., Alako, B., Amid, C., Cerdeño-Tárraga, A., Cleland, I., Holt, S., Hussein, A., Jayathilaka, S., Kay, S., Keane, T., Leinonen, R., Liu, X., Martínez-Villacorta, J., Milano, A., Pakseresht, N., Rajan, J., Reddy, K., Richards, E., Rosello, M., Silvester, N., Smirnov, D., Toribio, A.-L., Vijayaraja, S. & Cochrane, G. The European Nucleotide Archive in 2018. *Nucleic Acids Research* **47,** D84–D88 (2019).

101. Bauer, Y., Tedrow, J., de Bernard, S., Birker-Robaczewska, M., Gibson, K. F., Guardela, B. J., Hess, P., Klenk, A., Lindell, K. O., Poirey, S., Renault, B., Rey, M., Weber, E., Nayler, O. & Kaminski, N. A Novel Genomic Signature with Translational Significance for Human Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory Cell and Molecular Biology* **52,** 217–231 (2015).

102. Kim, S., Herazo-Maya, J. D., Kang, D. D., Juan-Guardela, B. M., Tedrow, J., Martinez, F. J., Sciurba, F. C., Tseng, G. C. & Kaminski, N. Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* **16,** 924 (2015).

103. Peng, X., Moore, M., Mathur, A., Zhou, Y., Sun, H., Gan, Y., Herazo-Maya, J. D., Kaminski, N., Hu, X., Pan, H., Ryu, C., Osafo-Addo, A., Homer, R. J., Feghali-Bostwick, C., Fares, W. H., Gulati, M., Hu, B., Lee, C.-G., Elias, J. A. & Herzog, E. L. Plexin C1 deficiency permits synaptotagmin 7–mediated macrophage migration and enhances mammalian lung fibrosis. *The FASEB Journal* **30,** 4056–4070 (2016).

104. Anathy, V., Lahue, K. G., Chapman, D. G., Chia, S. B., Casey, D. T., Aboushousha, R., van der Velden, J. L. J., Elko, E., Hoffman, S. M., McMillan, D. H., Jones, J. T., Nolin, J. D., Abdalla, S., Schneider, R., Seward, D. J., Roberson, E. C., Liptak, M. D., Cousins, M. E., Butnor, K. J., Taatjes, D. J., Budd, R. C., Irvin, C. G., Ho, Y.-S., Hakem, R., Brown, K. K., Matsui, R., Bachschmid, M. M., Gomez, J. L., Kaminski, N., van der Vliet, A. & Janssen-Heininger, Y. M. W. Reducing protein oxidation reverses lung fibrosis. *Nature Medicine* **24,** 1128–1135 (2018).

105. Yang, I. V., Coldren, C. D., Leach, S. M., Seibold, M. A., Murphy, E., Lin, J., Rosen, R., Neidermyer, A. J., McKean, D. F., Groshong, S. D., Cool, C., Cosgrove, G. P., Lynch, D. A., Brown, K. K., Schwarz, M. I., Fingerlin, T. E. & Schwartz, D. A. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* **68,** 1114–1121 (2013).

106. Kadara, H., Fujimoto, J., Yoo, S.-Y., Maki, Y., Gower, A. C., Kabbout, M., Garcia, M. M., Chow, C.-W., Chu, Z., Mendoza, G., Shen, L., Kalhor, N., Hong, W. K., Moran, C., Wang, J., Spira, A., Coombes, K. R. & Wistuba, I. I. Transcriptomic Architecture of the Adjacent Airway Field Cancerization in Non–Small Cell Lung Cancer. *JNCI: Journal of the National Cancer Institute* **106,** dju004 (2014).

107. Kabbout, M., Garcia, M. M., Fujimoto, J., Liu, D. D., Woods, D., Chow, C.-W., Mendoza, G., Momin, A. A., James, B. P., Solis, L., Behrens, C., Lee, J. J., Wistuba, I. I. & Kadara, H. ETS2 Mediated Tumor Suppressive Function and MET Oncogene Inhibition in Human Non–Small Cell Lung Cancer. *Clinical Cancer Research* **19,** 3383–3395 (2013).

108. Sanchez-Palencia, A., Gomez-Morales, M., Gomez-Capilla, J. A., Pedraza, V., Boyero, L., Rosell, R. & Fárez-Vidal, M. E. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. *Clinical Cancer Research* **129,** 355–364 (2011).

109. Vuga, L. J., Ben-Yehudah, A., Kovkarova-Naumovski, E., Oriss, T., Gibson, K. F., Feghali-Bostwick, C. & Kaminski, N. WNT5A Is a Regulator of Fibroblast Proliferation and Resistance to Apoptosis. *American Journal of Respiratory Cell and Molecular Biology* **41,** 583–589 (2009).

110. Xu, Y., Mizuno, T., Sridharan, A., Du, Y., Guo, M., Tang, J., Wikenheiser–Brokamp, K. A., Perl, A. T., Funari, V. A., Gokey, J. J., Stripp, B. R. & Whitsett, J. A. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1,** e90558 (2017).

111. Königshoff, M., Kramer, M., Balsara, N., Wilhelm, J., Amarie, O. V., Jahn, A., Rose, F., Fink, L., Seeger, W., Schaefer, L., Günther, A. & Eickelberg, O. WNT1-inducible signaling protein–1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis. *The Journal of Clinical Investigation* **119,** 772–787 (2009).

112. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14,** 417–419 (2017).

113. Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K. J., Gall, A., Girón, C. G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., Kay, M., Laird, M. R., Lavidas, I., Liu, Z., Loveland, J. E., Marugán, J. C., Maurel, T., McMahon, A. C., Moore, B., Morales, J., Mudge, J. M., Nuhn, M., Ogeh, D., Parker, A., Parton, A., Patricio, M., Abdul Salam, A. I., Schmitt, B. M., Schuilenburg, H., Sheppard, D., Sparrow, H., Stapleton, E., Szuba, M., Taylor, K., Threadgold, G., Thormann, A., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Frankish, A., Hunt, S. E., Kostadima, M., Langridge, N., Martin, F. J., Muffato, M., Perry,

E., Ruffier, M., Staines, D. M., Trevanion, S. J., Aken, B. L., Yates, A. D., Zerbino, D. R. & Flicek, P. Ensembl 2019. *Nucleic Acids Research* **47,** D745–D751 (2019).

114. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2017).

115. RStudio Team. *RStudio: Integrated Development Environment for R* RStudio, Inc. (Boston, MA, 2015).

116. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 1; peer review: 2 approved]. *F1000Research* **4,** 1–17 (2015).

117. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 550 (2014).

118. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20,** 307–315 (2004).

119. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43,** e47 (2015).

120. Ulke, H. M., Mutze, K., Lehmann, M., Wagner, D. E., Heinzelmann, K., Günther, A., Eickelberg, O. & Königshoff, M. The Oncogene ECT2 Contributes to a Hyperplastic, Proliferative Lung Epithelial Cell Phenotype in Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory Cell and Molecular Biology* **61,** 713–726 (2019).

121. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12,** 35 (2011).

122. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25,** 25–29 (2000).

123. Consortium, T. G. O. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **47,** D330–D338 (2019).

124. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28,** 27–30 (2000).

125. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **47,** D590–D595 (2019).

126. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H. & D'Eustachio, P. The Reactome Pathway Knowledgebase. *Nucleic Acids Research* **46,** D649–D655 (2018).

127. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research* **18,** 623–632 (2019).

128. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J. & von Mering, C. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43,** D447–D452 (2015).

129. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13,** 2498–2504 (2003).

130. Subhash, S. & Kanduri, C. GeneSCF: a real-time based functional enrichment tool with support for multiple organisms. *BMC Bioinformatics* **17,** 365 (2016).

131. Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C. L., Haase, J., Janes, J., Huss, J. W. & Su, A. I. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biology* **10,** R130 (2009).

132. Wu, C., MacLeod, I. & Su, A. I. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Research* **41,** D561–D565 (2013).

133. Wu, C., Jin, X., Tsueng, G., Afrasiabi, C. & Su, A. I. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Research* **44,** D313–D316 (2016).

134. Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R. & Hogenesch, J. B. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 6062–6067 (2004).

135. Lê, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* **25,** 1–18 (2008).

136. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102,** 15545–15550 (2005).

137. Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D. & Groop, L. C. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34,** 267–273 (2003).

138. Corti, M., Brody, A. R. & Harrison, J. H. Isolation and primary culture of murine alveolar type II cells. *American Journal of Respiratory Cell and Molecular Biology* **14,** 309–315 (1996).

139. Mutze, K., Vierkotten, S., Milosevic, J., Eickelberg, O. & Königshoff, M. Enolase 1 (ENO1) and protein disulfide-isomerase associated 3 (PDIA3) regulate Wnt/β-catenin-driven trans-differentiation of murine alveolar epithelial cells. *Disease Models & Mechanisms* **8,** 877–890 (2015).

140. Smirnova, N. F., Schamberger, A. C., Nayakanti, S., Hatz, R., Behr, J. & Eickelberg, O. Detection and quantification of epithelial progenitor cell populations in human healthy and IPF lungs. *Respiratory Research* **17,** 83 (2016).

141. Pearson, K. F. R. S. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2,** 559–572 (1901).

142. Ringnér, M. What is principal component analysis? *Nature Biotechnology* **26,** 303–304 (2008).

143. Robbie, H., Daccord, C., Chua, F. & Devaraj, A. Evaluating disease severity in idiopathic pulmonary fibrosis. *European Respiratory Review* **26,** 170051 (2017).

144. Cao, H. J., Hogg, M. G., Martino, L. J. & Smith, T. J. Transforming growth factor-beta induces plasminogen activator inhibitor type-1 in cultured human orbital fibroblasts. *Investigative Ophthalmology & Visual Science* **36,** 1411–1419 (1995).

145. Kohl, P., Crampin, E. J., Quinn, T. A. & Noble, D. Systems Biology: An Approach. *Clinical Pharmacology & Therapeutics* **88,** 25–33 (2010).

146. Bruggeman, F. J. & Westerhoff, H. V. The nature of systems biology. *Trends in Microbiology* **15,** 45–50 (2007).

147. Tavassoly, I., Goldfarb, J. & Iyengar, R. Systems biology primer: the basic methods and approaches. *Essays in Biochemistry* **62,** 487–500 (2018).

148. Maleki, F., Ovens, K., Hogan, D. J. & Kusalik, A. J. Gene Set Analysis: Challenges, Opportunities, and Future Research. *Frontiers in Genetics* **11,** 654 (2020).

149. Spek, C. A. & Duitman, J. Is idiopathic pulmonary fibrosis a cancer-like disease? Transcriptome analysis to fuel the debate. *ERJ Open Research* **5,** 00157–2018 (2019).

150. Leng, D., Yi, J., Xiang, M., Zhao, H. & Zhang, Y. Identification of common signatures in idiopathic pulmonary fibrosis and lung cancer using gene expression modeling. *BMC Cancer* **20,** 986 (2020).

151. Shi, L., Jones, W. D., Jensen, R. V., Harris, S. C., Perkins, R. G., Goodsaid, F. M., Guo, L., Croner, L. J., Boysen, C., Fang, H., Qian, F., Amur, S., Bao, W., Barbacioru, C. C., Bertholet, V., Cao, X. M., Chu, T.-M., Collins, P. J., Fan, X.-h., Frueh, F. W., Fuscoe, J. C., Guo, X., Han, J., Herman, D., Hong, H., Kawasaki, E. S., Li, Q.-Z., Luo, Y., Ma, Y., Mei, N., Peterson, R. L., Puri, R. K., Shippy, R., Su, Z., Sun, Y. A., Sun, H., Thorn, B., Turpaz, Y., Wang, C., Wang, S. J., Warrington, J. A., Willey, J. C., Wu, J., Xie, Q., Zhang, L., Zhang, L., Zhong, S., Wolfinger, R. D. & Tong, W. The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics* **9,** S10 (2008).

152. Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G. & Stahl, F. Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology* **24,** 22–30 (2013).

153. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLOS ONE* **9,** e78644 (2014).

154. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10,** 57–63 (2009).

155. Angelidis, I., Simon, L. M., Fernandez, I. E., Strunz, M., Mayr, C. H., Greiffo, F. R., Tsitsiridis, G., Ansari, M., Graf, E., Strom, T.-M., Nagendran, M., Desai, T., Eickelberg, O., Mann, M., Theis, F. J. & Schiller, H. B. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nature Communications* **10,** 963 (2019).

156. Tamayo, P., Steinhardt, G., Liberzon, A. & Mesirov, J. P. The limitations of simple gene set enrichment analysis assuming gene independence. *Statistical Methods in Medical Research* **25,** 472–487 (2016).

157. Goeman, J. J. & Bühlmann, P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* **23,** 980–987 (2007).

158. Maciejewski, H. Gene set analysis methods: statistical models and methodological differences. *Briefings in Bioinformatics* **15,** 504–518 (2014).

159. Nam, D. & Kim, S.-Y. Gene-set approach for expression pattern analysis. *Briefings in Bioinformatics* **9,** 189–197 (2008).

160. Moore, B. B. & Hogaboam, C. M. Murine models of pulmonary fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **294,** L152–L160 (2008).

161. Tashiro, J., Rubio, G. A., Limper, A. H., Williams, K., Elliot, S. J., Ninou, I., Aidinis, V., Tzouvelekis, A. & Glassberg, M. K. Exploring Animal Models That Resemble Idiopathic Pulmonary Fibrosis. *Frontiers in Medicine* **4,** 118 (2017).

162. Jenkins, R. G., Moore, B. B., Chambers, R. C., Eickelberg, O., Königshoff, M., Kolb, M., Laurent, G. J., Nanthakumar, C. B., Olman, M. A., Pardo, A., Selman, M., Sheppard, D., Sime, P. J., Tager, A. M., Tatler, A. L., Thannickal, V. J. & White, E. S. An Official American Thoracic Society Workshop Report: Use of Animal Models for the Preclinical Assessment of Potential Therapies for Pulmonary Fibrosis. *American Journal of Respiratory Cell and Molecular Biology* **56,** 667–679 (2017).

163. Peng, R., Sridhar, S., Tyagi, G., Phillips, J. E., Garrido, R., Harris, P., Burns, L., Renteria, L., Woods, J., Chen, L., Allard, J., Ravindran, P., Bitter, H., Liang, Z., Hogaboam, C. M., Kitson, C., Budd, D. C., Fine, J. S., Bauer, C. M. T. & Stevenson, C. S. Bleomycin Induces Molecular Changes Directly Relevant to Idiopathic Pulmonary Fibrosis: A Model for "Active" Disease. *PLOS ONE* **8,** e59348 (2013).

164. Flaherty, K. R., Wells, A. U., Cottin, V., Devaraj, A., Walsh, S. L. F., Inoue, Y., Richeldi, L., Kolb, M., Tetzlaff, K., Stowasser, S., Coeck, C., Clerisme-Beaty, E., Rosenstock, B., Quaresma, M., Haeufel, T., Goeldner, R.-G., Schlenker-Herceg, R. & Brown, K. K. Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. *New England Journal of Medicine* **381,** 1718–1727 (2019).

165. Reyfman, P. A. & Gottardi, C. J. Idiopathic Pulmonary Fibrosis and Lung Cancer: Finding Similarities within Differences. *American Journal of Respiratory Cell and Molecular Biology* **61,** 667–668 (2019).

166. Blackwell, T. S., Tager, A. M., Borok, Z., Moore, B. B., Schwartz, D. A., Anstrom, K. J., Bar-Joseph, Z., Bitterman, P., Blackburn, M. R., Bradford, W., Brown, K. K., Chapman, H. A., Collard, H. R., Cosgrove, G. P., Deterding, R., Doyle, R., Flaherty, K. R., Garcia, C. K., Hagood, J. S., Henke, C. A., Herzog, E., Hogaboam, C. M., Horowitz, J. C., King, T. E., Loyd, J. E., Lawson, W. E., Marsh, C. B., Noble, P. W., Noth, I., Sheppard, D., Olsson, J., Ortiz, L. A., O'Riordan, T. G., Oury, T. D., Raghu, G., Roman, J., Sime, P. J., Sisson, T. H., Tschumperlin, D., Violette, S. M., Weaver, T. E., Wells, R. G., White, E. S., Kaminski, N., Martinez, F. J., Wynn, T. A., Thannickal, V. J. & Eu, J. P. Future Directions in Idiopathic Pulmonary Fibrosis Research. An NHLBI Workshop Report. *American Journal of Respiratory and Critical Care Medicine* **189,** 214–222 (2014).

167. Hironaka, M. & Fukayama, M. Pulmonary fibrosis and lung carcinoma: A comparative study of metaplastic epithelia in honeycombed areas of usual interstitial pneumonia with or without lung carcinoma. *Pathology International* **49,** 1060–1066 (1999).

168. Chilosi, M., Poletti, V., Murer, B., Lestani, M., Cancellieri, A., Montagna, L., Piccoli, P., Cangi, G., Semenzato, G. & Doglioni, C. Abnormal Re-epithelialization and Lung Remodeling in Idiopathic Pulmonary Fibrosis: The Role of ΔN-p63. *Laboratory Investigation* **82,** 1335–1345 (2002).

169. Wistuba, I. I. & Gazdar, A. F. Lung Cancer Preneoplasia. *Annual Review of Pathology: Mechanisms of Disease* **1,** 331–348 (2006).

170. Eymin, B. & Gazzeri, S. Role of cell cycle regulators in lung carcinogenesis. *Cell Adhesion & Migration* **4,** 114–123 (2010).

171. Horowitz, J. C., Osterholzer, J. J., Marazioti, A. & Stathopoulos, G. T. "Scarcinoma": viewing the fibrotic lung mesenchymal cell in the context of cancer biology. *European Respiratory Journal* **47,** 1842–1854 (2016).

172. Tzouvelekis, A., Gomatou, G., Bouros, E., Trigidou, R., Tzilas, V. & Bouros, D. Common Pathogenic Mechanisms Between Idiopathic Pulmonary Fibrosis and Lung Cancer. *CHEST* **156,** 383–391 (2019).

173. Selvarajah, B., Azuelos, I., Platé, M., Guillotin, D., Forty, E. J., Contento, G., Woodcock, H. V., Redding, M., Taylor, A., Brunori, G., Durrenberger, P. F., Ronzoni, R., Blanchard, A. D., Mercer, P. F., Anastasiou, D. & Chambers, R. C. mTORC1 amplifies the ATF4-dependent de novo serine-glycine pathway to supply glycine during TGF-β1-induced collagen biosynthesis. *Science Signaling* **12,** eaav3048 (2019).

174. O'Leary, E. M., Tian, Y., Nigdelioglu, R., Witt, L. J., Cetin-Atalay, R., Meliton, A. Y., Woods, P. S., Kimmig, L. M., Sun, K. A., Gökalp, G. A., Mutlu, G. M. & Hamanaka, R. B. TGF-β Promotes Metabolic Reprogramming in Lung Fibroblasts via mTORC1-dependent ATF4 Activation. *American Journal of Respiratory Cell and Molecular Biology* **63,** 601–612 (2020).

175. Zhu, W., Ding, Q., Wang, L., Xu, G., Diao, Y., Qu, S., Chen, S. & Shi, Y. Vitamin D3 alleviates pulmonary fibrosis by regulating the MAPK pathway via targeting PSAT1 expression in vivo and in vitro. *International Immunopharmacology* **101,** 108212 (2021).

176. Marmai, C., Sutherland, R. E., Kim, K. K., Dolganov, G. M., Fang, X., Kim, S. S., Jiang, S., Golden, J. A., Hoopes, C. W., Matthay, M. A., Chapman, H. A. & Wolters, P. J. Alveolar epithelial cells express mesenchymal proteins in patients with idiopathic pulmonary fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **301,** L71–L78 (2011).

177. Zhang, Y., Scoumanne, A. & Chen, X. G Protein-Coupled Receptor 87: a Promising Opportunity for Cancer Drug Discovery. *Molecular and cellular pharmacology* **2,** 111–116 (2010).

178. Heinzelmann, K., Hu, Q., Hu, Y., Dobrinskikh, E., Ansari, M., Melo-Narváez, M. C., Ulke, H. M., Leavitt, C., Mirita, C., Trudeau, T., Saal, M. L., Rice, P., Gao, B., Janssen, W. J., Yang, I. V., Schiller, H. B., Vladar, E. K., Lehmann, M. & Königshoff, M. Single-cell RNA sequencing identifies G-protein coupled receptor 87 as a basal cell marker expressed in distal honeycomb cysts in idiopathic pulmonary fibrosis. *European Respiratory Journal* **59,** 2102373 (2022).

List of Figures

List of Tables

Appendix

Indicated parts of the appendix have been previously published in Ulke *et al.* [120]. Adapted and reprinted with permission of the American Thoracic Society (p. III).

Table 1: Complete list of eigenvalues and variance for the PCA in section 4.2.1.

	Eigenvalue	**Variance (%)**	**Cumulative variance (%)**
PC 1	70.25049569	50.70414838	50.70414838
PC 2	14.48757326	10.45658194	61.16073032
PC 3	8.702191229	6.280912198	67.44164252
PC 4	5.90068254	4.25888928	71.7005318
PC 5	4.428190431	3.196100219	74.89663202
PC 6	3.204893545	2.313170836	77.20980285
PC 7	2.496660336	1.801994917	79.01179777
PC 8	2.334185591	1.684726797	80.69652457
PC 9	2.16778493	1.564625099	82.26114966
PC 10	1.739052936	1.255182575	83.51633224
PC 11	1.511036918	1.090609245	84.60694149
PC 12	1.501548648	1.083760972	85.69070246
PC 13	1.209759347	0.8731585	86.56386096
PC 14	1.158564654	0.836208108	87.40006906
PC 15	1.097886177	0.792412681	88.19248175
PC 16	1.008233923	0.727705078	88.92018682
PC 17	0.944060463	0.681387104	89.60157393
PC 18	0.911155519	0.657637561	90.25921149
PC 19	0.784233348	0.566029942	90.82524143
PC 20	0.770446589	0.556079182	91.38132061
PC 21	0.732712957	0.528844475	91.91016509
PC 22	0.659757072	0.476187679	92.38635277
PC 23	0.636935767	0.459716125	92.84606889
PC 24	0.577747722	0.416996435	93.26306533
PC 25	0.547928308	0.39547391	93.65853924
PC 26	0.488356023	0.35247689	94.01101613
PC 27	0.470589072	0.339653377	94.3506695
PC 28	0.456582691	0.329544102	94.6802136
PC 29	0.414868338	0.299436261	94.97964987
PC 30	0.403982965	0.291579611	95.27122948

