In vitro modelling of respiratory infections in idiopathic pulmonary fibrosis (IPF) using human induced pluripotent stem cell (hiPSCs)-derived alveolar epithelial type II cells

Submitted by:

Ana Lilia Serna Valverde

June 2023

School of Medicine
Biodiscovery Institute
The University of Nottingham

Principal supervisor: Dr Nicholas R. F. Hannan
Secondary supervisors: Dr Amanda Tatler and Dr Sara Cuevas Ocaña

Thesis submitted to The University of Nottingham for the Degree of Doctor in Philosophy

Acknowledgements

First, I would like to thank Dr Nicholas R.F. Hannan, my principal supervisor, for his dedication and unconditional support throughout my PhD journey. His invaluable insight, positivity, and guidance were vital in sustaining my confidence in my abilities and the project, particularly during the challenging times posed by the pandemic. Additionally, I would like to express my gratitude to Dr Amanda Tatler, my secondary supervisor, as well as Professor Gisli Jenkins and Dr Chitra Joseph for their scientific expertise and training. Their support was essential in making the viral work of this thesis possible. I would also like to thank Dr Sara Cuevas Ocaña for the gene-editing work and for giving me that extra push whenever I needed it, especially in finding my scientific voice.

I would like to thank the staff and technicians of the Biodiscovery Institute who provided me with help and support and made the work possible, especially Paula Olsen and Craig Dicken, who always gave the extra mile to help students. I would also like to acknowledge the members of the flow cytometry facility Dr David Onion, Dr Nicola Croxall, and Maria Haig, for their help and patience on the long sorting days.

To the past and current members of the Hannan Lab, who made this journey and the long days more enjoyable. Big thank you to Dr Peggy Cho Kiu Lo for giving me the entry point into the lab and the hours of training that were the foundation for my project. I would like to thank Carlos Sainz Zuñiga for stepping in to analyse the RNAseq data of this thesis with the patience to guide me through the process. A special thanks to my partner in crime, Liam Reed, the first to embark with me on the mesmerising journey of lung differentiations and who celebrated every achievement or suffered every frustration with me; never a dull day.

Thank you to those friends I made along the way, who unexpectedly became one of the best parts of this journey: Paulina Durczak, Liam Reed, Peggy Lo, Kate Vo, Emma Collins, Maria Eleftheriou, Sara Bodbin, and Sara Pijuan-Galito. My moments with them will always be treasured, and I will be forever grateful for their friendship. I would also like to thank my Mexican best friends on both sides of the world, whose spark was always present and helped me take a breath to enjoy the journey. I especially thank the family I made in the UK, my housemates, Moni, Mayra, Miros, Karo and Angel. I am fortunate to be surrounded by such a loving group of people. To Jane, for keeping me sane; I could not have made it through this journey without her.

To all my family, I would like to say a big thank you for their unconditional love and support. Especially my parents, Ricardo and Mireya, for giving me every opportunity imaginable and encouraging me to be the best version of myself. To my sister, Yeya, for being endlessly supportive while helping me keep perspective, no matter what, and of course, for giving me the joy of being an aunt to Victoria and Valentina, even with the distance. Lastly, to my Abue Raul and Abue Licha, whom I constantly think of making proud, I know how meaningful this milestone would have been to him.

Declaration

I hereby declare that this thesis has been composed by myself and has not been submitted for any other degree previously. This project was supervised by Dr Nicholas R.R. Hannan, Dr Amanda Tatler and Dr Sara Cuevas Ocaña. Acknowledgements of specific procedures not performed by myself are stated; otherwise, the work described is my own.

Signature:

Ana Lilia Serna Valverde

Date:

30 June 2023

Abstract

Idiopathic pulmonary fibrosis (IPF) is a chronic, interstitial lung disease estimated to affect 4.5 individuals per 10,000[1]. It is characterised by the accumulation of extracellular matrix in the lung interstitium, which disrupts the normal architecture and impairs lung function resulting in severe respiratory failure and death. The rapid progression of IPF and the lack of curative treatments is often accompanied by a median survival of 2-3 years [2]. Its aetiology is unclear; however, environmental and genetic risk factors indicate that alveolar epithelial type 2 cells (AT2s), the facultative progenitor of the alveoli, play a pivotal role [3, 4]. Mutations in surfactant protein (SFTPC) genes compromise AT2s function and viability, reducing its capacity to regenerate damaged alveolar epithelial tissue. This function is further compromised by infection-induced lung injury. Integrating such environmental stimuli and genetic predisposition in IPF patients results in an aberrant fibrotic response driven by AT2s [5, 6].

The study of this host-pathogen interaction has been hampered by the lack of adequate human IPF models.[6]. Therefore, the differentiation of patient-derived human induced pluripotent stem cells (hiPSC) into alveolar organoids represents a valuable tool for studying genotype-phenotype relationships of IPF in vitro. For this, an organoid platform using IPF patient-derived hiPSCs carrying a BRICHOS-domain SFTPC mutation was generated. Furthermore, using base editing techniques, the platform was also adapted to be used with corrected isogenic hiPSCs [7]. A thoroughly optimised directed differentiation protocol was generated to differentiate lung progenitors in monolayer cultures. These cells were further enriched and transferred to 3D Matrigel droplets in maturation media to obtain a population of AT2 cells expressing SFTPC that was further enriched using a surrogate surface marker [8]. Transcriptome profiling of differentiating cells at several stages of the protocol confirmed the expression of key markers at the definitive endoderm, lung progenitor, and alveolar stages of the differentiation.

Finally, the SFTPC-mutant and control AT2 cells were enriched and exposed to H1N1 influenza A virus to evaluate their capacity to mount a response to immune stimuli and serve as an in vitro model for human respiratory infections. The infected and uninfected samples were analysed using RNA sequencing. Upon infection, it was found that AT2 cells induced an inflammatory and antiviral response led by the activation of interferon signalling pathways. Furthermore, overrepresented processes in mutant IPF AT2 cells included “extracellular matrix organisation", "extracellular matrix disassembly", and

"collagen catabolic process" when compared with the wild-type cells. The analysis also revealed differentially expressed genes linked to IPF, such as MMP7, IL1B, MUC5B and TERC [9–11]. These data validate IPF-patient-derived AT2 cells as a robust in vitro model to better understand the connection between genotype and environmental interactions in IPF disease progression.

Table of contents

List of figures

List of tables

Abbreviations

Abbreviation	Description
%	Percentage
°C	Degree Celcius
AA2P	L-Ascorbic Acid 2-phosphate
ABCA3	ATP Binding Cassette Subfamily A Member 3
ActA	Activin A
AE	Acute Exacerbations
AFE	Anterior Foregut Endoderm
AFP	Alpha Foetal Protein
ALB	Albumin
ANOVA	Analysis of Variance
AQP5	Aquaporin 5
AT2	Alveolar epithelial type 2 cell
ATF4	Activating Transcription Factor 4
ATF6	Activating Transcription Factor 6
ATP	Adenosine triphosphate
BADJ	Bronchioalveolar Duct Junction
BAL	Bronchoalveolar lavage
BASC	Bronchioalveolar stem cells
BiP	Binding Immunoglobulin Protein
BMP	Bone morphogenic protein
BMP4	Bone Morphogenetic Protein 4
bp	Base Pairs
BPE	bovine pituitary extract
BSA	Bovine Serum Albumin
cAMP	Cyclic adenosine monophosphate
Cas9	CRISPR Associated Protein 9
cDNA	Complementary Deoxyribonucleic acid
CDX2	Caudal type homeobox 2
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CHIR	CHIR99021
cKIT	Tyrosine-protein Kinase Kit
CMV	Cytomegalovirus
COL1A1	Collagen Type I Alpha 1 Chain
COL3A1	Collagen Type III Alpha 1 Chain
COPD	Chronic obstructive pulmonary disease
CPE	Cytopathic Effect
CPM	Carboxypeptidase-M
CRISPR	Clustered regularly interspaced short palindromic repeats
CXCR4	C-X-C chemokine receptor type 4
DAPI	Di-Amino Phenyl-Indole
DE	Definitive Endoderm
DEGs	Diferentilally Expressed Genes
DEXA	Dexamethasone
dgDNA	Direct Genomic DNA

DMEM	Dulbecco's (Modified Eagle's) Minimal Essential Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide Triphosphate
DPBS	Dulbecco's Phosphate Buffered Saline
DPPC	Dipalmitoylphosphatidylcholine
DS	Dorsomorphin
DSB	Double-Stranded Break
DTT	Dithiothreitol
E8	Essential 8 Medium
EBs	Embroyid bodies
EBV	Epstein-Barr virus
ECAD	E-cadherin
ECM	Extracellular matrix
EDEM	ER-degradation-enhancing-α-manidose like protein
EDTA	Ethylene Diamine Tetra-Acetic Acid
EGF	Epidermal Growth Factor
ELISA	Enzyme-Linked Immunosorbent Assay
EMEM	Eagle's minimum essential medium
EMT	Epithelial-Mesenchymal Transition
eNOS	Endothelial Nitric Oxide Synthase
EpCAM	Epithelial Cellular Adhesion Molecule
ER	Endoplasmic Reticulum
ERAD	ER-associated degradation
ESCs	Embryionic stem cells
FACS	Fluorescence-Activated Cell Sorting
FBS	Foetal bovine serum
FDR	False Discovery Rate
FG	Foregut
FGF	Fibroblast Growth Factor
FOXA2	Forkhead Box Protein A2
FOXJ1	Forkhead Box J1
GCDR	Gentle Cell Dissociation Reagent
gDNA	Genomic DNA
GFP	Green fluorescent protein
GO	Gene Ontology
gRNA	Guide RNA
GRP94	glucose-regulated protein 94
GSK3β	Glycogen Synthase Kinase-3 β
GWAS	Genome-Wide Association Study
HA	Haemagglutinin
HB E8	Homebrew Essential 8
HBSS	Hanks' balanced salt solution
HDF	Human Dermal Fibroblast
hESC	Human embryonic stem cell
HHV	Human herpes viruses
hiPSC	Human induced pluripotent stem cell

HNF4α	hepatocyte Nuclear Factor 4 α
hPCLS	human precision-cut lung slices
IBMX	3-isobutyl-1-methylxanthine
ICC	Immunocytochemistry
IgG	Immunoglobulin G
iHBECs	Immortalised Human Bronchial Epithelial Cells
ILD	Interstitial lung disease
IMDM	Iscove's Modified Dulbecco's Medium
INS	Insulin
IPF	Idiopathic Pulmonary Fibrosis
IRE-1	Inositol-requiring Protein 1
IRES	Internal Ribosome Entry Site
ITS	Insulin-Transferrin-Selenium
kb	Kilobases
KGF	Fibroblast Growth Factor 7
KRT5	Keratin 5
KSFM	Keratinocyte serum-free medium
LAMA1	Laminin Subunit Alpha 1
LAP	Latency Associated Peptide
LB	Lysogeny Broth
LPA	Lysophosphatidic acid
LY	LY294002
MDCKs	Madin Darby Canine Kidney Cells
MEM	Eagle's Minimum Essential Medium
mESC	Mouse embryonic stem cell
MFI	Mean fluorescence intensity
MIXL1	Mix Paired-Like Homeobox
MOI	Multiplicity of Infection
mRNA	Messenger RNA
MT	Matrigel
MTG	α-Monothioglycerol
MUC5AC	Mucin 5B
MUC5B	Mucin 5AC
MUT	Mutant
MVB	Multi vesicular body
NaPi2b	Sodium-dependent phosphate transport protein 2b
NEAA	Non-Essential Amino Acids
NFkB	Nuclear Factor Kappa B Subunit 1
NFW	Nuclease-Free Water
NKX2.1	NK2 Homeobox 1
NR	Neutral Red
NTC	No Template Control
OCT4	Octamer-binding Transcription Factor 4
OWB	Organoid Wash Buffer
PAI-1	Plasminogen Activator Inhibitor- 1
PBGD	Porphobilinogen Deaminase
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis

PCR	Polymerase Chain Reaction
PDGF	Platelet-derived growth factor
PDX1	Pancreatic and Duodenal Homeobox 1
Pen-Strep	Penicilin/Streptomycin
PERK	Protein Kinase RNA-like ER kinase
PG	phosphatidylglycerol
PNECs	pulmonary neuroendocrine cells
Poly I:C	Polyinosinic:polycytidylic acid
qPCR	Quantitative Real-Time Polymerase Chain Reaction
RA	Retinoic Acid
RBCs	Red blood cells
RLU	Relative Luminescence Units
RNA	Ribonucleic acid
RNA-Seq	Ribonucleic acid sequencing
ROS	Reactive Oxygen Species
Rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute Medium
RT	Room temperature
SB	SB431542
SCGB1A1	Secretoglobin Family 1A Member 1
SCGB3A2	Secretoglobin Family 3A Member 2
SD	Standard Deviation of The Mean
SEM	Standard error of the mean
SeV	Sendai-virus
SFTPA	Surfactant Protein A
SFTPB	Surfactant Protein B
SFTPC	Surfactant Protein C
SFTPD	Surfactant Protein D
SHH	Sonic Hedgehog
SLC	Small Latent Complex
SLC34A2	Solute Carrier Family 34 Member 2/ NaPi2b encoding gene
SM	Small Molecule
SNP	Single Nucleotide Polymorphism
SOX17	SRY-Box Transcription Factor 17
SOX2	SRY-Box Transcription Factor 2
SOX9	SRY-Box Transcription Factor 9
TCID	Tissue Culture Infective Dose
TERC	Telomerase RNA Component
TERT	Telomerase Reverse Transcriptase
TGF-β	Transforming Growth Factor Beta
TLR	Toll-Like Receptor
TLR4	Toll-Like Receptor 4
TMLCs	Transformed Mink Lung Epithelial Cells
TNF-α	Tumour Necrosis Factor Alpha
TPCK	L-1-tosylamido-2-phenylethyl chloromethyl ketone
TRP63	Tumor Protein P63
TTR	Transthyretin
TUBB3	Tubulin Beta 3 Class III

UIP	Usual interstitial pneumonia
UPR	Unfolded Protein Response
WFI	Water-For-Injection for Cell Culture
WNT	Wingless/INT protein
WT	Wild type
XBP1	X-box binding protein 1
Y-27632	Rho-Associated Kinase Inhibitor (ROCKi)
YAP1	Yes1 Associated Transcriptional Regulator

Chapter 1 Introduction

1.1 Aims of the thesis

This thesis aimed to develop an in vitro model of Idiopathic Pulmonary Fibrosis (IPF) using alveolar type 2 (AT2) cells derived from mutant human induced pluripotent stem cells (hiPSCs). These cells would act as a platform to gain a better understanding of the role of genetic risk factors and infections in lung fibrogenesis affecting IPF patients.

The hypotheses of this work were:

I. Optimised direct differentiation protocol generates mature patient derived AT2 cells providing a reliable source for disease modelling.
II. AT2 cells from gene-edited mutant hiPSCs present higher levels of ER stress and inflammation.
III. Influenza infection-induced response is exaggerated in mutant AT2 cells.

The work performed in this thesis is contextualised throughout this introduction chapter. The second chapter details the methodology employed to obtain the results discussed in Chapters 3, 4, and 5. Chapter 6 is dedicated to discussing the findings within the context of the current literature.

Chapter 3 outlines human skin fibroblasts' derivation and reprogramming to hiPSCs from IPF donors with different heterozygous mutations in surfactant proteins A and C. The generated hiPSC lines are then subjected to an established in vitro differentiation protocol to produce lung progenitor cells to identify the cell line with the best lung differentiation potential [12]. This chapter further highlights the genetic modification of the selected hiPSC line ($SFTPC^{WT/MUT}$) using base editing technology, resulting in the creation of a surfactant protein C homozygous mutant cell line ($SFTPC^{MUT/MUT}$) alongside a homozygous corrected counterpart ($SFTPC^{WT/WT}$) to serve as a healthy control.

Chapter 4 is dedicated to evaluating a series of strategies aimed at enhancing the efficiency of the alveolar differentiation protocol using IPF-patient-derived hiPSC lines. This optimisation process begins with refining the definitive endoderm (DE) stage to yield higher quantities of DE. An additional step involving anterior foregut ventralisation is incorporated before generating and enriching the lung progenitor cells. The chapter also presents the adaptation of the differentiation platform into 3D organoids to induce

the differentiation and maturation of AT2 cells, providing a comprehensive characterisation. Finally, the chapter portrays an additional optimisation strategy designed to increase the yields of lung progenitors obtained with the differentiation platform.

The final results chapter, Chapter 5, establishes the in vitro infection model of hiPSC-derived AT2 cells. As a proof of concept, the AT2 cells are exposed to Pseudomonas aeruginosa LPS to demonstrate the platform's utility for infection modelling. Subsequently, the quantification of influenza A (H1N1) viral particles, a virus relevant to IPF pathogenesis, is conducted in order to infect the AT2 cells generated in this thesis. Following optimisation, the enriched alveolar cells obtained from the SFTPC$^{WT/MUT}$, SFTPC$^{MUT/MUT}$, and SFTPC$^{WT/WT}$ hiPSC lines are infected with the H1N1 virus. The chapter culminates with a global downstream analysis that compares the transcriptomic profiles of infected and uninfected cells, offering insights into the physiological characteristics of mutant and wild-type AT2 cells in response to infections. This could enable the identification of disease-triggering factors as potential therapeutic targets or diagnostic biomarkers.

A representation of the summarised workflow employed to attain the objectives described is illustrated in **Figure 1.1**.

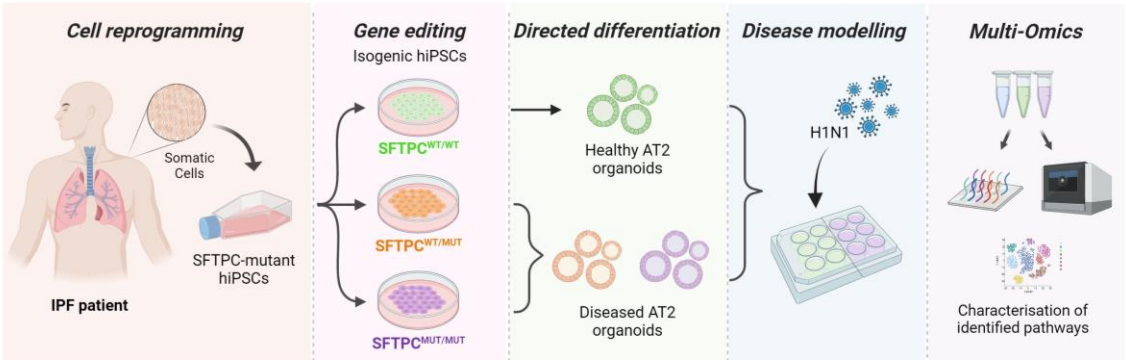


Figure 1.1 PhD workflow

Human skin biopsy was obtained from an IPF patient carrying a heterozygous SFTPC mutation. The derived dermal fibroblasts were then reprogrammed into hiPSCs generating the parental SFTPC$^{WT/MUT}$ cell line. The parental cell line was gene-edited using base editors' technology to generate the homozygous mutant and corrected hiPSC cell lines (SFTPC$^{MUT/MUT}$ and SFTPC$^{WT/WT}$). Through directed differentiation, AT2 cells were grown and matured in 3D organoid cultures. The cells covering the whole spectrum of the SFTPC genotype were enriched and transferred to 2D culture conditions to be infected with H1N1 influenza virus. Downstream analysis was done by comparing transcriptomics of infected and uninfected cells to evaluate the pro-inflammatory and pro-fibrotic pathways impacted by the viral exposure and identify targets for future drug studies. Created with BioRender.com.

1.2 The lung

The respiratory system is responsible for facilitating the exchange of gases between the bloodstream and the external environment. Oxygen is taken up by red blood circulating in the blood vessels, while carbon dioxide diffuses into the gases in the alveoli sacs to be exhaled. To perform this primary function, the lungs have a unique architecture with various anatomical regions composed of remarkably heterogeneous types of cells, including epithelial, vascular, mesenchymal, and immune cells. Each of these cell types plays a crucial role in the proper functioning of their respective compartments. The respiratory tree, from the trachea to the respiratory bronchioles, is lined by secretory, ciliated, neuroendocrine, and basal cells that help condition the inhaled air and maintain the integrity of the airway epithelium. In the alveoli, where the gas exchange occurs, the lining is composed of alveolar epithelial cells that secrete surfactant and regenerate the alveoli or interact closely with the epithelial cells to facilitate efficient gas exchange [13].

1.2.1 Development and Anatomy

Embryological development of the lung

The lung development is a complex process that occurs both in utero and postnatally. It is one of the last organs to form during embryonic development and undergoes maturation long after birth [14]. Its morphogenesis involves a series of sequential changes within individual cells and cell-cell interactions. First, during gastrulation, anterior cells of the primitive streak give rise to the definitive endoderm (DE). The DE then undergoes morphogenesis, forming the embryonic gut tube that is patterned along the anterior-posterior axis forming the foregut and hindgut domains [15]. Around the 4th gestational week, cells in the anterior foregut endoderm (AFE) become specified to the pulmonary epithelial lineage by expressing the transcription factor NK2 Homeobox 1 (NKX2.1) [16]. This marks the beginning of the embryonic stage of lung development (3-7 weeks of gestation). The series of stages comprising lung development are depicted in Figure 1.2 Stages of the development of the human lung.**Figure 1.2.**

During the embryonic stage, the ventral NKX2.1^{+} progenitors evaginate giving rise to a respiratory diverticulum that is connected dorsally to the oesophagus. Prior to the pseudo glandular stage (5-17 weeks of gestation), the diverticulum splits to form the trachea and lung buds to continue with the branching morphogenesis. This stage is characterised by sprouting epithelial tubes that branch into the surrounding lung

mesenchyme until the formation of the terminal bronchioles. The formation of the highly arborised airway tree through branching morphogenesis is a tightly controlled process to ensure the proximal-distal patterning of the lungs [17, 18]. This is followed by the canalicular stage (16-26 weeks of gestation), where the respiratory bronchioles give rise to alveolar ducts. These ducts contain alveolar epithelial cells with lamellar bodies capable of surfactant production, representing primitive alveoli [19]. Additionally, the surrounding capillary network becomes more prominent, indicating the potential for respiration at approximately 25 weeks of gestation. The last in-utero stage is the saccular stage, which occurs from week 26 of gestation until birth, and it is characterised by the development of terminal alveolar sacs separated by primary septae. During the saccular stage, the continuous expansion of the pulmonary parenchyma leads to a decrease in the thickness of the connective tissue between airspaces and promotes the maturation of the surfactant system, laying the foundation for further alveolar development [20]. Upon birth, the alveoli are not fully developed, and the lung is still structurally immature. During the alveolar stage, which spans from birth to approximately the first three years of life, significant increases in gas exchange surface area occur through septation and capillary remodelling. These processes and further alveolar progenitors' maturation contribute to the lungs acquiring their adult structure and morphology [17, 21].

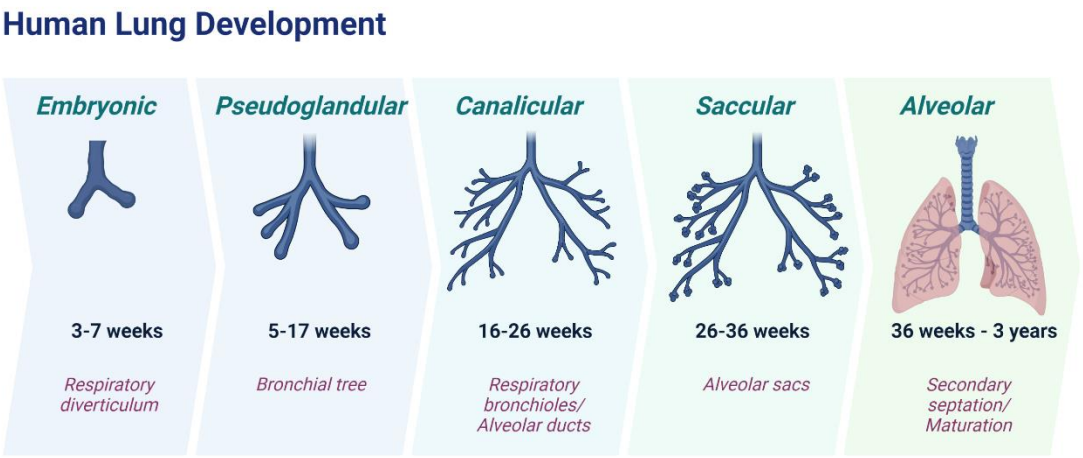


Figure 1.2 Stages of the development of the human lung.

The process of lung development initiates with a group of multipotent NKX2.1^{+} progenitors that proliferate and differentiate through a series of 5 stages to yield more restricted and specialised lung epithelial cells. Adapted from [22]. Created with BioRender.com.

The development of the lung involves several molecular mechanisms that contribute to epithelial differentiation. During early embryonic development, cells from the

proximal-posterior region of the epiblast migrate through the Nodal-expressing primitive streak to form the germ layers. As these cells migrate through the primitive streak, they are exposed to varying levels of Nodal ligands, resulting in a transient mesendoderm fate. Cells that emerge from the anterior end of the primitive streak encounter higher levels of Nodal and adopt a DE state [23]. Nodal activity, through Smad2 phosphorylation, induces the expression of key transcription factors in the endoderm like SRY-Box Transcription Factor 17 (Sox17) and Forkhead Box Protein A2 (Foxa2). Studies using animal models have demonstrated that β-catenin, part of the canonical Wnt signalling pathway, also plays a significant role in specifying and patterning the DE by maintaining elevated levels of Nodal [24].

Following gastrulation, the DE transforms into the primitive gut tube, which exhibits distinct regions along the anterior-posterior axis. Anteriorly, SRY-Box Transcription Factor 2 (Sox2) is expressed in the foregut region, while Pancreatic and Duodenal Homeobox 1 (Pdx1) is expressed in the midgut, and Caudal type homeobox 2 (Cdx2) in the hindgut [23]. During this stage, the mesenchyme surrounding the anterior part of the embryo expresses Wnt ligands which are essential for signalling the specification of lung endoderm progenitors within the ventral anterior foregut endoderm (AFE). Wnt signalling initiates the expression of NKX2.1+ through the activity of β-catenin, which translocates to the nucleus following stimulation, having previously been sequestered in the cytoplasm [16, 25]. Along with Wnt, bone morphogenic protein (BMP) signalling also plays a role in promotes of NKX2.1+ progenitors and limits the expression of SOX2, which was initially present throughout the entire AFE, to the dorsal AFE [26].

Upon the emergence of the respiratory diverticulum from the AFE, the NKX2.1 cells push ventrally and elongate to form the trachea. Proper septation between the trachea and the oesophagus relies on the correct balance of Wnt and BMP activity [27]. The epithelial-mesenchymal interactions play a vital role in promoting the growth and differentiation of lung buds. These interactions are regulated by the expression of Yes1 Associated Transcriptional Regulator (YAP1) and Sonic Hedgehog (SHH) in the lung epithelium, as well as FGF10 from the mesenchyme [28, 29]. Retinoic acid (RA) also plays a role in regulating lung bud outgrowth by coordinating Fibroblast Growth Factor (FGF) and Wnt signalling pathways [30]. These interactions are crucial for the distal-proximal patterning of the airway tree during branching morphogenesis. The expression of SOX2 is distinct to the proximal airways, while SRY-Box Transcription Factor 9 (SOX9) and SOX2 are coexpressed in the distal lung. SOX2 is essential for the differentiation of basal cells in the proximal airways into multiciliated and secretory

club cells, with Notch signalling playing a role in balancing the differentiation into secretory cells [31]. On the other hand, SOX9 is essential for normal lung branching and the differentiation of the alveoli [32].

The process of differentiation and maturation of the distal epithelium during the postnatal alveolar stage remains an area of ongoing research. As the peripheral lung tips expand, they give rise to alveoli sacs lined by alveolar epithelial type 1 (AT1) and type 2 (AT2) cells, and further septation occurs. It has been established that platelet-derived growth factor (PDGF) is necessary for the alveolarisation process as it promotes the formation of "secondary crest myofibroblasts" [33]. Additionally, BMP signalling acts as a factor that promotes the differentiation of AT1 cells, while Transforming Growth Factor Beta (TGFβ) signalling promotes their flattening and growth, thereby increasing the coverage of the gas-exchange surface area [34, 35]. Similarly, Wnt signalling pathways contribute to AT2 cell proliferation and numbers, while glucocorticoids aid in their maturation [36–38]. Nonetheless, further research is required to fully comprehend the intricate mechanisms underlying the differentiation and maturation of the alveoli epithelium during postnatal stages.

Lung regions

Lung morphogenesis gives rise to the formation of two distinct regions: the proximal conductive airways and the distal airways containing the alveoli responsible for gas exchange (**Figure 1.3**).

Proximal airways

The proximal airways are characterised by a pseudostratified epithelium and comprise different cell types with specific functions. Multiciliated cells are crucial for airway clearance due to the presence of numerous cilia on their surface. Genes such as Forkhead Box J1 (FOXJ1) and Tubulin Beta 3 Class III (TUBB3) regulate the development and function of these multiciliated cells [39]. Secretory cells, also known as Club cells, have a protective and immunomodulatory role, and genes such as Secretoglobin Family 1A Member 1 (SCGB1A1) and Secretoglobin Family 3A Member 2 (SCGB3A2) are associated with their function. Goblet cells produce mucus to capture and remove inhaled particles, and Mucin 5AC (MUC5AC) and Mucin 5B (MUC5B) are essential genes involved in their function [39]. Ionocytes are involved in regulating mucus production and ion transport, and they express high levels of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), a protein important for ion regulation. Other cell types in the proximal airway include pulmonary neuroendocrine

cells (PNECs), which influence smooth muscle tone and participate in neural regulation, and brush cells, which are believed to regulate allergic responses. Additionally, the stem cell of the proximal airways, the basal cell, has the ability to regenerate the airway lining, playing a critical role in maintaining its integrity. Key genes associated with basal cells include Tumor Protein P63 (TRP63) and Keratin 5 (KRT5) [40, 41].

Distal airways

AT1 and AT2 cells are spatially restricted to the distal airway. The flattened squamous cell bodies of AT1 cells and their proximity to endothelial cells allows them to form a surface for gas exchange in the alveolus. They cover 95% of the alveolar surface area but only constitute about 8% of the total alveolar cell population [42]. This is due to the increase in size by more than 10-fold that AT1 cells undergo during the postnatal secondary septa formation and only a 2-fold increase in AT1 cell numbers [43].

On the other hand, AT2 cells are larger and have more complex metabolic and functional characteristics. Even though the quantity of AT2 cells is approximately double that of AT1 cells, they account for less than 5% of the total surface area of the alveoli [42]. Throughout the postnatal alveolar stage, there is a notable nearly six-fold increase in the population of AT2 cells. However, their impact on the overall growth of the alveoli is relatively limited [43]. These cells are responsible for synthesising and secreting the aqueous-active material that helps maintain a low surface tension at the alveoli surface and exhibits bacteriostatic properties [41, 44, 45]. Additionally, a subset of AT2 cells is considered the facultative stem cell of the alveoli. AT2 cells have the ability to re-enter the cell cycle following injury, allowing the renewal of the alveolar epithelium through clonal expansion and differentiation into AT1 cells, as observed in post-pneumonectomy cases and lineage-tracing experiments [38, 46].

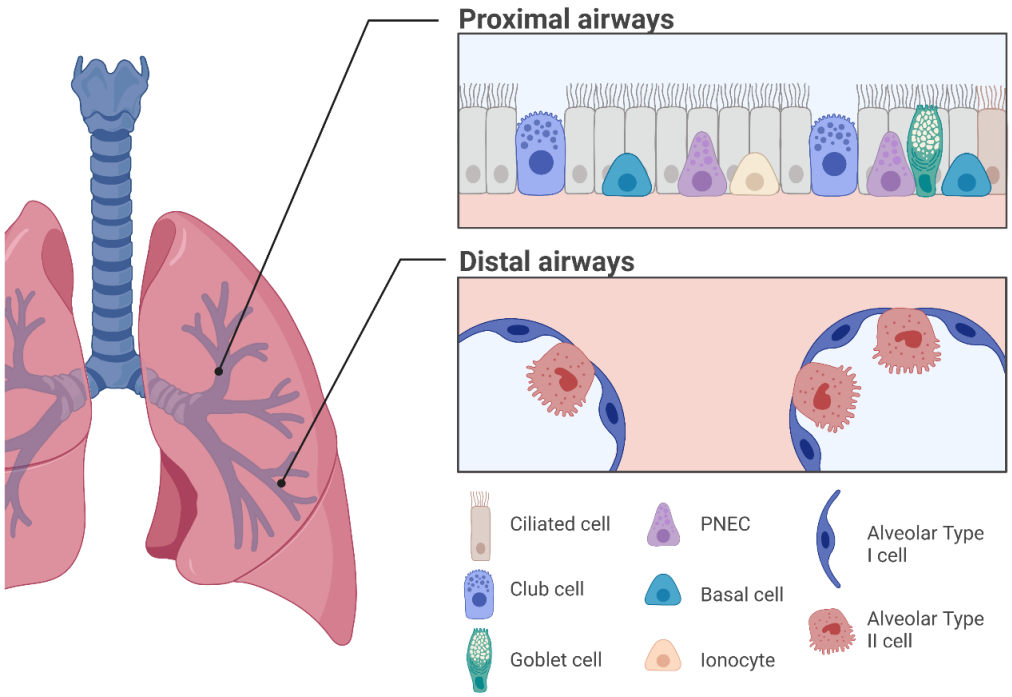


Figure 1.3 Epithelium of proximal and distal airways.

The cellular composition of the lung epithelium varies regionally. In humans, the proximal airways are lined by a pseudostratified epithelium consisting of different types of cells, including basal, secretory, ciliated, pulmonary neuroendocrine cells (PNECs) and ionocytes to form a pseudostratified epithelium. The distal airways connect to the alveolar compartment and are lined with AT1 and AT2 cells. The transition between the distal airways and alveoli occurs through the respiratory bronchioles, which have a simply cuboidal epithelium. Image created using a modified Biorender template.

1.2.2 Pulmonary surfactant

Pulmonary surfactant is a vital component of lung function, responsible for reducing surface tension within the alveoli preventing the adhesion of the alveolar surfaces and facilitating efficient gas exchange. When surfactant is absent or its function is impaired, it can cause the collapse of the alveoli during exhalation, leading to the formation of atelectasis [47]. It is composed of a complex mixture of lipids and proteins, with phospholipids accounting for 90% of its composition, predominantly phosphatidylcholine, and the remaining 10% consisting of surfactant proteins A, B, C and D (SFTPA, SFTPB, SFTPC, and SFTPD) [48]. SFTPB and SFTPC are hydrophobic proteins essential for the dynamic properties of phospholipids in the pulmonary surfactant. They promote the transition between the storage form and the functional surface film, alter lipid packing and stabilise the lipid layers during the respiratory cycle [49]. SFTPA and SFTPD, structurally related members of the collectins family, contribute to the host defence against viral, fungal, and bacterial

pathogens, and play roles in the reuptake of surfactant components [50]. The synthesis of surfactant lipids and proteins occurs within AT2 cells and is stored in lamellar bodies before being secreted onto the alveolar surface. These organelles are large storage granules within AT2 cells that contain tightly packed lipid bilayers organised into concentric layers [51].