	Eigenvalue	Variance (%)	Cumulative variance (%)
PC 31	0.373466088	0.269553685	95.54078316
PC 32	0.339639873	0.245139203	95.78592236
PC 33	0.309039764	0.223053203	96.00897557
PC 34	0.289442653	0.208908749	96.21788432
PC 35	0.286877832	0.207057559	96.42494188
PC 36	0.257131886	0.185588061	96.61052994
PC 37	0.253716287	0.183122811	96.79365275
PC 38	0.240532616	0.173607336	96.96726008
PC 39	0.221959365	0.160201867	97.12746195
PC 40	0.207536188	0.149791764	97.27725372
PC 41	0.203643365	0.146982071	97.42423579
PC 42	0.185844442	0.134135483	97.55837127
PC 43	0.176177177	0.127158017	97.68552929
PC 44	0.168274533	0.121454188	97.80698347
PC 45	0.15263517	0.110166288	97.91714976
PC 46	0.150694247	0.108765403	98.02591517
PC 47	0.140783567	0.101612249	98.12752741
PC 48	0.137139739	0.098982272	98.22650969
PC 49	0.135080274	0.097495828	98.32400551
PC 50	0.130274445	0.094027163	98.41803268
PC 51	0.120020484	0.086626242	98.50465892
PC 52	0.117637416	0.084906234	98.58956515
PC 53	0.11396021	0.082252165	98.67181732
PC 54	0.11342309	0.081864492	98.75368181
PC 55	0.104380192	0.075337671	98.82901948
PC 56	0.094942203	0.068525688	98.89754517
PC 57	0.090939526	0.065636707	98.96318188
PC 58	0.088206945	0.063664434	99.02684631
PC 59	0.083102992	0.059980594	99.0868269
PC 60	0.079643946	0.057483985	99.14431089
PC 61	0.076172274	0.054978264	99.19928915
PC 62	0.073378034	0.052961487	99.25225064
PC 63	0.07077662	0.051083885	99.30333453
PC 64	0.064630628	0.046647941	99.34998247
PC 65	0.061702273	0.044534365	99.39451683
PC 66	0.05932446	0.04281815	99.43733498
PC 67	0.057710246	0.041653071	99.47898805
PC 68	0.054533553	0.039360254	99.51834831
PC 69	0.051248819	0.036989457	99.55533776
PC 70	0.049720068	0.035886063	99.59122383
PC 71	0.046254952	0.033385074	99.6246089

	Eigenvalue	Variance (%)	Cumulative variance (%)
PC 72	0.042421133	0.030617968	99.65522687
PC 73	0.040927555	0.02953996	99.68476683
PC 74	0.039417829	0.028450297	99.71321712
PC 75	0.038239874	0.027600094	99.74081722
PC 76	0.03614222	0.026086086	99.7669033
PC 77	0.034284396	0.024745179	99.79164848
PC 78	0.03150146	0.022736561	99.81438504
PC 79	0.028287802	0.020417065	99.83480211
PC 80	0.028008735	0.020215645	99.85501775
PC 81	0.026014838	0.018776525	99.87379428
PC 82	0.024534135	0.01770781	99.89150209
PC 83	0.021534439	0.015542743	99.90704483
PC 84	0.020649813	0.014904253	99.92194909
PC 85	0.019167242	0.01383419	99.93578327
PC 86	0.017554071	0.012669864	99.94845314
PC 87	0.015318827	0.01105655	99.95950969
PC 88	0.014299401	0.010320766	99.96983046
PC 89	0.012774248	0.009219968	99.97905042
PC 90	0.011442814	0.00825899	99.98730941
PC 91	0.009867428	0.007121936	99.99443135
PC 92	0.007715354	0.005568651	100

Table 2: Full results of the REACTOME pathway analysis (Table 4.4). (Data was partially published in Ulke *et al.* [120] - *modified.*)

Stable identifier	Pathway	FDR value	*p*-value	Genes
R-HSA-69278	Cell Cycle, Mitotic	4.06E-10	1.27E-12	*BUB1, BUB1B, CCNA2, CCNB1, CCNB2, CDC45, CDK1, CENPE, CENPF, FOXM1, HMMR, KIF20A, MYBL2, NUF2, PLK1, RRM2, TOP2A, TPX2, TYMS*
R-HSA-1640170	Cell Cycle	6.73E-10	4.19E-12	*BUB1, BUB1B, CCNA2, CCNB1, CCNB2, CDC45, CDK1, CENPE, CENPF, EXO1, FOXM1, HMMR, KIF20A, MYBL2, NUF2, PLK1, RRM2, TOP2A, TPX2, TYMS*
R-HSA-156711	Polo-like kinase mediated events	5.88E-08	5.49E-10	*CCNB1, CCNB2, CENPF, FOXM1, MYBL2, PLK1*
R-HSA-2500257	Resolution of Sister Chromatid Cohesion	2.43E-07	3.03E-09	*BUB1, BUB1B, CCNB1, CCNB2, CDK1, CENPE, CENPF, NUF2, PLK1*

Stable identifier	Pathway	FDR value	*p*-value	Genes
R-HSA-69273	Cyclin A/B1/B2 associated events during G2/M transition	3.37E-07	5.25E-09	*CCNA2, CCNB1, CCNB2, CDK1, FOXM1, PLK1*
R-HSA-69620	Cell Cycle Checkpoints	3.87E-07	7.24E-09	*BUB1, BUB1B, CCNA2, CCNB1, CCNB2, CDC45, CDK1, CENPE, CENPF, EXO1, NUF2, PLK1*
R-HSA-69275	G2/M Transition	8.13E-07	1.84E-08	*CCNA2, CCNB1, CCNB2, CDK1, CENPF, FOXM1, HMMR, MYBL2, PLK1, TPX2*
R-HSA-453274	Mitotic G2-G2/M phases	8.13E-07	2.03E-08	*CCNA2, CCNB1, CCNB2, CDK1, CENPF, FOXM1, HMMR, MYBL2, PLK1, TPX2*
R-HSA-2980767	Activation of NIMA Kinases NEK9, NEK6, NEK7	4.73E-06	1.33E-07	*CCNB1, CCNB2, CDK1, PLK1*
R-HSA-68877	Mitotic Prometaphase	7.76E-06	2.42E-07	*BUB1, BUB1B, CCNB1, CCNB2, CDK1, CENPE, CENPF, NUF2, PLK1*
R-HSA-453279	Mitotic G1-G1/S phases	1.73E-05	5.91E-07	*CCNA2, CCNB1, CDC45, CDK1, MYBL2, RRM2, TOP2A, TYMS*
R-HSA-162658	Golgi Cisternae Pericentriolar Stack Reorganization	3.21E-05	1.20E-06	*CCNB1, CCNB2, CDK1, PLK1*
R-HSA-1442490	Collagen degradation	4.71E-05	1.91E-06	*MMP1, MMP11, MMP12, MMP7, PRSS2*
R-HSA-69478	G2/M DNA replication checkpoint	0.0001	5.12E-06	*CCNB1, CCNB2, CDK1*
R-HSA-68886	M Phase	0.0001	5.61E-06	*BUB1, BUB1B, CCNB1, CCNB2, CDK1, CENPE, CENPF, KIF20A, NUF2, PLK1*
R-HSA-141444	Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	0.0001	6.59E-06	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-141424	Amplification of signal from the kinetochores	0.0001	6.59E-06	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-176417	Phosphorylation of Emi1	0.0001	7.65E-06	*CCNB1, CDK1, PLK1*

Stable identifier	Pathway	FDR value	*p*-value	Genes
R-HSA-1474244	Extracellular matrix organization	0.0002	1.10E-05	*COL10A1, COL3A1, ITGA11, MMP1, MMP11, MMP12, MMP7, PRSS2, SPP1*
R-HSA-1538133	G0 and Early G1	0.0002	1.18E-05	*CCNA2, CDK1, MYBL2, TOP2A*
R-HSA-539107	Activation of E2F1 target genes at G1/S	0.0002	1.52E-05	*CDC45, CDK1, RRM2, TYMS*
R-HSA-69205	G1/S-Specific Transcription	0.0002	1.52E-05	*CDC45, CDK1, RRM2, TYMS*
R-HSA-69618	Mitotic Spindle Checkpoint	0.0002	1.55E-05	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-1474228	Degradation of the extracellular matrix	0.0002	1.63E-05	*MMP1, MMP11, MMP12, MMP7, PRSS2, SPP1*
R-HSA-1592389	Activation of Matrix Metalloproteinases	0.0003	2.42E-05	*MMP1, MMP11, MMP7, PRSS2*
R-HSA-5663220	RHO GTPases Activate Formins	0.0003	2.59E-05	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-2514853	Condensation of Prometaphase Chromosomes	0.0004	3.26E-05	*CCNB1, CCNB2, CDK1*
R-HSA-176814	Activation of APC/C and APC/C:Cdc20 mediated degradation of mitotic proteins	0.0004	3.51E-05	*BUB1B, CCNA2, CCNB1, CDK1, PLK1*
R-HSA-69206	G1/S Transition	0.0004	3.81E-05	*CCNA2, CCNB1, CDC45, CDK1, RRM2, TYMS*
R-HSA-176408	Regulation of APC/C activators between G1/S and early anaphase	0.0005	4.43E-05	*BUB1B, CCNA2, CCNB1, CDK1, PLK1*
R-HSA-6811434	COPI-dependent Golgi-to-ER retrograde traffic	0.0005	5.24E-05	*CENPE, KDELR3, KIF11, KIF20A, KIF4A*
R-HSA-174143	APC/C-mediated degradation of cell cycle proteins	0.0006	6.15E-05	*BUB1B, CCNA2, CCNB1, CDK1, PLK1*
R-HSA-453276	Regulation of mitotic cell cycle	0.0006	6.15E-05	*BUB1B, CCNA2, CCNB1, CDK1, PLK1*
R-HSA-983189	Kinesins	0.0006	6.87E-05	*CENPE, KIF11, KIF20A, KIF4A*

Stable identifier	Pathway	FDR value	*p*-value	Genes
R-HSA-1362300	Transcription of E2F targets under negative control by p107 (RBL1) and p130 (RBL2) in complex with HDAC1	0.0008	8.54E-05	*CCNA2, CDK1, MYBL2*
R-HSA-2980766	Nuclear Envelope Breakdown	0.0011	0.0001	*CCNB1, CCNB2, CDK1, PLK1*
R-HSA-176412	Phosphorylation of the APC/C	0.0012	0.0001	*CCNB1, CDK1, PLK1*
R-HSA-2467813	Separation of Sister Chromatids	0.0014	0.0002	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-8852276	The role of GTSE1 in G2/M progression after G2 checkpoint	0.0017	0.0002	*CCNB1, CCNB2, CDK1, PLK1*
R-HSA-68882	Mitotic Anaphase	0.0018	0.0002	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-2555396	Mitotic Metaphase and Anaphase	0.0018	0.0002	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-8856688	Golgi-to-ER retrograde transport	0.0018	0.0002	*CENPE, KDELR3, KIF11, KIF20A, KIF4A*
R-HSA-8854518	AURKA Activation by TPX2	0.0030	0.0004	*CDK1, HMMR, PLK1, TPX2*
R-HSA-176409	APC/C:Cdc20 mediated degradation of mitotic proteins	0.0034	0.0005	*BUB1B, CCNA2, CCNB1, CDK1*
R-HSA-499943	Interconversion of nucleotide di- and triphosphates	0.0038	0.0005	*AK4, RRM2, TYMS*
R-HSA-70614	Amino acid synthesis and interconversion (transamination)	0.0044	0.0006	*PSAT1, PYCR1, SERINC2*
R-HSA-3301854	Nuclear Pore Complex (NPC) Disassembly	0.0047	0.0007	*CCNB1, CCNB2, CDK1*
R-HSA-69481	G2/M Checkpoints	0.0050	0.0008	*CCNB1, CCNB2, CDC45, CDK1, EXO1*
R-HSA-2565942	Regulation of PLK1 Activity at G2/M Transition	0.0052	0.0008	*CCNB1, CCNB2, CDK1, PLK1*

Stable identifier	**Pathway**	**FDR value**	***p*-value**	**Genes**
R-HSA-2299718	Condensation of Prophase Chromosomes	0.0074	0.0012	*CCNB1, CDK1, PLK1*
R-HSA-195258	RHO GTPase Effectors	0.0090	0.0015	*BUB1, BUB1B, CENPE, CENPF, NUF2, PLK1*
R-HSA-6791312	TP53 Regulates Transcription of Cell Cycle Genes	0.0099	0.0017	*CCNA2, CCNB1, CDK1*
R-HSA-2132295	MHC class II antigen presentation	0.0099	0.0018	*CENPE, KIF11, KIF20A, KIF4A*
R-HSA-6811442	Intra-Golgi and retrograde Golgi-to-ER traffic	0.0103	0.0019	*CENPE, KDELR3, KIF11, KIF20A, KIF4A*
R-HSA-68875	Mitotic Prophase	0.0105	0.0019	*CCNB1, CCNB2, CDK1, PLK1*
R-HSA-194315	Signaling by Rho GTPases	0.0130	0.0025	*BUB1, BUB1B, CENPE, CENPF, ECT2, NUF2, PLK1*
R-HSA-2022090	Assembly of collagen fibrils and other multimeric structures	0.0138	0.0028	*COL10A1, COL3A1, MMP7*
R-HSA-983231	Factors involved in megakaryocyte development and platelet production	0.0227	0.0047	*CENPE, KIF11, KIF20A, KIF4A*
R-HSA-174184	Cdc20:Phospho-APC/C mediated degradation of Cyclin A	0.0237	0.0049	*BUB1B, CCNA2, CDK1*
R-HSA-179419	APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfation of the cell cycle checkpoint	0.0242	0.0051	*BUB1B, CCNA2, CDK1*
R-HSA-69473	G2/M DNA damage checkpoint	0.0277	0.0061	*CCNB1, CDK1, EXO1*
R-HSA-5633007	Regulation of TP53 Activity	0.0302	0.0069	*CCNA2, CDK1, EXO1, TPX2*
R-HSA-1430728	Metabolism	0.0302	0.0069	*AK4, AKR1B10, CYP24A1, GPX2, HMMR, HS6ST2, PDK1, PSAT1, PYCR1, RPL39L, RRM2, SERINC2, SLC2A1, SLC44A5, TCN1, TDO2, TNFRSF21, TYMS*

Stable identifier	Pathway	FDR value	*p*-value	Genes
R-HSA-3700989	Transcriptional Regulation by TP53	0.0323	0.0075	*CCNA2, CCNB1, CDK1, EXO1, GPX2, TPX2*
R-HSA-1474290	Collagen formation	0.0378	0.0090	*COL10A1, COL3A1, MMP7*
R-HSA-6804756	Regulation of TP53 Activity through Phosphorylation	0.0396	0.0095	*CCNA2, EXO1, TPX2*
R-HSA-15869	Metabolism of nucleotides	0.0499	0.0124	*AK4, RRM2, TYMS*

Table 3: Full results of the GO term analysis (Table 4.6). (Data was partially published in Ulke *et al.* [120] - *modified.*)

Accession number	Term name	FDR value	Genes
GO:0000278	Mitotic cell cycle	1.08E-13	*HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, MELK, PLK1, TPX2, PBK, BUB1, TYMS, CDKN3, FOXM1, HIST1H4K, RRM2, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDC45, TOP2A*
GO:1903047	Mitotic cell cycle process	1.08E-13	*HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, MELK, PLK1, TPX2, PBK, BUB1, TYMS, CDKN3, FOXM1, RRM2, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDC45, TOP2A*
GO:0007067	Mitotic nuclear division	1.34E-11	*MYBL2, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, PLK1, TPX2, PBK, BUB1, CENPF, ASPM, CEP55, KIF4A, CDK1*
GO:0022402	Cell cycle process	1.34E-11	*FAP, ECT2, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, MELK, TPX2, PBK, BUB1, TYMS, CDKN3, FOXM1, RRM2, CENPF, ASPM, MKI67, CEP55, STIL, KIF4A, KIF20A, CDC45, TOP2A*
GO:0051301	Cell division	1.51E-11	*ECT2, HELLS, CCNB1, KIF11, CENPE, ANLN, NUF2, CCNA2, BUB1B, CCNB2, PLK1, TPX2, BUB1, CENPF, ASPM, CEP55, KIF4A, KIF20A, CDK1, TOP2A*

Accession number	Term name	FDR value	Genes
GO:0000280	Nuclear division	2.83E-11	*MYBL2, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, TPX2, PBK, BUB1, CENPF, ASPM, MKI67, CEP55, KIF4A, CDK1, TOP2A*
GO:0007049	Cell cycle	4.37E-11	*FAP, ECT2, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, MELK, TPX2, PBK, BUB1, EXO1, TYMS, CDKN3, FOXM1, HIST1H4K, RRM2, CENPF, ASPM, MKI67, CEP55, STIL, KIF4A, KIF20A, CDC45, TOP2A*
GO:0005819	Spindle	1.21E-09	*ECT2, DLGAP5, CCNB1, KIF11, CENPE, BUB1B, PLK1, TPX2, CKAP2L, CENPF, ASPM, TTK, KIF4A, KIF20A, CDK1*
GO:0044772	Mitotic cell cycle phase transition	8.75E-08	*MYBL2, CCNB1, CCNA2, BUB1B, CCNB2, MELK, PLK1, TYMS, CDKN3, FOXM1, RRM2, CENPF, CDK1, CDC45*
GO:0030496	Midbody	1.33E-07	*ECT2, CENPE, PLK1, CENPF, ASPM, CEP55, KIF4A, KIF20A, CDK1, SLC2A1*
GO:0051726	Regulation of cell cycle	1.92E-07	*FAP, MYBL2, ECT2, DLGAP5, CCNB1, KIF11, CENPE, ANLN, CCNB2, KIAA0101, PLK1, TPX2, BUB1, CDKN3, FOXM1, CENPF, ASPM, TTK, STIL, CDC45, TOP2A*
GO:0000940	Condensed chromosome outer kinetochore	3.63E-07	*CCNB1, BUB1B, PLK1, BUB1, CENPF*
GO:0030071	Regulation of mitotic metaphase/anaphase transition	1.15E-06	*DLGAP5, CENPE, BUB1B, PLK1, BUB1, CENPF, TTK*
GO:0010564	Regulation of cell cycle process	4.06E-06	*FAP, ECT2, DLGAP5, CCNB1, KIF11, CENPE, ANLN, CCNA2, PLK1, TPX2, BUB1, FOXM1, CENPF, TTK, CDC45*
GO:0000778	Condensed nuclear chromosome kinetochore	4.96E-06	*CCNB1, BUB1B, PLK1, BUB1*
GO:0015630	Microtubule cytoskeleton	6.60E-06	*ECT2, DLGAP5, CCNB1, KIF11, CENPE, BUB1B, CCNB2, PLK1, TPX2, CKAP2L, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDK1, CDC45*
GO:1901990	Regulation of mitotic cell cycle phase transition	8.15E-06	*DLGAP5, CCNB1, CENPE, ANLN, CCNA2, BUB1B, PLK1, BUB1, CENPF, TTK, CDC45*
GO:0000086	G2/M transition of mitotic cell cycle	9.10E-06	*MYBL2, CCNB1, CCNA2, CCNB2, MELK, PLK1, FOXM1, CENPF, CDK1*

Accession number	Term name	FDR value	Genes
GO:0051983	Regulation of chromosome segregation	9.77E-06	*ECT2, DLGAP5, CCNB1, PLK1, BUB1, CENPF, TTK*
GO:0000922	Spindle pole	1.37E-05	*DLGAP5, CCNB1, KIF11, PLK1, TPX2, CKAP2L, CENPF, ASPM*
GO:0044430	Cytoskeletal part	1.92E-05	*ECT2, DLGAP5, CCNB1, KIF11, CENPE, ANLN, BUB1B, CCNB2, PLK1, TPX2, CKAP2L, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDK1, CDC45, SLC2A1*
GO:0007346	Regulation of mitotic cell cycle	3.90E-05	*DLGAP5, CCNB1, KIF11, CENPE, ANLN, CCNA2, PLK1, TPX2, BUB1, CENPF, TTK, CDC45, TOP2A*
GO:0007088	Regulation of mitotic nuclear division	4.03E-05	*DLGAP5, KIF11, CENPE, ANLN, PLK1, BUB1, CENPF, TTK*
GO:0016043	Cellular component organization	6.02E-05	*MYBL2, TOX3, HELLS, COL10A1, DLGAP5, CCNB1, SULF1, MMP7, CDH3, CENPE, ANLN, NUF2, CCNA2, CCNB2, KIAA0101, GREM1, TPX2, PBK, BUB1, COL3A1, ATP10B, ETV4, MMP1, ITGA11, CTHRC1, HIST1H4K, RRM2, CENPF, ASPM, CRABP2, MKI67, TTK, CEP55, KIF4A, GJB2, GOLM1, KIF20A, CDK1, CDC45, TOP2A, SLC2A1*
GO:0000281	Mitotic cytokinesis	6.33E-05	*ANLN, PLK1, CEP55, KIF4A, KIF20A*
GO:0007059	Chromosome segregation	6.33E-05	*DLGAP5, CCNB1, KIF11, CENPE, NUF2, BUB1, CENPF, TTK, TOP2A*
GO:0000942	Condensed nuclear chromosome outer kinetochore	6.71E-05	*CCNB1, PLK1, BUB1*
GO:0051439	Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	7.42E-05	*CCNB1, BUB1B, PLK1, BUB1, CENPF, TTK, CDK1*
GO:0051302	Regulation of cell division	7.42E-05	*ECT2, DLGAP5, KIF11, CENPE, ANLN, PLK1, BUB1, CENPF, ASPM, TTK*
GO:0032434	Regulation of proteasomal ubiquitin-dependent protein catabolic process	7.77E-05	*DLGAP5, CENPE, BUB1B, PLK1, PBK, BUB1, CENPF, TTK*
GO:0000780	Condensed nuclear chromosome, centromeric region	8.37E-05	*CCNB1, BUB1B, PLK1, BUB1*