The secretion of the pulmonary surfactant is a tightly regulated process, involving the transport of its components from the endoplasmic reticulum (ER) to lamellar bodies in AT2 cells and their subsequent translocation across the lamellar body membrane [52]. Newly synthesised surfactant proteins, pro-SFTPB and pro-SFTPC, undergo processing to their mature peptides in the Golgi and multi vesicular body (MVB) before fusion with lamellar bodies. Similarly, lipid transfer proteins facilitate the direct transport of surfactant phospholipids, such as dipalmitoylphosphatidylcholine (DPPC) and phosphatidylglycerol (PG), from the ER directly to the lamellar bodies. The large aggregate surfactant, contained in the lamellar bodies is then secreted into the alveolar surface where they form the tubular myelin, creating a phospholipid-rich film at the air-liquid interface of the alveolus [52]. The turnover of alveolar surfactant is managed through various mechanisms. Alveolar macrophages contribute to the clearance of surfactant lipids and proteins, while AT2 cells play a role in recycling surfactant components back to lamellar bodies for their processing via the MVB. Some surfactant components are degraded in lysosomes. This intricate metabolism and turnover of lung surfactant ensure the maintenance of its optimal levels and functionality in the alveoli [53]. An overview of the processing of the pulmonary surfactant is illustrated in **Figure 1.4.**

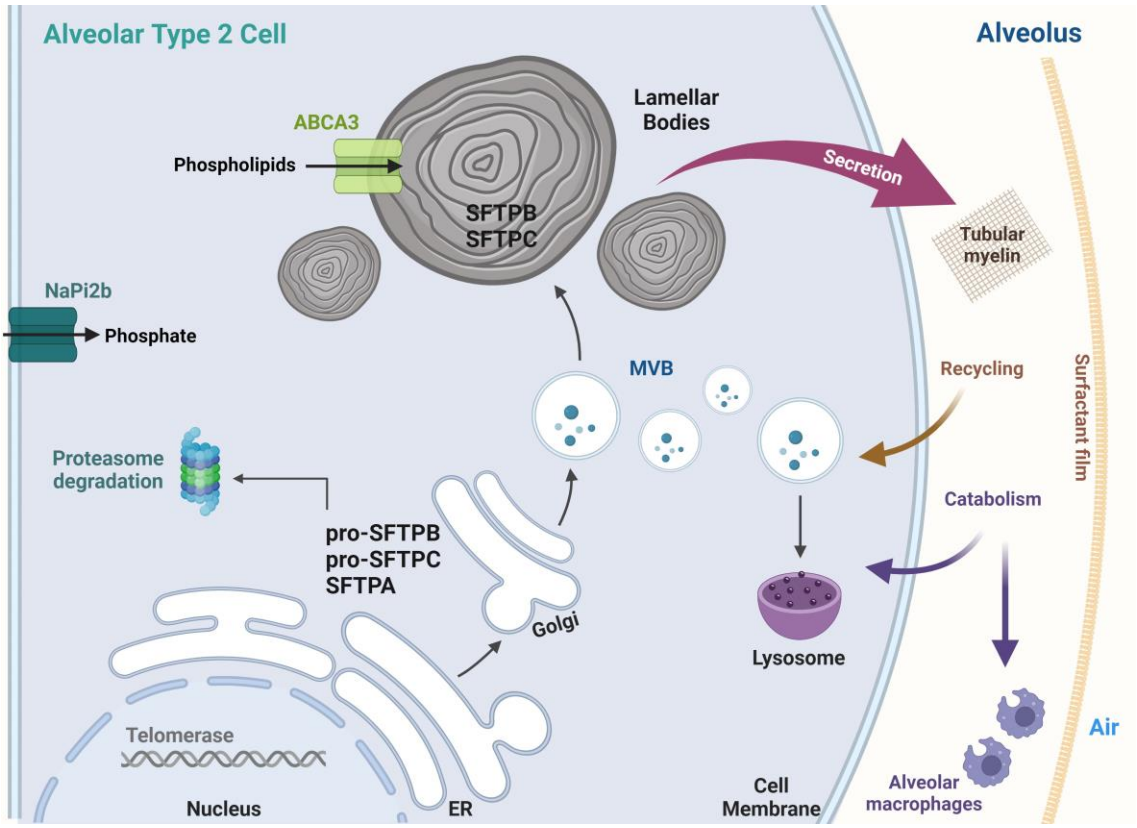


Figure 1.4 Surfactant synthesis and secretion.

Lamellar bodies release surfactant proteins and lipids into the alveolar aqueous subphase to form the complex tubular myelin structure that prevents alveolar collapse. The secreted surfactant is cleared through alveolar phagocytosis, while the remainder is recycled by AT2 cells via Multi Vesicular Body (MVB) or degraded in lysosomes. This process is essential to maintain the surfactant pool. Adapted from [54]. Image created using a Biorender template with modifications.

Disorders related to the surfactant system can cause significant lung diseases. Some disorders disrupt the functions of proteins critical for surfactant homeostasis, such as mutations in SFTPB, which affects the biosynthesis of lamellar bodies and can impact other surfactant proteins [55]. Other disorders involve alveolar cell injury mediated by protein misfolding or toxic gain of function. Among these, mutations affecting SFTPC, SFTPA, ATP Binding Cassette Subfamily A Member 3 (ABCA3), and Solute Carrier Family 34 Member 2 (SCL34A2) genes can disrupt the metabolism of surfactant and contribute to interstitial lung disease. These mutations can lead to the accumulation of misfolded surfactant proteins, altered lamellar body structure, leading to increased proteosome-mediated clearance, cytotoxicity, and microliths in the alveoli [53]. Understanding the complex mechanisms and interactions within the surfactant system is essential for diagnosing and managing these surfactant-related disorders.

1.3 Idiopathic pulmonary fibrosis (IPF)

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive lung disease. It remains a rare disease, with an estimated incidence and prevalence in the range of 0.09–1.30 and 0.33–4.51 per 10,000 persons, respectively [1]. The incidence reports per gender have been controversial, with many arguing a higher incidence in male patients, while others claim no statistically significant difference between groups [2]. The disease is characterised by interstitial fibrosis, leading to an irreversible alveolar architecture distortion [56]. The histological features of IPF have been widely described as the usual interstitial pneumonia (UIP) pattern (**Figure 1.5**). The typical architectural distortion with "honeycomb changes" is due to progressive areas of hyperplastic AT2 cells lining fibroblastic foci containing myofibroblasts and abundant extracellular matrix interspersed with areas of normal lung tissue [57]. These foci are mainly identified to the subpleural parenchyma giving the disease a periphery-to-centre progression [58].

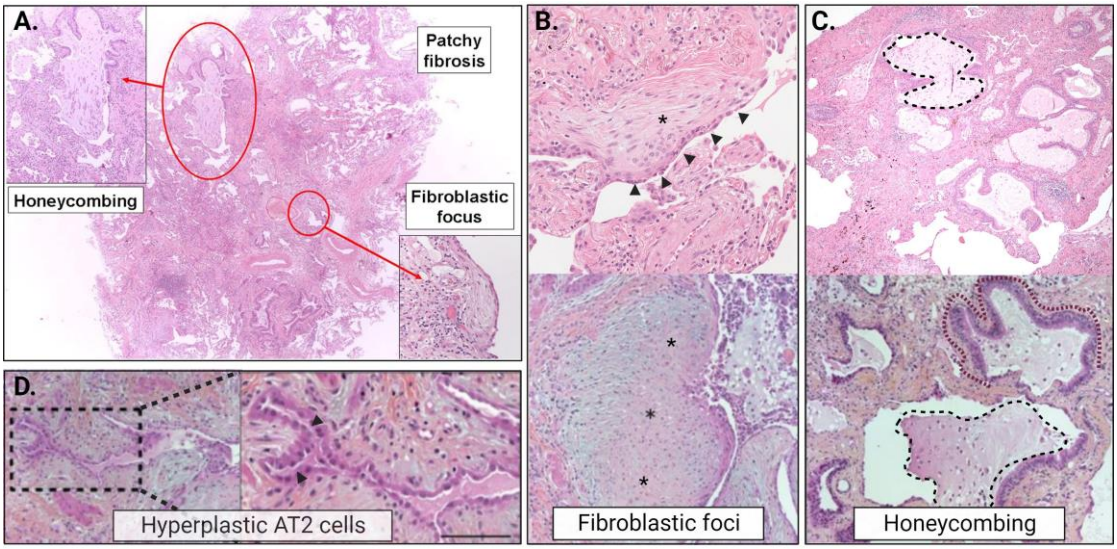


Figure 1.5 Histological hallmarks of IPF.

Haematoxylin eosin staining of lung paraffin sections showing the following: A) Transbronchial biopsy at low magnification, displaying histological features consistent with usual interstitial pneumonia (UIP). These features include fibrotic areas juxtaposed to normal lung tissue, forming a patchwork pattern (top right) and circled in red the typical honeycombing and fibroblastic focus. Image obtained from [59]. b) Fibroblastic foci, consisting of dome-shaped aggregates (asterisk) of spindle-shaped cells arranged in parallel, embedded in a pale matrix beneath hyperplastic alveolar lining cells (arrowheads). c) Honeycombing, observed as enlarged alveolar spaces lined by bronchiolar metaplastic epithelium (red dashed line) and filled with mucus and inflammatory cells (black dashed line). d) Alveolar septa lined by hyperplastic alveolar epithelial type 2 (AT2) cells; the image on the right is a magnification of the dashed box. Scale bar: 80 µm. Images obtained from [60].

The resultant destruction of the gas exchange area inevitably leads to respiratory failure characterised by progressive dyspnoea and ultimately death [61]. The dismal prognosis of the disease leads to a median survival of 2-3 years from the time of diagnosis [62, 63]. This outcome is in part due to the limited effect of the available treatments (**Figure 1.6**). In the past, treatment was aimed at minimising inflammation in order to slow its progression to fibrosis; however, the underlying lesions in IPF do not present prominent inflammation features but rather fibrotic [64]. Nowadays, the therapeutic approach relies on the anti-fibrotic agent, Pirfenidone and the tyrosine-kinase inhibitor, Nintedanib. These compounds help with the downregulation of TGFβ, procollagens, and inflammatory mediators, as well as the reduction of extracellular matrix deposition. The use of these therapeutic agents can only slow down the decline in lung function; however, they do not reverse or prevent the progression of the fibrotic process [65]

Most of IPF cases are sporadic and predominantly affect adults above the age of 50 years. However, familial forms can affect younger individuals [66]. Familiar IPF occurs when at least two first-degree relatives present the condition. This form of the disease accounts for 3% to 20% of the total IPF cases [67, 68]. The clinical presentation and histological characteristics of these two forms of IPF are believed to be indistinguishable from each other, leading to the hypothesis of a common underlying disease mechanism [69, 70].

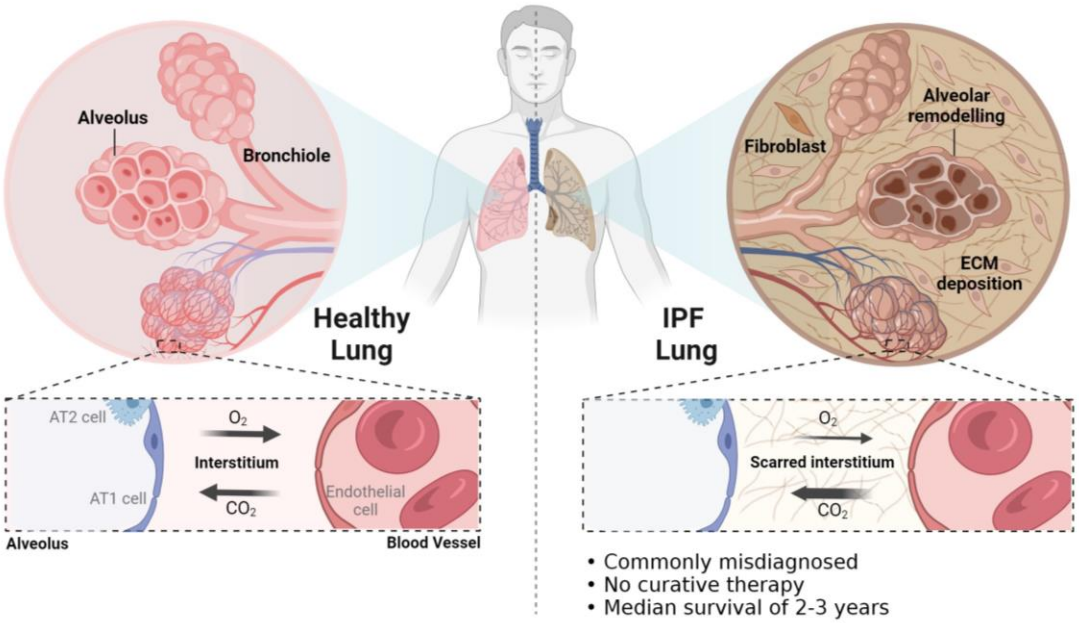


Figure 1.6 Idiopathic pulmonary fibrosis (IPF).

Chronic progression and accumulation of ECM deposition leads to irreversible interstitial fibrosis and respiratory failure. ECM: extracellular matrix, AT2: alveolar epithelial type 2 cell, AT1: alveolar epithelial type 1 cell. Created with BioRender.com.

1.3.1 Pathogenesis and genetics of IPF

The cause of IPF remains unclear. Even though there has been progress in understanding the widely heterogeneous pathogenesis of this disease, no specific cause has been identified and the critical factors for the initial development and progression of the disease remain undetermined. It is believed that epithelial injury and impaired repair play a central role in the development of parenchymal changes associated with IPF [71]. Furthermore, studies suggest that the accumulation of multiple hits is necessary to disrupt the normal repair response of the alveolar epithelium leading to the initiation and progression of fibrosis [72]. The dysfunction has been associated with the activation and expansion of fibroblasts, which are responsible for the excessive production and deposition of extracellular matrix (ECM) proteins characteristic of fibrosis [73]. Therefore, sustained alveolar epithelial cell injury in genetically predisposed individuals has been putatively accepted as the main trigger for IPF (**Figure 1.7**) [74, 75]. In response to the initial injury, the alveolar epithelium releases various soluble mediators, inflammatory cytokines, and pro-remodelling factors that contribute to the pathogenesis of IPF. These factors further perpetuate the fibrotic process and contribute to disease progression [76]. Hence, investigating

inherited forms of the disease may offer insights into the early pathogenic events leading to sporadic IPF cases.

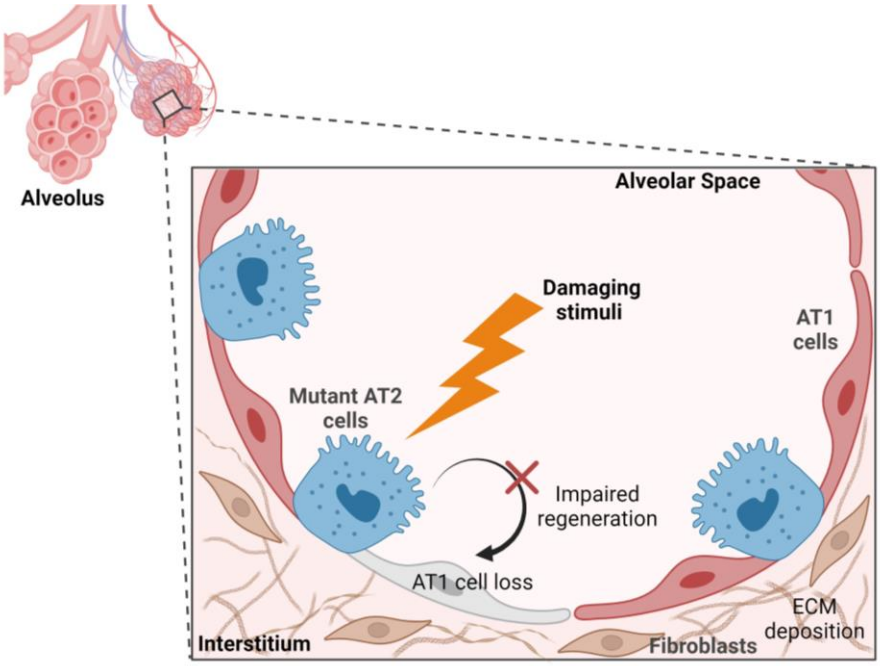


Figure 1.7 Pivotal role of the alveolar epithelium in IPF pathogenesis.

Repeated microinjury to genetically abnormal AT2 cells disrupts the normal repair response and leads to epithelial dysfunction accompanied by the release of soluble mediators, inflammatory chemokines and pro-remodelling factors [74, 77]. Created with BioRender.com.

Familial forms of IPF are associated with inheritable mutations affecting surfactant and other AT2 cell structural components, as well as intrinsic quality control systems [78]. The concept of epithelial injury and dysregulated repair by AT2 cells causing fibrotic lung disease was proposed by Haschek and Witschi in 1979 and has since been supported by experimental and clinical evidence [64]. Translational studies have shown that protein aggregation, ER stress, altered autophagy, and apoptosis observed in model systems expressing surfactant and ABCA3 mutations are also present in the lungs and AT2 epithelia of both sporadic and familial IPF patients. Molecular signatures of ER stress and the unfolded protein response (UPR) have been identified in the lungs of IPF patients, further supporting the role of chronic ER stress in the pathogenesis of interstitial lung disease [3, 79].

Familial cases have identified mutations in over 25 genes and genetic loci associated with IPF risk, with surfactant production and telomere maintenance genes being common targets [80]. Mutations in these genes can disrupt molecular pathways and increase the vulnerability of alveolar epithelium to injury, leading to profibrotic changes

and irreversible fibrotic remodelling in the lungs. The most common genetic variant associated to IPF correspond to genes important for the surfactant production and telomere maintenance, turning the focus point to AT2 cells. About 8-15% of the individuals with familial IPF present with pathogenic variants in SFTPC, SFTPA, Telomerase Reverse Transcriptase (TERT), Telomerase RNA Component (TERC) or MUC5B [3, 4]. Furthermore, genome-wide association studies have identified new genes associated with increased susceptibility to IPF, as well as the previously detected MUC5B, TERT, Desmoplakin (DSP), and KAT8 Regulatory NSL Complex 1/ Microtubule Associated Protein Tau (KANSLI/MAPT) [11, 81, 82]. Mutations on these genes may underlie the perturbations in several molecular pathways that increase the vulnerability of the alveolar epithelium to different injurious stimuli. Eventually, the activation of these pathways triggers profibrotic molecular changes originating within AT2 cells, decreasing their proliferation and contributing to irreversible fibrotic remodelling in the lungs [83].

1.3.2 Infections in IPF

Underlying genetic susceptibility in conjunction with environmental insults is proposed to trigger a chronic inflammatory response, leading to the activation of the fibrotic cascade in the lung parenchyma of IPF patients. Among these external stimuli, infections, particularly from viruses, have been postulated to contribute to the development of pulmonary fibrosis [84]. The lungs are frequently exposed to pathogens, making them susceptible to infections that can contribute to the development of pulmonary fibrosis. Various studies have shown that airborne transmitted infections cause increased AT2 cell death, secretion of pro-fibrotic cytokines, and accelerated deposition of extracellular matrix in the lungs [85–87]. Furthermore, research has suggested that changes in lung microbiome are associated with the progression of IPF with patients experiencing acute worsening of fibrosis after contracting infectious pneumonia [84].

Several studies have shed light on the association between bacterial burden and the progression of idiopathic pulmonary fibrosis (IPF). Molyneaux et al. discovered that patients with IPF had a significantly higher bacterial burden in their bronchoalveolar lavage (BAL) samples compared to healthy individuals and those with chronic obstructive pulmonary disease (COPD) [88]. The bacterial burden at baseline was found to be a predictor of functional decline and increased risk of death. Building on these findings, O'Dwyer et al. confirmed that subjects with progressive IPF had a higher

bacterial burden at the time of diagnosis, as measured using digital droplet PCR. This strengthens the link between bacterial burden and survival in IPF [89] .

In addition to bacterial involvement, human herpes viruses (HHV) have attracted attention as potential causative factors due to their ability to establish latent infections in alveolar epithelial cells and reactivate later in life. HHV-infected epithelial cells from IPF patients exhibit signs of endoplasmic reticulum stress and apoptosis, providing a potential mechanistic link between viral infection and the development of IPF [90]. This association is supported by growing evidence from human tissue and animal models, indicating a role for airborne viruses in the initiation and progression of IPF. Human herpes viruses, including cytomegalovirus (CMV), Epstein-Barr virus (EBV), HHV-7, and HHV-8, have shown consistent associations with IPF, with viral DNA detected in the lung tissue of 97% of IPF patients compared with only 36% of controls [91]. Other respiratory viruses including Influenza H1N1 and H5N1 have also been implicated in promoting the development of pulmonary fibrosis [92, 93]. Experimental studies have demonstrated that viral-induced lung injury exacerbates lung fibrosis in aged mice [94, 95]. Retrospective observational cohort studies have also linked viral pneumonia to an increased risk of developing post inflammatory pulmonary fibrosis (PIPF), with patients who had prior viral pneumonia experiencing earlier onset of PIPF and at a younger age [96]

1.4 Disease modelling of IPF

Understanding the role of AT2 cells in IPF pathogenesis has been hampered by the lack of suitable animal models that reliably recapitulate the heterogeneous pathogenesis of human IPF. Additionally, the limited access to primary tissue and complicated primary airway epithelium culture makes it difficult to study the initial stages of the disease [6]. Furthermore, established AT2 cell lines fail to faithfully recapitulate the morphologic, functional, and molecular markers of the in vivo AT2 phenotype [97].

1.4.1 In vivo and ex vivo models

Mammalian models, including rodents, ferrets, pigs, and sheep, have been utilized to study genetic and acquired diseases of the alveolar epithelium due to their similarities in pulmonary physiology to humans [98, 99]. Among these models, mice have been particularly valuable in advancing our understanding of alveolar repair. Furthermore, mouse models have played a significant role in understanding the pathogenesis of genetic and inflammatory lung disorders and in developing novel therapies [100–103].

Although mouse models of bleomycin-induced pulmonary fibrosis have been employed, they have limitations in directly translating to human IPF [104]. Bleomycin is a chemotherapeutic antibiotic known to induce DNA damage by generating DNA strand breaks [105]. While its primary clinical use is as an antitumor drug for various carcinomas and lymphomas, one of its most notorious side effects is pulmonary fibrosis [106]. This undesirable side effect has led to the utilisation of bleomycin in animal models for the study of lung fibrosis, making the bleomycin mouse model a widely used experimental model [107]. The administration of bleomycin induces lung epithelial damage with a localized inflammatory and fibrotic response from the underlying mesenchyme, which, in turn, results in the activation and accumulation of myofibroblasts, playing a pivotal role in the deposition of fibrotic tissue [108]. The popularity of this model stems from its accessibility, reproducibility, and ease of use, along with histological characteristics similar to those seen in human IPF [109]. However, it is crucial to note that the fibrotic changes induced by bleomycin in the mouse model are self-limiting and do not progress to a chronic fibrotic state. Instead, myofibroblasts are cleared, and there is spontaneous resolution of fibrosis, typically within three months failing to fully recapitulate the progressive nature of human IPF [104, 110].

Conditional gene targeting systems have been developed in mice to investigate lung formation, and function, and identify lung progenitor cells and their fate [36, 111] . Genetic mouse models often require a second environmental trigger to exhibit a phenotype analogous to the human disease, such as the combination of an Sftpc mutation and bleomycin exposure to induce fibrosis [87]. Recent studies have demonstrated the generation of mice models that mimic the spatial-temporal characteristics of fibrosis progression in human IPF by knocking out Cdc42 or Sin3a in adult mouse AT2 cells. The fibrosis induced in these models does not resolve and does not need a second trigger, rendering it a potentially superior model for studying fibrosis compared to the commonly used bleomycin injury model [112, 113].

The murine and human lung share the architectural structure of the alveoli and the cellular populations within the alveoli are also similar. However, significant differences exist between mouse and human lungs. In mice, the transition point from conducting airways to the alveolar unit is known as the Bronchioalveolar Duct Junction (BADJ), housing a rare population of progenitor cells called bronchioalveolar stem cells (BASC) [114]. In contrast, humans have a distinct distal airway compartment that includes respiratory bronchioles instead of a BADJ. The differences mentioned do not decrease

the value of the information gathered by using these models; however, they limit the extent of knowledge that can be gained solely from studies conducted on mice. Therefore, alternative models should have the capacity to fulfil one or multiple principles outlined in the 3Rs guidelines to replace the use of animal models whenever possible, reduce the number of animals used in experiments, and refine the methodology employed in animal testing [115]. This highlights the importance of developing models that more accurately recapitulate human diseases.

A valuable human model for studying lung physiology ex vivo uses human precision-cut lung slices (hPCLS) [116–118]. This technique involves obtaining thin slices of human lung tissue, which can be maintained and cultured in a controlled laboratory environment. hPCLS provide a platform for investigating the complex interactions between various cell types within the lung, including epithelial cells, mesenchymal cells, and immune cells, in a more physiologically relevant context, as the slices retain the native architecture of the lung [119]. Using hPCLS, researchers have examined these cell populations' behaviour and responses to better understand fibrotic lung diseases [120, 121]. Alsafadi et al. established a model of IPF by exposing hPCLS derived from patients without interstitial lung disease (ILD)/IPF to a mixture of profibrotic growth factors, proinflammatory cytokines, and signalling molecules, including TGF-β, tumour necrosis factor-α (TNF-α), platelet-derived growth factor (PDGF)-AB, and lysophosphatidic acid. This mixture induced early fibrosis-like changes in the tissue explants, contributing to understanding the pathological mechanisms of IPF [120]. Similarly, a model of lung fibrogenesis that recapitulates the TGF-β1-mediated events in IPF using hPCLS offers the potential to predict drug efficacy and accelerate the characterisation of IPF therapeutics [121, 122]. Nevertheless, access to human tissue is limited, and despite the optimisation of their culture conditions, hPCLS are only viable from 3 to 14 days in culture, resulting in variations depending on the tissue source for each set of experiments, further limiting the possible applications for the model [123].

Due to the limitations presented by these models, there is an increasing need to develop human-specific models that recapitulate cell-intrinsic disease characteristics, such as hiPSC-derived platforms, to provide insights into human mechanisms in alveolar repair.

1.4.2 In vitro models

Efforts to maintain the phenotype of primary AT2 cells in vitro have posed challenges, particularly in 2D conditions where their gene expression and morphology are difficult to sustain. Studies on rat and mouse AT2 cells in 2D culture have shown a loss of SFTPC expression when cultured in vitro, highlighting the limitations of this approach [124, 125]. Recent advancements in 3D culture techniques have improved the maintenance of these cells in vitro. Primary human AT2 cells formed organoids or alveolospheres when cultured in Matrigel and successfully formed alveolar-like structures composed of polarised epithelial cells surrounding a central lumen [126]. Furthermore, these alveolospheres exhibited characteristics of differentiated AT2 cells, including the production of surfactant and tubular myelin; however, the use of feeder cells to maintain the AT2 phenotype, rendered the model unviable for many downstream applications [127–129]. New approaches from multiple research groups have shown promising results developing protocols for maintaining and expanding primary AT2 cells using chemically defined media in feeder-free conditions [130–132]. One limitation of these protocols is the difficulty to maintain the cultures in long-term, giving the reduce self-renewing capacity of primary AT2 cells in vitro. However, these advancements hold promise for improving our understanding of AT2 cell biology and their role in lung physiology and disease.

The directed differentiation of pluripotent stem cells has proven to be one of the most effective methods for studying AT2 cells in vitro. Through this approach, researchers have successfully established protocols for differentiating stem cells into fully functional airway and alveolar cells. These rely on the controlled addition of growth factors and small molecules at specific concentrations and culture stages to replicate the signalling pathways involved in the in vivo development of AT2 cells. This process involves the specification of definitive endoderm and the subsequent patterning of the endoderm to induce Nkx2-1^{+} lung progenitors. These progenitors are then induced to differentiate into alveolar epithelial cells [18, 133–135].

Initial studies reported the successful differentiation of murine embryonic stem cells (ESCs) into mature SFTPC-producing AT2 cells containing lamellar bodies. Their approach consisted of the initial generation of embryoid bodies (EBs) to be later transferred to serum-free small airway growth medium for the derivation of AT2 cells. However, their cultures exhibited considerable heterogeneity and low efficiency, resulting in only small proportions of mature AT2 cells being generated [136, 137]. Later, Longmire et al. achieved the differentiation of lung progenitor cells with a

relatively high degree of purity without the generation of EBs . They employed a mouse NKX2.1 reporter ESC line, along with a well-established protocol that mimicked the initial phases of lung development. Additionally, they implemented a cell sorting technique to refine the population of cells [134].

Building on the earlier breakthroughs, researchers were also able to generate pure populations of AT2 cells derived from human ESCs (hESCs). Wang et al. utilised a transfected hESC line containing an SFTPC promoter-neomycin transgene, which was maintained in mouse embryonic fibroblast conditioned hESC medium to eliminate the need for feeder cells [138]. Subsequently, they cultivated the cells in differentiation media to obtain lung epithelial cells expressing SFTPC, SFTPA, cystic fibrosis transmembrane conductance regulator, and lamellar bodies. Later, other research groups also demonstrated the successful generation of lung epithelial cells expressing NKX2.1, SFTPC, and Aquaporin 5 (AQP5) when cultured with a defined differentiation medium under air-liquid interface conditions [139, 140]. However, the initial accomplishments with hESCs were mostly replaced by more recent human induced pluripotent stem cell (hiPSC)-based techniques due to various obstacles that hindered their scalability, such as political, ethical, and practical limitations.

1.5 Human induced pluripotent stem cell-derived alveolar epithelium

In the recent years, human induced pluripotent stem cell (hiPSC) technology has offered the potential for modelling human diseases. hiPSCs are derived from the reprogramming of human somatic cells, such as skin fibroblasts, by the transient overexpression of four transcription factors, Oct3/4, Sox2, Klf4, and c-Myc [141]. These reprogrammed cells acquire an embryonic-like state that phenotypically resembles that of human ESCs with the ability to differentiate into any specific cell type of the three germ layers [141]. This source of stem cells offers the potential to develop patient-specific lung epithelial cells for creating in vitro models specific to diseases and patients. The utilization of hiPSCs-based IPF models provide a valuable platform for investigating the impact of various physiological and pathological factors, including genetic mutations, infections, and toxins, on the biology of airway epithelial cells in IPF. One significant advantage of employing this model over traditional mouse models and immortalized cell lines is the ability to observe human-specific responses and avoid the phenotypic drift often observed in immortalized cells.

Various platforms have been developed for the differentiation of hiPSCs into alveolar cells, with the ultimate goal of achieving a disease-relevant phenotype. Successful

induction of hiPSCs into functional AT2 cells relies on following in vitro protocols that mimic key stages of lung development by exposing the cells to specific growth factors, hormones, and inhibitors. These protocols start with early embryonic stages, progressing through definitive endoderm characterised by the expression of SOX17, C-X-C chemokine receptor type 4 (CXCR4) and Tyrosine-protein Kinase Kit (cKIT), followed by the foregut stages with cells expressing FOXA2. These cells lead to the generation of NKX2.1 positive lung progenitors, which ultimately differentiate into distal lung epithelial AT2 cells, expressing markers such as pro-SFTPC, pro-SFTPA, and ABCA3 [142, 143]. The markers expected to be expressed at each stage of the alveolar cell differentiation are shown in **Figure 1.8**. Along with their molecular profile, these alveolar cells exhibit functional characteristics similar to in vivo distal alveolar cells, including surfactant processing in lamellar bodies. As a result, these cells serve as a suitable in vitro model for studying lung diseases that specifically affect the distal alveolar cell population [144, 145].

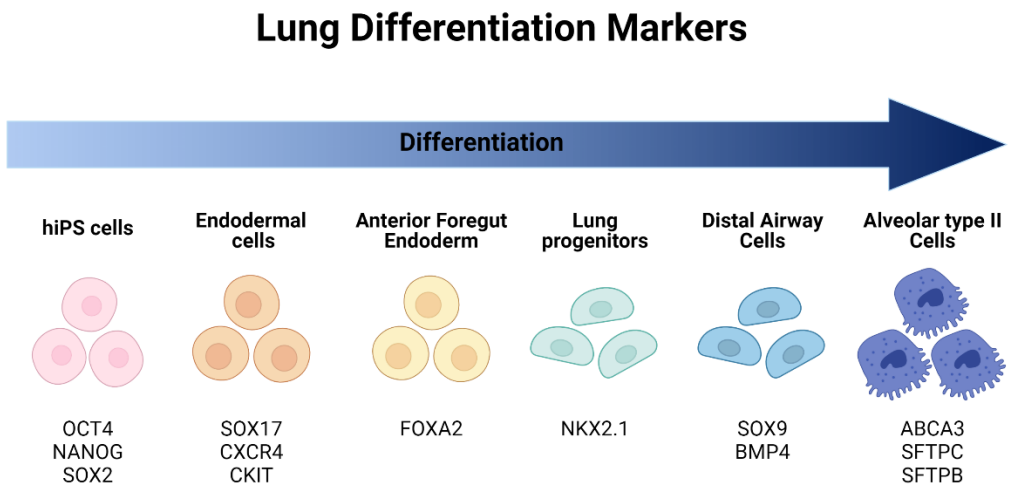


Figure 1.8 AT2 cell differentiation markers.

Markers that are expected to be expressed during the in vitro differentiation of cells serve as indicators of successful induction along the in vivo developmental pathways. Created with BioRender.com.

The cell culture conditions, which under these alveolar cells are generated, are in part based on discoveries made from mouse lung development studies [21]. For instance, to derive definitive endoderm (DE), it is necessary to activate Nodal signalling to replicate the embryonic induction of gastrulation and the formation of the anterior primitive streak using high-dose Activin A [146, 147]. Moreover, the formation of the primitive streak from pluripotent stem cells also relies on the involvement of BMP and

Wnt signalling [148–151]. To further specify the DE to anterior foregut endoderm (AFE), the additional dual inhibition of BMP and TGFβ signalling has proven to be essential [133, 143, 152]. It has been demonstrated that WNT signalling is necessary for the specification of lung progenitors expressing the lung/thyroid marker NKX2.1 within the anterior foregut endoderm [153]. Furthermore, members of the FGF family are necessary to promote the outgrowth and differentiation of the emerging NKX2.1+ lung bud, along with BMP4, to promote the initiation of the proximal-distal patterning [154]. Additionally, retinoic acid (RA) is also implicated in this early lung developmental stage as it upregulates FOXP2, a lung bud marker, and suppresses the expression of other thyroid markers (HHEX). Therefore, combinations of WNT signalling activators, FGF, BMP4 and RA have been widely used to induce the expression of NKX2.1 in differentiating lung progenitors [134, 142, 144, 155].

Further maturation of these NKX2-1+ lung progenitors has shown their capacity to differentiate into SFTPC+ distal lung epithelial cells when stimulated by trophic factors. Research has shown that WNT signalling activators, members of the FGF family, Cyclic adenosine monophosphate (cAMP), 3-isobutyl-1-methylxanthine (IBMX), and corticosteroid signalling can promote alveolar differentiation [12, 156] The maturation process of these differentiating cells has been significantly improved by adopting 3D culture conditions, as evidenced by studies demonstrating the enhanced development of a mature AT2 program phenotype [145, 157, 158]. Overall, the combination of these defined factors and signalling pathways plays a crucial role in driving the differentiation and maturation of lung progenitor cells into functional alveolar epithelial cells.

Lung epithelial and alveolar organoids

In vitro generation of lung organoids derived from human pluripotent stem cells has proven to be a valuable source of functional AT2 cells. These approaches rely on 3D culture conditions, with cells typically embedded in Matrigel, leading to the emergence of a mature AT2 phenotype that includes the production of surfactant proteins and phospholipids [145, 158, 159]. A research group used hiPSC-derived foregut spheroids to generate two distinct types of lung organoids: human lung organoids containing airway-like cells and mesenchymal cells, and bud tip progenitor organoids containing mainly undifferentiated progenitor cells [160]. These last ones were highly proliferative and could be serially needle passaged to enrich the bud tip progenitors, obtaining ~88% positivity for SOX2/SOX9 and a similar molecular profile to the bud tips in the human foetal lung. However, if organoids in culture were left intact, the bud tip structure underwent bifurcation events forming patterned lung organoids. This

patterning produced an interior region with cells positive for SOX2 and mucus-producing cells, whereas the bud tip region maintained SOX9+, pro-SFTPC cells [160]. Both organoid types contained a mixture of proximal and distal airway cells positive for NKX2.1, which could be maintained and expanded for the long term under the influence of FG7, CHIR-99021 and RA [161] .

In a study conducted by Jacob et al. (2017), a different approach was employed to study the emergence of cell populations during differentiation [145]. They utilised reporter cell lines modified to express the fluorescent reporters Green fluorescent protein (GFP) and tdTomato when positive for NKX2.1 and SFTPC, respectively. This strategy allowed them to isolate single and double-positive cell populations. The differentiation to lung progenitors involved inducing NKX2.1 expression in the anterior foregut endoderm using CHIR-99021, BMP4, and RA. The NKX2.1 GFP^{+} cells were then isolated and cultured in 3D Matrigel droplets with the presence of CHIR-99021, Fibroblast Growth Factor 7 (KGF), and lung maturation factors such as cAMP, dexamethasone (DEXA) and IBMX. The combination of these factors, along with their differentiation media, resulted in a strong expression of $SFTPC^{+}$ cells. To date, this protocol has been widely used for disease modelling purposes with AT2 cells. Images of organoids generated using these peer-reviewed protocols are shown in **Figure 1.9**.

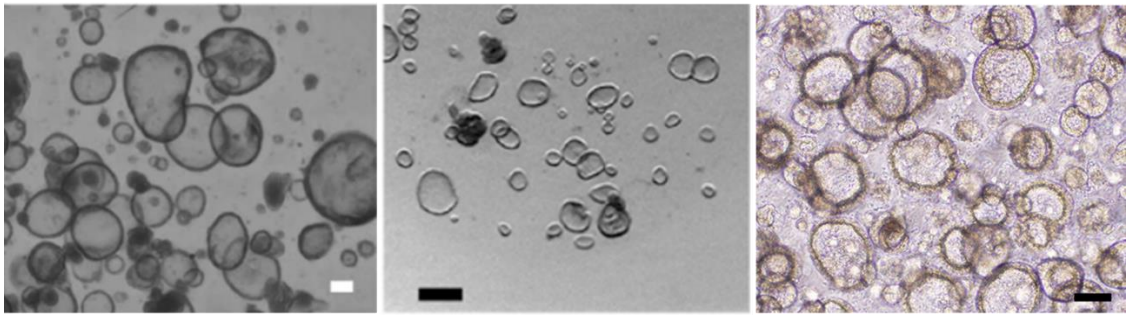


Figure 1.9 In vitro generation of lung organoids.

Brightfield images of lung organoids embedded in Matrigel droplets. A. Alveolospheres. Scale bar, 200 µm. B. Lung bud tip progenitor organoids maintained by needle passaging. Scale bar, 200 µm. C. Carboxypeptidase M (CPM)-positive 'ventralised' anterior foregut endoderm-derived spheroids. Scale bar, 100 µm. Images obtained from [159, 160, 162] respectively.

Despite of the methodology followed for the AT2 cell generation, cells used as a reliable disease-modelling platform commonly possess certain characteristics. Among these is a mature surfactant system capable of the biosynthesis and complete posttranslational processing of pro-SFTPB and pro-SFTPC proteins, and secretion of surfactant phospholipids. Cellular ultrastructure, including lamellar bodies and lysosomal-like organelles is also necessary for these cells to fulfil their functional characteristics. [97].

Furthermore, cells should exhibit self-renewal and pluripotency, as well as maintained expression of AT2 cell genes (SFTPB, SFTPC, ABCA3) and exclusion of AT1 (AQP5) and proximal airway cell (P63) markers [128]. The use of this AT2 cells derived from hiPSCs to generate disease-oriented in vitro models represent a promising platform to translate discoveries made on murine models, perform complex assays or high-throughput drugs screening for generating personalized and definitive IPF therapies.

Chapter 2 Materials and Methods

2.1 Cell culture

Cells were cultured in a type II Biological Safety Cabinet and maintained in a humidified incubator at 37°C with 5% CO2. All culture media and reagents were warmed to 37°C in a water bath before use unless otherwise stated.

2.1.1 Culture media

Basal media preparations were stored at 4˚C for up to two weeks. On the day of use, it was supplemented with small molecules and/or growth factors and warmed up to 37˚C prior to cell feeding. The composition of the media used in this thesis is listed in **Table 2.1**.

Table 2.1 Composition of culture media.

Medium	Component	% (vol/vol) or Final Conc.	Manufacturer	Catalogue #
HDF medium: Human dermal fibroblasts growth	DMEM	85%	Gibco	21969-035
	Antibiotic-Antimycotic (100X)	2%	Gibco	15240-062
	Foetal bovine serum (FBS)	10%	Gibco	10270106
	GlutaMAX™ Supplement	1%	Gibco	35050-038
	Non-Essential Amino Acids (NEAA, 100X)	1%	Gibco	11140-050
	Penicillin-Streptomycin (10,000 U/mL; Pen/Strep)	1%	Gibco	15140122
TeSR E7 medium: reprogramming	TeSR E7 (Basal medium)	96%	Stemcell	5919
	TeSR E7 (25X Supplement)	4%	Stemcell	5915
TeSR E8 medium: reprogramming	TeSR E8 (Basal medium)	96%	Stemcell	5991
	TeSR E8 (25X Supplement)	4%	Stemcell	5992
Home-brew E8 (HBE8) medium: hiPSC culture (adapted from [163])	DMEM/F12 50/50 mix	99.85%	Corning	10-092-CM
	L-ascorbic acid 2-phosphate tri-sodium salt	64 µg/ml	Sigma	49752
	Heparin sodium salt	100 ng/ml	Sigma	H3149
	Sodium selenite	14 ng/ml	Sigma	S5261
	Transferrin	5 µg/ml	Sigma	T3705
	Insulin	20 µg/ml	Sigma	11376497001
	FGF2 (basic)	100 ng/ml	Peprotech	100-18B
	TGF-β1	2 ng/ml	Peprotech	100-21
	RPMI 1640	96%	Gibco	21875034

Basal differentiation-1 (DIFF1) medium (adapted from [164]	B27 supplement (50X)	2%	Gibco	17504044
	NEAA, 100X	1%	Gibco	11140-050
	Pen/Strep	1%	Gibco	15140122
Definitive endoderm (DE) medium (adapted from [164])	Basal DIFF-1 medium			
	Activin A	100 ng/ml	Peprotech	120-14P
	Wnt 3a	50 ng/ml	R&D	5036-WN
Anterior foregut endoderm (AFE) medium (adapted from [165])	Basal DIFF-1 medium			
	Activin A	50 ng/ml	Peprotech	120-14P
Lung progenitors' medium	Basal DIFF-1 medium			
	CHIR99021	1µM	Sigma	SML1046
	RA	250nm	Sigma	R2625
	BMP4	5ng/ml	R&D	314-BP
Lung maturation medium	Basal DIFF-1 medium			
	CHIR99021	3 µM	Sigma	SML1046
	FGF10	10 ng/mL	PeproTech	100-26
	Dexamethasone	50 nM	Sigma	D1756
	8 Bromoadenosine 3 5 '- cyclic monophosphate (cAMP)	0.1 mM	Sigma	B7880
	3-isobutyl-1-methylxanthine (IBMX)	0.1 mM	Sigma	I5879
	Recombinant Human KGF (FGF7)	10 ng/mL	PeproTech	100-19
Basal differentiation-2 (DIFF2) medium (adapted from [162])	IMDM	72.5%	Gibco	21056-023
	Ham's F12	23.5%	Corning	10-080-CV
	N-2 Supplement (100X)	0.5%	Gibco	17502-048
	B27 supplement (50X)	1%	Gibco	17504044
	GlutaMAX™ Supplement	1%	Gibco	35050-038
	Bovine Albumin Fraction V (7.5% solution)	0.6%	Gibco	15260037
	α-Monothioglycerol (1.3%)	40µl/L	Sigma	M6145
	L-Ascorbic Acid 2-phosphate (AA2P)	50 µg/mL	Sigma	A8960
	Pen/Strep	0.5%	Gibco	15140122
Optimised DE medium	Basal DIFF-1 medium			
	Activin A	100 ng/ml	Peprotech	120-14P
	CHIR99021	3 µM	Sigma	SML1046-5MG

	LY294002	10 µM	Selleck	S1105
Anteriorisation medium	Basal DIFF-2 medium			
	SB431542	10 µM	Selleck	S1067
	Dorsomorphin	2 µM	Tocris	3093
Lung progenitors medium_2	Basal DIFF-2 medium			
	CHIR99021	3 µM	Sigma	SML1046
	RA	100 nM	Sigma	R2625
	BMP4	10ng/mL	R&D	314-BP
Lung maturation medium_2	Basal DIFF-2 medium			
	CHIR99021	3 µM	Sigma	SML1046
	FGF10	10 ng/mL	PeproTech	100-26
	Dexamethasone	50 nM	Sigma	D1756
	CAMP	0.1 mM	Sigma	B7880
	IBMX	0.1 mM	Sigma	I5879
	FGF7	10 ng/mL	PeproTech	100-19
Basal differentiation-3 (DIFF3) medium (adapted from [166])	IMDM	70%	Gibco	21056-023
	Ham's F12	23.5%	Corning	10-080-CV
	PVA	1 mg/ml	Sigma	P8136
	ITS-X	2%	Corning	51500056
	Lipids	0.1%	Gibco	11905031
	α-Monothioglycerol (1.3%)	40µl/L	Sigma	M6145
	AA2P	64mg/L	Sigma	A4544
	GlutaMAX™ Supplement	1%	Gibco	35050-038
	Non-Essential Amino Acids	1%	Gibco	11140-050
	Pen/Strep	0.5%	Gibco	15140122
iHBEC medium	KSFM	99.5%	Gibco	11590526
	Bovine Pituitary Extract (BPE	25µg/ml	Gibco	11590526
	Epidermal Growth Factor (EGF)	0.2 ng/ml	Gibco	11590526
	Geneticin (G418 sulfate)	25µg/ml	Sigma	A1720
	Puromycin dihydrochloride	250ng/ml	Sigma	P8833
TMLC medium	DMEM	90%	Gibco	21969-035
	FBS	10%	Gibco	A5256701
	L-glutamine	4 mM	Gibco	25030081
	Geneticin (G418 sulfate)	250µg/ml	Sigma	A1720

2.1.2 Derivation of human skin fibroblasts from IPF patients

Human skin punch biopsies from 6 patients with Familial Idiopathic Pulmonary Fibrosis (IPF) carrying a heterozygous mutation in surfactant proteins C and A were donated by Dr Coline van Moorsel from the Division of Heart and Lungs, University Medical Centre Utrecht, The Netherlands. The patients formed part of a cohort diagnosed at the St. Antonius Hospital in Nieuwegein, The Netherlands, by a multidisciplinary team

and the process was overseen by clinicians, Joanne J. van der Vis and Jan C. Grutters. The diagnostic criteria applied were in accordance with the guidelines set by the American Thoracic Society and the European Respiratory Society [167].

For the DNA sequence analysis, blood samples were donated by the IPF patients. These samples were then utilized for DNA extraction, and the coding regions of the SFTPC and SFTPA2 genes were sequenced [168, 169]. Subsequently, the patients who exhibited an SFTPC mutation were subjected to further sequencing to investigate the ABCA3 coding regions. No consequences on protein function were predicted by the analysis using the Sorting Tolerant From Intolerant' (SIFT) algorithm for ABCA3 identified variants [170]. Additionally, it was observed that all ABCA3 variants detected were also present in the control subjects. The information about the IPF patients from whom the biopsy samples were obtained and utilised in this thesis, can be found in **Table 2.2**.

Upon arrival at The University of Nottingham, biopsy samples were processed in the quarantine laboratory's class 2 biosafety cabinet. Using a scalpel, the ~2mm skin punch biopsies were dissected into 12 smaller pieces, which were then transferred individually into a well of uncoated 6-well tissue culture plates. Once plated, a coverslip was placed on each piece to ensure better tissue attachment and promote cell outgrowth. Tissue pieces were cultured in human dermal fibroblast (HDF) medium at 37˚C in 5%CO2, with media changes every 2-3 days. Once the fibroblasts reached the borders of the coverslip, the confluent cells were trypsinised using 0.25% Trypsin-EDTA (Gibco #25200072) for 3 min at 37°C. Following this, cells and remaining tissue pieces were passaged into a T75 tissue culture flask. Cell lines were further expanded and passaged in a 1:3 to 1:6 ratio. All cell lines were tested for mycoplasma contamination using EZ-PCR Mycoplasma Test Kit (Biological Industries, #20-700-20) prior to reprogramming and cryopreservation for future experiments. The hiPSC lines used for the experimental work in this thesis were derived from patients IPF1, IPF2 and IPF5.

Table 2.2 Idiopathic Pulmonary Fibrosis patient's details.

	IPF1	IPF2	IPF3	IPF4	IPF5	IPF6
Gene	SFTPC	SFTPC	SFTPC	SFTPC	SFTPC	SFTPA2
Mutation	Y113C	M71V	I73T	I73T	Y113C	N171Y
DNA notation	c.338A>G	c.211A>G	c.218 T>C	c.218 T>C	c.338A>G	c.511A>T
Protein domain	BRICHOS domain	Non-BRICHOS Linker domain	Non-BRICHOS Linker domain	Non-BRICHOS Linker domain	BRICHOS domain	NA
Age	64	42	47	44	30	39
Sex	Female	Male	Male	Female	Female	Female
Treatment	Prednisolone	Pirfenidone	No treatment	Azathioprine and prednisone	Pirfenidone	Prednisolone
Lung transplantation	Yes	No	No	Yes	No	Yes

2.1.3 Establishment of human induced pluripotent stem cell (hiPSC) lines

The human induced pluripotent stem cell (hiPSC) clones employed in this thesis were reprogrammed using non-integrating RNA viral (Sendai) vectors in feeder-free conditions. The vector preparations encoding for the Yamanaka reprogramming factors OCT3/4, KLF4, c-MYC and SOX2 were introduced into the dermal fibroblasts using the CytoTuneTM-iPS 2.0 Sendai Reprogramming kit (Invitrogen #A16517) following the manufacturer's recommendations (**Figure 2.1**). Fibroblasts were seeded at 20,000 cells/well density in a 6-well plate. The following day, transduction was performed with the Sendai virus vectors at an MOI of 5:5:3 (hKOS:hc-Myc: hklf4) in antibiotic-free HDF medium. 24h later, the transduction medium was replaced with fresh HDF medium. Cells were cultured in HDF medium for expansion until 80-90% confluency was reached, usually between 3 to 7 days post-transduction. Once confluent, cells were dissociated with 0.05% Trypsin (Gibco #25300062), resuspended in TeSR-E7 medium (Stemcell #5919, 5915) and passaged into growth factor-reduced Matrigel (Corning, #354230)-coated 6-well plates in a 1:3 ratio. TeSR-E7 media changes were made daily to induce hiPSCs formation. Around days 13 to 22 post-transduction, when defined colonies started emerging, the medium was changed to TeSR-E8 (Stemcell #5991, 5992) to promote the growth of the tightly packed cells within the colonies. The growing hiPSCs colonies reached a large enough diameter to be isolated from day 13 to day 45 post-transduction.

hiPSC colonies with a diameter of 500-1000µm (10X) were ready to be isolated using either EDTA (Invitrogen # 15575020) or ReLeSR (Stemcell #05872). To isolate cells using EDTA, cells were washed with Dulbecco's Phosphate Buffered Saline, no calcium, no magnesium (DPBS; Gibco #14190169) and then incubated for 3 min with 1ml 0.5mM EDTA in DPBS at 37°C. Next, EDTA solution was removed and replaced with TeSR-E8 medium, allowing individual hiPSC colonies to be picked. If ReLeSR was used instead, 1ml was added to the wells and aspirated within 1 min to ensure the cells were only exposed to a thin film of liquid. According to the manufacturer's recommendations, cells were ready to be picked after 7-9 minutes of incubation at room temperature with ReLeSR. When ready, hiPSC colonies were carefully picked using a P200 pipette tip under sterile conditions, using a microscope inside a sterile biological safety cabinet. Both enzyme-free cell dissociation reagents, EDTA and ReLeSR, enabled the selective detachment of hiPSC colonies using a pipette, reducing the necessity for labour-intensive manual dissection and mechanical disaggregation of the colonies derived from the hiPSC lines utilized in this thesis [171]. Each picked colony was transferred onto a Matrigel-coated well of a 48-well plate containing 200ul of TeSR-E8 medium supplemented with 10µM Y-27632 Rho-associated kinase inhibitor (Y-27632; Tocris, #1254) and dissociated mechanically by carefully pipetting up and down, promoting colony attachment to the well and expansion. To ensure hiPSC establishment, approximately 36-96 colonies per reprogrammed cell line were picked. The selected colonies were cultured in TeSR-E8 medium and incubated at 37°C in 5% CO2 with daily TeSR-E8 medium changes.

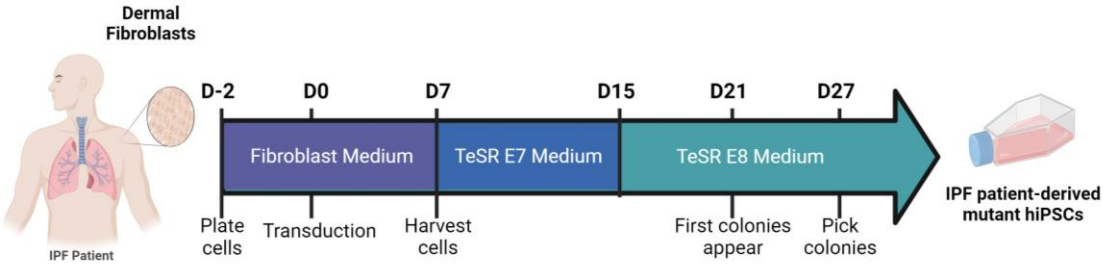


Figure 2.1 Schematic timeline of the protocol used to generate hiPSCs by transduction of Sendai virus vectors.

Dr Nicholas Hannan and Dr Peggy Cho Kiu Lo applied this method to clonally derive the hiPSC lines IPF1, IPF2, IPF3, IPF4, IPF5 and IPF6 from the reprogramming of dermal fibroblasts of six Familial IPF-patients carrying a heterozygous surfactant protein mutation. In addition, REBL-PAT and LOPCK hiPSC lines were reprogrammed from healthy donors' human dermal fibroblasts using Sendai virus vectors by Dr Gary Duncan and Dr Peggy Cho Kiu Lo, respectively. All the reprogramming mentioned above was performed at The University of Nottingham.

2.1.4 hiPSC culture and maintenance

REBL-PAT, LOPCK and IPF hiPSC lines were cultured using Essential 8 (E8) medium supplemented with 1% penicillin/streptomycin (P/S). E8 medium used was either commercially available (TeSR-E8) or made in the Hannan Lab (homebrew E8, HBE8) as described by Chen et al. (2011) with minor modifications shown in **Table 2.1** [163]. For maintenance, hiPSCs were cultured on Matrigel-coated culture vessels at a constant coating density of 34.7µg/cm^2. The diluted Matrigel in cold Dulbecco's Modified Eagle Medium (DMEM; Gibco, #21969-035) was added and distributed evenly on the plasticware surface and left to polymerise for 1 hour to overnight at 37˚C before seeding the cells.

hiPSCs were passaged when 75-90% confluent, usually every 3-4 days, using TrypLE Express (Gibco, #12604021). The culture medium was aspirated; cells were washed once with DPBS and then incubated with TryPLE for 3-5 minutes at RT until cells were visibly detached. TrypLE was carefully aspirated without disrupting the cells before gently tapping the culture vessel to disassociate the cells. Any remaining dissociating reagent was neutralised by resuspending the cells in DMEM, obtaining a single-cell suspension transferred to a 15ml conical tube. This was centrifuged at 1.200 rpm for 3 min, forming a tight cell pellet. The remaining supernatant was discarded, and the cells were resuspended in E8 medium in the presence of 10µM Y-27632 to prevent dissociation-induced apoptosis. Cells were then seeded at a split ratio of 1:10 to 1:20 or a seeding density of 20,000 cells/cm^2. The following day, the medium was replaced with fresh E8 medium without Y-27632; this was repeated daily until confluent.

2.1.5 Directed differentiation of hiPSCs into lung progenitors

hiPSCs were dissociated using TrypLE and seeded into 12-well culture plates (ThermoFisher) pre-coated with Matrigel (Corning #354230). Calculating the total number of cells required for seeding, cells were plated using a cell density of 10,000 to 30,000 cells/cm^2 in HBE8 medium supplemented with 10µM Y-27632 (Selleck #S1049). Cells were maintained in HBE8 medium for 48hrs to allow proliferation, replacing the medium 24 hours after passaging to avoid toxicity by Y-27632.

Original differentiation method

The first differentiation method employed in the experiments of this thesis consisted of the sequential induction of hiPSCs into definitive endoderm (DE), followed by its patterning into anterior foregut endoderm (AFE) and lung progenitor cells, finalising with their maturation into distal lung epithelial cells, as shown in **Figure 2.2**. All the

stages within this differentiation methodology were performed in the same culture vessel with supplemented basal Differentiation-1 (DIFF1) medium, a serum-free basal differentiation medium comprising RPMI 1640 medium (Gibco #21875034) supplemented with Non-Essential Amino Acids (NEAA; Gibco #11140-050), B27 supplement (Gibco, #17504044), and 1% penicillin/streptomycin (P/S; Gibco #1514-122), with daily media changes.

Definitive endoderm differentiation was induced 48hrs from passaging once 50-80% confluence was reached by supplementing the basal DIFF1 medium with 100 ng/mL recombinant human Activin A (ActA; R&D Systems #338-AC) and 50 ng/mL recombinant human Wnt3a (Wnt3a; R&D Systems #5036-WN) for three days. After DE induction, cells were exposed to foregut medium consisting of basal DIFF1 medium supplemented with 50 ng/mL ActA to promote anterior foregut endoderm specification for two days. Next, to promote the differentiation of anterior foregut endoderm into distal lung epithelial progenitors, basal DIFF1 medium was supplemented with 1 μM CHIR99021 (Sigma #SML1046), 5 ng/mL recombinant human Bone Morphogenetic Protein 4 (BMP4; R&D # 314-BP) and 250 nM all-trans retinoic acid (RA; Sigma #R2625) for 12 days.

For the experiments where maturation of the generated lung progenitors was attempted on 2D, following the 12-day lung progenitors' induction, cells were changed to basal DIFF1 medium supplemented with the maturation factors consisting of 3 μM CHIR99021, 50 nM dexamethasone (DEXA; Sigma # D1756), 0.1 mM 8 Bromoadenosine 3 5 '- cyclic monophosphate (cAMP; Sigma #B7880), 0.1 mM IBMX (Sigma #I5879) and 10 ng/ml recombinant human FGF10 (FGF10; PeproTech #100-26) for 16 days to stimulate cell functionality.

Optimised differentiation method

The optimised version of the protocol was designed based on previously published differentiation methods [157, 162]. While the differentiation also consisted of the induction of definitive endoderm, anterior foregut endoderm and lung progenitor cells, it included cell passaging and enrichment steps illustrated in **Figure 2.3**. For this optimised differentiation method, the DE specification was performed in basal DIFF1 medium, adding the relevant growth factors on the day of feeding. On day 0 of differentiation, 48hrs after seeding, the HBE8 medium was changed to basal DIFF1 medium supplemented with 3 μM CHIR99021, 100 ng/mL ActA and 10 μM LY294002 (Selleck #S1105). The next day, basal DIFF1 medium was supplemented with 100

ng/ml ActA and 10 μM LY294002. On days 2 and 3, basal DIFF1 medium was only supplemented with 100 ng/ml ActA. Endoderm yield efficiency was determined on day 4 of the differentiation by dissociating the monolayer and evaluating CXCR4 and c-KIT co-expression by flow cytometry (refer to the methods section for performing FACS on endoderm).

In order to proceed with the differentiation, cells must have shown at least 80% endoderm yield efficiency. Having confirmed a good yield efficiency, DE cells were ready for clump passaging and anterior foregut endoderm (AFE) induction. After that, the subsequent steps of the differentiation were performed in supplemented basal Differentiation-2 (DIFF2) medium, a serum-free differentiation medium comprising Iscove's Modified Dulbecco's Medium (IMDM; Gibco #12440053), Ham's F12 medium (Corning #10-080-CV), GlutaMAX (Gibco #35050061), B-27 supplement, 7.5% BSA (Gibco #15260037), N2 supplement (Gibco #17502-048), L-Ascorbic Acid 2-phosphate (AA2P, 50ug/ml; Sigma #A4544), diluted α-Monothioglycerol (MTG; Sigma #M6145) and 1% P/S with daily media changes.

First, 12 ml of anteriorisation medium was prepared by supplementing basal DIFF2 medium with 10% CloneR2 (Stemcell #100-0691), 2μM Dorsomorphin (DS; Tocris #3093), 10uM SB431542 (SB; Selleck #S1067) and 1μl/mL Y-27632. Next, cells were gently washed with 1 ml of DMEM/F12 (Gibco #11330-057), and 1 ml of Gentle Cell Dissociation Reagent (GCDR; Stemcell #100-0485) was added per well and incubated at RT for 2-3min. Before the cells started to detach, the GCDR was gently aspirated, trying not to disrupt the underlying monolayer. Then, using a P1000 and without resuspending, 1ml of DS/SB + Y-27632 supplemented DIFF2 medium was added to each of 4 wells to collect the detaching clumps of cells and transferred into a 30 ml universal tube. It was essential to preserve the cells as clumps; therefore, this step was carefully repeated without resuspending much until the 12ml of medium was used and the cells detached entirely from the plate surface. Collected cells were then plated on a Matrigel pre-coated 12-well plate at a 1:3 dilution making sure cells were homogenously distributed on the well. Next, cells were incubated in anteriorisation medium for 72 hrs, refreshing the medium every 24 hrs without Y-27632. On day 7 of the differentiation, to induce the specification of lung progenitor cells from AFE, cells were switched to basal DIFF2 medium supplemented with 3 μM CHIR99021, 10 ng/mL BMP4 100 nM RA and cultured with the same medium with daily changes until day 15 of the differentiation. At the end of the lung progenitors' induction, the culture was

purified for NKX2.1^{+} epithelial cells using FACS for the surrogate cell membrane marker CPM (refer to the methods section for performing FACS on lung progenitors).

The anteriorisation and lung progenitors' induction stages were also performed using supplemented basal Differentiation-3 (DIFF3) medium, comprising IMDM, Ham's F12, PVA (1mg/ml; Sigma #P8136), ITS-X (Corning #51500056), Lipids (Gibco #11905031), GlutaMAX, AA2P, MTG, NEAA, and 1% P/S with daily media changes. Growth factors and small molecules were added fresh on the day of feeding and the rest of the steps were performed as per the optimised differentiation methodology unless otherwise stated.

Throughout the differentiation and maturation induction, cells were fixed for immunocytochemistry, and total RNA was harvested at specific time points to be stored at -20°C until assayed. Gene expression was measured in cell lysates by reverse-transcriptase polymerase chain reaction (RT-qPCR) and/or bulk mRNA sequencing.

2.1.6 Generation and maintenance of alveolar organoids

Generation and passage of organoids as clumps

After cells were induced into lung progenitors with basal DIFF1 medium supplemented with CHIR99021 (1μM), BMP4 (5 ng/mL) and RA (250nM) for 9 days, cells were embedded in 3D Matrigel droplets to promote their maturation into distal lung epithelial organoids. First, cells were washed and collected in 1ml of DPBS using the flat tip of a 5ml stripette to scrape the monolayer from the bottom of the well carefully and a p1000 to transfer the cell clumps into a 1.5ml Eppendorf tube to be centrifuged at 2,000 rpm for 1 min. Next, the cell pellet was resuspended in cold basal DIFF1 medium supplemented with the maturation factors (CHIR99021, FGF10, DEXA, cAMP, IBMX) and Y-27632 to be embedded in a cold mix of 25% medium and 75% phenol red-free Matrigel (Corning #356231). Then, using a cold P1000 pipette tip, the cell suspension was gently mixed with the thawed phenol-red free Matrigel. Cells from a single 6-well-plate well generated eight droplets of approximately 50µl. Droplets of the Matrigel-media-cell mixture were dropped into the middle of the well of a pre-warmed 48-well plate and left to polymerise in the incubator at 37°C for 10 min. Once the droplets solidified, 300ul of warm lung maturation medium plus Y-27632 were added per well, replacing if after 24hrs to remove Y-27632. Following this, the medium was replaced every 48 hrs.

To passage the organoids as fragments, the Matrigel droplets were mechanically dissociated from the bottom of the wells using the tip of a 5 ml stripette, collected in a 15 ml conical tube containing 5 ml of cold PBS and centrifuged at 2000 rpm for 1 min. The organoid pellet was repeatedly washed in cold PBS to remove most of the Matrigel. After the last wash, using a p200, the remaining pellet was pipetted up and down to further fragment the organoids before re-embedding them in Matrigel droplets with maturation medium and Y-27632 following the abovementioned steps. With the optimisation, the addition of 10 ng/ml FGF7 (Peprotech #100-19) was adopted as part of the maturation factors used for organoid culture.

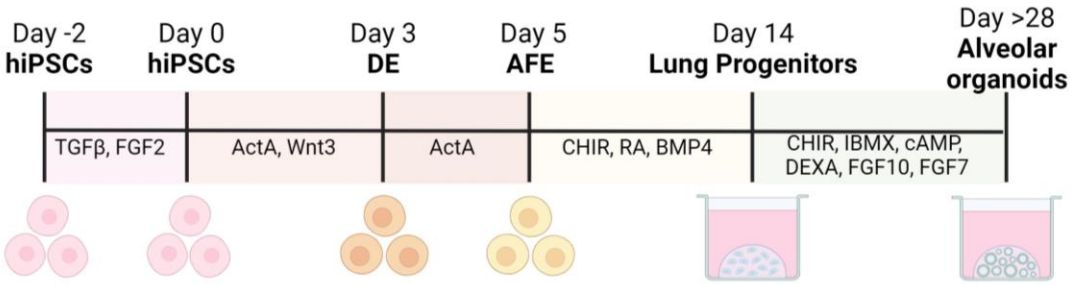


Figure 2.2 Schematic diagram of the original stepwise strategy to induce hiPSC differentiation into lung progenitors and their maturation into alveolar organoids.

Single cell-derived organoids

After collecting the CPM^{+high} lung progenitor cells or the $EpCAM^{+}/NaPi2b^{+}$ alveolar type II cells from the sorter, they were embedded in 3D Matrigel droplets and cultured with basal DIFF2 medium supplemented with 3 µM CHIR99021, 50 nM DEXA, 0.1 mM Camp, 0.1 mM IBMX, 10 ng/ml FGF7, 10 ng/ml FGF10 plus 10 µM Y-27632 and 10% ClonR2 to promote their survival and maturation into lung epithelial organoids.

The volume of medium and Matrigel required to embed the cells into droplets was calculated based on the number of cells recovered from the purification by FACS. Each droplet comprised a mixture of cells in 25% medium, and 75% phenol red-free Matrigel. Approximately 400 cells were seeded per ul of undiluted Matrigel droplet, giving a seeding density of 20,000 cells per 50ul droplet. After preparing the medium, an aliquot containing the desired medium volume to make up for 25% of the droplets was transferred to a 1.5 ml Eppendorf tube and placed in the fridge while the collected cells were centrifuged for 5 min at 300g. The cell pellet was resuspended in the aliquoted medium, and a cooled P1000 pipette tip was used to mix it with the desired volume of Matrigel carefully, ensuring no air bubbles were introduced. Next, a pre-warmed 48-well plate was used to plate 50ul droplets in the middle of the wells and transferred

carefully to a 37°C incubator for at least 10 min to allow the Matrigel to polymerise. Finally, 300ul of warm maturation medium was added per well. After 48 hours, the organoids were fed with fresh maturation medium minus Y-27632 and CloneR2. Thereafter, media was changed every other day, and the organoids were passaged when confluent, approximately every 10-14 days.

Passaging of organoids as single cells

To passage the organoids as single cells or to process them for further CPM^{+high} progenitors or EpCAM^{+}/NaPi2b^{+} AT2 cell enrichment, the 3D droplets and the containing organoids were dissociated enzymatically. Medium was removed from the wells, and 1ml of 2mg/ml Dispase II (Dispase; Gibco #17105-041) in DMEM/F12 was per well and incubated at 37˚C for 1 hour to release the lung organoids from the Matrigel, pipetting twice with a p1000 halfway through the incubation time. The organoids were collected, transferred to a 15 or 50-ml conical tube containing DMEM/F12, and centrifuged at 200g for 5min. The pellet was resuspended with 1ml of 0.05% Trypsin/EDTA (Gibco # 25-300-062) per well and incubated at 37˚C for 10-15min. Single-cell dissociation was verified under the microscope; if a single suspension was not obtained, cells were washed and incubated additional 5 min with 0.05% Trypsin/EDTA. Next, trypsinised single cells were washed with 10ml of Trypsin-stopping medium (5 ml of FBS HI (Gibco # 10500064) in 45 ml of DMEM (Gibco, #21969-035) to neutralise the enzyme. Following centrifugation at 300g for 5min, cells were washed once with DMEM/F12 before preparing suspension for cell counting using a haemocytometer. At this point, cells were either embedded in fresh Matrigel droplets as described in this section or prepared for purification via FACS (refer to the methods section for performing FACS).

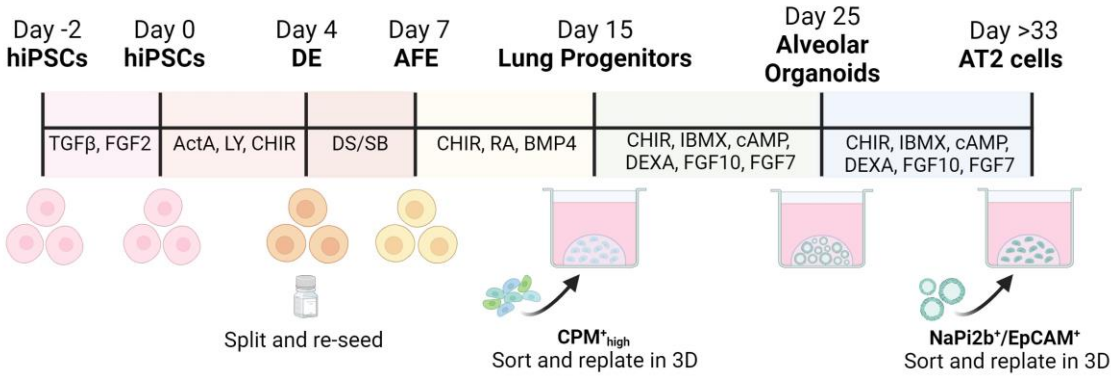


Figure 2.3 Schematic diagram of the optimised protocol to generate lung progenitors and their maturation into AT2 cells including enrichment steps.

2.1.7 Immortalised cell lines culture

Immortalised Human Bronchial Epithelial Cells (iHBECs)

iHBECs are immortalised cells derived from human bronchial epithelial cells that retain the ability to differentiate into ciliated, basal and goblet cells, as their primary counterpart (iHBECs were a gift from Prof. Jerry Shay, University of Texas, USA) [172]. To culture the cells, a vial containing $1x10^6$ cells was retrieved from liquid nitrogen and thawed at 37°C in a water bath. Afterwards, the cells were transferred to a T75 flask containing 15-20ml of keratinocyte serum-free medium (KSFM) supplemented with 0.2ng/ml epithelial growth factor (EGF; Gibco), 25µg/ml bovine pituitary extract (BPE; Gibco), 250ng/ml puromycin (Sigma) and 25µg/ml Gentecin (G418; Sigma), these last two were used for the positive selection of immortalised cells. The flask was then incubated at 37°C, in 5% CO2 in a humidified incubator, and then, 24 hours after, the media was aspirated off and replaced with fresh supplemented KSFM (KSFM+) to remove the remaining DMSO of the freezing media. Cells were cultured in KSFM+, changing the medium every three days and splitting when confluent. For experimentation purposes, 24 hours before the experiment, cells were growth arrested by changing the medium to supplement-free KSFM.