Accession number	Term name	FDR value	Genes
GO:0000775	Chromosome, centromeric region	9.07E-05	*HELLS, CCNB1, CENPE, NUF2, PLK1, BUB1, CENPF, MKI67*
GO:0032435	Negative regulation of proteasomal ubiquitin-dependent protein catabolic process	9.12E-05	*BUB1B, PLK1, PBK, BUB1, CENPF, TTK*
GO:0007094	Mitotic spindle assembly checkpoint	1.37E-04	*BUB1B, PLK1, BUB1, CENPF, TTK*
GO:0000793	Condensed chromosome	1.42E-04	*CCNB1, CENPE, NUF2, PLK1, BUB1, CENPF, MKI67, TOP2A*
GO:1903050	Regulation of proteolysis involved in cellular protein catabolic process	1.49E-04	*DLGAP5, CCNB1, CENPE, BUB1B, PLK1, PBK, BUB1, CENPF, TTK, CDK1*
GO:0000075	Cell cycle checkpoint	1.50E-04	*CCNA2, BUB1B, PLK1, BUB1, CENPF, TTK, CDK1, CDC45, TOP2A*
GO:1902589	Single-organism organelle organization	1.60E-04	*MYBL2, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, KIAA0101, TPX2, PBK, BUB1, HIST1H4K, CENPF, ASPM, MKI67, TTK, CEP55, STIL, KIF4A, KIF20A, CDK1, TOP2A*
GO:0007093	Mitotic cell cycle checkpoint	1.74E-04	*CCNA2, BUB1B, PLK1, BUB1, CENPF, TTK, CDK1, TOP2A*
GO:0051782	Negative regulation of cell division	1.84E-04	*BUB1B, PLK1, BUB1, CENPF, ASPM, TTK*
GO:0010965	Regulation of mitotic sister chromatid separation	2.92E-04	*DLGAP5, PLK1, BUB1, CENPF, TTK*
GO:0007052	Mitotic spindle organization	3.06E-04	*MYBL2, KIF11, TTK, STIL, KIF4A*
GO:0000226	Microtubule cytoskeleton organization	3.06E-04	*MYBL2, CENPE, KIAA0101, PLK1, ASPM, TTK, STIL, KIF4A, KIF20A, CDK1*
GO:0044710	Single-organism metabolic process	3.13E-04	*MMP11, CYP24A1, RAD51AP1, COL10A1, CCNB1, TCN1, SULF1, MMP7, PDK1, BUB1B, TNFRSF21, KIAA0101, PLK1, COL3A1, AK4, MMP1, PYCR1, PPAP2C, FOXM1, HIST1H4K, AKR1B10, UBE2T, CRABP2, SLC44A5, SERINC2, PSAT1, GPX2, FUT2, HMMR, CDK1, GCNT3, CDC45, TOP2A, SLC2A1, CFB, TDO2*

Accession number	Term name	FDR value	Genes
GO:0051340	Regulation of ligase activity	3.13E-04	*CCNB1, BUB1B, PLK1, BUB1, CENPF, TTK, CDK1*
GO:0000910	Cytokinesis	3.57E-04	*ECT2, ANLN, PLK1, CEP55, KIF4A, KIF20A*
GO:0007051	Spindle organization	3.77E-04	*MYBL2, KIF11, ASPM, TTK, STIL, KIF4A*
GO:0000777	Condensed chromosome kinetochore	4.68E-04	*CCNB1, CENPE, NUF2, PLK1, BUB1, CENPF*
GO:0000779	Condensed chromosome, centromeric region	5.78E-04	*CCNB1, CENPE, NUF2, PLK1, BUB1, CENPF*
GO:0051338	Regulation of transferase activity	6.01E-04	*ECT2, CCNB1, CENPE, CCNA2, BUB1B, PLK1, GREM1, TPX2, BUB1, CDKN3, CENPF, TTK, STIL, SERINC2, CDK1*
GO:0005856	Cytoskeleton	6.34E-04	*ECT2, DLGAP5, TNS4, CCNB1, KIF11, CENPE, ANLN, BUB1B, CCNB2, PLK1, TPX2, CKAP2L, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDK1, CDC45, SLC2A1*
GO:0005737	Cytoplasm	7.50E-04	*FAP, MMP11, CYP24A1, ECT2, DLGAP5, TNS4, CCNB1, SULF1, KIF11, CDH3, CENPE, ANLN, NUF2, CCNA2, CXCL13, BUB1B, CCNB2, RPL39L, STEAP1, MELK, KIAA0101, PLK1, TPX2, BUB1, COL3A1, CKAP2L, CDCA7, ATP10B, TYMS, AK4, PYCR1, CTHRC1, CDKN3, FOXM1, AKR1B10, RRM2, CENPF, UBE2T, ASPM, RALGPS2, AIM2, CRABP2, MKI67, CEP55, STIL, KIF4A, PSAT1, GJB2, GOLM1, GPX2, FUT2, HMMR, KIF20A, SPP1, CDK1, GCNT3, KDELR3, CDC45, TOP2A, HS6ST2, TDO2*
GO:1902850	Microtubule cytoskeleton organization involved in mitosis	7.62E-04	*MYBL2, KIF11, CENPE, KIF4A*
GO:0008283	Cell proliferation	7.82E-04	*FAP, HELLS, DLGAP5, PDK1, BUB1B, MELK, PLK1, TPX2, BUB1, CENPF, ASPM, MKI67, STIL*
GO:0006950	Response to stress	8.39E-04	*RAD51AP1, ECT2, CCNB1, MMP7, KIF11, CDH3, CENPE, CCNA2, LCN2, PDK1, CXCL13, MELK, KIAA0101, PLK1, COL3A1, EXO1, MMP1, PYCR1, FOXM1, UBE2T, AIM2, MKI67, KIF4A, GJB2, GPX2, SPP1, KDELR3, CDC45, TOP2A, SLC2A1, CFB*

Accession number	Term name	FDR value	Genes
GO:0005694	Chromosome	9.33E-04	*MYBL2, HELLS, CCNB1, CENPE, NUF2, PLK1, BUB1, HIST1H4K, CENPF, MKI67, KIF4A, CDC45, TOP2A*
GO:0000083	Regulation of transcription involved in G1/S transition of mitotic cell cycle	0.0010	*TYMS, RRM2, CDK1, CDC45*
GO:0007017	Microtubule-based process	0.0010	*MYBL2, DLGAP5, CENPE, KIAA0101, PLK1, ASPM, TTK, STIL, KIF4A, KIF20A, CDK1*
GO:1901991	Negative regulation of mitotic cell cycle phase transition	0.0011	*CCNA2, BUB1B, PLK1, BUB1, CENPF, TTK, CDK1*
GO:0006915	Apoptotic process	0.0011	*FAP, TOX3, ECT2, TNS4, SULF1, LCN2, PDK1, BUB1B, TNFRSF21, GREM1, TPX2, BUB1, CDCA7, AIM2, CDK1, TOP2A*
GO:0050896	Response to stimulus	0.0013	*CYP24A1, RAD51AP1, ECT2, CCNB1, SULF1, MMP7, KIF11, GPR87, CENPE, NUF2, CCNA2, LCN2, PDK1, CXCL13, BUB1B, TNFRSF21, KIAA0101, PLK1, GREM1, PBK, BUB1, CST1, ETV4, AK4, MMP1, ITGA11, PYCR1, CTHRC1, HIST1H4K, AKR1B10, CENPF, UBE2T, RALGPS2, CRABP2, MKI67, STIL, KIF4A, GJB2, GPX2, SPP1, GCNT3, KDELR3, CDC45, TOP2A, DEPDC1, SLC2A1, CFB*
GO:0007275	Multicellular organismal development	0.0013	*FAP, MMP11, ECT2, HELLS, COL10A1, CCNB1, SULF1, CDH3, CENPE, ANLN, CCNA2, CXCL13, CCNB2, TNFRSF21, MELK, GREM1, EXO1, TYMS, ETV4, AK4, ITGA11, CTHRC1, FOXM1, CENPF, ASPM, CRABP2, MKI67, STIL, KIF4A, GJB2, SPP1, CDK1, GCNT3, TOP2A*
GO:0051297	Centrosome organization	0.0013	*KIF11, KIAA0101, PLK1, STIL, CDK1*
GO:0048731	System development	0.0015	*FAP, ECT2, COL10A1, CCNB1, SULF1, CDH3, ANLN, CCNA2, CXCL13, CCNB2, TNFRSF21, MELK, GREM1, EXO1, TYMS, ETV4, AK4, ITGA11, CTHRC1, FOXM1, CENPF, ASPM, CRABP2, MKI67, STIL, KIF4A, GJB2, SPP1, CDK1, GCNT3, TOP2A*
GO:0022617	Extracellular matrix disassembly	0.0018	*MMP11, COL10A1, MMP7, COL3A1, MMP1, SPP1*

Accession number	Term name	FDR value	Genes
GO:0006996	Organelle organization	0.0018	*MYBL2, TOX3, HELLS, DLGAP5, CCNB1, CENPE, ANLN, NUF2, CCNA2, CCNB2, KIAA0101, TPX2, PBK, BUB1, HIST1H4K, CENPF, ASPM, MKI67, TTK, CEP55, STIL, KIF4A, GOLM1, KIF20A, CDK1, CDC45, TOP2A*
GO:0030162	Regulation of proteolysis	0.0020	*DLGAP5, CCNB1, CENPE, BUB1B, PLK1, PBK, BUB1, CST1, CENPF, AIM2, TTK, CDK1, CFB*
GO:0030574	Collagen catabolic process	0.0023	*MMP11, COL10A1, MMP7, COL3A1, MMP1*
GO:0048519	Negative regulation of biological process	0.0023	*FAP, MMP11, TOX3, CEACAM5, HELLS, CCNB1, SULF1, GPR87, CDH3, CCNA2, CXCL13, BUB1B, TNFRSF21, PLK1, GREM1, PBK, BUB1, COL3A1, CST1, ETV4, PYCR1, CTHRC1, CDKN3, FOXM1, HIST1H4K, CENPF, ASPM, AIM2, TTK, STIL, SPP1, KDELR3, TOP2A, DEPDC1*
GO:0045786	Negative regulation of cell cycle	0.0027	*FAP, CCNA2, BUB1B, PLK1, BUB1, CDKN3, CENPF, TTK, CDK1, TOP2A*
GO:0006928	Movement of cell or subcellular component	0.0028	*FAP, DLGAP5, KIF11, CENPE, ANLN, CXCL13, GREM1, COL3A1, ETV4, MMP1, ITGA11, CTHRC1, ASPM, KIF4A, KIF20A, SPP1, CDK1*
GO:0072686	Mitotic spindle	0.0030	*ECT2, CENPE, ASPM, CDK1*
GO:0045120	Pronucleus	0.0030	*CCNA2, CENPF, SLC2A1*
GO:0033554	Cellular response to stress	0.0031	*RAD51AP1, ECT2, CCNB1, CCNA2, PDK1, MELK, KIAA0101, PLK1, EXO1, PYCR1, FOXM1, UBE2T, MKI67, GJB2, SPP1, KDELR3, CDC45, TOP2A, SLC2A1*
GO:0048523	Negative regulation of cellular process	0.0033	*FAP, MMP11, TOX3, CEACAM5, HELLS, CCNB1, SULF1, GPR87, CCNA2, CXCL13, BUB1B, TNFRSF21, PLK1, GREM1, PBK, BUB1, COL3A1, CST1, ETV4, PYCR1, CTHRC1, CDKN3, FOXM1, HIST1H4K, CENPF, ASPM, AIM2, TTK, STIL, SPP1, TOP2A, DEPDC1*
GO:0000819	Sister chromatid segregation	0.0037	*DLGAP5, CCNB1, CENPE, PLK1, TOP2A*
GO:0006259	DNA metabolic process	0.0038	*RAD51AP1, HELLS, KIAA0101, FOXM1, HIST1H4K, RRM2, CENPF, UBE2T, MKI67, CDK1, CDC45, TOP2A*
GO:0033044	Regulation of chromosome organization	0.0038	*DLGAP5, CCNB1, PLK1, BUB1, CENPF, TTK, CDC45*