When confluent, cells were split by trypsinisation. First, the media was aspirated, and cells were washed with sterile PBS. Afterwards, 5 ml of 0.5% trypsin/EDTA were added to the flask and placed in the incubator for 3-5 minutes. Once the cells were detached from the flask, the trypsin was neutralised by adding 5 ml of 100% Foetal Bovine Serum (FBS). Cells were transferred to a sterile tube and centrifuged for 5 minutes at 1500 rpm. The supernatant was aspirated, and the remaining pellet was resuspended in KSFM+. Finally, cells were seeded in a range of 1:3 to 1:10 ratio by transferring the desired amount of cell suspension to a new T75 flask containing 15-20ml of prewarmed KSFM+. When setting up for an experiment, the cells were instead seeded in the desired multi-well plate format, changing the medium every three days until confluence.

Transformed Mink Lung Epithelial Cells (TMLCs)

TMLCs are a reporter cell line transfected with a TGFβ responsive portion of the plasminogen activator inhibitor (PAI-1) promoter fused to the firefly luciferase reporter gene and a neomycin resistance gene. Based on the ability of TGFβ to induce PAI-1 expression, this cell line can be used as a quantitative bioassay for TGFβ activity [173]. TMLCs were a gift from Prof. Dan Rifkin, University of New York, USA. Cells were

cultured in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% FBS, supplemented with 4mM L-glutamine and 50µg/ml G418 and incubated at 37 °C in 5% CO_2, changing the medium every three days and splitting when confluent as previous.

Madin Darby Canine Kidney Cells (MDCKs)

MDCK cells are an immortalised epithelial cell line derived from the kidney of a healthy adult Cocker Spaniel. These cells are commonly used as epithelial models due to their clear apico-basal polarity, tight junctions and rapid growth rate. They are mainly used in virology for titration of viral stocks and influenza virus isolation as they are permissive and support the growth of this virus, maintaining its infectivity [174]. MDCKs were maintained in culture in Eagle's Minimum Essential Medium (MEM) supplemented with 10% FBS and incubated at 37°C in 5% CO2 in a humidified incubator. When confluent, cells were split by trypsinisation, as above, adding 2 x PBS washing steps before adding the trypsin/EDTA. For their use in plaque assays, cells were growth arrested in MEM without FBS for 24 hours prior to the experimental set-up.

All cell lines were tested to be free of mycoplasma infection before liquid nitrogen storage and while in culture.

2.1.8 Flow cytometry analysis and fluorescence activated cell sorting (FACS)

Flow cytometry analysis was performed on an FC500 Flow Cytometer (Beckman Coulter), ID7000 Spectral Flow Cytometer (Sony Biotechnology) or ImageStreamX MkII Imaging Flow Cytometer (Luminex) and cell sorting was performed on a Moflo Astrios EQ Cell Sorter (Beckman Coulter) at the University of Nottingham Flow Cytometry Core Facility with the assistance of David Onion and Nicola Croxall. Data analysis was completed using Kaluza Analysis Software 2.1 or IDEAS® Image Data Exploration and Analysis Software. All analysis and/or sorting gates were established based on the unstained control to eliminate background levels of fluorescence and on the isotype/ secondary-only controls to eliminate non-specific signals. Sorting gates were used to exclude debris (side scatter (SSC) versus forward scatter (FSC)) and doublets (pulse width versus FSC), to include live cells (Propidium iodide versus FSC) and finally, to include positively stained cells (marker dependant).

Definitive endoderm FACS

Flow cytometry analysis on day 4 of definitive endoderm was performed to ensure the induction efficiency was above 80% via expression of C-KIT and CXCR4. Culture medium was aspirated from 2 wells, and 500µl of DMEM/F12 was used to wash each

well before adding 500ul of Gentle Cell Dissociation Reagent (GCDR; Stemcell Technologies #07174.) After a 4-5 min incubation at 37˚C, the GCDR containing floating endodermal cells was collected and transferred to a 15ml conical tube. Any remaining cells were also collected by washing the plate with 1ml of DMEM/F12. Cells were pelleted by centrifugation for 5min at 300g and resuspended in 600ul of staining buffer (1ml of FBS in 49ml of PBS, no calcium, no magnesium; Gibco, #14190144). The cell suspension was divided equally into five 1.5ml Eppendorf tubes (samples: C-KIT-PE, CXCR4-APC, and C-KIT-PE/CXCR4-APC; -PE and -APC isotypes; and unstained control) to be stained with 2ul of antibody or isotype (refer to **Table 2.3** for details). Tubes were finger-flicked to mix cells properly and incubated on ice for 30min in the dark, flicking the tubes halfway through the incubation. Cells were washed twice with 1ml of staining buffer and centrifuged for 5 minutes at 300g. The supernatant was removed, and the cells were resuspended in 350ul of staining buffer to perform FACS analysis.

Lung progenitors and AT2 cells FACS

To assess the efficiency of the lung induction stage and to further purify the CPM^{+high} lung progenitors' population, a single-cell suspension of lung day eight cells was processed for FACS. Cells were washed with PBS, and 500ul of warm 0.05% trypsin-EDTA was added per well, followed by the gentle scratching of the monolayer in a crosshatch pattern using a p200 pipette tip and incubation at 37°C for 10 min. Cells were pipetted two to three times and visualised under the microscope to ensure cells were dissociated as single-cell; if not, cells were incubated for an additional 5 min. After 15 min, 500ul of Trypsin-stopping medium was added per well, and trypsinised cells were pipetted and collected into a 15ml conical tube. Following centrifugation at 300g for 5min, the pellet was resuspended in DMEM/F12, and single cells were counted. The required number of cells in suspension were transferred to each of three tubes (sample, secondary-only and unstained control), and washed again with 10ml of DMEM/F12 before being stained.

For the isolation of CPM^{+high} lung progenitors or $EpCAM^{+}/NaPi2b^{+}$ AT2 cells populations from 3D lung organoids, the embedding Matrigel droplets were dissociated, as per the passaging methodology, by incubating with dispase for 60 minutes followed by single-cell dissociation by incubation for 10-15 min with 0.05% Trypsin/EDTA. The obtained suspension was counted for single cells under the microscope. The desired number of cells was distributed into conical tubes in preparation for the staining of the sample and the secondary-only and unstained

controls. Cells were washed one last time with 10ml of DMEM/F12 before proceeding with the staining.

Anti-human CPM or EpCAM/NaPi2b primary antibodies diluted in 500-1000ul of filtered staining buffer were added to the cell pellet and incubated for 15min at 4˚C (dilutions used are indicated in **Table 2.3**). Cells were washed once with 1 ml of 1% FBS/PBS and centrifuged for 5 minutes at 300g. The respective fluorochrome-labelled secondary antibodies were added to the sample cells and the secondary-only control at a dilution of 1:300 in staining buffer for 15min at 4˚C away from the light. Stained cells were washed once by centrifugation and resuspended in 500-1500ul of filtered sorting buffer (200 µl of FBS HI in 10 ml of HBSS (Hank's Buffered Saline Solution, no calcium, no magnesium, no phenol red; Gibco #14175095) containing freshly added Y-27632 and live/dead stain, propidium iodide (PI; Sigma #P4170), at a 1:1000 dilution. Finally, each cell suspension was filtered through a 40µm filter into round-bottom polypropylene tubes to sort CPM^{+high} or $EpCAM^{+}/NaPi2b^{+}$ cells.

Table 2.3 Details of primary and secondary antibodies used in this thesis.

Target protein	Species	Dilution	Manufacturer	Catalogue #	Use
Primary Antibodies					
NKX2.1/ TTF-1	Rabbit	3 µg/mL	Abcam	ab76013	ICC
NKX2.1/ TTF-1	Rabbit	1.70 µg/mL	Abcam	ab227652	ICC
NKX2.1/ TTF-1	Mouse	2 µg/mL	Fisher Scientific	MS-699	FC
PRO-SFTPC	Rabbit	20 µl/mL	Abcam	ab170699	ICC
PRO-SFTPC	Rabbit	1 µg/mL	Abcam	ab90716	ICC
PRO-SFTPC	Mouse	2.6 µg/mL	Santa Cruz	sc-518029	ICC
NANOG	Goat	4 µg/mL	R&D	MABF983	ICC
OCT4	Mouse	10 µg/mL	R&D	MABF984	ICC
GATA4	Mouse	10 µg/mL	Santa Cruz	sc-25310	ICC
GATA6	Mouse	4 µg/mL	R&D	MAB1700	ICC
SOX2	Mouse	4 µg/mL	R&D	MAB2018	ICC
SOX17	Mouse	4 µg/mL	R&D	AF1924	ICC
SOX9	Goat	2.6 µg/mL	Santa Cruz	sc-166505	ICC
FOXA2	Goat	4 µl/mL	R&D	AF2400	ICC
ECAD	Goat	0.8 µl/mL	R&D	AF748	ICC
CXCR4	Mouse	0.8 µl/mL	Sigma	MABF981	ICC
CD63	Mouse	10 µg/mL	Santa Cruz	SC-5275	ICC
CALNEXIN	Mouse	10 µg/mL	Santa Cruz	SC-23954	ICC
TGN38	Mouse	8 µg/mL	Novus	NB300-575SS	ICC
ALBUMIN	Mouse	0.52 µg/mL	R&D	MAB1455	ICC
TTR	Mouse	0.6ul/1x106 cells	R&D	MAB7505	ICC

AFP	Rabbit	0.5 µg/mL	Abcam	ab133617	ICC
CPM	Mouse	2.5 µl/mL	Fujifilm Wako Chemicals	014-27501	FC/ICC
CPM	Mouse	0.8 µl/1x106 cells	Origene	TA807364	FC/ICC
NaPi2b (SLC34A2)	Rabbit	1 µl/1x106 cells	Cell Signaling	66445	ICC
NaPi2b (SLC34A2)	Mouse	1.25 µl/1x106 cells	Ludwig Institute for Cancer Research, NY	MX35	FC/ICC
NaPi2b (SLC34A2)	Rabbit	2 µl/100 µl	NSJ Bioreagents	RQ4878	FC/ICC
EpCAM	Goat	2 µl/100 µl	R&D	AF960	FC
C-KIT (CD117)-APC	Mouse	2 µl/100 µl	Life Technologies	CD11705	FC
CXCR4 (CD184)-PE	Mouse	2 µl/100 µl	Life Technologies	MHCXCR404	FC
IgG1 isotype-PE	Mouse	3 µg/mL	BioLegend	400113	FC
IgG1 isotype-APC	Mouse	1.70 µg/mL	Life Technologies	MA5-18093	FC
Secondary Antibodies					
Anti-Mouse IgG AF 647	Donkey	4.6 µg/mL	Jackson Immunoresearch	715-605-150	FC
Anti-Mouse IgG AF 647	Donkey	2 µl/mL	Invitrogen	A31571	ICC
Anti-Mouse IgG AF 488	Donkey	2 µl/mL	Invitrogen	A21202	ICC
Anti-Rabbit IgG AF 647	Donkey	2 µl/mL	Invitrogen	A31573	ICC
Anti-Rabbit IgG AF 488	Donkey	3.3 µl/mL	Invitrogen	A32790	FC
Anti-Rabbit IgG AF 488	Donkey	2 µl/mL	Invitrogen	A21206	ICC
Anti-Goat IgG AF 488	Donkey	2 µl/mL	Invitrogen	A11055	FC
Anti-Goat IgG AF 647	Donkey	2 µl/mL	Invitrogen	A21447	ICC

Base editing of hiPSCs

2.1.9 Guide RNA and base editors' constructs

Base editing tools consist of a base editor plasmid and an independent single guide RNA plasmid. Base editing was used to modify IPF-patient-derived iPSCs containing a heterozygous surfactant protein C mutation c.338A>G (Y113C). To perform base editing, two different systems were used: pCMV_AncBEmax-P2A-GFP to correct the mutation and generate the wild-type cell line $SFTPC^{WT/WT}$ (G>A (only A)), and pCMV_ABEmax-P2A-GFP to introduce the mutation in the other allele and generate the mutant cell line $SFTPC^{MUT/MUT}$ (A>G (only G)). The advantage of these systems

over the standard CRISPR-Cas9 technology is that they allow for the direct conversion of one base to another without creating double-strand breaks in the DNA.

The oligo1 and oligo2 for each gRNAs,1 and 2 (details in **Table 2.4**), targeting the specific DNA sequences, were annealed, and cloned into a U6-gRNA backbone (~5kb) containing an ampicillin resistance cassette driven by an Amp promoter and a puromycin resistance cassette driven by a PGK promoter. The base editor plasmids contain the corresponding deaminase preceded by a CMV promoter and followed by an eGFP cassette. The base editors also contain an ampicillin resistance functional cassette.

Table 2.4 Guide RNAs and oligo targeting sequences.

gRNAs (name)	Sequence (5' --> 3')	Made by
gRNA1 BE4corrY113-Oligo1	**CCGG**Gcttgcaggcgatcagcagct	gRNA1 BE4corrY113-Oligo1guide gRNA1 BE4corrY113-Oligo2
gRNA1 BE4corrY113-Oligo2	**AAAC**agctgctgatcgcctgcaagC	-
gRNA2 ABEintrY113C-Oligo1	**CCGG**Gatcgcctacaagccagcccc	gRNA2 ABEintrY113C-Oligo1guide gRNA2 ABEintrY113C-Oligo2
gRNA2 ABEintrY113C-Oligo2	**AAAC**ggggctggcttgtaggcgatC	-

For plasmid transformation, NEB® 5-alpha Competent E. coli (High efficiency) kit was used (NEB; #C2987H). 50ng of plasmid DNA was incubated with 50ul of E. Coli on ice for 30min, heat-shocked at 42°C for 90sec in a heat block, followed by a cold shock on ice for 5min. Next, 250ul of lysogeny broth (LB) medium was added and the tube was placed in a shaking incubator at 37°C for 1 hr x 225rpm. After incubation, 250µl medium was plated onto LB agar plates with 100µg/ml ampicillin. The plate was incubated at 37°C overnight, and viable independent colonies were selected. Plasmid DNA was extracted and eluted in endotoxin-free water, and the concentration was measured using a NanoDrop-1000 spectrophotometer (Thermo-Fisher).

2.1.10 hiPSCs transfection optimisation and antibiotic enrichment

IPF-patient-derived hiPSC IPF1 was nucleofected using Amaxa™ 4D-Nucleofactor and P3 primary cell kit (Lonza; #V4XP-3024) or Ingenio® electroporation solution. Nucleofection mixtures were prepared with P3 buffer or Ingenio® electroporation solution containing the plasmids of interest (eGFP plasmid for transfection optimisation and/or transfection control and pguideRNA and pCMV_ABEmax_P2A_GFP or pCMV_AncBE4max_P2A_GFP were used for the experiments) in 20µl total volume.

On the day of nucleofection, confluent cells were dissociated into single cells using TrypLE and centrifuged at 300 x g for 4 min. Cell pellet was gently resuspended in the nucleofection mixture and transferred to a Nucleocuvette™ nucleofection cuvette. Each transfection condition used 0.6x106 cells. CA-137, CB-150, CD-118, CE-118, CM-113, DC-100, and DN-100 nucleofection programs were applied to the cuvettes in the 4D-Nucleofactor™ X unit as previously optimised by Sara Cuevas Ocaña [175]. Transfected cells were then gently plated onto Matrigel® a pre-coated well (24 well plates) prepared with pre-warmed E8 media containing 10μM Y-27632 and incubated at 37°C. 24 hours post-nucleofection, cells were fed with fresh pre-warmed medium. Transfection efficiency was monitored by GFP visualisation with fluorescent microscopy at 24h and 48h post nucleofection and by flow cytometry at 48h post nucleofection.

The transfected hiPSCs were subjected to transient puromycin selection 24 hours post-nucleofection using 0.7μg/ml of puromycin dihydrochloride (Gibco, #A1113802) to enrich transfected cells. Following enrichment, the remaining cells were dissociated into single cells and seeded at clonal density. Once the single-cells generated visible defined colonies (~500μm), each colony was manually dissected using a stem cell knife (Invivogen, #14601), using a portion of the colony for clonal expansion and the rest for direct genomic DNA (dgDNA) extraction for further PCR-based genotyping.

2.1.11 Direct genomic DNA extraction and genotyping

Direct genomic DNA (dgDNA) was extracted for PCR-based screening of human iPSCs colonies using the Phire Tissue Direct PCR Master Mix (ThermoFisher, #F170S) as previously optimised [7]. Briefly, a small colony fragment was collected with 10ul of media in a PCR tube. 20ul of dilution buffer and 0.5ul of DNA release mix were added per sample, mixed by vortexing for 30sec and incubated at room temperature (RT) for 2min. Sample was then heated at 98C for 3min using a thermocycler, spun for 30sec, and stored at -20°C until further use. Prior to using the gDNA for PCR, each stored gDNA sample was properly mixed by pipetting 5 times, and sample was extracted from the centre of the mixture to be used in the PCR reaction.

PCR genotyping was performed using the Phire master mix (MM) included in the Phire Tissue Direct PCR Master Mix kit. The PCR reaction was set up by combining 1μl of gDNA sample with 0.5μl of each primer, 5μl of the Phire Master Mix, and 3μl of nuclease-free ddH2O. Primers used for genotyping are listed in **Table 2.1**. The PCR programme used for genotyping was as follows: initial denaturation at 98°C for 5min;

denature template at 98°C for 5sec, primer annealing at 60-71.4°C (depending on primers) for 15sec and extension at 72°C for 1 min (40 cycles); final extension at 72°C for 1 min; hold at 10°C. After DNA amplification, DNA samples were loaded in 1% (w/v) agarose gels (Sigma-Aldrich, #A7431) stained with 0.5ug/ml ethidium bromide (Invitrogen; #15585011). Agarose gel electrophoresis was carried out in 1X TAE buffer at 130V for 35 min, and LAS-4000 imaging system (Fujifilm Life Science) was used for UV visualisation of the gels.

Purified PCR products and plasmids were sent to Source BioScience, Nottingham for Sanger sequencing. Purified plasmids were provided at a final concentration of 100ng/µl, in 5µl of sample, and the PCR products were provided at a final concentration of 1ng/µl per 100 bp in 5µl of sample as per manufacturer recommendations. The sequencing results were analysed using SnapGene Viewer 6.2.2.

All the base editing of hiPSCs mentioned above was performed at The University of Nottingham by Dr. Sara Cuevas Ocaña according to the previously described work [7].

Table 2.5 Primers' sequences used for genotyping base-edited hiPSC cell lines.

Primers (name)	Size (bp)	Sequence (5'→ 3')
Fw_SFTPC	20bp	aaccaggcagcaacccagct
U6_Fw	23bp	tacgatacaaggctgttagagag

2.2 In vitro infection

2.2.1 LPS infection of hiPSC-derived alveolar epithelial cells

Presumable AT2 cells were exposed to Lipopolysaccharides from Pseudomonas aeruginosa to model a distal airway infection. AT2 cells grown in the air-liquid interface cultured in maturation medium were infected by aspirating the media from the bottom well and replacing it with RPMI medium containing 1µl/ml of LPS (Sigma-Aldrich) for 30, 60, 120 minutes, 24 and 48 hours before lysing them.

Whole-mount AT2 organoids embedded in Matrigel droplets were infected by aspirating the maturation media, washing the wells once with 300ul of PBS and adding 300ul of RPMI medium with 1µl/ml of LPS for 30, 60, and 120 minutes. Uninfected controls and infected organoids were harvested by scratching the Matrigel droplet with a p1000 pipette tip and resuspending it in cold PBS in a 15ml Falcon tube to help dissolve the Matrigel. Cells were shortly centrifuged; the supernatant was removed, and the remaining pellet was lysed for RNA processing.

Gene expression of pro-inflammatory mediators and ECM components were measured in cell lysates by reverse-transcriptase polymerase chain reaction (RT-qPCR), and protein phosphorylation downstream TLR4 was assessed using a membrane-based antibody array.

2.2.2 H1N1 Influenza virus propagation and titration

A series of methods were followed in order to assess influenza A H1N1 virus' viability and infectivity. To do so, the virus concentration in the generated stocks (viral titre) was measured using the haemagglutination and plaque assays and the Viral Tox-Glo™ assay.

Replication of H1N1 virus in cell culture

MDCK cells were grown to confluency in MEM media supplemented with 10% FBS and 2 mM L-glutamine at 37° C. The medium was removed for the experiment, and the cells were washed with PBS. 25 ml of serum-free MEM supplemented with 1 µg/ml TPCK-trypsin was added to the flask. 100 µl of 5 HAU/µl influenza A virus obtained from Public Health England (A/PR/8/34 H1N1) was added to the flask and swirled to ensure uniform distribution. The flask was incubated for three days at 37° C without media changes to allow the virus to infect the MDCK cells and expand until cytopathic effect (CPE) was visible. After 72hrs, the supernatant was harvested, aliquoted into 1.5ml tubes and labelled adequately before placing it at -80°C for storage. Viral stocks were always kept at -80°C, and freeze/thaw cycles were avoided as this could reduce the infectivity of the virus.

Haemagglutination assay

An aliquot of viral stock was thawed on ice for viral titre determination by HA assay. To avoid any effects of the anticoagulant solution contained in the 5% chicken red blood cells stock (Innovative Research), the required amount was washed twice with PBS before diluting it. 0.5%, 1% and 2% dilutions of chicken red blood cells in PBS were prepared.

Using a V-bottom 96-well plate, 50µl of PBS was dispensed into each well of the experimental layout. After flicking the viral stock tube to mix it, 50µl was added to the first well of each row of column 1. The contents of the wells were mixed by pipetting up and down slowly, avoiding the generation of bubbles. Using a multichannel pipette, 50µl from the first well was transferred to the second well making a two-fold serial dilution; this step was repeated across the entire plate until column 12, dispensing each

dilution in triplicates. Next, 50µl of red blood cells' dilution (0.5%, 1% or 2%) was added to each well, including control wells that contained only PBS and red blood cells. The plate was covered and gently tapped on the sides to mix before incubating at 4° C for 45 minutes. The plate was then examined for the formation of red 'buttons' or faint lattices.

Plaque assay

For the plaque assay, a confluent monolayer of MDCK cells was infected with serial dilutions of H1N1 of unknown starting concentration. In preparation for the assay, a confluent T-150 flask of MDCK cells grown in 10% FBS and 2 mM L-glutamine in MEM medium were trypsinised, the enzyme was neutralised with FBS and gently centrifuged. Cells were counted and resuspended in maintenance medium to prepare a suspension at a density of $2.25x10^5$ cells/ml. MDCK cells were seeded into 3 x 6-well cell culture plates at a density of $4.5x10^5$ cells/well by adding 2 ml of the suspension to each well. The plates were then incubated at 37° C in a 5% CO2 incubator to achieve confluency. Infection was performed once cells had reached 95% to 100% confluency (usually between 48 and 72 hours after seeding).

Once cells had grown into a monolayer, 10-fold dilutions of the virus stock were prepared in serum-free MEM containing 1ug/ml of TPCK-trypsin. Starting with 4mL of a 1:100 dilution, the viral stock was serially diluted 10-fold by transferring 400uL of the diluted virus to 3.6 mL of sterile medium, mixing well by vortexing, and transferring 400 uL of the dilution to the subsequent tube. The range of dilutions tested was 10-2 to 10-6. Cells were inoculated using three wells for each dilution of viral stock being tested. After aspirating the medium from the cell monolayer and washing each well with 1mL of PBS, 1mL of the designated viral dilution was added per well. This critical step needed to be performed rapidly to prevent the monolayer from drying. The virus was then allowed to attach to the cells incubating the plates at 37° C for 2 hours, rocking every 15 min.

During this incubation period, the solid agarose overlay was prepared using a 125mL bottle to mix 2% (w/v) Sea plaque low melting temperature agarose (Lonza #50101) in DPBS. The bottle was autoclaved to sterilise the agarose and left to cool down to ~60°C. Just before using it, the liquid agarose was mixed with an equal volume of pre-warmed EMEM (Eagle's minimum essential medium) and 1ug/ml of TPCK-trypsin, enough to cover each well with 3 mL. The overlay was then swirled to obtain an even mixture and cooled to about 40–42 °C. When the 2hr-incubation was over, cells were

removed from the incubator, and viral inoculum was aspirated. Immediately after, 3mL of the solid overlay was gently added down the edge of each well, making sure to cover the entire monolayer and avoid the formation of bubbles. The plates were left for ~30 min at room temperature to let the agarose set and then placed at 37°C to incubate for 3-7 days.

During this period, the plates were checked daily for cytopathic effects and for developing plaques that appeared as zones of cell death surrounded by an intact monolayer. When these were detected, in order to enhance the contrast between the plaques/ zones of cell death and the monolayer, 2 mL of a secondary overlay with staining was added to the wells. The remaining autoclaved 2% agarose in DPBS was microwaved to make this overlay. Once in liquid form, it was mixed with an equal amount of warm EMEM with 0.01% neutral red (diluted from a previously prepared stock of 0.33% (w/v) neutral red (ThermoFisher Scientific #AC415491000) in 100 ml distilled water). The overlay was allowed to solidify before incubating the plates at 37°C overnight. The next day, using a lightbox to aid in visualising the contrast, the plates were examined for clear plaques in the staining, as the neutral red was meant to be taken up by the healthy/ uninfected in the monolayer and not the dead cells. Considering the dilution factor and plating volume, the count of these plaques could be used to calculate the titer of the original viral stock.

Viral Tox-Glo assay

MDCK cells maintained in MEM with 2 mM L-glutamine and 10% FBS at 37° C in a 5% CO2 incubator were harvested from a flask using trypsin-EDTA. Cells were passaged as needed, counted, and resuspended in MEM at a density of $3x10^5$ cells/ml. Using an opaque-walled, clear-bottom 96-well cell culture plate (Greiner bio-one #655090), $3x10^5$ MDCK cells were seeded per well by adding 100µl of the cell suspension to each well. The cells were then incubated at 37° C for 24 hours for attachment and recovery from seeding stress.

The next day, virus stock was diluted in MEM for a final stock dilution of 1:10 and 1:100. Internal controls consisted of untreated wells containing cells only and blank wells containing media only. Using a multichannel pipette, cells in column 1 were infected with 46ul of the initial viral dilution in triplicates, 1:10 or 1:100; this volume was mixed by pipetting five times. After changing tips to avoid viral carryover, 46ul were aspirated from column 1, transferred to column 2 and appropriately mixed. This step was repeated all the way through column 10 to perform half-log (3.16-fold) serial dilutions.

After mixing the contents in column 10, 46ul were removed and dispensed to waste. Column 11 was used as the no-virus control, and column 12 served as the no-cell control. Cells were incubated with the virus at 37° C in a 5% CO2 incubator for 72 hours.

After the incubation period, the ATP detection reagent from the Viral ToxGlo Assay (Promega #G8942) was prepared per the manufacturer's recommendations. The thawed ATP detection buffer was transferred into the ATP detection substrate's bottle and gently vortexed to create a homogeneous solution. Being careful not to introduce bubbles, 100ul of ATP detection reagent was added to each well. The plate was placed on an orbital shaker for 30 minutes before measuring the luminescence on the plate reader (BMG Labtech FLUOstar Omega).

2.2.3 TMLC and iHBECs co-culture assay

In order to measure TGFβ activation in iHBECs, these cells were co-cultured with reporter TMLCs. In preparation for the assay, iHBECs grown in KSFM with L-glutamine supplemented with EGF, BPE, Puromycin and G418 (KSFM+) were trypsinised for 5 minutes, and FBS was used to neutralise the trypsin-EDTA. Cells were then centrifuged and gently resuspended in KSFM+ to be seeded in a 96-well cell culture plate at a density of $1x10^5$ cells/ml. Once confluence, usually three days after, the medium was changed to supplement-free KSFM to growth arrest the cells for 24 hours before the experiment.

On the day of the experiment, TMLCs cultured in DMEM with 10% serum, L-glutamine, pen/strep, amphotericin B and G418 were harvested by trypsinisation, counted and resuspended at a density of $0.5x10^6$ cells/ml. 100ul of this cell suspension was plated directly on top of iHBECs wells. Cells were incubated at 37°C with 5% CO2 for 1 hour to allow TMLCs' adherence. Meanwhile, the following stimuli (**Table 2.6**) were prepared using double concentrations as they would be diluted 1:2 once added to the cells.

Table 2.6 Co-culture assay stimuli preparation

	Final Concentration	Doubled concentration
TFGβ (10ug/ml)	1ng/ml	2ng/ml
	0.5ng/ml	1ng/ml
	0.250ng/ml	0.5ng/ml
	0.125ng/ml	0.250ng/ml
	0.0625ng/ml	0.125ng/ml
LPA (13mM)	100µM	200µM
	50µM	100µM

	25µM	50µM
	12.5µM	25µM
Poly I:C (10mg/ml)	20ug/ml	40ug/ml
	15ug/ml	30ug/ml
	10ug/ml	20ug/ml

After incubation, iHBECs+TMLC cells were stimulated with 100ul per triplicate well of each concentration of TGF-β and TGF-β agonists (Lysophosphatidic acid (LPA) and Polyinosinic:polycytidylic acid (Poly I:C)) to produce their standard curves. A triplicate was left unstimulated to be used as a control for basal TGF-β activation. Cells were incubated at 37°C for 16 hours. The next day, culture media was removed, and the wells were washed with PBS before adding 50ul of 1x lysis buffer per well. The cells were then placed at -20°C for at least one hour to promote cell lysis. Following the freezing period, cell lysates were defrosted and transferred to a 96-well luma plate. Next, the plate was placed in a luminometer (BMG Labtech FLUOstar Omega), previously loaded with a vial of luciferase substrate, for the addition of 100ul luciferase substrate per well and measurement of the emitted luminescence. The data from the unstimulated TMLC wells was subtracted from the stimulated ones to obtain the TGF-β-dependent luciferase activity values.

2.2.4 H1N1 infection of hiPSC-derived alveolar epithelial cells

The infection of AT2s was performed in 2D culture conditions. As described previously, alveolar organoids growing in 3D-matrigel droplets were dissociated to obtain single cells using enzymatic digestion. The obtained AT2s were then transferred to a Matrigel-coated 12-well culture plate with a seeding density of $1.5x10^5$. Cells were seeded in maturation media supplemented with 10% CloneR2. The cells were incubated for 24hrs before being examined for attachment and changing the media carefully without disrupting the cells underneath. 48hrs post seeding, the cells were inoculated with 300ul of initial viral stock per well, to obtain an MOI of 1. The inoculum was supplemented with 1ug/ml of Trypsin TPCK to aid the virus to entry the cells. The cells were inoculated for 2 hours before aspirating the virus-containing media and changing it for maturation media. Cells were lysed 24- and 48-hours post-infection, collecting technical duplicates in 300ul of RLT buffer (**Figure 2.4**). The supernatant was also collected, spun down and added to the lysate. Samples were analysed using qPCR or sent for bulk RNA sequencing.

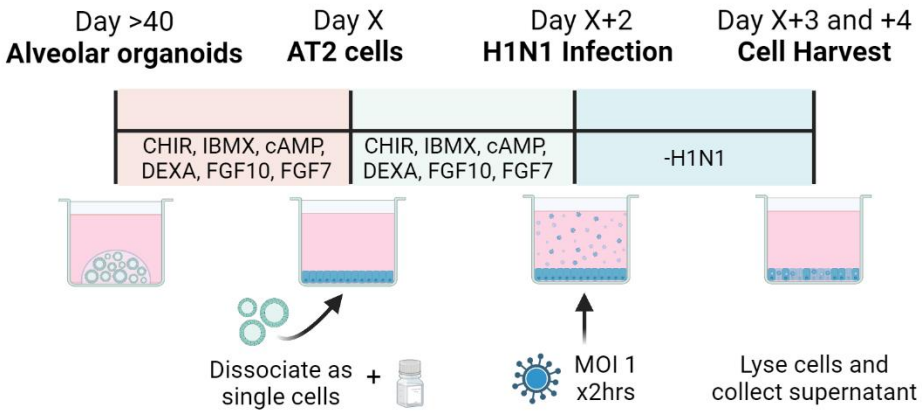


Figure 2.4 Schematic representation of the protocol used to infect AT2 cells with Influenza A virus.

2.3 Molecular biology

2.3.1 RNA extraction and cDNA synthesis

Total RNA was isolated from whole well, or sorted cells using RNeasy® Mini Kit (Qiagen, 74106), including the on-column DNase digestion step. Collected cells were washed once with DPBS and lysed using RLT buffer (Qiagen #79216) to be stored at -20˚C until ready for RNA extraction following the manufacturer's protocol. NanoDrop-1000 spectrophotometer (Thermo-Fischer) was used to determine the concentration and purity of the eluted RNA before proceeding with the reverse transcription (RT) reaction. cDNA was synthesised using 500ng of total RNA in a total reaction volume of 20µl. Briefly, 500ng of RNA were made up to a total volume of 11.875µl with nuclease-free water and mixed with 0.5µl random primers (Promega #C1181) and 1µl of dNTP's (Promega, #U1511) per RT reaction to obtain cDNA. Samples were first incubated at 65°C for 5 minutes and then snap-cooled on ice to prevent the re-formation of secondary structures. Next, 6.625µl of master mix containing 4µl 1st strand buffer (5x), 2µl 0.1M dithiothreitol (DTT), 0.5µl RNase OUT (Invitrogen, #18054071) and 0.125µl SuperScript II (Invitrogen, #18064071) was added to each RNA sample. RT reaction was performed by incubating the samples using the thermocycler settings of 25°C for 10 minutes, followed by 42°C for 50 minutes, and finally 70°C for 15 minutes.

2.3.2 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

For RT-qPCR, each cDNA sample was diluted at a ratio of 1:30 with nuclease-free water (NFW) prior to use for qPCR, giving a final volume of 600µl. Each reaction for qPCR contained 5µl of cDNA and 10µl of master mix containing 7.5µl SensiMix™ SYBR® Low-ROX Kit (Bioline #QT625-05), 1.3µL of NFW, 0.6µl of forward and reverse

primers (5μM). qPCR reactions were performed using an Applied Biosystems® 7500 Fast Real-Time PCR Systems. All qPCR experiments were carried out at a minimum of three technical triplicates, running triplicates for each sample in same 96-well plate. The amplification settings used consisted of a three-step cycling programme, starting with a single cycle at 95°C for 5 minutes, followed by 40 cycles at 95°C for 15 seconds, 60°C for 30 seconds and 72°C for 30 seconds, finalising with a one-minute cycle at 95°C. Relative gene expression was normalised to porphobilinogen deaminase (PBGD) internal reference gene and fold change over the control sample (mainly undifferentiated hiPSCs) was calculated using 2-ΔΔCt method [176]. Presented qPCR data shows the mean of CT values ± standard error of the mean (SEM), unless otherwise stated. **Table 2.7** lists the primer sequences used to perform RT-qPCR gene expression analysis in this thesis. The primers targeting pluripotency, definitive endoderm, foregut, and lung markers were according to previously established literature references. For the inflammatory and fibrotic markers, the Hannan Lab custom-designed primers based on the mRNA sequence, selecting the longest isoform and aiming for a 100-150bp size with an optimal melting temperature of 60°C while spanning an exon-exon junction. The design process involved utilizing the Basic Local Alignment Search Tool (BLAST) provided by the National Center for Biotechnology Information (NCBI). To verify the specificity of the used primers, dissociation curves were examined to confirm the presence of a singular peak, ensuring that a single target was being amplified. Representative images of these dissociation curves for the primers essential for lung characterization are available in **Annex. 1** for reference.

Table 2.7 List of primer sequences used for qPCR.

Gene	Forward	Reverse	Ref
PBGD	GGAGCCATGTCTGGTAACGG	CCACGCGAATCACTCTCATCT	[165]
OCT 4	TGGATGTCAGGGCTCTTTGTC	ACCTTCCCAAATAGAACCCCC	[165]
NANOG	CATGAGTGTGGATCCAGCTTG	CCTGAATAAGCAGATCCATGG	[165]
SOX2	TGGACAGTTACGCGCACAT	CGAGTAGGACATGCTGTAGGT	[165]
CXCR4	CACCGCATCTGGAGAACCA	GCCCATTTCCTCGGTGTAGTT	[147]
C-KIT	ATTCCCAGAGCCCACAATAG	ACCACTAGCTTTCCAAACGG	[177]
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC	[165]
GATA4	TCCCTCTTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG	[165]
FOXA2	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA	[147]
SOX9	GCTCTGGAGACTTCTGAACGA	CCGTTCTTCACCGACTTCCT	[142]
NKX2.1	CTCATGTTCATGCCGCTC	GACACCATGAGGAACAGCG	[161]
CPM	GCGCTGGATTTCAACTACCAC	TCCCGCCCAACAGTCTCAT	[178]
SFTPC	AGCAAAGAGGTCCTGATGGA	CGATAAGAAGGCGTTTCAGG	[161]
SFTPB	CAGCACTTTAAAGGACGGTGT	GGGTGTGTGGGACCATGT	[161]

ABCA3	GCCCTCTTTACACTCAGTTTTCA	GACGAGCAGTTGTCGTACCTAAT	[179]
SLC34A2	TCGCCACTGTCATCAAGAAG	CTCTGTACGATGAAGGTCATGC	[8]
CDX2	GGCAGCCAAGTAAAACCAG	TTCCTCTCCTTTGCTCTGCG	[142]
BRACHYURY	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTTCCTTTGCATCAAG	[165]
MIXL1	GGTACCCCGACATCCACTTG	TAATCTCCGGCCTAGCCAAA	LAB
TTR	ACCGGTGAATCCAAGTGTCC	GGTTTTCCCAGAGGCAAATGG	LAB
AFP	AAACTATTGGCCTGTGGCGA	TTTTGTCCCTCTTCAGCAAAGC	LAB
ALB	CTCGGCTTATTCCAGGGGTG	AAAGGCAATCAACACCAAGGC	LAB
HNF4α	ACTCTCCAAAACCCTCGTCG	CCCTTGGCATCTGGGTCAAA	LAB
PDX1	CCCATGGATGAAGTCTACC	GTCCTCCTCCTTTTTCCAC	LAB
TLR4	CTCGGTCAGACGGTGATAGC	CTGCCTAAATGCCTCAGGGG	LAB
NFkB	CAACAGATGGCCCATACCTTC	GCTCTTTTTCCCGATCTCCCA	LAB
INFγ	TGACCAGAGCATCCAAAAGA	CTCTTCGACCTCGAAACAGC	[180]
TNFα	CCCATGTTGTAGCAAACCCTC	TATCTCTCAGCTCCACGCCA	LAB
TGFβ1	ACCTGCCACAGATCCCCTAT	CCGGTAGTGAACCCGTTGAT	LAB
IL8	CTGTGTGAAGGTGCAGTTTTGC	CAACCCTCTGCACCCAGTTT	LAB
IL1B	CCAGCTACGAATCTCCGACC	AGAACACCACTTGTTGCTCCA	LAB
IL6	AGACAGCCACTCACCTCTTCAG	TTCTGCCAGTGCCTCTTTGCTG	[181]
IL10	AAGACCCAGACATCAAGGCG	AATCGATGACAGCGCCGTAG	LAB
IL18	AAAACCTGGAATCAGATTACTTTGG	GTCCGGGGTGCATTATCTCT	LAB
IL23	GACAACAGTCAGTTCTGCTTGC	TGGAGGCTGCGAAGGATTTT	LAB
LAMA1	GTCAGCGACTCAGAGTGTTTG	AACTTGGGTGAAAGATCGTCAG	LAB
COL1A1	GTTCGTGACCGTGACCTCG	TCTTGTCCTTGGGGTTCTTGC	LAB
COL3A1	TCCAACTGCTCCTACTCG	CGAGTCCTCCTACTGCTA	LAB
FN	CCCATCACAGGGTACAGAATAG	CGGTGTTGTAAGGTGGAATAGA	LAB

2.3.3 Immunocytochemistry and image acquisition

2D-culture

Along the differentiation protocol, cell monolayers were fixed in 4% paraformaldehyde (PFA; Sigma #P6148) in PBS for 20 min at 4°C. The volumes required for each cell culture vessel utilised are listed in **Table 2.8**. Cells were washed twice with PBS and left in the last wash. The plate was wrapped using parafilm (SLS #FIL1024) to be stored at 4°C until ready for staining.

Table 2.8 Volumes of reagents used for immunostaining.

Plate	PBS (µl)	4% PFA (µl)	0.1% Triton X-100 (µl)	1% Donkey Serum (µl)	Antibody/ DAPI incubation (µl)
12 well	1000	500	500	400	300
24 well	500	250	250	200	200
48 well	300	200	200	150	150

96 well	200	100	100	50	50

First, cells were blocked to prevent unspecific antibody binding and treated for membrane permeabilisation to stain intracellular proteins using 0.1% Triton X-100 (Sigma #9036-19) in 10% donkey serum (Sigma #D9663) diluted in PBS for 30 minutes at room temperature. Cells were then incubated overnight at 4°C with primary antibody diluted in 1% donkey serum in PBS. The following day, cells were washed three times for 15 min using 1% donkey serum in PBS. Next, cells were incubated with the species-specific secondary antibody diluted in 1% donkey serum in PBS for 1 hr at RT in the dark.

The secondary antibody solution was removed, and cells underwent three more 15 min-washes in 1% donkey serum in PBS to remove any unbound antibodies. For co-staining experiments, this procedure was repeated using additional primary and secondary antibodies, ensuring the antibodies' host species did not overlap. Finally, nuclei counterstaining was performed by incubating the cells with 0.5µg/ml DAPI (Sigma, #D9542) diluted in 1% donkey serum in PBS for 5 min at RT in the dark. Once stained, samples were washed and stored in PBS at 4˚C until image acquisition. Cells were imaged using Automated Operetta™ High Content Imagining System (PerkinElmer), and the obtained images were analysed using Columbus™ software (PerkinElmer) or FIJI-ImageJ with fluorescence intensity values determined by software's algorithms. All bright-field images of live and fixed cells were captured using a Nikon Eclipse TE2000-S inverted microscope.

3D-culture

For 3D imaging, generated lung organoids were processed and stained following a literature-based protocol with some minor modifications, as described by Dekkers et al. (2019) [182]. Briefly, after removing the culture medium and washing with PBS, organoids were isolated from the 3D Matrigel matrix by incubation with ice-cold cell recovery solution (Corning #354253) on a horizontal shaker (60 RPM) at 4 °C for 60 min. The organoids were resuspended 5-10 times using a 1% BSA in PBS-coated 1ml tip and collected in a 15ml conical tube precoated with 1% BSA in PBS. Tubes were filled up to 10ml with cold PBS and centrifuged at 70 g for 3 min at 4 °C. The remaining pellet was fixated by resuspending the organoids in 4% PFA using a 1% BSA in PBS-coated 1ml tip and incubated at 4 °C for 45 min. Next, the organoids were mixed with cold 0.1% PBS-Tween (Sigma #P7949) and incubated at 4 °C for 10 min; these were

then centrifuged at 70 g for 5 min at 4 °C. The cell pellet was resuspended and blocked in cold organoid wash buffer (OWB) comprising Triton X-100, bovine serum albumin (BSA; Sigma #A3059) and PBS. Approximately, 200ul of this organoid suspension was transferred to each well of a low-adherence 24-well plate (ThermoFisher) per staining. For whole-mount staining, organoids were incubated in OWB with primary antibody (2x concentration) overnight at 4 °C on a horizontal shaker (60 RPM). Next day, samples were carefully washed three times with OWB, allowing organoids to settle at the bottom of the well between washes, and then incubated in OWB with secondary antibody (2x concentration) overnight at 4 °C on a horizontal shaker (60 RPM). After repeating the washing steps to get rid of any remaining secondary antibody, cell nuclei were stained with DAPI (1:500, Sigma) for 20 min and samples were washed with OWB. The organoids of each well were then transferred to a 1.5-ml Eppendorf tube to be cleared by resuspending them in 50ul of fructose-glycerol clearing solution (60% (vol/vol) glycerol and 2.5M fructose) and incubated at RT for 20 min. In preparation for imaging, 1 or 2 Imaging Spacers (SecureSeal™ 20mm diameter x 0.12mm Depth) were adhered to a glass slide. Using a 200ul tip ~20ul of organoids were placed in the middle of the spacers. A coverslip was carefully placed on top of the Imaging spacer and pressure was gently applied to firmly attach it and allow the fluid to spread out. The slides were then ready for confocal imaging. Cells were imaged using Zeiss LSM880 confocal laser scanning microscope with the assistance of Tim Self and Dr Robert Markus. The obtained images were analysed using ZEN 3.2 (blue edition) and FIJI-ImageJ.

A list of the primary and secondary antibodies used for this thesis is provided in **Table 2.3**. An collection of examples of their use in the field for western blotting, immunocytochemistry and immunohistochemistry can be located in the **Annex. 2-4.**

2.3.4 Karyotype analysis of hiPSCs

Cytogenetic analysis of hiPSCs was performed following a literature-based protocol with minor modifications [183]. The different hiPSC lines used in this thesis were passaged into two T25 flasks to the exponential growth of the cells. When the cultures reached a confluence of 80%, around 48hrs after, 100ng/ml of KaryoMAX™ Colcemid™ solution (Gibco, #15212012) was added to culture medium in each flask and then incubated for 1hr at 37˚C. Colcemid™-containing medium was collected in a conical tube. The remaining cells in the flask were harvested with TrypLE and transferred to the same conical tube to be centrifuged at 160 x g for 4 min. For the hypotonic treatment, 8-10ml of 0.6% tri-sodium citrate (Fisher Scientific #S/3320/53) in distilled water was slowly and carefully added to resuspend the cell pellet. This was

followed by incubation at RT for 30 min. The cell suspension was then centrifugated at 300 x g for 5 min, and the obtained cell pellet was carefully resuspended in 8-10ml of 16.7% acetic acid glacial (Fisher Scientific, #A/0400/PB17) in methanol for fixation. To improve the quality of the metaphases for analysis, the samples were washed twice in fixative solution before resuspending the pellet in 0.5-1ml of fixative and stored at -20˚C until ready for Giemsa-banding analysis. The cytogenetic analysis was performed at the Cytogenetics Department of Nottingham City Hospital either by Mr Nigel Smith or Ms Katherine Yates in accordance with International System for Human Cytogenetic Nomenclature (2013) [184]. Whenever possible, 30 metaphase spreads were examined per sample. Sample preparation for karyotype analyses was performed by Dr. Sara Cuevas Ocaña and Dr Peggy Cho Kiu

2.3.5 Enzyme-linked immunosorbent assay (ELISA)

Signalling pathways involved in the host-pathogen response of AT2 cells to Pseudomonas aeruginosa lipopolysaccharides exposure were identified using enzyme-linked immunosorbent assay (ELISA). Cell lysates from the infected AT2 cells after 0, 30, 60 and 120-min exposure were collected and stored at -20°C until ready for determination of the relative levels of protein kinase phosphorylation. ELISA analyses were performed using the Proteome Profiler™ Human PhosphoMAPK Array Kit (R&D Systems) per the manufacturer's recommendations. The spot pixel density determined the relative phosphorylation levels of each activated protein kinase. Densitometry was performed using Image J version 1.52e software.

2.3.6 Bulk RNA sequencing and bioinformatics data analysis

RNA Sequencing was outsourced to GeneWiz NGS Laboratory and performed using the Ilumina HISeq platform. Bioinformatic analysis was performed by Carlos Sainz Zuñiga using RStudio for the hit-counts deliverable by GeneWiz. The limma package from Bioconductor was used to statistically analyse bulk mRNA expression dataset [185]. Data was normalised to $\log_2$ (0.5 + Dataset). Gene hit count dataset was tested for differential gene expression using limma to fit linear models. Empirical Bayes moderation of the standards errors towards a common value were done for each gene between samples for differential expression analysis, using a threshold p-value of 0.05. Benjamini-Hochberg procedure (False Discovery Rate (FDR)) was used for multiple hypothesis test correction. PCA plots, volcano plots, bar graphs, and heatmaps were generated in R. Gene Ontology (GO) analysis was performed using the GO.db package from Bioconductor [186–188]. Disease Ontology (DO) was performed using

the clusterProfiler package from Bioconductor [189]. Reads were mapped to the GRCh38.p13 Ensembl human genome.

2.4 Statistical analysis

Presented data in text and figures are shown as mean ± standard error of the mean (mean ± SEM) unless otherwise stated. The statistical analysis was performed using GraphPad Prism Software version 8. The statistical significance was evaluated using the most relevant statistical test; an unpaired student's t-test was performed when only two groups were being compared; if comparing more than two groups, a standard one-way analysis of variance (ANOVA) was performed, followed by Tukey's multiple comparisons test to compare the mean of each group with the mean of every other group. A two-way ANOVA followed by Sidak's post hoc test was used to compare data grouped by two factors. Statistically significant p-values were represented as * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ and **** $p \leq 0.0001$. P-values >0.05 were classified as not significant (ns).

Chapter 3 Generation of IPF patient-derived hiPSCs for in vitro lung differentiation

3.1 Introduction

3.1.1 Genotype-phenotype relationship in IPF pathogenesis

When the lung epithelium with genetic abnormalities affecting AT2 cells is subjected to repeated exposure to environmental damage, the capacity of these cells to activate efficient repair mechanisms becomes compromised. This impairment in their ability to respond to and address injuries can lead to further damage and hinder the overall regenerative process of the alveoli [36]. Mutations in SFTPC pose a significant disruption to the correct protein folding and secretion of pulmonary surfactant by AT2 cells. This, combined with cell injury or damage, can trigger the activation of intrinsic endoplasmic reticulum (ER) stress and senescent signalling pathways [71]. Notably, these mutations in SFTPC have been extensively associated with the development of IPF, significantly contributing to its pathogenesis. A thorough understanding of these genetic abnormalities and their impact on this progressive lung disease is crucial for devising effective therapeutic interventions for individuals affected by IPF [190].

Surfactant protein C

Surfactant protein C (SFTPC) is the only surfactant protein synthesised and secreted exclusively by AT2 cells. The mature form of SFTPC is a 3.7 kDa protein encoded on chromosome 8p21.3 by a 6-exon gene. Its precursor is 197-aminoacid ER integral membrane protein [191, 192]. The pro-SFTPC is structurally organised into four distinct regions illustrated in **Figure 3.1** [193]. The targeting domain or N-terminal segment (residues 1–23) is a short region located on the cytosolic side, playing a crucial role in intracellular trafficking. The transmembrane (TM) region (residues 24–58) constitutes the main portion of the mature SFTPC protein. This region is eventually secreted into the alveoli along with phospholipids. In environments rich in phospholipids, the mature protein forms a highly stable and rigid α-helix structure [194]. The next region is the linker domain (residues 59–89) which serves as a connector, facilitating the docking of the BRICHOS domain to the TM segment. The BRICHOS domain (residues 90–197), located at the COOH-terminus, is localised to the ER lumen [49]. Its structure comprises two α-helices that enclose a central five-stranded β-sheet. The BRICHOS domain plays an important role in safeguarding the metastable mature SFTPC domain,

which is prone to β-sheet formation. Therefore, BRICHOS is considered to act like a chaperon, preventing the formation of amyloid fibrils [195]

Through a series of four endoproteolytic cleavages, involving the removal of both C-terminal and N-terminal propeptides, the precursor SFTPC is transformed into its mature form. This cleaved protein is subsequently secreted into the alveolar surface with the rest of the surfactant components [190]. When SFTPC is secreted into the alveolar surface, it is inserted into the phospholipid membranes, disrupting their lipid packing. This promotes the movement of lipids between the membrane sheets and increases the rate of adsorption of lipids into the surface film. This biophysical activity of SFTPC is mainly dependent on its helical structure [196].

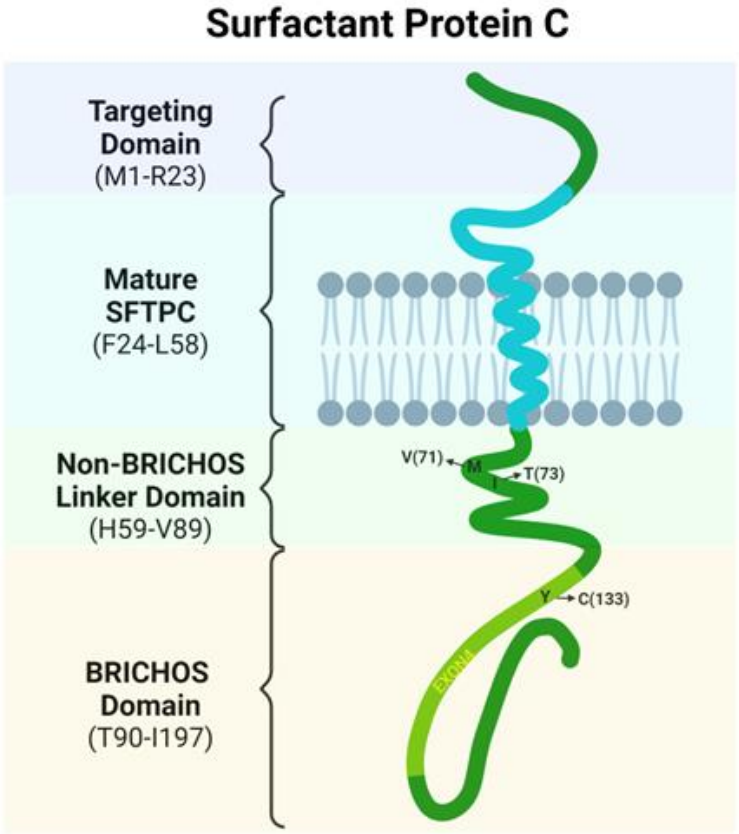


Figure 3.1 Pro-surfactant protein C structure.

The SFTPC mutations of the IPF-patient derived cell lines used in this thesis are noted in black. Two of the mutations, M71V and I73T, are located to the non-BRICHOS linker domain and Y113C is located in the BRICHOS domain. Adapted from [193]. Created with BioRender.com.

More than 50 SFTPC mutations have been identified since the first description of two interstitial lung disease-associated mutations, SP-CΔExon4 and SP-CL188Q. The vast majority of these mutations are located in the linker and BRICHOS domains, leading

to post-translational protein processing alterations [197]. Of these mutations, the linker mutation I73T is among the most prevalent [198]. The mutations in the linker domain do not induce cytosolic aggregation of the pro-peptide, but instead result in its mistrafficking [199]. This leads to its aberrant subcellular localization with accumulation in the plasma membrane, multi vesicular bodies, and in early and late endosomes. In contrast, the SFTPC mutations in the BRICHOS domain turns the protein prone to misfolding, leading to the accumulation of the mutant pro-SFTPC protein forming aggregates in the ER lumen of AT2 cells [5]. This in turn results in ER Stress, which if unresolved leads to the secretion of cytokines and apoptosis [4]. Furthermore, mutations in SFTPC also make the AT2 cell more susceptible to injurious external stimuli, as shown when the severity of bleomycin-induced lung fibrosis was increased in the lungs of mice expressing the SFTPC mutation L188Q. This mutation was associated to greater AT2 cell death as detected by TUNEL assay and increased caspase-12 and caspase-3 activation [6].

ER STRESS and UPR

ER is an organelle essential for the processing of newly synthesized proteins. In order to maintain proper function, there must be a balance between the ER protein load and the ER folding capacity. However, under certain physiological conditions and pathological stimuli, such as environmental and genetic factors, this ER homeostasis can be disrupted. Mutations that alter the ability of a protein to fold correctly result in the accumulation of the unfolded/misfolded protein in the ER. This condition is known as ER stress, and it induces an ER-to-nucleus signal pathway called the Unfolded Protein Response (UPR) [200]. A diagram of the UPR signalling pathways is shown in **Figure 3.2**. Upon ER stress, the immunoglobin binding protein (BiP) dissociates from the luminal domain of three ER transmembrane proteins (Inositol Requiring 1 (IRE1), PKR-like ER kinase (PERK), and Activating Transcription Factor 6 (ATF6), resulting on their activation [201]. This initiates the UPR cascade of downstream signalling events which increases transcription of chaperones that promote protein folding. In addition, translation of mRNA is attenuated, effectively slowing entry of newly synthesized protein into the ER. In addition, it increases the ratio of chaperone to unfolded protein. Furthermore, the proteins that fail to fold under these conditions are diverted to the ER-associated degradation (ERAD) pathway, in which misfolded protein is translocated to the cytosol for degradation by the proteasome.

If successful, UPR restores the ER homeostasis promoting cell survival and adaptation. However, if these systems fail to eliminate the misfolded proteins, it can lead to formation of cytotoxic aggregates and activate cell death pathways [202] .

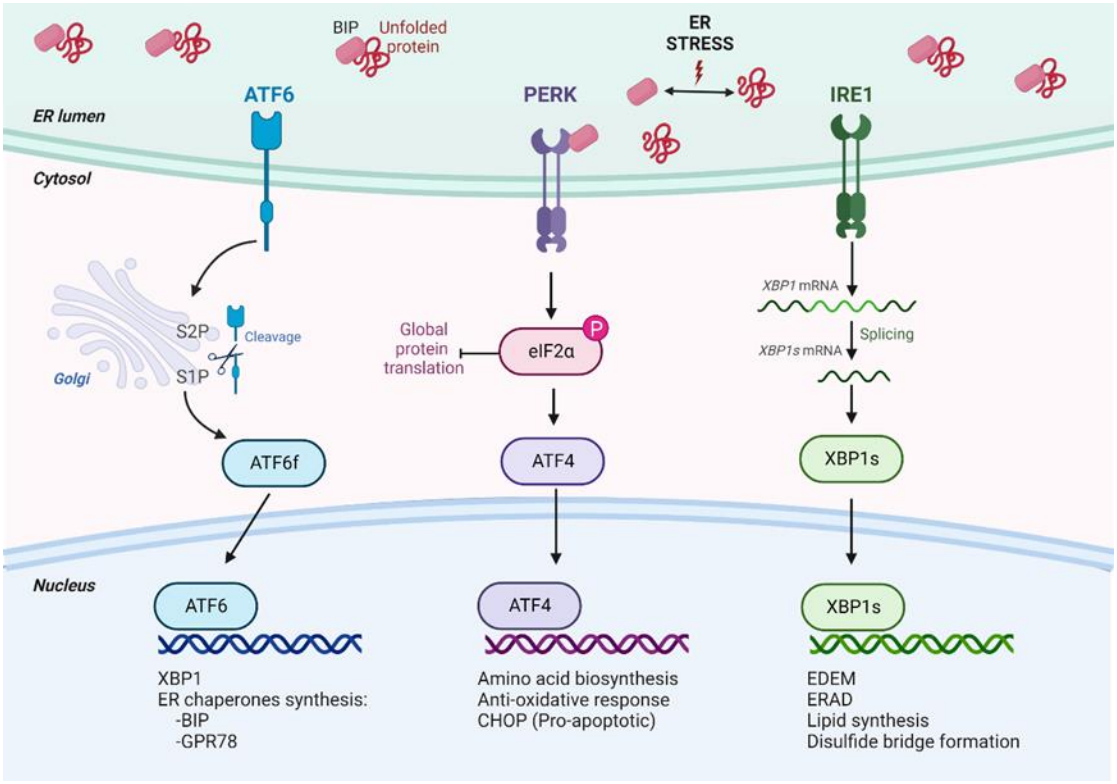


Figure 3.2 ER stress and UPR signalling pathways.

The unfolded protein response (UPR) is a cellular stress response that promotes proper protein folding in the endoplasmic reticulum (ER). Its activation involves three stress sensors: IRE1α, PERK, and ATF6. Upon detection of ER stress, activated ATF6 is transported to the Golgi apparatus where it is cleaved to release its cytosolic domain, ATF6f, a transcription factor that upregulates genes associated with protein degradation and XBP1. When activated, PERK phosphorylates eIF2α to promote the translation of ATF4 mRNA, which controls genes related to autophagy, apoptosis, amino acids, and redox metabolism. Upon ER stress, dimerisation and phosphorylation activate IRE1α, leading to its RNAase activity to produce spliced XBP1 (XBP1s), a transcription factor that regulates genes involved in protein folding and quality control. Adapted from [200]. Created with BioRender.com.

Studies have reported that markers of ER stress-UPR activation (BiP, EDEM, XBP-1) are also elevated in AT2 cells of patients with sporadic forms of IPF. This suggests that ER stress response is a common feature for familial and sporadic cases [9,10]. A common non-BRICHOS domain mutation (SFTPC variant I73T), found in both sporadic and familial IPF cases, has been associated with mis-trafficked SFTPC. Using a SFTPC^{I73T} knock-in mouse model for a dose-dependent toxic gain of function, a study showed that instead of accumulating in the ER, the mutant SFTPC protein was trafficked inside of endosomes directly to the plasma membrane [203]. The mutation

also led to alterations in AT2 cell quality control and chemokines profile, increasing the secretion of fibrotic mediators. These changes led to a dysfunctional cell phenotype causing acute alveolitis with aberrant remodelling, collagen deposition and AT2 hyperplasia [11]. A second study using a knock-in mouse model generated AT2 cells expressing a SFTPC BRICHOS domain mutation which spontaneously induced lung inflammation and fibrotic remodelling in vivo through activation of the UPR [12]. Regardless of this evidence, whether ER stress-UPR activation is sufficient to lead to the development of lung fibrosis remains unclear. For instance, there is data supporting the concept that ER stress produces a dysfunctional AT2 cell phenotype that, when subjected to a second external stimulus, is more prone to activation of profibrotic pathways, enhancing the fibrotic remodelling of the lungs [13].

To gain deeper insights into the role of SFTPC mutations in the pathogenesis of idiopathic pulmonary fibrosis, human induced pluripotent stem cells (hiPSCs) present a promising avenue to investigate the underlying molecular mechanisms driving this condition. By employing hiPSCs derived from IPF patients and recreating in vivo lung developmental milestones, functional AT2 cells can be generated. These patient-specific AT2 cells can be used as a valuable platform for studying the disease at a cellular level, offering the opportunity to manipulate them and explore the molecular pathways implicated in the development and progression of IPF.

3.1.2 Chapter aims and objectives

The aims and objectives of this results chapter are as follows:

- To generate and characterise hiPSC lines derived from IPF patients with a specific SFTPC mutation in the BRICHOS or linker domains. This involves the isolation of skin fibroblasts from IPF donors and their reprogramming into hiPSCs.
- To identify the hiPSC line with the highest potential for lung differentiation from the IPF-patient-derived cell lines. This involves subjecting the hiPSCs to an established in vitro differentiation process designed to guide them through stages of endodermal commitment, anterior foregut development, and the acquisition of lung progenitor and alveolar fates. This differentiation will enable the selection of the most suitable hiPSCs candidate for downstream experiments.
- To perform gene editing on the selected hiPSC line ($SFTPC^{WT/MUT}$) using base editing technology. This process will result in the development of a cell line bearing a homozygous mutation in the surfactant protein C gene ($SFTPC^{MUT/MUT}$) as well as the creation of a homozygous corrected counterpart ($SFTPC^{WT/WT}$). These genetically modified cell lines will serve as a valuable experimental model for examining the impact of the SFTPC mutation in the context of IPF.

3.2 Results

3.2.1 Generation of hiPSC lines derived from IPF patient's skin fibroblasts.

Isolation of fibroblasts from human skin explants

Human skin biopsies were obtained from six individuals carrying different heterozygous mutations in surfactant proteins A and C who were diagnosed with Familial Idiopathic Pulmonary Fibrosis. Among the patients, two female individuals had the BRICHOS domain mutation c.338A>G, resulting in an A>G single nucleotide variation (SNV) in exon 4 of the surfactant protein C (SFTPC) gene. This mutation led to the substitution of Tyrosine (Tyr) with Cysteine (Cys) at protein position 113 (Y113C). Additionally, three patients exhibited SFTPC Non-BRICHOS linker domain mutations. One of these mutations was identified as c.211 A>G (Met71Val), while the other one was identified as c.218 T>C (Ile73Thr). The remaining patient carried the surfactant protein A (SFTPA2) gene mutation c.511 A>T in exon 6, resulting in an Asparagine (Asn) to Tyrosine (Tyr) substitution at protein position 171 (N171Y). More details about the patients' genetic variations can be found in **Table 2.2**.

To establish human dermal fibroblast cultures from each biopsy, the samples were dissected to remove subcutaneous fat and then transferred into 6-well plates for culture in human dermal fibroblast (HDF) media. Fibroblast outgrowths emerged from the biopsy fragments over a period of approximately 10-24 days. Within the first 7-10 days of culture, keratinocytes started migrating out from the tissue samples. After an additional 7-10 days, fibroblasts began to appear following the initial outgrowth of keratinocytes (around day 14-20). From day 20-30, cells were trypsinised and transferred from 6-well plates to a T75 flask in order to isolate the dermal fibroblasts from the heterogeneous cell population. The cells began growing in a monolayer and were subsequently passaged at a 1:2 ratio into two T175 flasks once they reached ~80% confluency. The cultured fibroblasts were further expanded for cryopreservation or reprogramming into hiPSCs. In total, six fibroblast cell lines were derived, comprising five $SFTPC^{WT/MUT}$ lines and one $SFTPA2^{WT/MUT}$ line.

hiPSC Sendai-virus (SeV) based reprogramming in feeder-free conditions.

The IPF-patient-derived fibroblast cell lines were reprogrammed into human induced pluripotent stem cells (hiPSCs) by transducing the cells with the Yamanaka factors (Oct3/4, Sox2, Klf4, c-Myc). For the reprogramming process, Sendai virus vectors were utilised due to their non-integrating properties and their demonstrated ability to reprogram adult fibroblasts with an efficiency of 1% [204]. All cell lines were

transduced with CytoTune-iPS 2.0 Sendai reprogramming vectors between passages 3-7. On the day before the transduction, the skin fibroblasts were plated in an uncoated 6-well plate at a seeding density of $2.0x10^5$ cells per well, allowing them to grow and reach approximately 60% confluency before the transduction.

Following 48 hours post-transduction with the Yamanaka factors, the cells underwent the initial morphological changes, transitioning from the elongated shape typical of fibroblasts to a more rounded cellular morphology. Between days 7 and 14, loose clusters of cells began to appear, forming small colonies. These colonies gradually became tighter and more compact, typically around day 15-30, although the timing varied among the different cell lines. The newly formed colonies exhibited morphological characteristics reminiscent of human embryonic stem cells (hESCs), with well-defined borders and a high nucleus-to-cytoplasm ratio. These features were indicative of the reprogrammed cells acquiring a pluripotent state similar to hESCs (**Figure 3.3**). In the subsequent days, after the formation of well-defined hiPSC colonies, the colonies were treated with EDTA or ReLeSR to facilitate their manual selection using a p200 pipette tip. The isolated colonies were then individually seeded, and subpopulations of each cell line were expanded for further characterisation and cryopreservation.

From the IPF-patient-derived fibroblast cell lines, three $SFTPC^{WT/MUT}$ hiPSC lines were successfully established and named IPF1 (Y113C), IPF2 (M71V) and IPF5 (Y133C), and four clones were generated from the latter one, named IPF5#3, IPF5#4, IPF5#5 and IPF5#6. Fibroblasts derived from patient IPF6, bearing an SFTPA mutation (N171Y), were omitted from further experimental procedures due to the increased correlation of SFTPC mutations with the development of IPF, emphasising the significance of focusing on SFTPC-related mechanisms in this thesis [168]. Similarly, fibroblasts from IPF3 and IPF4 patients were excluded, as they harbour an SFTPC-linker domain mutation (I73T). This specific mutation has well-established associations with infant forms of interstitial lung disease, and it is noteworthy that the resulting defect affects epithelial cell function through alternative pathways that do not involve misfolding, aggregation, or induction of ER stress [192, 205]. Instead, the I73T mutation leads to abnormal SFTPC trafficking [198].

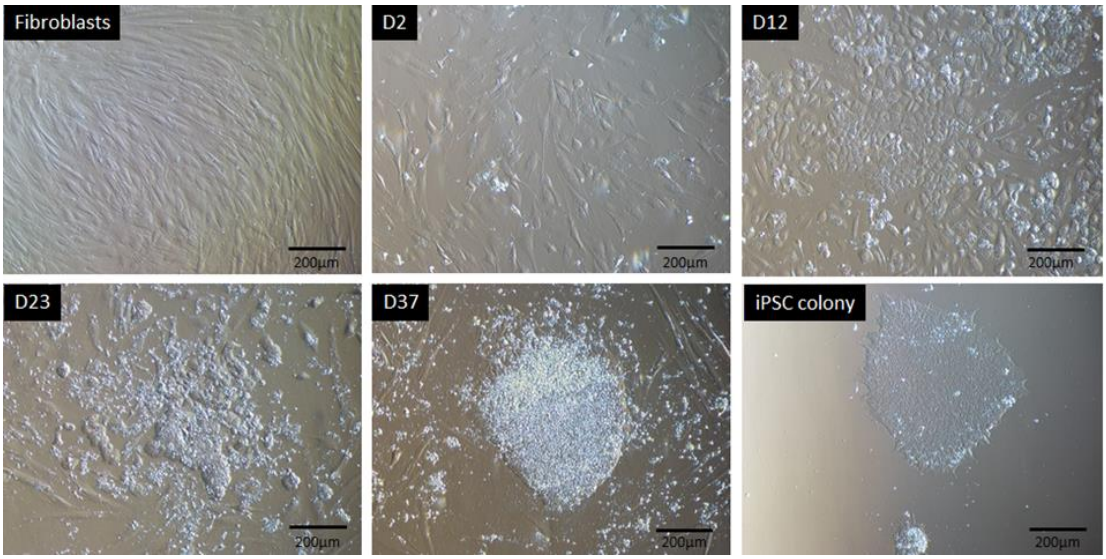


Figure 3.3 Reprogramming of SFTPC-mutant hiPSCs from IPF patient's fibroblasts.
Representative brightfield images of dermal fibroblasts after transduction with Sendai-virus (SeV) reprogramming vectors and the sequential emergence of a patient-derived hiPSC colony. D2, D12, D23, D37: days post-transduction. Scale bars= 200µm

The reprogrammed hiPSC lines and their selected clones were characterised to ensure cells were genomically stable and corroborate their pluripotency status. This was done via chromosomal analysis by karyotyping, assessment of pluripotency markers and trilineage differentiation (data not shown), confirming the hiPSC lines were suitable for directed differentiation and downstream experiments. All cell lines were expanded and tested for mycoplasma before cryopreservation for long-term storage. Dr Nicholas Hannan and Dr Peggy Cho Kiu Lo performed the work above-mentioned.

3.2.2 Characterisation and selection of reprogrammed hiPSC lines for lung differentiation protocol

Heterogeneous generation of alveolar progenitors using 17-day lung differentiation protocol.

Determining the cell line that best produced alveolar epithelial type 2 (AT2) cells was essential for future experiments. Therefore, different SFTPC-mutant cell lines were differentiated following an established in vitro differentiation protocol to generate lung progenitor cells [12]. This protocol consisted of a 3-day endoderm induction using Activin A (ActA) and Wnt3A (WNT) signalling, followed by a 2-day exposure to ActA to drive the cells into foregut lineage and finalising with the induction of lung progenitors with the aid of CHIR99021 (CHIR), retinoic acid (RA), and Bone Morphogenetic Protein 4 (BMP4). All the stages were performed using the basal differentiation-1 (DIFF1)

medium that consisted of RPMI medium supplemented with B-27™ Supplement, non-essential amino acids (NEAA) and Pen-Strep.