Accession number	Term name	FDR value	Genes
GO:0044446	Intracellular organelle part	0.0041	*MMP11, CYP24A1, ECT2, HELLS, COL10A1, DLGAP5, CCNB1, SULF1, KIF11, CENPE, ANLN, NUF2, CCNA2, CCNB2, RPL39L, STEAP1, KIAA0101, PLK1, TPX2, BUB1, COL3A1, CKAP2L, CDCA7, EXO1, ATP10B, ETV4, AK4, PYCR1, FOXM1, HIST1H4K, RRM2, CENPF, UBE2T, ASPM, CRABP2, TTK, CEP55, STIL, KIF4A, FUT2, KIF20A, CDK1, GCNT3, KDELR3, CDC45, TOP2A, DEPDC1, SLC2A1, HS6ST2*
GO:0044707	Single-multicellular organism process	0.0044	*FAP, MMP11, CYP24A1, ECT2, HELLS, COL10A1, CCNB1, SULF1, MMP7, KIF11, CDH3, CENPE, CCNA2, CXCL13, CCNB2, TNFRSF21, MELK, GREM1, CST1, EXO1, TYMS, ETV4, AK4, MMP1, ITGA11, CTHRC1, FOXM1, AKR1B10, CENPF, ASPM, AIM2, CRABP2, MKI67, STIL, KIF4A, SPP1, CDK1, GCNT3, TOP2A*
GO:0044763	Single-organism cellular process	0.0047	*FAP, CYP24A1, TOX3, RAD51AP1, HELLS, COL10A1, DLGAP5, TNS4, CCNB1, TCN1, SULF1, MMP7, GPR87, CDH3, CENPE, NUF2, CCNA2, LCN2, PDK1, CXCL13, CCNB2, TNFRSF21, STEAP1, KIAA0101, GREM1, TPX2, PBK, BUB1, COL3A1, CDCA7, ATP10B, AK4, MMP1, ITGA11, PYCR1, PPAP2C, CTHRC1, CDKN3, HIST1H4K, AKR1B10, CENPF, UBE2T, ASPM, RALGPS2, AIM2, MKI67, SLC44A5, CEP55, STIL, KIF4A, PSAT1, FUT2, KIF20A, GCNT3, CDC45, DEPDC1, HS6ST2, TDO2*
GO:0051984	Positive regulation of chromosome segregation	0.0048	*DLGAP5, CCNB1, PLK1*
GO:0007077	Mitotic nuclear envelope disassembly	0.0050	*CCNB1, CCNB2, PLK1, CDK1*
GO:0050790	Regulation of catalytic activity	0.0054	*ECT2, CCNB1, GPR87, CDH3, CENPE, CCNA2, CXCL13, BUB1B, PLK1, GREM1, TPX2, BUB1, CST1, CDKN3, CENPF, RALGPS2, AIM2, TTK, STIL, SERINC2, CDK1, DEPDC1*
GO:0005871	Kinesin complex	0.0056	*KIF11, CENPE, KIF4A, KIF20A*
GO:0034508	Centromere complex assembly	0.0061	*HELLS, CENPE, HIST1H4K, CENPF*
GO:0051303	Establishment of chromosome localization	0.0061	*DLGAP5, CCNB1, CENPE, CENPF*

Accession number	Term name	FDR value	Genes
GO:0070013	Intracellular organelle lumen	0.0062	*MMP11, MYBL2, COL10A1, CCNB1, ANLN, CCNA2, BUB1B, CCNB2, KIAA0101, PLK1, TPX2, BUB1, COL3A1, CDCA7, EXO1, ETV4, AK4, PYCR1, FOXM1, HIST1H4K, RRM2, CENPF, UBE2T, CRABP2, MKI67, KIF4A, KIF20A, CDK1, CDC45, TOP2A, DEPDC1, HS6ST2*
GO:0044427	Chromosomal part	0.0062	*MYBL2, HELLS, CCNB1, CENPE, NUF2, PLK1, BUB1, HIST1H4K, CENPF, MKI67, CDC45*
GO:0051179	Localization	0.0063	*FAP, ECT2, DLGAP5, TNS4, CCNB1, TCN1, MMP7, CENPE, ANLN, CXCL13, BUB1B, STEAP1, PLK1, GREM1, COL3A1, ATP10B, MMP1, ITGA11, CTHRC1, CENPF, ASPM, AIM2, CRABP2, SLC44A5, CEP55, STIL, SERINC2, KIF4A, KIF20A, SPP1, CDK1, GCNT3, KDELR3*
GO:0022411	Cellular component disassembly	0.0063	*MMP11, COL10A1, CCNB1, MMP7, CCNB2, COL3A1, MMP1, SPP1, CDK1, TOP2A*
GO:0051225	Spindle assembly	0.0064	*MYBL2, KIF11, ASPM, KIF4A*
GO:0043232	Intracellular non-membrane-bounded organelle	0.0065	*MYBL2, ECT2, HELLS, DLGAP5, TNS4, CCNB1, KIF11, CENPE, ANLN, NUF2, CCNB2, RPL39L, PLK1, TPX2, BUB1, CKAP2L, TYMS, ETV4, HIST1H4K, CENPF, ASPM, TTK, CEP55, STIL, KIF4A, KIF20A, CDK1, CDC45, SLC2A1*
GO:0044421	Extracellular region part	0.0066	*FAP, MMP11, CEACAM5, COL10A1, TCN1, SULF1, MMP7, LCN2, CXCL13, GREM1, COL3A1, CST1, AK4, MMP1, CTHRC1, HIST1H4K, AKR1B10, CRABP2, CEP55, SERINC2, PSAT1, GOLM1, GPX2, FUT2, SPP1, CDK1, GCNT3, SLC2A1, CFB, HS6ST2*
GO:0007079	Mitotic chromosome movement towards spindle pole	0.0067	*DLGAP5, CENPE*

Accession number	Term name	FDR value	Genes
GO:0065007	Biological regulation	0.0073	*FAP, MMP11, CYP24A1, MYBL2, TOX3, CEACAM5, RAD51AP1, ECT2, HELLS, DLGAP5, CCNB1, SULF1, MMP7, KIF11, GPR87, CDH3, CENPE, ANLN, NUF2, CCNA2, LCN2, PDK1, CXCL13, TNFRSF21, STEAP1, KIAA0101, PLK1, GREM1, TPX2, PBK, COL3A1, CST1, CDCA7, ATP10B, MMP1, ITGA11, PYCR1, CTHRC1, CDKN3, FOXM1, HIST1H4K, AKR1B10, RRM2, CENPF, ASPM, RALGPS2, AIM2, TTK, STIL, SERINC2, KIF4A, GOLM1, SPP1, CDC45, SLC2A1, CFB*
GO:0060236	Regulation of mitotic spindle organization	0.0074	*CCNB1, PLK1, TPX2*
GO:0005813	Centrosome	0.0079	*DLGAP5, CCNB1, CCNB2, PLK1, CKAP2L, CEP55, STIL, CDK1, CDC45*
GO:0005654	Nucleoplasm	0.0082	*MYBL2, CCNB1, ANLN, CCNA2, CCNB2, KIAA0101, PLK1, TPX2, BUB1, CDCA7, EXO1, TYMS, FOXM1, HIST1H4K, RRM2, CENPF, UBE2T, CRABP2, KIF4A, KIF20A, CDK1, CDC45, TOP2A, DEPDC1, HS6ST2*
GO:0005815	Microtubule organizing center	0.0083	*DLGAP5, CCNB1, BUB1B, CCNB2, PLK1, CKAP2L, CEP55, STIL, CDK1, CDC45*
GO:0051233	Spindle midzone	0.0085	*CENPE, BUB1B, PLK1*
GO:0032501	Multicellular organismal process	0.0090	*FAP, MMP11, CYP24A1, ECT2, HELLS, COL10A1, CCNB1, SULF1, KIF11, CDH3, CENPE, CCNA2, CXCL13, CCNB2, RPL39L, TNFRSF21, MELK, GREM1, CST1, EXO1, TYMS, ETV4, AK4, MMP1, ITGA11, CTHRC1, FOXM1, AKR1B10, CENPF, ASPM, AIM2, CRABP2, MKI67, STIL, KIF4A, SPP1, CDK1, GCNT3, TOP2A*
GO:0051128	Regulation of cellular component organization	0.0098	*FAP, ECT2, DLGAP5, CCNB1, KIF11, CENPE, ANLN, PDK1, CXCL13, PLK1, GREM1, TPX2, BUB1, MMP1, FOXM1, CENPF, AIM2, CRABP2, TTK, SPP1, CDC45*
GO:0007264	Small GTPase mediated signal transduction	0.0098	*ECT2, CENPE, NUF2, CCNA2, BUB1B, PLK1, BUB1, HIST1H4K, CENPF, RALGPS2, CDK1*

Accession number	Term name	FDR value	Genes
GO:0051716	Cellular response to stimulus	0.0098	*CYP24A1, RAD51AP1, ECT2, CCNB1, SULF1, MMP7, GPR87, CDH3, CENPE, NUF2, CCNA2, PDK1, CXCL13, BUB1B, TNFRSF21, KIAA0101, PLK1, GREM1, PBK, BUB1, EXO1, ITGA11, PYCR1, CTHRC1, FOXM1, HIST1H4K, AKR1B10, CENPF, UBE2T, RALGPS2, CRABP2, MKI67, STIL, GJB2, SPP1, KDELR3, CDC45, TOP2A, DEPDC1, SLC2A1*
GO:0090307	Mitotic spindle assembly	0.0098	*MYBL2, KIF11, KIF4A*
GO:0005615	Extracellular space	0.0098	*FAP, TCN1, SULF1, MMP7, LCN2, CXCL13, SPINK1, GREM1, COL3A1, CST1, CTHRC1, GOLM1, SPP1, SLC2A1, CFB*
GO:0000082	G1/S transition of mitotic cell cycle	0.0109	*CCNB1, TYMS, CDKN3, RRM2, CDK1, CDC45*
GO:0009894	Regulation of catabolic process	0.0121	*FAP, DLGAP5, CCNB1, CENPE, PDK1, BUB1B, PLK1, PBK, BUB1, CENPF, TTK, CDK1*
GO:0050789	Regulation of biological process	0.0121	*FAP, MMP11, CYP24A1, MYBL2, TOX3, CEACAM5, RAD51AP1, ECT2, HELLS, DLGAP5, CCNB1, SULF1, MMP7, KIF11, GPR87, CDH3, CENPE, ANLN, NUF2, CCNA2, PDK1, CXCL13, CCNB2, TNFRSF21, KIAA0101, PLK1, GREM1, TPX2, PBK, COL3A1, CST1, CDCA7, TYMS, MMP1, ITGA11, PYCR1, CTHRC1, CDKN3, FOXM1, HIST1H4K, AKR1B10, RRM2, CENPF, ASPM, RALGPS2, AIM2, TTK, STIL, SERINC2, GOLM1, SPP1, CDC45, SLC2A1, CFB*
GO:0000794	Condensed nuclear chromosome	0.0130	*CCNB1, BUB1B, PLK1, BUB1*
GO:0030198	Extracellular matrix organization	0.0134	*COL10A1, SULF1, MMP7, GREM1, COL3A1, MMP1, ITGA11, SPP1*
GO:0031400	Negative regulation of protein modification process	0.0136	*CCNB1, BUB1B, SPINK1, PLK1, GREM1, PBK, BUB1, FOXM1, CENPF, TTK*
GO:0070507	Regulation of microtubule cytoskeleton organization	0.0143	*ECT2, CCNB1, KIF11, PLK1, TPX2*
GO:0031329	Regulation of cellular catabolic process	0.0145	*DLGAP5, CCNB1, CENPE, PDK1, BUB1B, PLK1, PBK, BUB1, CENPF, TTK, CDK1*