Four clones from the IPF patient-derived cell line IPF5, carrying an SFTPC BRICHOS domain mutation, were subjected to the endoderm, foregut and lung induction stages. **Figure 3.4** provides an overview of critical time points along the generation of lung progenitors with clones 3, 4, 5, and 6. The undifferentiated hiPSCs seeded at 15,000 cells/cm^2 looked indistinguishable from each other until the end of the endoderm induction (definitive endoderm induction day 3, ENDO D3) with the formation of a monolayer. Under the sole influence of ActA for two days (foregut induction day 2, FG D2), clone 4 had markedly more cell debris than the other clones, and some gaps on the cell monolayer could be observed. After four days of CHIR, RA, and BMP4 exposure (lung induction day 4, LD4), evident morphological differences appeared between clones. While clones 3 and 5 maintained the cobblestone epithelial morphology, swirls delimitating certain areas of the monolayer started appearing in clones 4 and 6. Four days later (lung induction day 8, LD8), these swirls became more prominent, and the curved edges' endpoints started closing themselves, forming annular structures, especially in clone 4. By the end of the lung progenitors' induction (lung induction day 12, LD12), the monolayer of clone 4 was invaded by annular structures, and a few were still present in clone 6. Clones 3 and 5 did not show any of these structures; however, areas within the monolayer were very compacted.

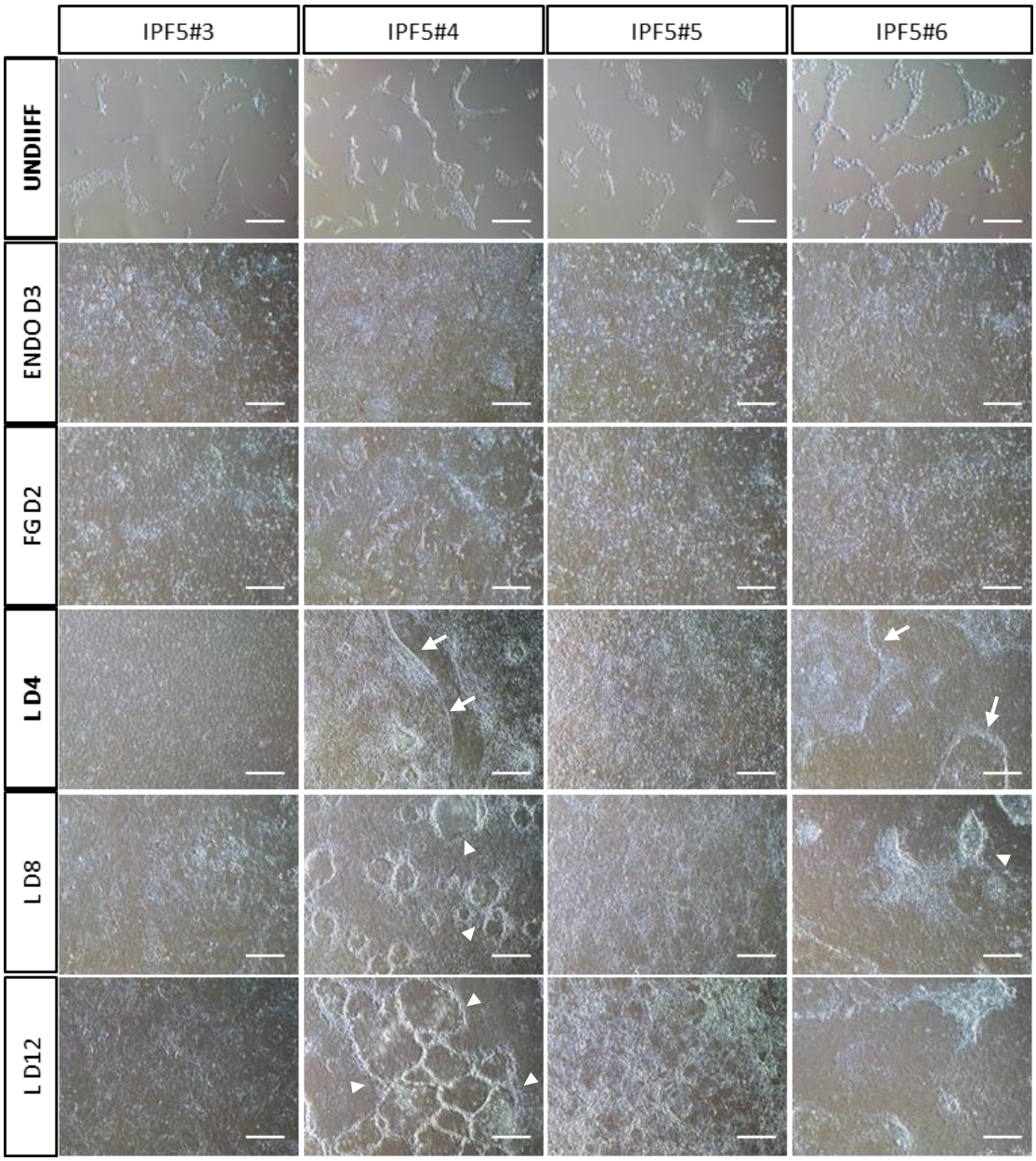


Figure 3.4 IPF patient-derived hiPSC in vitro differentiation into lung progenitors.

Representative brightfield images comparing IPF5 clones 3, 4, 5, and 6 at critical stages of lung differentiation. Note the presence of swirls (arrows) and annular structures (arrow heads) in clones 4 and 6 throughout the lung induction stages. UNDIFF: hiPSCs 24hrs after seeding, ENDO: definitive endoderm, FG: foregut endoderm, LD4, LD8, LD12: days of lung induction. Scale bars= 200µm

To investigate the gene expression profile of the generated lung progenitor cultures, RNA was harvested at key time points along the differentiation of the four IPF5 clones to perform qPCR (**Figure 3.5**). Pluripotency markers OCT4, SOX2 and NANOG, were included to evaluate the efficiency of the early stages of the differentiation. The first two markers decreased significantly during the endodermal and foregut induction

stages; however, the expression of NANOG did not diminish significantly until the beginning of the lung induction stage. This gene expression made evident that some pluripotent cells remained in the cultures, potentially influencing the efficiency of later stages of the differentiation. Furthermore, mesendoderm (Mix Paired-Like Homeobox (MIXL1), BRACHYURY) and intestinal (CDX2) markers were also included to discard the presence of cells that failed to follow the differentiation cues, contaminating the culture with cells from other lineages. The high expression levels of MIXL11 and BRACHYURY at the initial stages of the protocol suggest the presence of bipotent mesendoderm progenitors, which decrease with the induction of lung progenitors. Nevertheless, CDX2 levels significantly increased after LD4, indicating that some of the endoderm cells within the culture were differentiating into intestinal cells rather than lung cells.

Finally, NKX2.1, SOX17 and SOX9 were measured to corroborate the expression of lung progenitors' markers. Interestingly, NKX2.1 was significantly upregulated at the end of the foregut endoderm timepoint on clones 3 and 5. However, they did not maintain this expression towards the end of the lung progenitors' differentiation. Clone 4 was the only clone that showed the expected NKX2.1 significant increase at LD12. Similarly, SOX17 was only significantly upregulated by clone 4 from LD4 onwards. The expression trend of SOX9 was not as expected either, but it was upregulated on clones 4 and 5 after LD8.

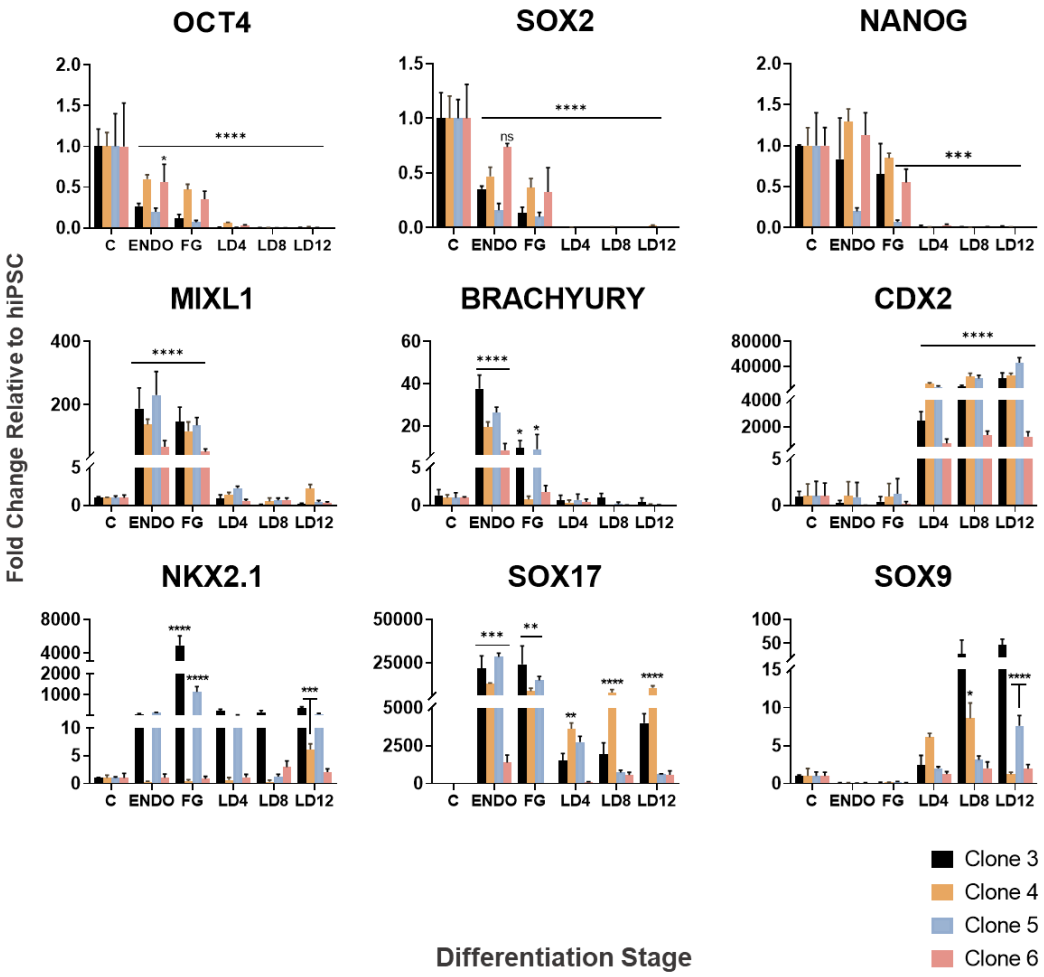


Figure 3.5 Gene expression of IPF patient-derived hiPSCs along the lung differentiation protocol.

IPF5 clones 3, 4, 5 and 6 qPCR showing expression of pluripotency (OCT4, SOX2, NANOG), mesendoderm (MIXL1, BRACHYURY), intestinal (CDX2) and lung epithelial (NKX2.1, SOX17 and SOX9) markers. C: undifferentiated hiPSCs, ENDO: definitive endoderm, FG: foregut endoderm, LD4, LD8, LD12: days of lung induction. Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to undifferentiated hiPSCs. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

These results were concerning as they showed mesendoderm progenitors were contaminating the cultures, cells were also differentiating into other tissues, and the levels of lung progenitors were lower than desirable. In order to increase the probability of finding a suitable cell line for the generation of AT2 cells, two additional IPF patient-derived cell lines, IPF1 and IPF2 (carrying an SFTPC BRICHOS and a non-BRICHOS domain mutation, respectively), were differentiated into lung progenitors along with the healthy control cell lines, REBLPAT and LOPCK (**Figure 3.6**). The four hiPSCs were also seeded at 15,000 cells/cm^2, and as with IPF5, no distinguishable morphological

differences were observed between cell lines until the foregut endoderm stage. In fact, IPF2, REBLPAT and LOPCK look indistinguishable from each other throughout the differentiation protocol showing an epithelial cobblestone morphology. It was only IPF1 that showed evident changes during the lung induction stage, with the emergence of swirls and annular structures similar to the ones presented on IPF5 clone 4.

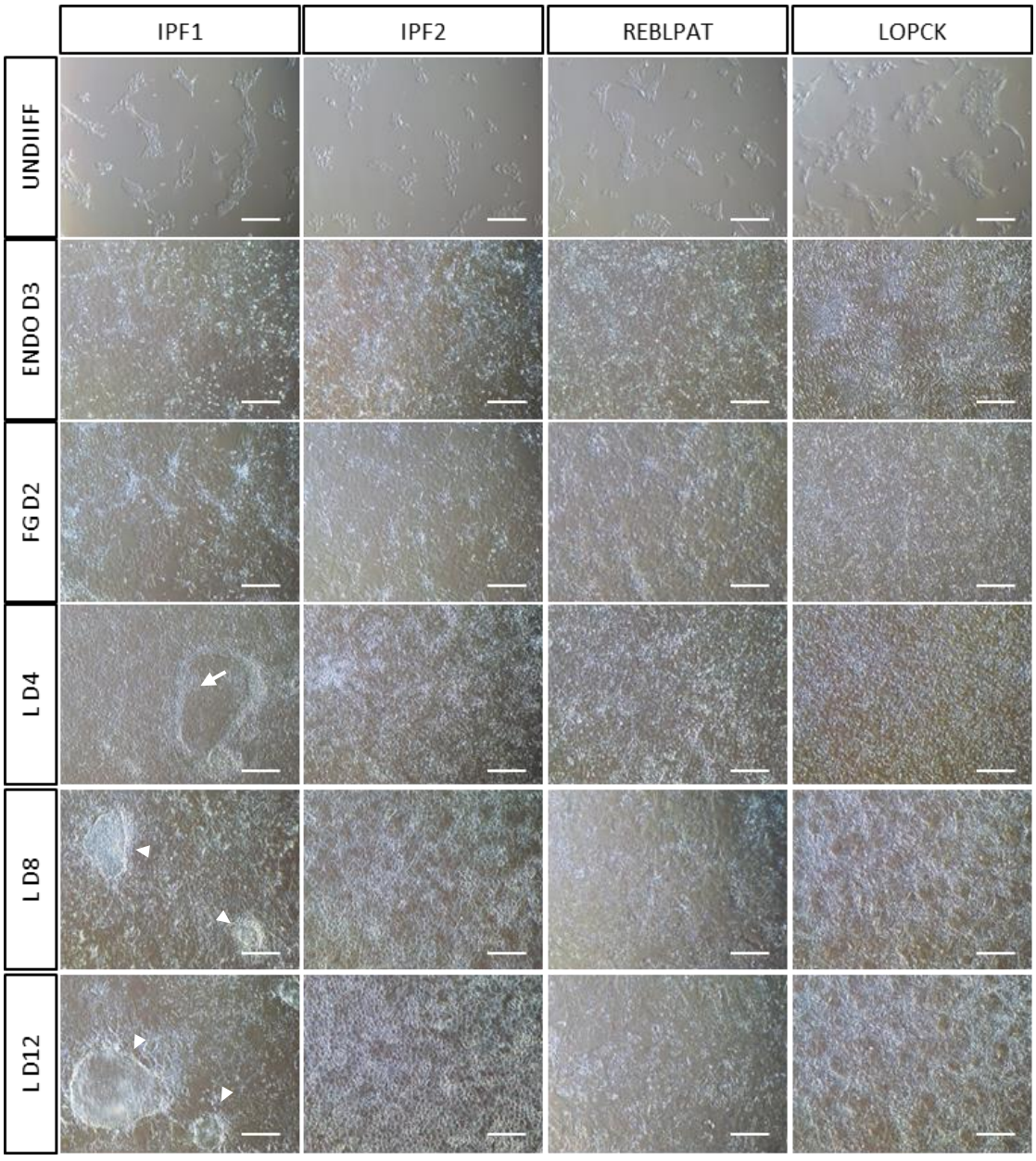


Figure 3.6 Healthy and additional IPF patient-derived hiPSCs in vitro differentiation into lung progenitors.

Representative brightfield images of IPF patient-derived cell lines, IPF1 and IPF2, and healthy cell lines, REBLPAT and LOPCK, at key stages of lung differentiation. Note the presence of swirls (arrow) and annular structures (arrow heads) in IPF1 throughout the lung induction stages. UNDIFF: hiPSCs 24hrs after seeding, ENDO: definitive endoderm, FG: foregut endoderm, LD4, LD8, LD12: days of lung induction. Scale bars= 200µm

In order to elucidate which cell lines were able to produce NKX2.1+ lung progenitors and ultimately generate cells capable of secreting SFTPC, immunocytochemistry was performed to assess the expression of these markers on IPF patient-derived cell lines and a control cell line, REBLPAT. Clone 6 was excluded from subsequent experiments due to the undetectable levels of lung progenitor markers found on the qPCR analysis.

Initially, cells were stained for NKX2.1 throughout the lung induction protocol on lung days 4, 8 and 12. Unfortunately, the NKX2.1 antibody used (mouse anti-NKX2.1, Thermofisher, MS-699) did not work. Not even the healthy control was stained positively. Therefore, the characterisation was repeated using a second NKX2.1 antibody (abcam, ab76013; rabbit) which positively stained the nucleus of the generated lung progenitors. **Figure 3.7** shows the presumptive lung progenitors on lung day 12, the last day of the induction protocol, generated from IPF5 clones 4 and 5. These two clones had shown to grow with very distinct morphological characteristics, clone 4 formed cystic structures, and clone 5 grew in a monolayer. Surprisingly, the immunofluorescence images in **Figure 3.7A** revealed that the cystic structures formed by clone 4 were clearly positive for NKX2.1 nuclear staining, whereas the surrounding monolayer was scarcely stained. Contrastingly, the entire clone's 5 monolayer was positive for NKX2.1, as observed in **Figure 3.7B**.

In addition to confirming the presence of lung progenitors in the cultures, it was necessary to assess the expression of SFTPC to detect presumable AT2 cells. **Figure 3.7** shows that IPF5 clones were positively stained for SOX9, confirming the generation of distal lung progenitors. Furthermore, cells were stained for SFTPC, protein expressed exclusively by AT2 cells, using an antibody aimed to detect the residues 1-100 of the SFTPC pro-peptide (abcam, ab90716). However, it unexpectedly stained the nuclei, perinuclear area and plasma membrane. The aberrant pro-SFTPC staining in IPF5 clone 4 can be observed in **Figure 3.7A**. Like the expression of NKX2.1, the staining of pro-SFTPC was stronger on the spherical structures. However, the pattern in which the antibody stained the cells differed in the cysts and the monolayer. In the monolayer, the cells appeared to be positively stained to the nucleus, and the cysts' staining looked peripheral. This expression pattern was confirmed in clone 5, as the lack of 3D-like structures made it easier to visualise it with clarity. **Figure 3.7B** shows the green fluorescence of pro-SFTPC localised to the nucleus and the plasma membrane of the cells in the monolayer, with faint cytoplasmic staining. Thus, a different pro-SFTPC antibody needed to be used to characterise the differentiated AT2 cells in the cultures.

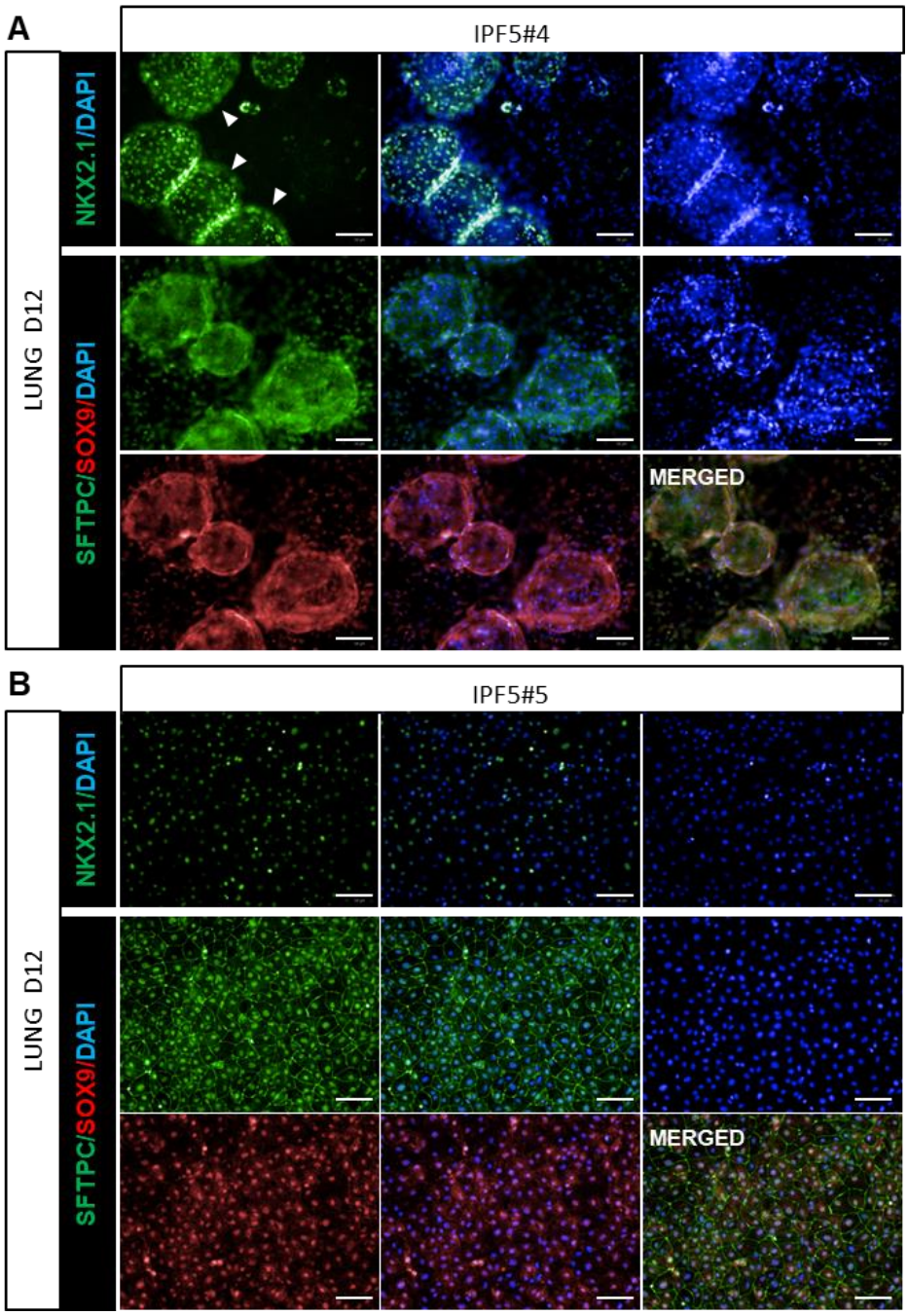


Figure 3.7 Characterisation of NKX2.1 positive lung progenitors.

Representative immunofluorescence microscopy with antibodies against NKX2.1 (green), SOX9 (red) and pro-SFTPC (green) in day 12 of the lung induction protocol. A. Shows the annular structures generated by IPF5 clone 4 with a positive NKX2.1 nuclear staining (arrow heads). These structures were also positively stained for SOX9 and SFTPC. B. Shows IPF5 clone 5 with NKX2.1 nuclear staining throughout the entire monolayer. These cells were also positive for SOX9; note that the staining for SFTPC is clearly localised to the plasma membrane and the nuclei. Nuclei were stained with DAPI (blue). Scale bars= 100µm

The other cell line forming clear spherical structures during the lung induction protocol was IPF1. This cell line was also stained for the lung marker NKX2.1 to elucidate if

these structures shared the pattern of expression with the ones generated by IPF5 clone 4. Additionally, IPF2, another cell line growing as a monolayer, was differentiated and stained alongside IPF1 to be able to compare them. This time the epithelial marker E-cadherin (ECAD) was included for co-staining.

The results obtained when imaging these IPF patient-derived cells were intriguing. As with IPF5 clone 4, the spherical structures generated by IPF1 were difficult to image with clarity. Nevertheless, it was possible to distinguish the green nuclear NKX2.1 in these epithelial structures that were clearly delimited by the positive E-cadherin red membrane staining. What was not expected was the positive cytoplasmic staining for NKX2.1 on the surrounding monolayer (**Figure 3.8**A). This correlated with the NKX2.1 expression pattern in IPF2, where the epithelial monolayer, stained in red for E-cadherin, also showed an NKX2.1 staining that was limited to the cytoplasm and excluded from the nucleus (**Figure 3.7B**).

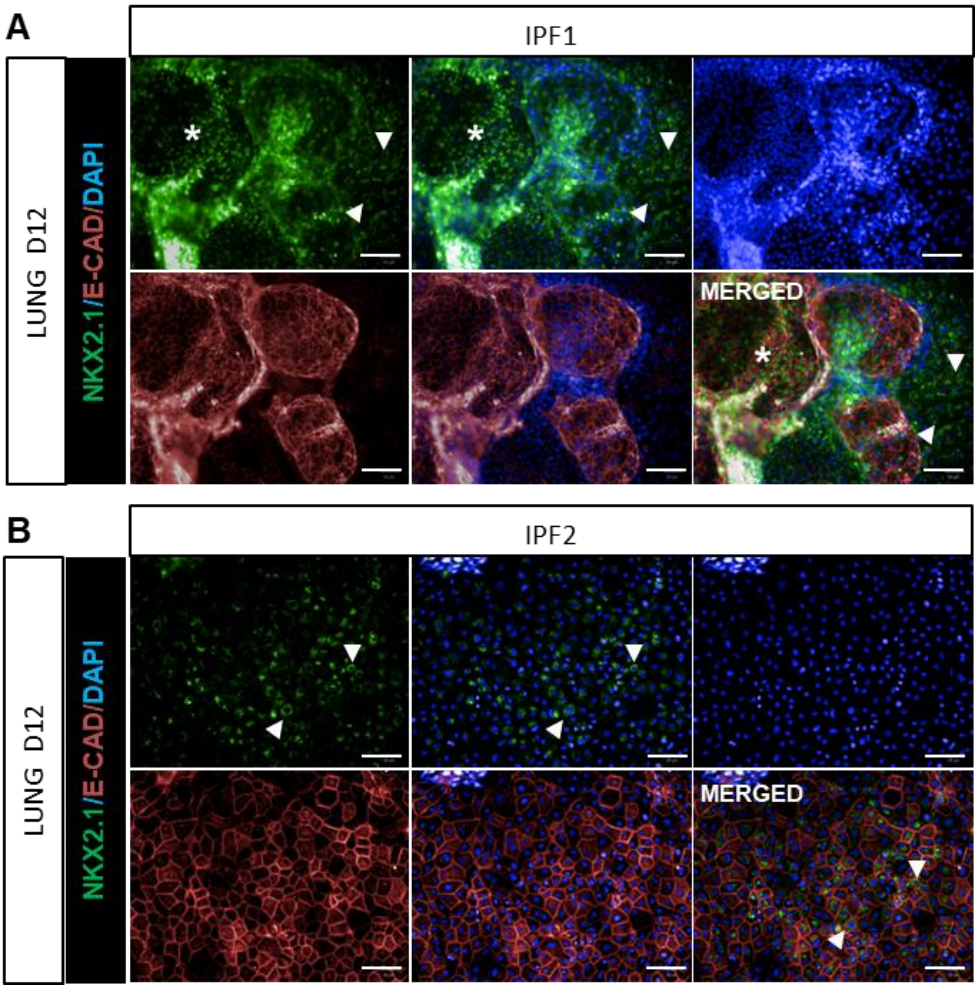


Figure 3.8 Differences in NKX2.1 immunostaining expression pattern within cell lines.
Representative immunocytochemistry images on day 12 of lung induction showing staining for the lung marker NKX2.1 (green) and the epithelial marker E-cadherin (red). A. IPF1 lung differentiation generates cystic structures that are clearly visible with the plasma membrane E-cadherin staining and a nuclear NKX2.1 staining (asterisk). The surrounding monolayer shows an aberrant NKX2.1 cytoplasmic staining (arrowheads). B. IPF2 differentiates mainly as a monolayer positive for E-cadherin, which also stains positively for NKX2.1 with a cytoplasmic pattern (arrowheads. Nuclei stained in blue (DAPI). Scale bars= 100µm

Characterisation with immunocytochemistry was repeated with all the cell lines being used for the lung differentiation protocol to make a fair comparison. The same NKX2.1 antibody was employed to detect lung progenitors emerging at the earlier lung induction day eight. To detect the SFTPC-producing cells at the end of the differentiation, a second pro-SFTPC antibody was employed, aimed for the first 30 residues of the peptide's precursor (abcam, ab170699).

Figure 3.9A shows the positive nuclear expression of NKX2.1 on lung day eight on all the cell lines; note the expression on IPF5 clones 3 and 5, and REBLPAT was weak. Again, IPF5 clone 4 and IPF1 presented an NKX2.1 expression pattern that coincided with the annular structures observed in the brightfield images. The rest of the monolayer appeared to be negative for this marker in clone 4; however, IPF1 monolayer was positive, but this time the staining for NKX2.1 was nuclear. The images for the cells stained with the new pro-SFTPC on lung day 12 showed the expected cytoplasmic pattern. By the end of the lung induction, all the differentiated cell lines had areas in the culture that were positive for pro-SFTPC expression. Notably, like with NKX2.1, IPF5 clone 4 annular structures were positively stained for pro-surfactant protein.

As the mean fluorescence intensity (MFI) values presented in **Figure 3.9**B show, the expression of NKX2.1 was not homogeneous among the different IPF patient-derived cultures growing in 2D-culture conditions ($p \leq 0.0001$); IPF5 clone 4, IPF1 and IPF2 had higher values than the rest of the differentiated cell lines. However, these values did not correlate directly with the expression of SFTPC. IPF5 clones 4 and 5 and IPF2 did not present any statistically significant difference in the MFI values of SFTPC.

The seeding, differentiation and staining of IPF5 clone 3 had to be repeated several times as the cells seemed to grow in a thin monolayer that made them prone to detachment during the lung induction stages. Despite the promising expression of lung markers by this clone, its maintenance in culture was laborious, and it was not possible to attain reproducibility. Due to this, clone 3 was also excluded from future experiments.

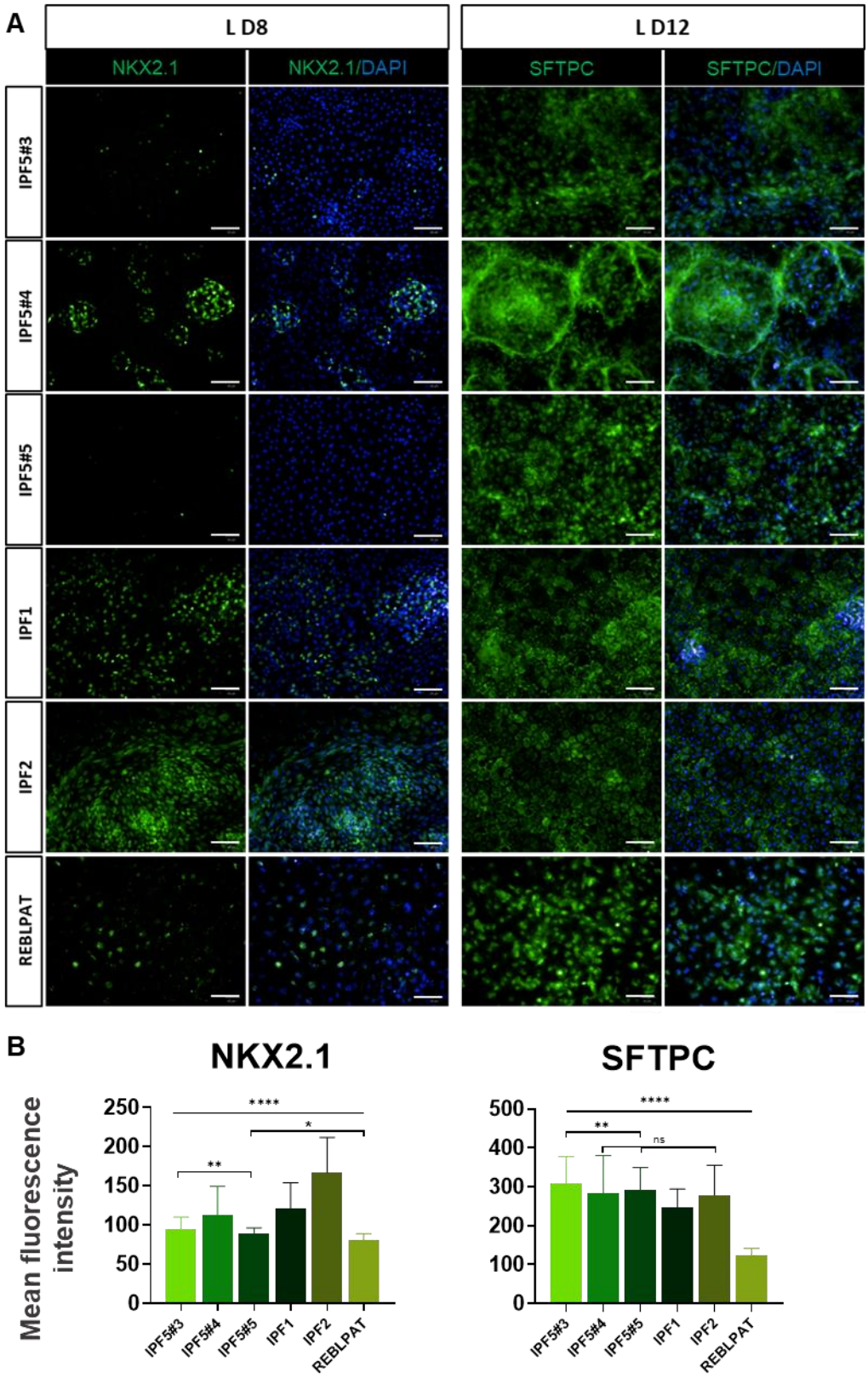

A
L D8
L D12
NKX2.1
NKX2.1/DAPI
SFTPC
SFTPC/DAPI
IPF5#3
IPF5#4
IPF5#5
IPF1
IPF2
REBLPAT
B
NKX2.1
SFTPC
Mean fluorescence intensity
250
200
150
100
50
0
500
400
300
200
100
0
****
*
**
ns
IPF5#3
IPF5#4
IPF5#5
IPF1
IPF2
REBLPAT

Figure 3.9 Differentiated lung epithelial cells produce SFTPC.

A. Representative immunocytochemistry images of IPF5 clones 3, 4 and 5, IPF1, IPF2 and REBPLAT on lung day 8 (LD8) and day 12 (LD12) of lung induction. Images show nuclear staining of the early lung bud marker NKX2.1 (green) and a positive cytoplasmic stain for the AT2 cell marker SFTPC (green). Nuclei stained in blue (DAPI). Scale bars= 100µm B. Mean fluorescence intensity (MFI) values ± standard error of the mean for NKX2.1 on LD8 and SFTPC on LD12. MFI was quantified using Columbus software. One-way-ANOVA statistical analysis for significance was performed in comparison to unstained control *p≤0.05, ** p≤0.01, ****p≤0.0001, and ns= no significant.

Together, these results demonstrated that the generation of lung progenitors using the 17-day differentiation protocol varied significantly between cell lines. Moreover, even if distal lung progenitors were present, the cultures were not homogeneously producing SFTPC. Thus, a maturation stage was needed to generate robust SFTPC-producing AT2 cells.

Maturation conditions in 2D culture increased SFTPC expression.

At the end of the lung progenitors' induction, after 17 days of initiating the directed differentiation, the presumptive $NKX2.1^{+}/SFTPC^{+}$ cells were cultured with a lung maturation cocktail using the basal DIFF-1 media used for the previous stages (RPMI + B-27™ Supplement + NEAA + Pen-Strep). This cocktail consisted of CHIR, DEXA, IBMX, cAMP, and FGF10. The IPF patient-derived cell lines and a control cell line REBLPAT were maintained under these conditions for 16 days to promote their AT2 cell phenotype.

The morphological changes observed between cell lines were similar to the ones presented during the first stages of the lung differentiation. IPF5 clone 4 and IPF1 showed similar annular structures by the fourth day of maturation conditions, whereas the rest of the cell lines were growing as an epithelial monolayer, as shown in **Figure 3.10**. These circular structures disappeared as the maturation protocol progressed and could not be found by day 16. Nonetheless, these structures seemed to have been replaced by tortuous structures and condensed cell areas on clone 4 and IPF1. Even the other cell lines that were previously growing as monolayers developed these highly packed areas surrounded by a monolayer.

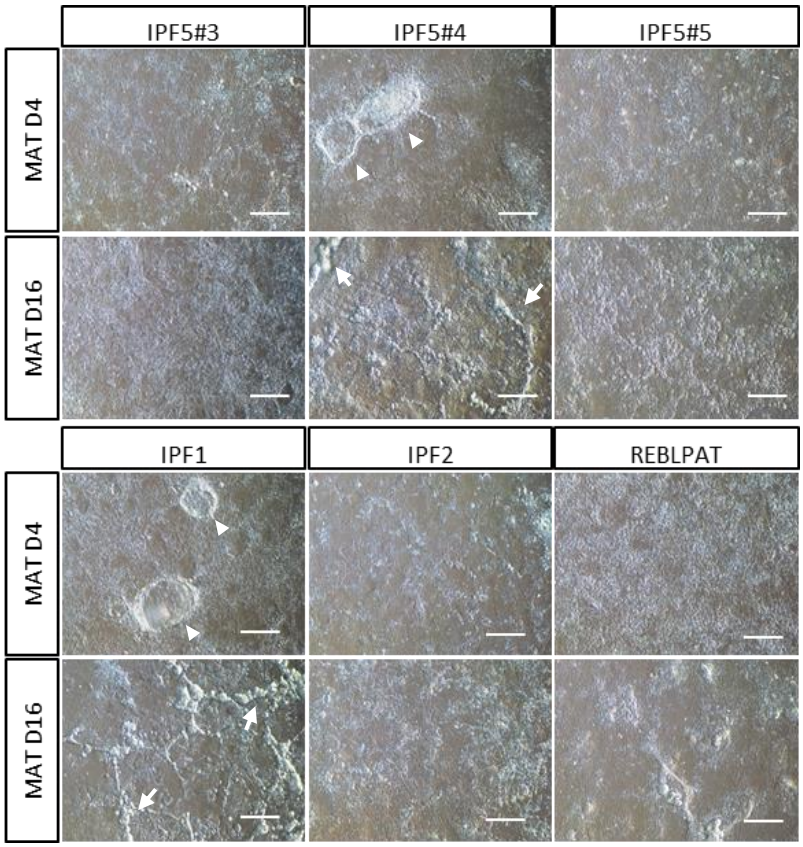


Figure 3.10 Morphological changes of healthy and IPF patient-derived hiPSC during lung maturation 2D-protocol.

Representative brightfield images of IPF5 clones 3, 4 and 5, IPF1, IPF2 and REBPLAT cultured with CHIR, Dexamethasone, IBMX, cAMP, and FGF10 for 16 days after the 12-day lung differentiation protocol. Note the annular structures (arrowheads) and the tortuous structures (arrows) formed on MAT D4 and MAT D16, respectively, for clone 4 and IPF1. MAT D4, MAT D16: days of lung maturation. Nuclei stained in blue (DAPI). Scale bars= 200µm

To assess if the maturation culture conditions improved the expression of SFTPC, culture samples on lung day 12 were fixed and stained with pro-SFTPC and compared to those at maturation days 4 and 16. Results showed that after 4 days of maturation, the IPF patient-derived cell lines and REBLPAT had already improved their expression of SFTPC; however, the staining was not homogenous within the cultures (**Figure 3.11A**). The pro-SFTPC staining seemed to be more intense on the rim of the annular structures of IPF5 clone 4 and IPF1 and the areas with higher cell density on the rest of the cell lines. This expression pattern became more evident on maturation day 16-time point with the emergence of bulky clusters of cells, where it was easier to appreciate that the rest of the monolayer was not stained as intensely for pro-SFTPC. To corroborate this increase in SFTPC production, the MFI of pro-SFTPC was measured at the three time points and compared between cell lines (**Figure 3.11B**). The maturation conditions significantly increased ($p < 0.01$) the expression of SFTPC

on maturation day 4 compared to lung day 12 on IPF5 clones, IPF1 and REBLPAT. IPF2's expression decreased by this time point but not significantly. To be noted, the maintenance of the cells under maturation conditions for an extra 12-day period did not enhance the production of SFTPC any further.

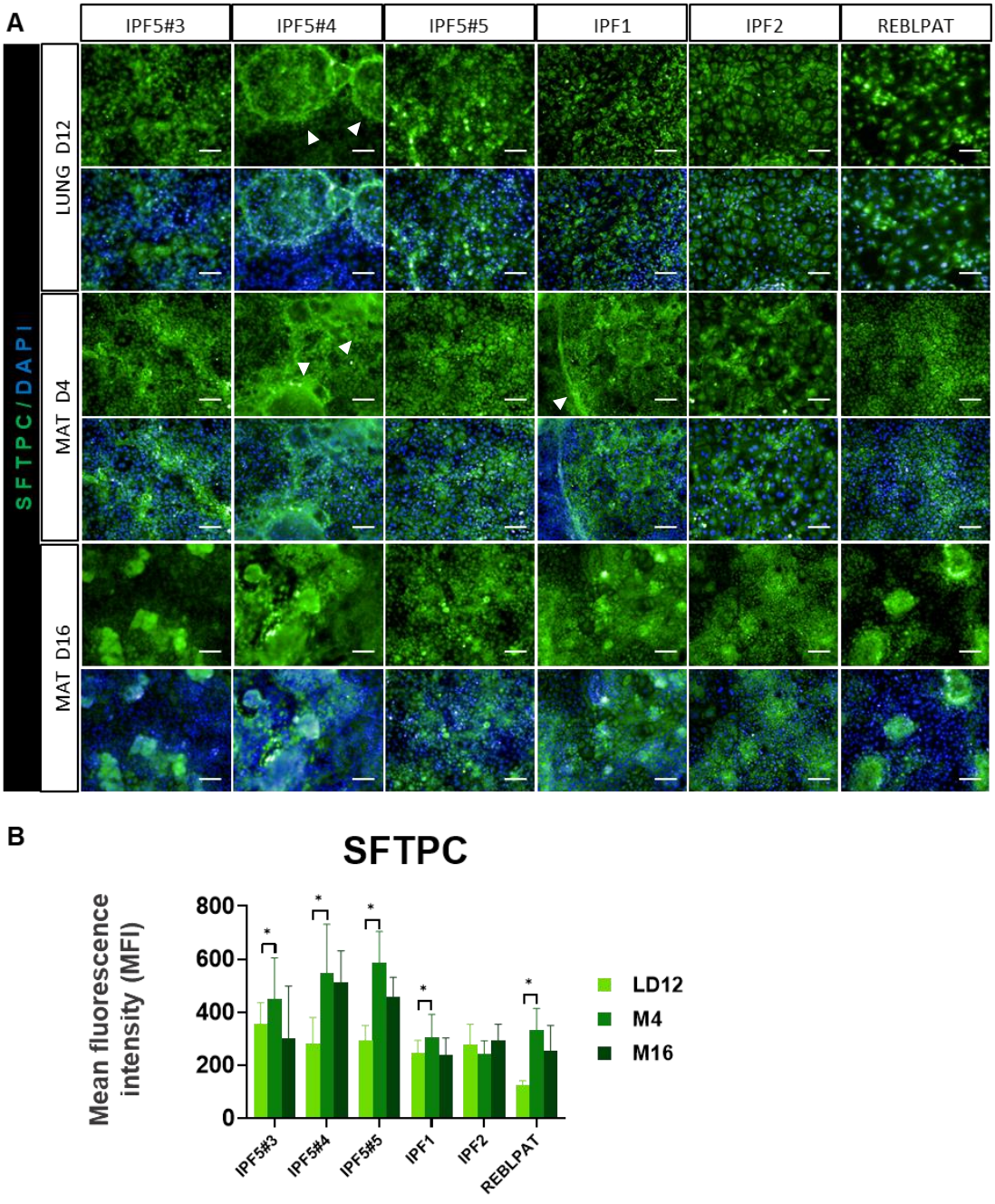


Figure 3.11 2D maturation conditions increased the expression of SFTPC.

A. Representative immunocytochemistry images at the beginning and end of the maturation protocol of IPF5 clones 3, 4 and 5, IPF1, IPF2 and REBPLAT. Images show presumptive hiPSC-derived AT2 cells' SFTPC cytoplasmic (green) expression. Note the annular structures (arrowheads) formed by IPF5 clone 4 and IPF1. Scale bars= 100µm B. Mean fluorescence

intensity (MFI) shows significantly increased expression ($p< 0.01$) of SFTPC by maturation day 4 in comparison with lung day 12 and a decrease on maturation day 16, n=1. MFI ± standard error of the mean was quantified using Columbus software. One-way-ANOVA statistical analysis for significance was performed in comparison to unstained control *$p\leq0.05$.

As AT2 cells mature, they increase their synthesis of phospholipids and proteins required to produce the surface tension-reducing surfactant. It is known that disruption to this biosynthetic function results in the accumulation of misfolded proteins and organelle dysfunction [79]. If the endoplasmic reticulum (ER) stress imparted by this disruption is not resolved with the activation of the unfolded protein response (UPR) downstream pathways, the ultimate outcome would be cell death. Given that mutations in the BRICHOS domain of SFTPC can lead to the accumulation of this misfolded protein, it was hypothesised that IPF patient-derived AT2 cells would have higher levels of ER-stress/ UPR markers upon maturation, in comparison to healthy control-derived AT2 cells. On that account, the gene expression of key components of the UPR pathway was measured on IPF5#5 and the control cell line REBLPAT (**Figure 3.12**).

The UPR protective mechanism against misfolded protein-induced ER damage is mediated by chaperone proteins within the ER lumen, like the binding immunoglobulin protein (BiP) and glucose-regulated protein 94 (GRP94). The detected levels of the chaperone BiP were significantly upregulated in IPF5#5 and REBLPAT along the maturation; however, they were significantly higher in REBLPAT compared to IPF5#5. Contrastingly, the other chaperone measured, GRP94, was also significantly upregulated but only on the SFTPC-mutant IPF5#5. When unstressed, these stress-sensing molecules are bound to three transmembrane signalling activators, inositol-requiring protein 1 (IRE-1), activating transcription factor 6 (ATF6) and protein kinase RNA-like ER kinase (PERK).

When ER stress is detected, BiP dissociates from the transmembrane proteins leading to their activation. In the maturation samples of both cell lines, the expression of IRE-1 and ATF-6 showed a statistically significant increasing trend towards the maturation day 16. However, IRE-1 was significantly higher on IPF5#5 and ATF-6 on REBLPAT. Similarly, the activating transcription factor 4 (ATF4), the downstream mediator of PERK, was also upregulated in both cell lines by maturation day 5. Still, only REBLPAT maintained a significant increase for the later maturation stages.

Furthermore, the expression of X-box binding protein 1 (total, tXBP1; unspliced, usXBP1), the downstream mediator of IRE-1 that is activated upon splicing, also showed a significant upregulating trend as the maturation progressed. The activation

of this pathway leads to the upregulation of other UPR target genes like the ER-degradation-enhancing-α-manidose like protein (EDEM), which aids with the degradation of misfolded proteins by the ubiquitin-proteasome system. This gene was found to be significantly upregulated in the maturation samples of both healthy and diseased cell lines without significant differences between them.

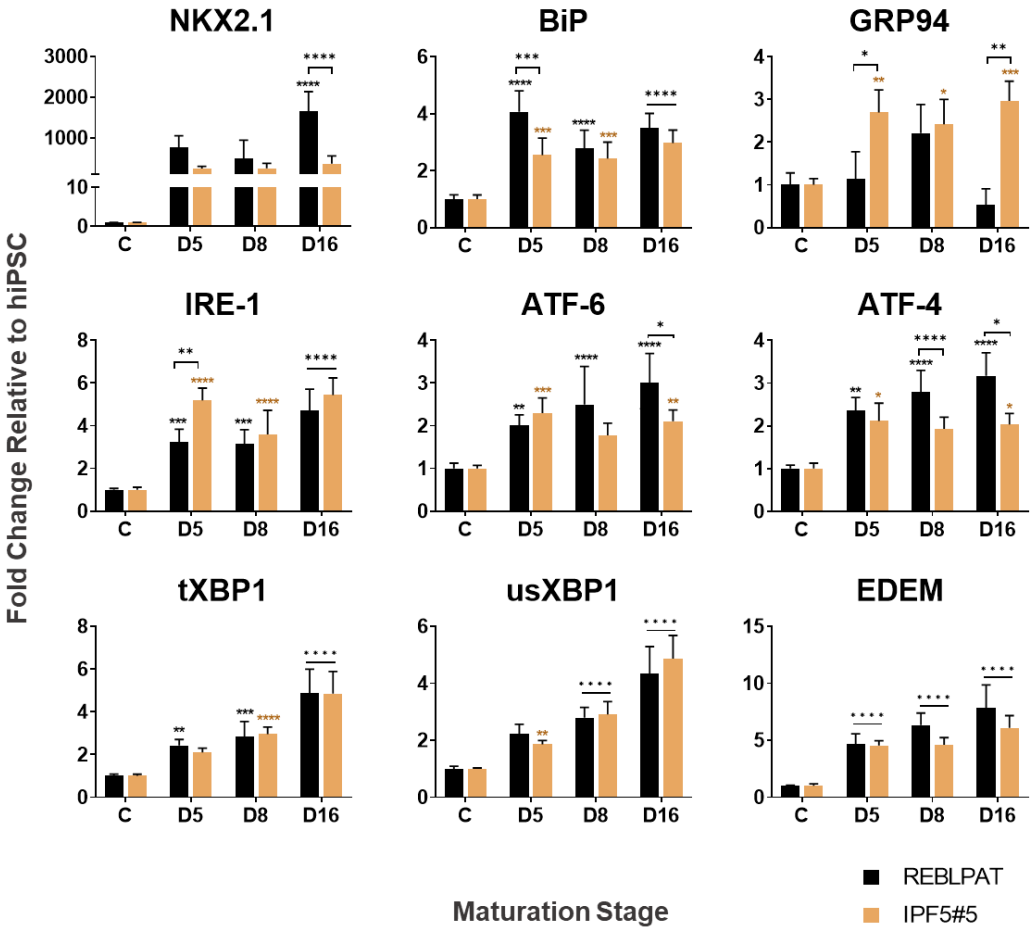


Figure 3.12 In vitro maturation of lung epithelial cells activates unfolded protein response.

qPCR shows the changes in gene expression of key molecules on the downstream pathway of the ER stress-UPR along the maturation protocol of healthy (REBLPAT) and SFTPC-mutant (IPF5#5) hiPSC-derived lung epithelial AT2 cells. C: undifferentiated hiPSCs, D5, D8, D16: days of lung maturation. Data shows the mean CT values ± standard error of the mean. Statistical analysis for significance was performed between sample sets using one-way ANOVA in comparison to undifferentiated hiPSCs. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Overall, these results suggest that the increased production of SFTPC induced by the maturation protocol was not enough to trigger a significant aggravation in ER stress response from mutant cells compared to the control cell line. Nevertheless, the

maturation was sufficient to activate the UPR on the AT2 cells in culture, indicating the cells were trying to attenuate the burden of the ER to avoid its damage, regardless of their genetic background. Therefore, a second hit must be needed to trigger the disease phenotype with ER stress due to the SFTPC BRICHOS domain mutation present in the IPF cell lines.

Transfer of selected cell lines to alternative 3D culture maturation conditions.

Even if suboptimal, the generation of lung progenitors with the employed protocol showed promising results; however, the introduction of a maturation step was deemed necessary to seek the generation of a more robust population of SFTPC-producing AT2 cells. Nevertheless, the maturation of lung progenitors using a 2D approach demonstrated the presence of a heterogeneous population within the cultures. This heterogeneity was especially evident in IPF1, and IPF5#4, where the annular structures spontaneously generated during the 2D-differentiation from the lung induction day eight, expressed higher levels of SFTPC and NKX2.1 than the surrounding monolayer. For this reason, it was believed that the isolation of these annular structures or the promotion of their emergence could enhance the purity of AT2 cell population.

Several approaches were considered for this, including the manual picking of the cyst-like structures, the scratching and disruption of the surrounding monolayer or the overlaying of the monolayer with a Matrigel coating to form a “sandwich” culture allowing the cells to grow in different directions. However, maintaining these cultures would have been laborious, and reproducibility would have been challenging. Therefore, the selected method to further promote the maturation of AT2 cells was embedding 2D-generated lung progenitors in Matrigel droplets. To do so, monolayers of lung progenitors on lung induction day 9 (LD9), after the emergence of the SFTPC/NKX2.1 positive areas, were scratched from the bottom of the culture plates, collected, and resuspended in a 75% Matrigel suspension in maturation media.

Figure 3.13 shows the progression of the formation of organoids after transferring LD9 cells from the IPF patient-derived cell lines, IPF5#4 and IPF1, and two healthy control cell lines, LOPCK and REBLPAT. The emergence of 3D-like structures was evident in the IPF cell lines within the first 24hrs of being embedded in Matrigel. These structures continued to grow and proliferate until the droplet was fully confluent, and the organoids needed to be passaged. Matrigel droplets were mechanically disintegrated with the help of cold PBS, and the organoids were split and reseeded as clumps approximately

every eight days. These steps were repeated with the mutant cell lines during five passages. Unfortunately, the control cell lines did not adapt well to the 3D environment. After six days of being immersed in the Matrigel droplets, the first sphere-like structures started to rise from the cell aggregates of REBLPAT and LOPCK, but they did not increase in size any further. When cells had to be split because of the fragility of the droplets, the forming organoids did not survive the reseeding process, and the cultures were discarded.

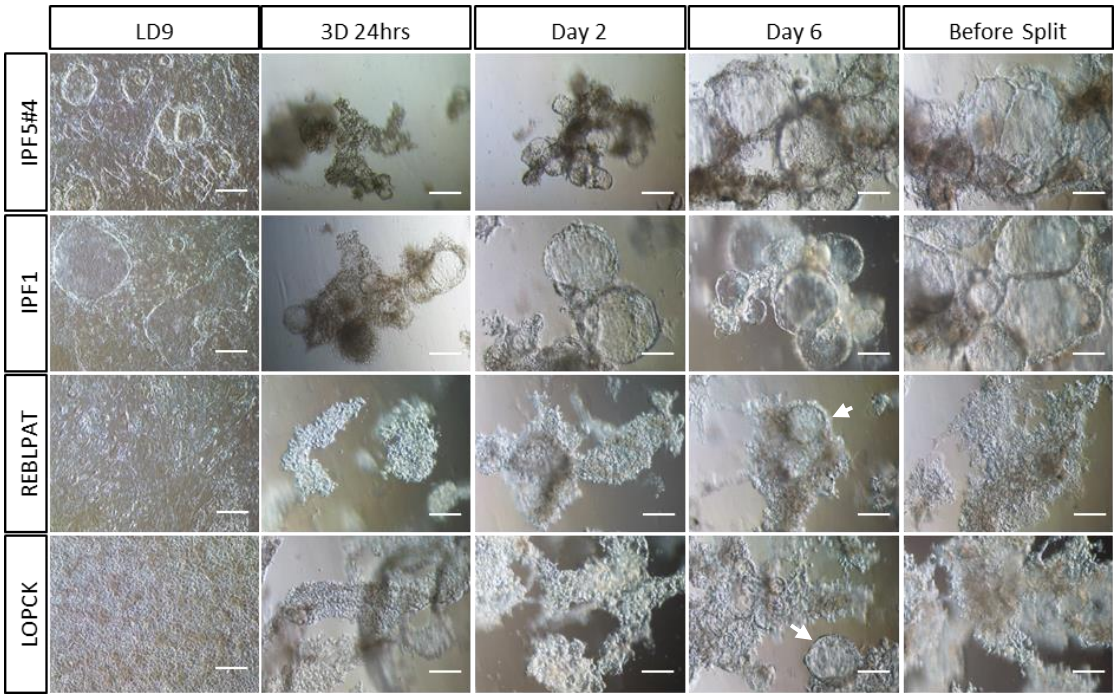


Figure 3.13 Transfer of lung progenitors to 3D culture conditions to generate alveolar organoids.

Brightfield images show the transfer of IPF5 clone 4, IPF1, REBLPAT, and LOPCK lung progenitors on lung day 9 to 3D Matrigel droplets. The formation of thriving 3D structures on IPF cell lines can be observed after 24hrs of being embedded in Matrigel, the few spheres formed on the healthy cell lines (arrows) did not recover after the split. Scale bars= 200µm

Maintenance and expansion of alveolar organoid in culture.

In an attempt to improve the formation of 3D organoids with healthy control cell lines REBLPAT and LOPCK, FGF7 was added to the maturation media. FGF7 has been reported to promote the initial expansion of human epithelial bud tip progenitors in vitro, so including it in our cultures could potentially enhance the emergence of spheroids [206]. The methodology above was repeated with the IPF and control cell lines, LD9 monolayers were transferred to Matrigel droplets, but this time the maturation media was supplemented with FGF7.

IPF5#4 and IPF1 formed organoids when transferred to 3D with both media compositions; however, after passaging them once, the morphological changes due to the addition of FGF7 were conspicuous, as shown in **Figure 3.14A**. IPF5#4 organoids maintained only with maturation media grew slower, and as the cultures progressed to passage 5, the complexity of their structure increased with invaginations and thickening of their walls. The FGF7-supplemented counterpart also developed thicker walls when passaged once, but their lumens were larger, and no folding or invaginations were observed. By passage 5, the organoids had lost their sphere-like shape and developed few indentations. The changes observed in IPF1 were not as pronounced, apart from the proliferation and growth of the organoids. The addition of FGF7 to LOPCK and REBLPAT cultures made the establishment of a 3D culture possible. **Figure 3.14B** demonstrates how ineffective the formation of organoids using the original maturation media for the control cell lines was and how it improved with the presence of FGF7. The supplemented cultures formed spheres immediately after embedding the LD9 monolayer in Matrigel droplets and thrived after every passage.

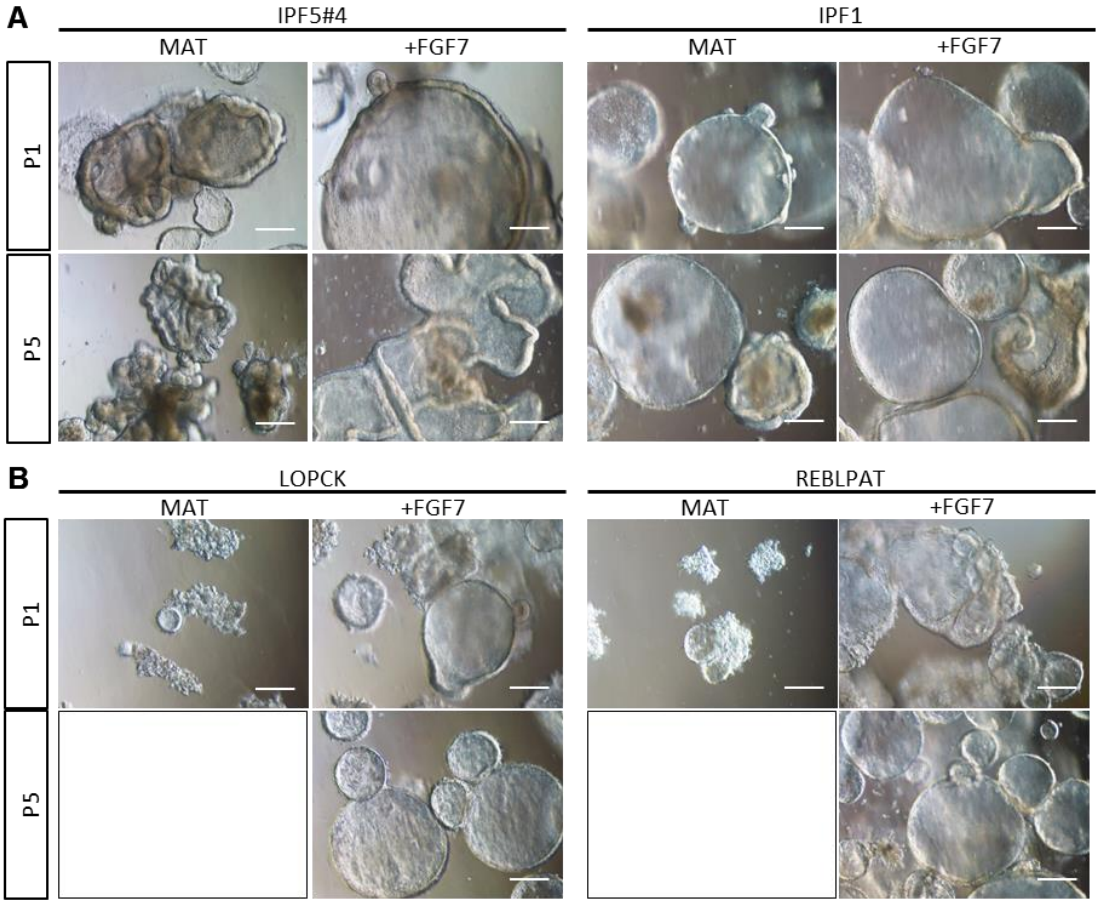


Figure 3.14 Addition of FGF7 promotes organoids' growth and proliferation.

Representative brightfield images of 3D organoids under distinct maturation conditions. A. IPF5#5 and IPF1 organoids on their 1st and 5th passage cultured with or without the presence of FGF7 in the maturation cocktail. The organoids growing without FGF7 are smaller in both cell lines. B. Images show the formation of organoids with the healthy cell lines LOPCK and REBELPAT upon FGF7 supplementation from passage 1 to passage 5. P1 and P5: passages in 3D Matrigel droplets. Scale bars= 200µm

To see if the two conditions induced similar gene expression profiles, IPF-patient-derived organoids growing in maturation media or maturation +FGF7 were collected and processed for qPCR analysis (**Figure 3.15**). The pluripotency marker OCT4 was included to ensure no undifferentiated progenitors were present in the 3D cultures. Both conditions significantly downregulated OCT4 when compared to the undifferentiated hiPSCs; however, the levels expressed in the maturation media were slightly higher than those of the FGF7-supplemented one. The data also revealed a statistically significant upregulation of the lung progenitor markers SOX9 and SOX17 in both media. The latter was expressed nearly 100-fold higher in the supplemented media than in the maturation media, though not statistically significant. The levels of NKX2.1, even if not significant, were also upregulated in both samples with no difference between them. In addition, the mesendodermal and intestinal markers, MIXL1 and CDX2, were included to evaluate the presence of contaminant populations. The expression of MIXL1 was significantly downregulated when organoids were cultured in both media, and as in OCT4, the addition of FGF7 induced a bigger reduction. Unexpectedly, the gene expression of CDX2 was significantly higher in both conditions; nevertheless, this marker was the only one with a statistically significant difference between media, with a significant decrease for the samples maintained in maturation +FGF7 media.

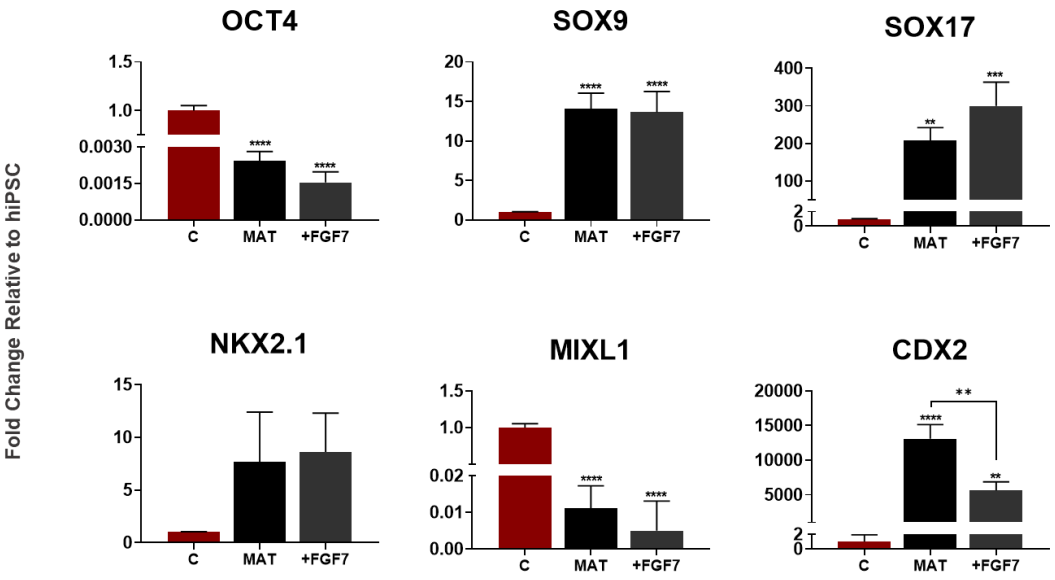


Figure 3.15 Gene expression of IPF-patient-derived organoids growing in different maturation conditions.

qPCR showing IPF1 organoids gene expression of pluripotency (OCT4), lung epithelial (NKX2.1, SOX17 and SOX9), mesendoderm (MIXL11) and intestinal (CDX2) markers when cultured in maturation (MAT) or maturation +FGF7 (+FGF) media. C: undifferentiated hiPSCs. Gene expression in organoids quantified as fold difference over levels in C. Data shows the mean CT values ± standard error of the mean. Statistical analysis for significance was performed between sample sets using one-way ANOVA in comparison to C. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

To assess whether the Bud Tip media reported by Miller A, and not only FGF7, could improve our culture conditions, some spheroids generated and maintained for eight passages in maturation +FGF7 media were split as fragments and seeded with Bud Tip media [206]. Their media composition consisted on a basal media supplemented with FGF7, monothioglycerol, ascorbic acid, RA and CHIR-99021. When using the decribed media, the morphology of the culture changed considerably after ~7 days. The fragments did not grow back into sphere-like organoids after being seeded in Matrigel; instead, they developed into compacted balls or fused budding structures with even ticker walls and no visible lumen (**Figure 3.16A**).

Organoids growing under both conditions were maintained for another three passages until passage 11, and mRNA samples were harvested at every split for qPCR analysis. **Figure 3.16B** shows a panel of lung-associated genes. SOX2, apart from being a pluripotency marker, is also present in the developing proximal airway epithelium. Its downregulation on REBLPAT and LOPCK spheroids growing in maturation and Bud Tip media suggested both induced the generation of distal airway epithelium. This

correlated with the significant upregulation of the distal airway epithelium marker SOX9 in both cell lines for both conditions. The lung bud marker SOX17 was also significantly higher on both cell lines, but only in the maturation media. Surprisingly, even if not statistically significant, the levels of NKX2.1 were lower than the undifferentiated control for both conditions in REBELPAT and most of LOPCK's. The same expression pattern was identified on the AT2 cell markers, SFTPC and SFTPB. Just LOPCK grown with maturation media on passage 11 showed a significant increase for SFTPC and on passage 10 for SFTPB but under Bud Tip media. To assess if contaminant lineages were also growing in the organoids as they did in the earlier lung specification stages, the mesendodermal marker MIXL1 and the intestinal marker CDX2 were analysed. This time, organoids maintained in maturation media +FGF7 at the low passage two were compared to the ones in passage eleven. The graphs in **Figure 3.16C** show the downregulation of MIXL1 on both samples for REBLPAT and LOCPK, contrary to CDX2, which was significantly upregulated on both cell lines. However, the trend between them differed; REBLPAT showed a higher CDX2 expression in the latest timepoint, whereas LOPCK was significantly higher in passage 2.

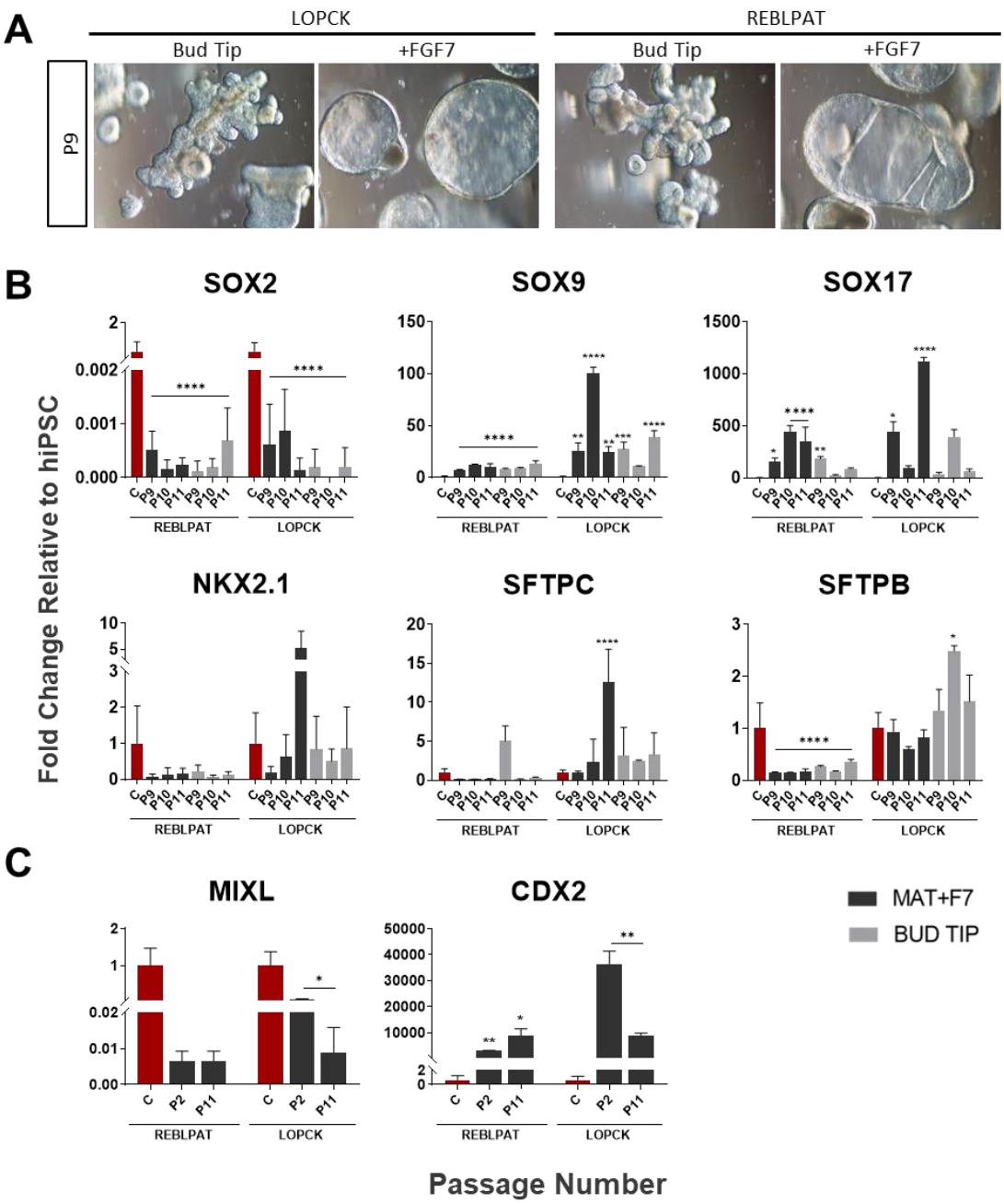


Figure 3.16 Bud tip media and long-term maintenance of healthy hiPSC-derived organoids did not improve their maturation.