Accession number	Term name	FDR value	Genes
GO:0031981	Nuclear lumen	0.0156	*MYBL2, CCNB1, ANLN, CCNA2, BUB1B, CCNB2, KIAA0101, PLK1, TPX2, BUB1, CDCA7, EXO1, ETV4, FOXM1, HIST1H4K, RRM2, CENPF, UBE2T, CRABP2, MKI67, KIF4A, KIF20A, CDK1, CDC45, TOP2A, DEPDC1, HS6ST2*
GO:0071822	Protein complex subunit organization	0.0163	*MMP11, MYBL2, ECT2, CCNB1, KIF11, CENPE, ANLN, GREM1, COL3A1, RRM2, CENPF, ASPM, TTK, STIL, KIF4A, SLC2A1*
GO:0051093	Negative regulation of developmental process	0.0169	*MMP11, CEACAM5, SULF1, CDH3, TNFRSF21, GREM1, COL3A1, ETV4, FOXM1, HIST1H4K, ASPM, SPP1*
GO:0001939	Female pronucleus	0.0171	*CCNA2, SLC2A1*
GO:0048518	Positive regulation of biological process	0.0176	*FAP, MYBL2, TOX3, ECT2, DLGAP5, CCNB1, SULF1, CDH3, CENPE, CCNA2, PDK1, CXCL13, TNFRSF21, MELK, PLK1, GREM1, TPX2, BUB1, COL3A1, ETV4, MMP1, CTHRC1, FOXM1, ASPM, RALGPS2, AIM2, CRABP2, TTK, STIL, SERINC2, SPP1, CDK1, CDC45, DEPDC1, CFB*
GO:0006997	Nucleus organization	0.0187	*CCNB1, CCNB2, PLK1, GOLM1, CDK1*
GO:0008152	Metabolic process	0.0187	*FAP, MMP11, CYP24A1, MYBL2, TOX3, RAD51AP1, HELLS, COL10A1, DLGAP5, TCN1, SULF1, KIF11, CENPE, PDK1, RPL39L, TNFRSF21, MELK, KIAA0101, PLK1, PBK, BUB1, COL3A1, CDCA7, ATP10B, ETV4, AK4, MMP1, PYCR1, PPAP2C, CDKN3, FOXM1, HIST1H4K, AKR1B10, CENPF, UBE2T, CRABP2, MKI67, TTK, SLC44A5, SERINC2, KIF4A, PSAT1, GPX2, FUT2, HMMR, KIF20A, CDK1, KDELR3, CDC45, TOP2A, DEPDC1, SLC2A1, CFB, TDO2*
GO:0050794	Regulation of cellular process	0.0189	*FAP, MMP11, CYP24A1, MYBL2, TOX3, CEACAM5, RAD51AP1, ECT2, HELLS, DLGAP5, CCNB1, SULF1, MMP7, KIF11, GPR87, CDH3, CENPE, ANLN, NUF2, CCNA2, PDK1, CXCL13, CCNB2, TNFRSF21, KIAA0101, PLK1, GREM1, TPX2, PBK, COL3A1, CST1, CDCA7, TYMS, MMP1, ITGA11, PYCR1, CTHRC1, CDKN3, FOXM1, HIST1H4K, AKR1B10, RRM2, CENPF, ASPM, RALGPS2, AIM2, TTK, STIL, SPP1, KDELR3, CDC45, SLC2A1*

Accession number	Term name	FDR value	Genes
GO:0043933	Macromolecular complex subunit organization	0.0193	*MMP11, MYBL2, TOX3, ECT2, HELLS, CCNB1, KIF11, CENPE, ANLN, GREM1, COL3A1, HIST1H4K, RRM2, CENPF, ASPM, TTK, STIL, KIF4A, CDC45, SLC2A1*
GO:0043168	Anion binding	0.0201	*CEACAM5, HELLS, MMP7, KIF11, CENPE, PDK1, CXCL13, BUB1B, MELK, PLK1, TPX2, PBK, BUB1, ATP10B, TYMS, AK4, UBE2T, CRABP2, MKI67, TTK, KIF4A, PSAT1, KIF20A, CDK1, TOP2A, TDO2*
GO:0005524	ATP binding	0.0201	*HELLS, KIF11, CENPE, PDK1, BUB1B, MELK, PLK1, TPX2, PBK, BUB1, ATP10B, AK4, UBE2T, MKI67, TTK, KIF4A, KIF20A, CDK1, TOP2A*
GO:0003824	Catalytic activity	0.0203	*FAP, MMP11, HELLS, DLGAP5, CCNB1, SULF1, MMP7, KIF11, CENPE, PDK1, BUB1B, MELK, PLK1, PBK, BUB1, EXO1, ATP10B, TYMS, AK4, MMP1, PYCR1, PPAP2C, CDKN3, HIST1H4K, AKR1B10, RRM2, UBE2T, TTK, KIF4A, PSAT1, GPX2, FUT2, KIF20A, CDK1, CDC45, TOP2A, CFB, HS6ST2, TDO2*
GO:0097367	Carbohydrate derivative binding	0.0203	*CEACAM5, HELLS, MMP7, KIF11, CENPE, PDK1, CXCL13, BUB1B, MELK, PLK1, TPX2, PBK, BUB1, ATP10B, AK4, UBE2T, MKI67, TTK, KIF4A, HMMR, KIF20A, CDK1, TOP2A*
GO:0042127	Regulation of cell proliferation	0.0212	*FAP, CCNB1, SULF1, MMP7, CDH3, CCNA2, CXCL13, TNFRSF21, GREM1, CDCA7, ETV4, CTHRC1, CDKN3, FOXM1, ASPM, TTK*
GO:0044712	Single-organism catabolic process	0.0212	*MMP11, CYP24A1, COL10A1, MMP7, BUB1B, PLK1, COL3A1, MMP1, AKR1B10, FUT2, HMMR, TDO2*
GO:0044767	Single-organism developmental process	0.0212	*FAP, MMP11, CYP24A1, HELLS, COL10A1, CCNB1, SULF1, MMP7, CDH3, CENPE, ANLN, CCNA2, CXCL13, CCNB2, TNFRSF21, MELK, GREM1, EXO1, TYMS, AK4, ITGA11, CTHRC1, CENPF, ASPM, CRABP2, MKI67, STIL, KIF4A, GJB2, SPP1, CDK1, GCNT3, TOP2A*
GO:0010631	Epithelial cell migration	0.0212	*FAP, ANLN, CXCL13, GREM1*
GO:0048513	Organ development	0.0214	*CCNB1, SULF1, CDH3, ANLN, CCNA2, CXCL13, CCNB2, MELK, GREM1, TYMS, ETV4, AK4, ITGA11, CTHRC1, FOXM1, CENPF, ASPM, MKI67, STIL, GJB2, SPP1, GCNT3, TOP2A*

Accession number	Term name	FDR value	Genes
GO:0065003	Macromolecular complex assembly	0.0223	*MYBL2, ECT2, HELLS, CCNB1, KIF11, CENPE, ANLN, GREM1, RRM2, CENPF, ASPM, KIF4A, CDC45, SLC2A1*
GO:0090132	Epithelium migration	0.0227	*FAP, ANLN, CXCL13, GREM1*
GO:0065009	Regulation of molecular function	0.0228	*ECT2, CCNB1, GPR87, CDH3, CENPE, CCNA2, CXCL13, BUB1B, PLK1, GREM1, TPX2, BUB1, CST1, CTHRC1, CDKN3, CENPF, RALGPS2, AIM2, TTK, STIL, SERINC2, CDK1, DEPDC1*
GO:0001932	Regulation of protein phosphorylation	0.0228	*ECT2, CCNB1, CENPE, CCNA2, SPINK1, PLK1, GREM1, TPX2, PBK, CDKN3, FOXM1, TTK, STIL, CDK1*
GO:0002376	Immune system process	0.0234	*HELLS, MMP7, KIF11, CENPE, ANLN, LCN2, CXCL13, CCNB2, TNFRSF21, MELK, EXO1, MMP1, AIM2, KIF4A, SPP1, CDK1, GCNT3, TOP2A, CFB*
GO:0000070	Mitotic sister chromatid segregation	0.0234	*DLGAP5, CCNB1, CENPE, PLK1*
GO:0045787	Positive regulation of cell cycle	0.0282	*FAP, ECT2, DLGAP5, CCNB1, PLK1, STIL, CDC45*
GO:0051174	Regulation of phosphorus metabolic process	0.0286	*ECT2, CCNB1, GPR87, CENPE, CCNA2, PDK1, SPINK1, PLK1, GREM1, TPX2, PBK, CDKN3, FOXM1, TTK, STIL, CDK1*
GO:0071704	Organic substance metabolic process	0.0294	*FAP, MMP11, CYP24A1, MYBL2, TOX3, RAD51AP1, HELLS, COL10A1, DLGAP5, TCN1, SULF1, PDK1, RPL39L, TNFRSF21, MELK, KIAA0101, PLK1, PBK, BUB1, COL3A1, CDCA7, ETV4, AK4, MMP1, PYCR1, PPAP2C, CDKN3, FOXM1, HIST1H4K, AKR1B10, CENPF, UBE2T, CRABP2, MKI67, TTK, SLC44A5, SERINC2, PSAT1, GPX2, FUT2, HMMR, CDK1, GCNT3, KDELR3, CDC45, TOP2A, DEPDC1, SLC2A1, CFB, TDO2*
GO:0034502	Protein localization to chromosome	0.0304	*BUB1B, PLK1, CDK1*
GO:2000026	Regulation of multicellular organismal development	0.0304	*ECT2, CCNB1, SULF1, CDH3, CXCL13, TNFRSF21, GREM1, COL3A1, ETV4, CTHRC1, HIST1H4K, CENPF, ASPM, CRABP2, SPP1, CDK1*
GO:0031572	G2 DNA damage checkpoint	0.0304	*CCNA2, PLK1, CDK1*

Accession number	Term name	FDR value	Genes
GO:0044699	Single-organism process	0.0305	*FAP, CYP24A1, TOX3, CEACAM5, RAD51AP1, HELLS, COL10A1, DLGAP5, TNS4, CCNB1, TCN1, SULF1, GPR87, CENPE, NUF2, CCNA2, LCN2, PDK1, CXCL13, RPL39L, TNFRSF21, STEAP1, KIAA0101, GREM1, TPX2, PBK, CST1, CDCA7, ATP10B, AK4, MMP1, ITGA11, PYCR1, PPAP2C, CTHRC1, CDKN3, HIST1H4K, AKR1B10, CENPF, UBE2T, ASPM, RALGPS2, AIM2, SLC44A5, CEP55, STIL, KIF4A, PSAT1, GPX2, FUT2, HMMR, KIF20A, SPP1, CDC45, DEPDC1, CFB, TDO2*
GO:0007091	Metaphase/anaphase transition of mitotic cell cycle	0.0317	*BUB1B, PLK1*
GO:0044238	Primary metabolic process	0.0317	*FAP, MMP11, CYP24A1, MYBL2, TOX3, RAD51AP1, HELLS, DLGAP5, SULF1, MMP7, PDK1, RPL39L, TNFRSF21, MELK, KIAA0101, PLK1, PBK, BUB1, COL3A1, CDCA7, ETV4, AK4, MMP1, PYCR1, PPAP2C, CDKN3, FOXM1, HIST1H4K, AKR1B10, CENPF, UBE2T, CRABP2, MKI67, TTK, SLC44A5, SERINC2, PSAT1, GPX2, FUT2, HMMR, CDK1, GCNT3, KDELR3, CDC45, TOP2A, DEPDC1, SLC2A1, CFB, TDO2*
GO:0007098	Centrosome cycle	0.0354	*KIF11, STIL, CDK1*
GO:0005874	Microtubule	0.0357	*KIF11, CENPE, TPX2, ASPM, KIF4A, KIF20A, CDK1*
GO:0033043	Regulation of organelle organization	0.0367	*ECT2, DLGAP5, CCNB1, KIF11, CENPE, ANLN, PDK1, PLK1, TPX2, BUB1, CENPF, TTK, CDC45*
GO:0044424	Intracellular part	0.0369	*FAP, MMP11, CYP24A1, TOX3, RAD51AP1, ECT2, HELLS, DLGAP5, TNS4, CCNB1, SULF1, KIF11, CDH3, CENPE, ANLN, NUF2, CXCL13, BUB1B, CCNB2, RPL39L, STEAP1, MELK, KIAA0101, PLK1, TPX2, PBK, BUB1, COL3A1, CKAP2L, CDCA7, EXO1, ATP10B, ETV4, AK4, PYCR1, CTHRC1, CDKN3, FOXM1, HIST1H4K, AKR1B10, RRM2, UBE2T, ASPM, RALGPS2, AIM2, CRABP2, TTK, CEP55, STIL, KIF4A, PSAT1, GJB2, GOLM1, GPX2, FUT2, HMMR, KIF20A, SPP1, CDK1, GCNT3, KDELR3, CDC45, DEPDC1, HS6ST2, TDO2*
GO:0007010	Cytoskeleton organization	0.0373	*MYBL2, CENPE, ANLN, KIAA0101, PLK1, ASPM, TTK, STIL, KIF4A, KIF20A, CDK1*

Accession number	Term name	FDR value	Genes
GO:0051321	Meiotic cell cycle	0.0375	*PLK1, EXO1, ASPM, MKI67, TOP2A*
GO:0070271	Protein complex biogenesis	0.0383	*MYBL2, ECT2, CCNB1, KIF11, CENPE, ANLN, GREM1, RRM2, CENPF, ASPM, KIF4A, SLC2A1*
GO:0006461	Protein complex assembly	0.0383	*MYBL2, ECT2, CCNB1, KIF11, CENPE, ANLN, GREM1, RRM2, CENPF, ASPM, KIF4A, SLC2A1*
GO:0031570	DNA integrity checkpoint	0.0383	*CCNA2, PLK1, CDK1, CDC45, TOP2A*
GO:0050793	Regulation of developmental process	0.0384	*MMP11, CEACAM5, ECT2, CCNB1, SULF1, CDH3, CXCL13, TNFRSF21, GREM1, COL3A1, ETV4, CTHRC1, FOXM1, HIST1H4K, CENPF, ASPM, CRABP2, SPP1, CDK1*
GO:0031399	Regulation of protein modification process	0.0398	*ECT2, CENPE, CCNA2, BUB1B, SPINK1, PLK1, GREM1, TPX2, PBK, BUB1, CDKN3, FOXM1, CENPF, TTK, STIL, CDK1*
GO:0051310	Metaphase plate congression	0.0398	*CCNB1, CENPE, CENPF*
GO:0051347	Positive regulation of transferase activity	0.0413	*ECT2, CCNB1, CENPE, PLK1, GREM1, TPX2, STIL, SERINC2, CDK1*
GO:0032155	Cell division site part	0.0415	*ECT2, ANLN, CEP55*
GO:0032153	Cell division site	0.0415	*ECT2, ANLN, CEP55*
GO:0044444	Cytoplasmic part	0.0415	*MMP11, CYP24A1, ECT2, CCNB1, SULF1, KIF11, CENPE, ANLN, NUF2, CXCL13, BUB1B, CCNB2, RPL39L, STEAP1, MELK, KIAA0101, PLK1, BUB1, COL3A1, ATP10B, TYMS, AK4, PYCR1, CDKN3, AKR1B10, RRM2, CENPF, AIM2, CRABP2, CEP55, STIL, KIF4A, PSAT1, GJB2, GOLM1, GPX2, FUT2, KIF20A, SPP1, CDK1, GCNT3, KDELR3, HS6ST2, TDO2*
GO:0044237	Cellular metabolic process	0.0425	*FAP, CYP24A1, MYBL2, TOX3, RAD51AP1, HELLS, DLGAP5, TCN1, SULF1, MMP7, PDK1, RPL39L, TNFRSF21, MELK, KIAA0101, PLK1, PBK, BUB1, COL3A1, CDCA7, ETV4, AK4, MMP1, PYCR1, PPAP2C, CDKN3, FOXM1, HIST1H4K, AKR1B10, CENPF, UBE2T, CRABP2, MKI67, TTK, SLC44A5, SERINC2, PSAT1, GPX2, FUT2, CDK1, GCNT3, KDELR3, CDC45, TOP2A, DEPDC1, SLC2A1, HS6ST2, TDO2*

Accession number	Term name	FDR value	Genes
GO:0000079	Regulation of cyclin-dependent protein serine/threonine kinase activity	0.0435	*CCNA2, PLK1, CDKN3, STIL*
GO:0030199	Collagen fibril organization	0.0456	*MMP11, GREM1, COL3A1*
GO:0090068	Positive regulation of cell cycle process	0.0456	*FAP, ECT2, DLGAP5, CCNB1, PLK1, CDC45*
GO:0032502	Developmental process	0.0456	*FAP, MMP11, CYP24A1, HELLS, COL10A1, CCNB1, SULF1, MMP7, CDH3, CENPE, ANLN, CCNA2, CXCL13, CCNB2, TNFRSF21, MELK, GREM1, EXO1, TYMS, AK4, ITGA11, CTHRC1, CENPF, ASPM, MKI67, STIL, KIF4A, GJB2, SPP1, CDK1, GCNT3, TOP2A*
GO:0045861	Negative regulation of proteolysis	0.0458	*BUB1B, PLK1, PBK, BUB1, CST1, CENPF, TTK*
GO:0045842	Positive regulation of mitotic metaphase/anaphase transition	0.0480	*DLGAP5, PLK1*

Acknowledgments

Ich möchte mich hiermit ganz besonders bei Frau Prof. Dr. Dr. Melanie Königshoff bedanken für die Chance eine experimentelle Doktorarbeit unter ihrer Betreuung durchzuführen. Ich bin sehr dankbar für ihre fortwährende Unterstützung in den vergangenen Jahren, die vielen Treffen, Diskussionen, die konstruktive Kritik und Motivation, welche maßgeblich zum Erfolg dieser Arbeit beigetragen haben.

Herrn Prof. Dr. Oliver Eickelberg, als ehemaligem Leiter des Comprehensive Pneumology Centers des Helmholtz Zentrums München, möchte ich danken für die Möglichkeit diese Promotion an seinem Institut realisiert zu haben.

Meiner direkten Betreuerin Frau Dr. Kathrin Mutze möchte ich ganz besonders danken für ihre ständige Hilfsbereitschaft, die exzellente Betreuung, die unzähligen Gespräche, ihre Ratschläge und Ideen.

Frau Dr. Katharina Heinzelmann möchte ich herzlich danken für ihre Unterstützung in der finalen Phase dieser Arbeit sowie das Korrekturlesen des Manuskripts.

Frau Dr. Darcy Wagner möchte ich danken für die Einführung in die Bioinformatik, ihre konstruktive Kritik und die zahlreichen Ratschläge.

Maria Magdalena Stein möchte ich danken für die ausgezeichnete Einarbeitung, die Einführung in grundlegende Arbeitstechniken der experimentellen Forschung und ihre unermüdliche Hilfsbereitschaft. Julia Kipp und Anastasia van den Berg danke ich für die Zellisolationen und technische Unterstützung.

Allen weiteren Mitarbeitern der MK-Arbeitsgruppe, insbesondere Cedric Thiel, Lara Buhl, Stephan Klee, Wioletta Skronska-Wasek, Aina Martin-Medina, Rita Costa, Florian Ciolek, Nadine Adam und Hoeke Baarsma danke ich für eine großartige Gemeinschaft im Labor und so manches Gespräch.

Mein größter Dank gebührt meiner Familie, die mich in allen Phasen meines Lebens unterstützt und begleitet hat. Abschließend möchte ich von ganzem Herzen meiner Frau Daniela danken für ihre Geduld, ihr Verständnis und die regelmäßige Motivation während meines Studiums und dieser Doktorarbeit.

Publications and presentations

1 Publications

Ulke, H. M., Mutze, K., Lehmann, M., Wagner, D. E., Heinzelmann, K., Günther, A., Eickelberg, O., and Königshoff, M.; The Oncogene ECT2 Contributes to a Hyperplastic, Proliferative Lung Epithelial Cell Phenotype in Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory Cell and Molecular Biology.* 61(6), 713-726 (2019) [120].

Heinzelmann, K., Hu, Q., Hu, Y., Dobrinskikh, E., Ansari, M., Melo-Narváez, M. C., Ulke, H. M., Leavitt, C., Mirita, C., Trudeau, T., Saal, M. L., Rice, P., Gao, B., Janssen, W. J., Yang, I. V., Schiller, H. B., Vladar, E. K., Lehmann, M., and Königshoff, M.; Single-cell RNA sequencing identifies G-protein coupled receptor 87 as a basal cell marker expressed in distal honeycomb cysts in idiopathic pulmonary fibrosis. *European Respiratory Journal.* 59(6), 2102373 (2022) [178].

2 Presentations

Poster presentation: Ulke, H. M., Mutze, K., Wagner, D. E., Stein, M. M., Lindner, M., Behr, J., Günther, A., and Königshoff, M.; *Non-Small Cell Lung Cancer (NSCLC) Gene Signature in Idiopathic Pulmonary Fibrosis (IPF)*; American Thoracic Society (ATS), May 2017, Washington, DC, USA.

Poster presentation: Ulke, H. M., Mutze, K., Lehmann, M., Wagner, D. E., Stein, M. M., Lindner, M., Behr, J., Günther, A., and Königshoff, M.; *A Non-Small Cell Lung Cancer (NSCLC) Associated Gene Set is Enriched in Human IPF Indicating Common Pathomechanisms in Both Diseases*; 15th ERS Lung Science Conference, March 2017, Estoril, Portugal.

Oral and poster presentation: Ulke, H. M., Mutze, K., Wagner, D. E., Stein, M. M., Kipp, J., Lindner, M., Behr, J., Günther, A., and Königshoff, M.; *Idiopathische Lungenfibrose und Lungenkrebs teilen ein signifikant überlappendes Muster an dysregulierten Genen*; 5th AIR Symposium Germany, February 2016, Mainz, Germany.

Oral presentation: Mutze, K., Ulke, H. M., Wagner, D. E., Eickelberg, O., and Königshoff, M.; *Oncogene ECT2 is concurrently upregulated in Idiopathic Pulmonary Fibrosis (IPF) and Non-Small Cell Lung Cancer (NSCLC) and contributes to enhanced proliferative capacity of hyperplastic epithelial cells*; Munich Pittsburgh Lung Conference, October 2015, Munich, Germany.