A. Brightfield images of LOPCK and REBLPAT organoids generated in +FGF7 media. These were transferred from +FGF7 media to Bud Tip media on passage 9, causing the thickening of the walls and increasing the complexity of the organoid's structure. Scale bars= 200µm B. qPCR analysis of REBELPAT and LOPCK organoids harvested on passages 9, 10 and 11 grown in maturation media supplemented with FGF7 from passage 0 and organoids transferred to Bud Tip media on passage 9. Graphs show the expression of genes encoding for lung bud and airway markers SOX2, SOX9, SOX17 and NKX2.1; and mature AT2 markers SFTPC and SFTPB C. Gene expression of mesendoderm (MIXL1) and intestinal (CDX2) markers on organoids harvested on passage 2 and 11 cultured with maturation media + FGF7 analysed by qPCR. Gene expression in organoids quantified as fold difference over levels in hiPSCs, "C". Data

shows the mean CT values ± standard error of the mean. Statistical analysis for significance was performed between sample sets using one-way ANOVA in comparison to undifferentiated hiPSCs. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

In order to discard that the low levels of lung markers on the spheres growing in Bud Tip media were not due to the organoids being initially generated in maturation +FGF7 media, new cultures were set up and transferred to Matrigel droplets on LD9 with maturation +FGF7 media or Bud Tip media. The differences in morphology induced by each condition are represented in **Figure 3.17**. The fragmented monolayer transferred to 3D in Bud Tip media grew into compacted clusters of cells; these did not change notoriously after being passaged once. For the second passage, LOPCK formed big aggregates of smaller clusters, whereas REBLPAT grew into multiple compacted balls without a lumen. The aspect of the organoids growing in maturation +FGF7 was comparable to the one previously presented, sphere-like structures with big lumens and thick walls.

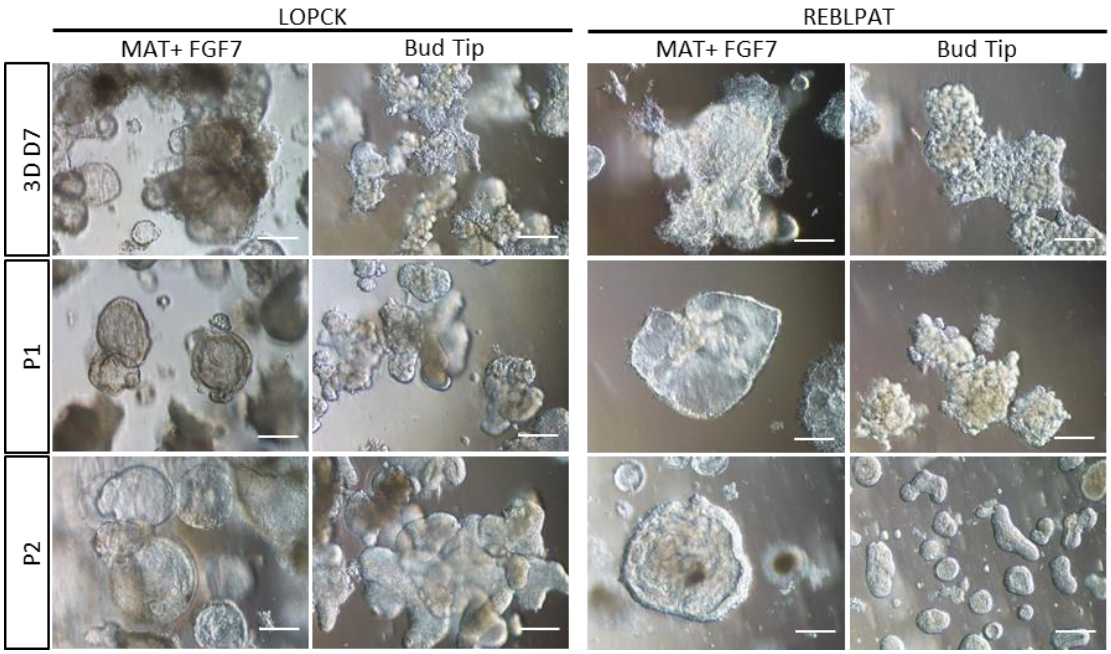


Figure 3.17 Generation of healthy control organoids cultured in Bud Tip media.

Brightfield images comparing the generation of 3D organoids with healthy hiPSCs using MAT+FGF7 or Bud Tip media. Images show LOPCK and REBLPAT differentiated lung progenitors seven days after being transferred to 3D droplets and the generated organoids on the first and second passage. Scale bars= 200µm

When analysing the samples harvested on these early passages by qPCR, the results did not improve in comparison to the organoids analysed before at later passages (**Figure 3.18**). The levels of NKX2.1 in REBLPAT matched those of the undifferentiated

hiPSCs in the organoids growing in both conditions. The upregulation of this marker in LOPCK was maintained after one passage, although not statistically significant. Furthermore, the expression of both surfactant proteins did not improve either, with SFTPC being lower than the control and SFTPB showing only slight increases that were not statistically significant.

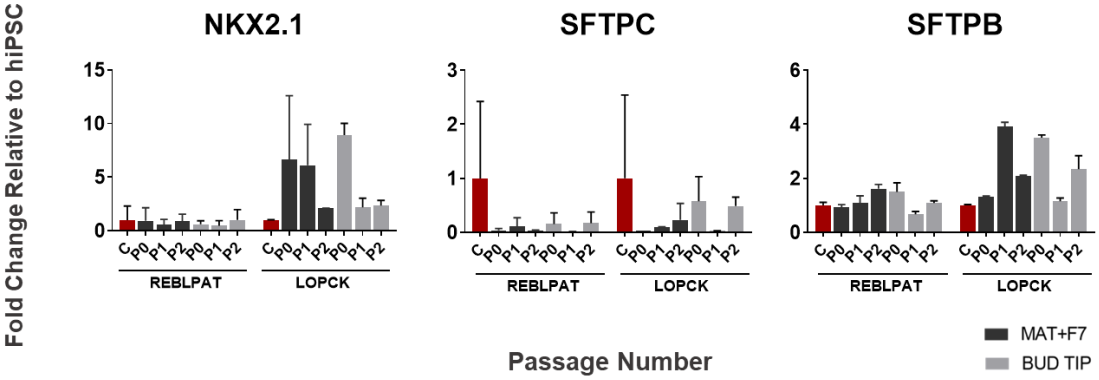


Figure 3.18 Healthy organoids generated in Bud Tip media do not upregulate AT2 cell's gene expression.

Gene expression of mature AT2 markers on REBELPAT and LOPCK organoids generated in maturation media supplemented with FGF7 or Bud Tip media. Samples were harvested on passages 0, 1 and 2 for qPCR analysis. Gene expression in organoids quantified as fold difference over levels in hiPSCs with no statistical significance found. Data shows the mean CT values ± standard error of the mean. Statistical analysis for significance was performed between sample sets using one-way ANOVA in comparison to undifferentiated hiPSCs, with no significant differences found. n=1.

Another prospect was that the primers used to evaluate the gene expression of these lung-specific markers were malfunctioning. A new set of primers was ordered and used to run the same samples, but the same results were obtained. Therefore, a total human adult lung RNA sample (BioChain) was processed and analysed to ensure the issue was not technical and to provide a second control that could be used for benchmarking the in vitro generated lung epithelial cells. From the qPCR graphs in **Figure 3.19**, it is apparent that the primers were working. The expression of the lung marker NKX2.1 was nearly 3,000-fold higher in the lung sample than in the undifferentiated control. Likewise, the fold changes in expression of the AT2 cell markers SFTPC and SFTPB were above $0.5x10^5$ and $4x10^6$, respectively, in the total lung sample relative to hiPSCs. AQP5, an AT1 cell marker, was also significantly higher in the lung sample compared to the control, but the expression level was lower than that of the AT2 cell markers. That was also the case with the lung bud/distal makers, SOX17 and SOX9. The only marker with a lower expression relative to the undifferentiated control was SOX2.

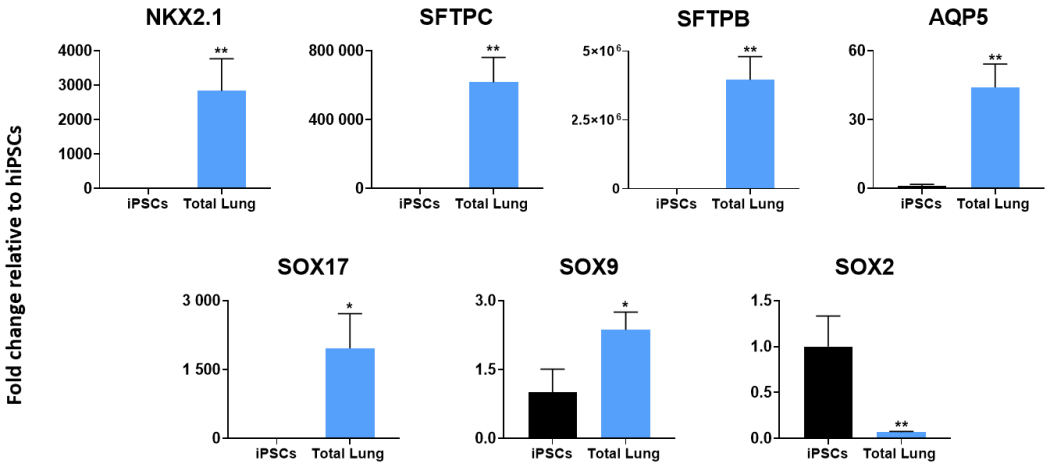


Figure 3.19 Expression levels of lung markers in a human total lung sample.

qPCR analysis shows gene expression of an array of lung-related markers in a human total lung sample (BioChain) compared to undifferentiated hiPSCs. Data shows mean CT values ± standard error of the mean. Unpaired student's t-test statistical analysis for significance was performed *p≤0.05, ** p≤0.01. n=1.

While the specificity of all antibodies used in this thesis was initially assessed in silico using the online tool BLAST NCBI to prevent any issues related to off-targets and secondary structures, it is crucial to perform experimental validation as well [207]. Validation methods, including gel electrophoresis, Sanger sequencing, endpoint PCR with restriction digest, and DNA sequencing, should be employed to minimise the risk of inaccurate results, leading to misleading conclusions [208]. Therefore, careful result interpretation is essential. Notably, qPCR results revealed distinct single peaks in the melt curves for all sets of primers used in this thesis, as exemplified in the **Annex. 1**. Taken together, these results were concerning as they indicate that the presumptive distal lung epithelial cells contained in the control cell lines' organoids were not differentiating effectively into AT2 cells, despite the time in culture or media conditions. Therefore, it was essential to assess whether these results were due to the cell lines being employed or the lung maturation protocol.

The same methodology was followed to address this matter with an IPF patient-derived cell line; differentiated LD9 lung progenitors were scrapped and transferred to Matrigel droplets with maturation +FGF7 media or Bud Tip media. Images of the 3D cultures in **Figure 3.20** show the emergence of spherical organoids with ample lumens and thin walls when the cells were cultured in maturation +FGF7 media. As with the control cell lines, the organoids formed by IPF5#4 in Bud Tip media had thicker walls. Instead of

forming compacted balls or clusters, they had budding and infolding areas, like the ones observed in intestinal organoids, forming a complex structure with large lumens.

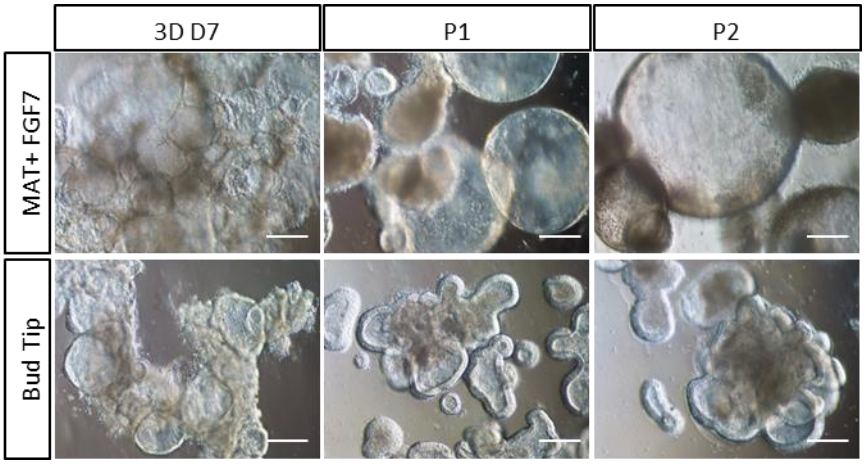


Figure 3.20 Generation of IPF patient-derived organoids cultured in Bud Tip media.
Brightfield images comparing the generation of 3D organoids with IPF5#5 using MAT+FGF7 or Bud Tip media. Images show IPF5#5 differentiated lung progenitors seven days after being transferred to 3D droplets and the generated organoids on the first and second passage. Scale bars= 200µm

The organoids were maintained in culture for five passages collecting RNA samples on every split for gene expression analysis. **Figure 3.21** shows the expression of the AT2 markers NKX2.1, SFTPC and SFTPB. These results were somewhat surprising as the three markers were significantly upregulated as the 3D culture progressed in maturation +FGF7 media. NKX2.1 was upregulated since the cells were transferred into the Matrigel droplets (P0), peaking after the first passage. The surfactant proteins had a rising trend, with a peak in passage three and in passage 5 for SFTPC. The samples growing in Bud Tip media showed a minimal increase in NKX2.1 and SFTPC on the first three passages, yet not statistically significant.

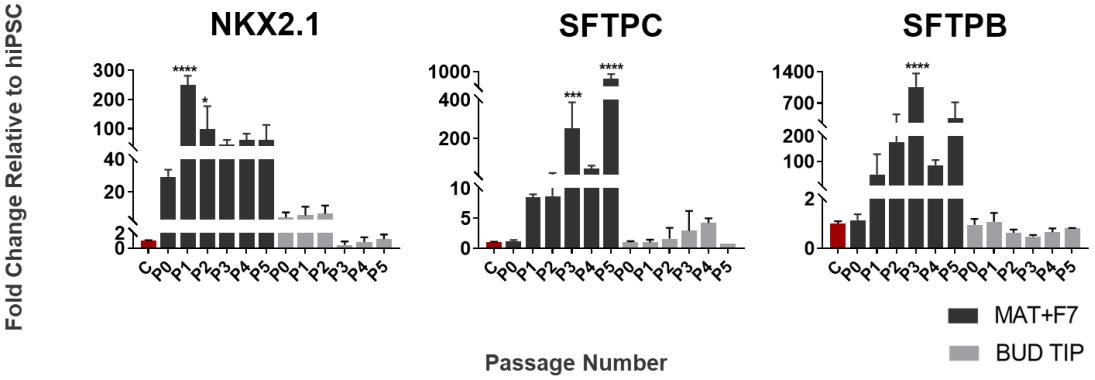


Figure 3.21 Maturation of IPF patient-derived organoids increased expression of AT2 cell markers.

Gene expression of mature AT2 cell markers on IPF5#5 organoids generated in maturation media supplemented with FGF7 or Bud Tip media. Samples were harvested on passages 0 to 5 for qPCR analysis. Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to C: undifferentiated hiPSCs. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

With this, it could be confirmed that the 3D maturation culture conditions using a maturation cocktail supplemented with FGF7 were suitable for the generation of AT2 cells using IPF-patient-derived hiPSCs and that the implementation of the Bud Tip media did not enhance the organoid culture. Due to the positive effects of FGF7 on the establishment and maintenance of the 3D cultures, it was adopted as part of the maturation media along with CHIR, Dexamethasone, IBMX, cAMP and FGF10. Furthermore, it was clear that REBLPAT and LOPCK's gene expression profiles did not match normal lung development, turning them unsuitable for generating hiPSCs-derived healthy alveolar organoid controls. Therefore, only genetically corrected isogenic IPF-patient-derived hiPSCs would be used to generate healthy controls.

Further characterisation of the AT2 cell-containing organoids was required to validate the expression of SFTPC. Immunocytochemistry was performed on whole-mount organoids at passage four, which were cultured in maturation media supplemented with FGF7. Organoids were fixed in the Matrigel droplets to be later stained and imaged. The purpose was to investigate the intracellular trafficking and processing of SFTPC by co-localisation with cellular structures; however, the quality of the images was not suitable for this purpose. **Figure 3.22A** shows how the fluorescence of calnexin, a marker of the endoplasmic reticulum, is glowing out of the sphere. Furthermore, the SFTPC staining in green was mainly localised in the organoid, but multiple bubbles within the surrounding Matrigel also got coloured in green. Only DAPI,

the nuclear dye used, seemed to have permeated the Matrigel droplet successfully. However, this was insufficient to discern the intracellular localisation of the structures stained in green and red. Nevertheless, these images helped confirm that these were epithelial organoids expressing the transmembrane cell adhesion protein E-cadherin on the organoid's outline. Furthermore, a few organoids were imaged with more clarity, and it was possible to observe that cells within the organoids showed the nuclear NKX2.1 staining, and a proportion of cells were also positive for what seemed to be cytoplasmic pro-SFTPC staining (**Figure 3.22B**).

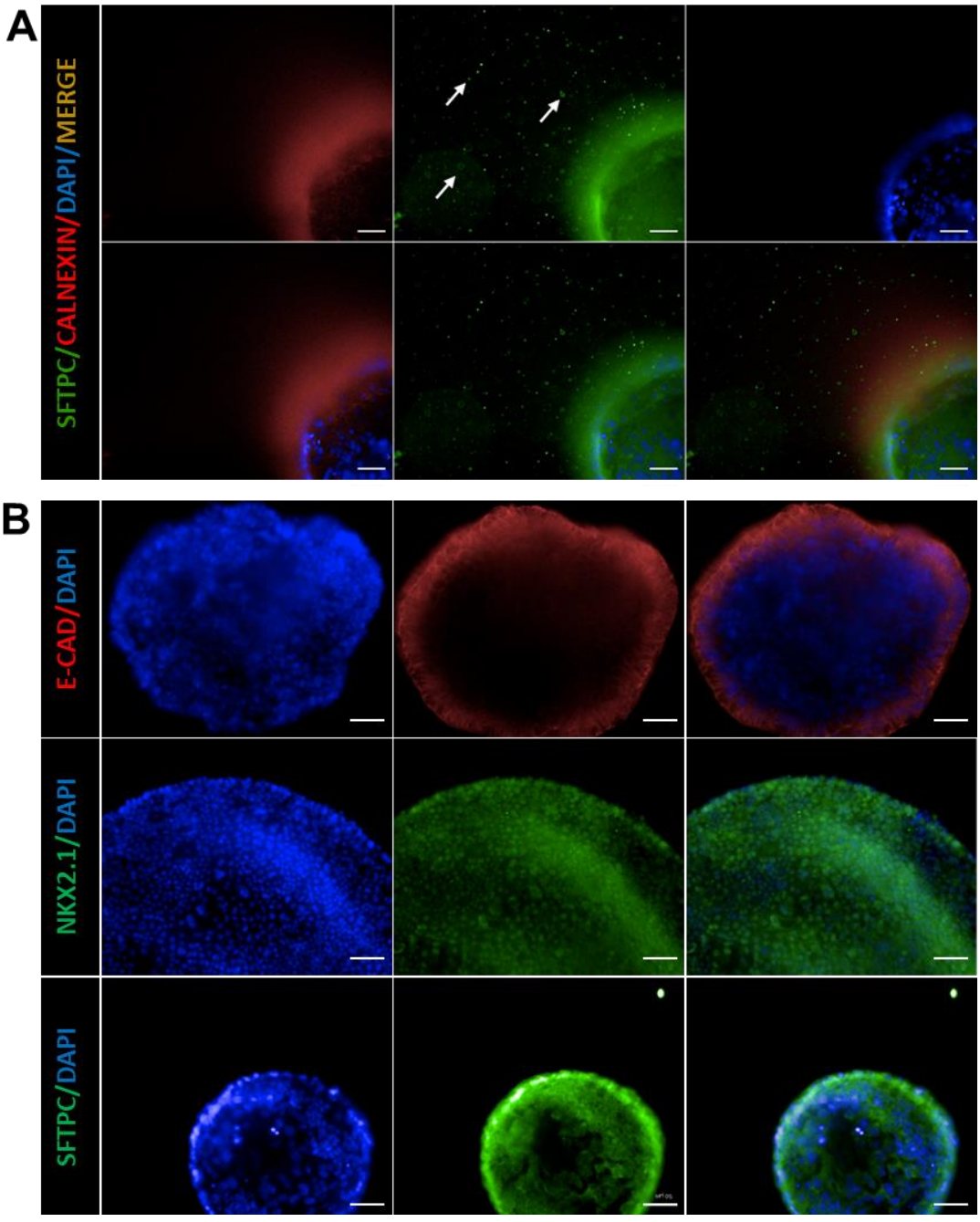


Figure 3.22 SFTPC-mutant hiPSC-derived organoids express main AT2 cell markers.

Immunocytochemistry images of whole mount IPF patient-derived organoids fixed four days after their third split. A. IPF1 organoids showing unspecific staining using calnexin (red) and pro-SFTPC (green) antibodies. White arrows point to bubbles in the surrounding Matrigel stained in green. B. Images show positive membrane staining for the epithelial marker E-cadherin (ECAD, red). The organoids also appear to be positively stained for NKX2.1 (nuclear) and pro-SFTPC (cytoplasmic) lung markers (green). Organoids were co-stained with the nuclear dye DAPI (blue). Scale bars= 50µm

In order to assess if the initial seeding density of the lung progenitors' differentiation could have an impact on the efficiency of the lung markers expression at the maturation stage, IPF-patient-derived hiPSCs were seeded at two different seeding densities, 20k and 30k cells/cm^2. At the end of lung induction day 9, the monolayers were scratched and embedded in Matrigel droplets to induce the formation of sphere-like organoids. **Figure 3.23** shows the changes in the morphology of IPF1 along the differentiation under the two seeding densities. Lung day four cells seeded at both densities presented a series of rosette-like structures that were more prominent in the higher density. These structures disappeared by day nine and were supplanted by annular structures with a prominent rim. Both conditions look similar after two days in 3D-culture; however, by the second week in the Matrigel droplets, the organoids coming from the differentiation seeded at 30k cells/cm^2 had accumulated considerable amounts of debris in the lumen, observed as dark brown areas.

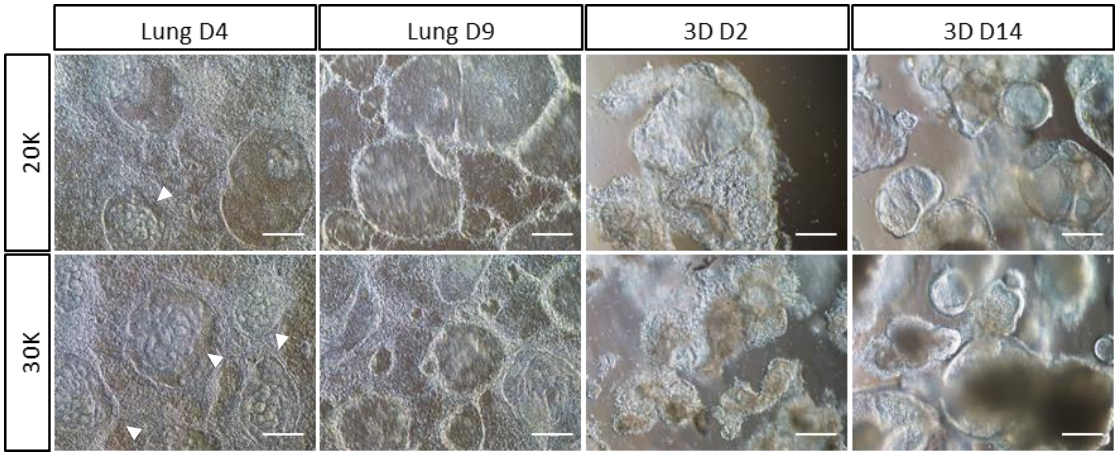


Figure 3.23 Seeding density affects the morphology of the alveolar organoids.

Brightfield images of IPF1 on day 4 and day 9 of the lung differentiation protocol and day 2 and day 14 after being transferred to 3D maturation conditions. Note the presence of rosette-like structures (arrowheads) in Lung D4. Scale bars= 200µm

Samples were harvested when the cells were transferred to the 3D culture (P0) and the organoids generated in the subsequent three passages. Gene expression of both culture conditions was analysed by qPCR for the AT2 cell markers NKX2.1, SFTPC

and SFTPB (**Figure 3.24**). It was encouraging to see the statistically significant rising trend in the levels of NKX2.1 as the organoids were maintained in culture. In passage three, the IPF1 organoids originated from the lung differentiation seeded at 20k cells/cm^2 reached a nearly 800-fold increase in NKX2.1 expression relative to the control, whereas the cells seeded at 30K cells/cm^2 reached a 500-fold increase. Interestingly, the mRNA levels of SFTPC did not correlate with those of NKX2.1. The organoids from the 20k cells/cm^2 differentiation presented a minor increase in their SFTPC expression along the culture. Conversely, the 30k cells/cm^2 condition revealed a 1200-fold increase in SFTPC expression at passage 3, which correlated nicely with the 1000-fold increase in SFTPB.

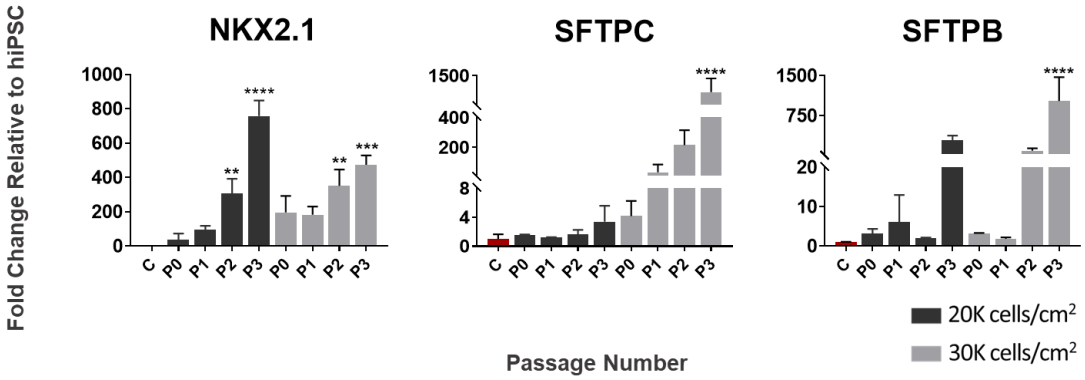


Figure 3.24 IPF patient-derived organoid gene expression is affected by lung differentiation's seeding density.

IPF1 hiPSC were seeded at different densities to generate lung progenitors to be transferred to 3D Matrigel droplets. Generated organoids were harvested on passages 0 to 3 to be analysed by qPCR for mature AT2 cell markers. Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to C: undifferentiated hiPSCs. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Thus far, the results indicate that moving from a 2D differentiation and maturation platform to a hybrid 2D-3D approach was essential to generate AT2 cells with the IPF-patient derived cell lines. The organoids generated via this approach were maintained in culture for more than 100 days in maturation culture conditions without evident changes in morphology. However, when analysing the gene expression profiles of the containing cells, the oscillating levels of surfactant proteins and other crucial markers were concerning. This pointed towards a heterogeneous population of cells growing in the organoids, together with the SFTPC-producing AT2 cells. Since the healthy cell lines used to attempt the generation of control AT2 cells for the disease modelling failed

to differentiate successfully, it was necessary to create an isogenic pair of controls derived from an IPF-derived hiPSC line known to differentiate properly into alveolar cells.

3.2.3 Gene editing of SFTPC-mutant hiPSCs to create homozygous mutant and corrected isogenic controls.

After characterising the lung differentiation potential of the different reprogrammed IPF patient-derived cell lines, the IPF1 hiPSC line was selected to be genetically modified. Using base editing technology, this SFTPC heterozygous mutant cell line ($SFTPC^{WT/MUT}$), was modified to create a healthy isogenic control ($SFTPC^{WT/WT}$), and a homozygous mutant ($SFTPC^{MUT/MUT}$) isogenic cell line. The generated genetically modified cell lines would provide the whole spectra of the SFTPC genotype, ideal for the modelling of genotype-phenotype correlations in IPF pathogenesis.

All gene editing work of IPF cell lines was performed by Sara Cuevas Ocaña at the University of Nottingham.

Optimisation of IPF hiPSC CRISPR base editing for high-efficiency enzyme targeting

Transfection of the parental SFTPC-mutant hiPSCs was first optimised using nucleofection and a GFP plasmid. 419ng of GFP plasmid were used to monitor transfection efficiency by fluorescent microscopy or flow cytometry 24 and 48hrs post-transfection. Cells were transfected using a nucleofection mixture containing P3 buffer or Ingenio® electroporation solution to identify the best transfection reagent for the cells. Furthermore, to achieve the highest transfection efficiency, the following transfection programs were screened: CA-137, CB-150, CD-118, CE-118, CM-113, DC-100, and DN-100.

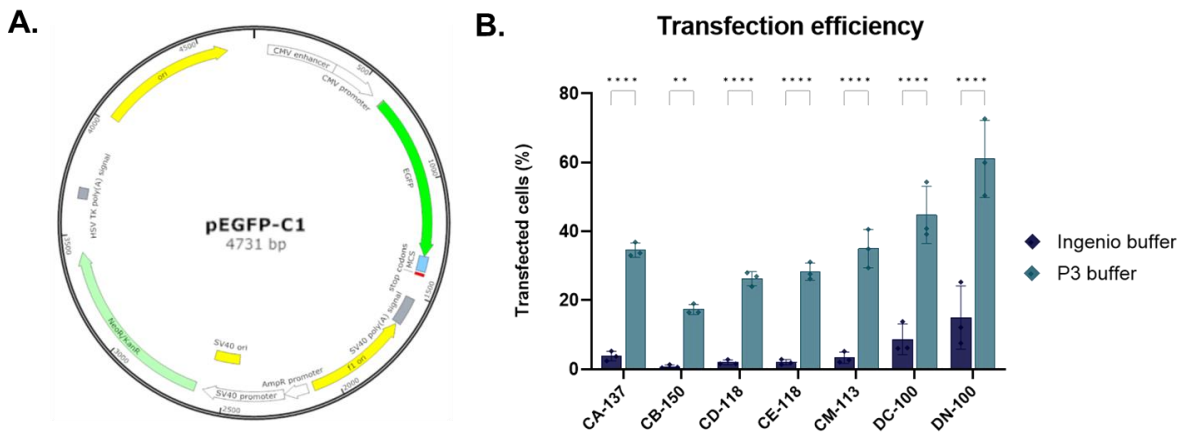


Figure 3.25 Determination of transfection efficiency.

A) Plasmid map of eGFP-C1 containing Kanamycin resistance cassette (light green) used for transfection purposes B) Flow cytometric quantification of GFP+ cells transfected using 419ng GFP. Data shows percentage ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance between samples sets. ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=3

Transient expression of GFP plasmid was checked 24 hours post-transfection with a fluorescent microscope and quantification of GFP-positive cells was performed with flow cytometry analysis 48hrs post-transfection (**Figure 3.25**). Nucleofection using the P3 buffer outperformed the results obtained by using Ingenio® electroporation solution as the transfection reagent, regardless of the programme used. The lowest percentage was achieved with the CB-150 programme and the highest with the DN-100. Therefore, P3 buffer and nucleofection programme DN-100 were chosen as the preferred conditions to achieve the maximum transfection efficiencies.

Once the optimal transfection conditions were determined, SFTPC heterozygous mutant cells were subjected to nucleofection using the different plasmid vectors designed to generate the SFTPC homozygous mutant and corrected cell lines. These comprised a base editor containing plasmid (pCMV_ABEmax_p2A_GFP or pCMV_AncBE4max_p2A_GFP) and an independent plasmid where a guide RNA was cloned into (**Figure 3.26**).

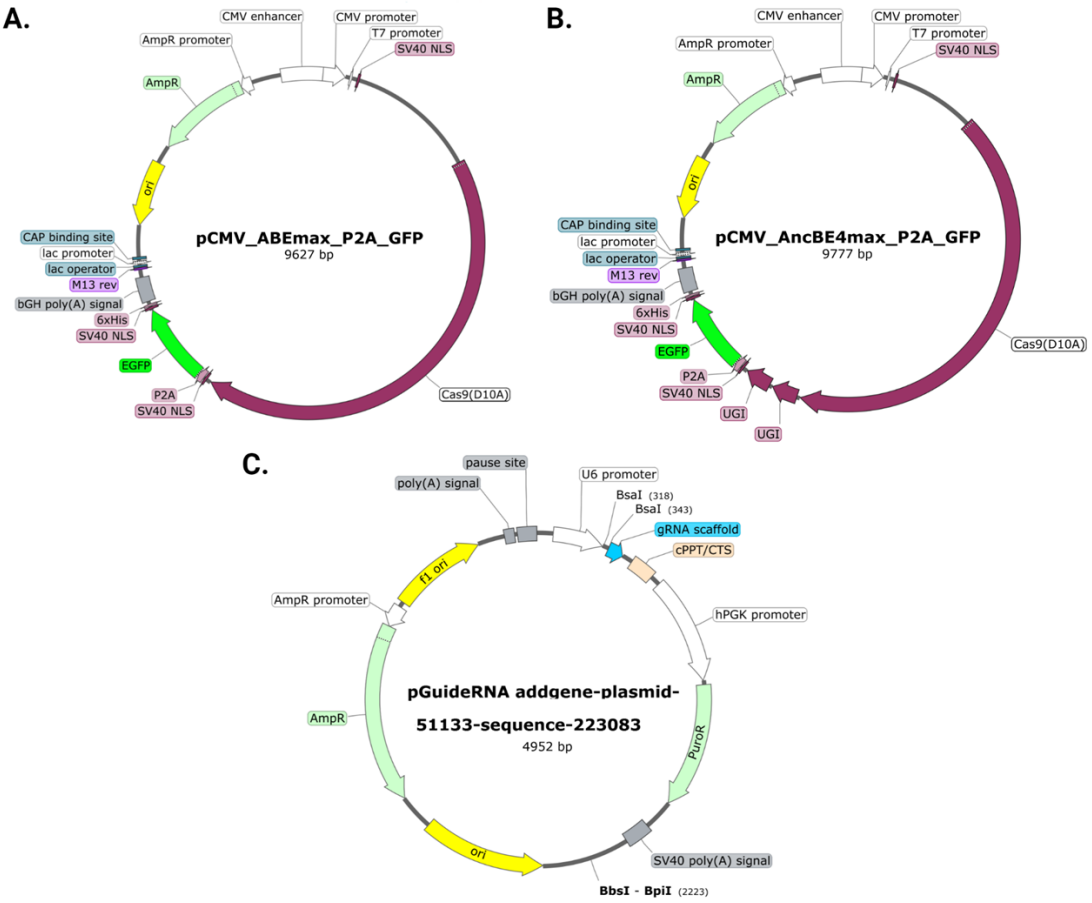


Figure 3.26 CRISPR/Cas9 plasmid maps used for base editing.

Plasmid maps detailing elements of the targeting vectors A) pCMV_ABEmax_P2A_GFP and B) pCMV_AncBE4max_P2A_GFP. C) Plasmid map of guide RNA pGuideRNAaddgene-plasmid-51133-sequence-223083.

Following nucleofection of these plasmids, cells were seeded into a well of a 24-well plate to allow them to recover and expand for 48 hrs. Cells transfected with 723ng of the GFP plasmid showed good viability after 48hrs by light microscopy and strong GFP expression as observed by fluorescence microscopy, in comparison to the untransfected control. When analysed by flow cytometry, transfected cells were nearly 80% GFP-positive (**Figure 3.27A, B**). The hiPSCs transfected with 1474ng of ABEmax plasmid showed 51% positive GFP cells by flow cytometry analysis versus 64% when enriched for transfection using transient antibiotic treatment with puromycin (0.7ug/ml). Morphologically, the puromycin-resistant colonies were clearly defined and compacted (**Figure 3.27C**). Following nucleofection and puromycin enrichment, the populations were dissociated into single cells and plated at clonal density. The HpyCH4V restriction digestion at the population level confirmed the base editing of the SFTPC heterozygous mutant cells. The AncBE4max_P2A_GFP corrected the G>A,

generating SFTPC wild-type cells (band size at 799bp) and the ABEmax_P2A_GFP introduced the mutation A>G, generating homozygous SFTPC mutant cells (band size at 218bp) as observed in **Figure 3.27D**.

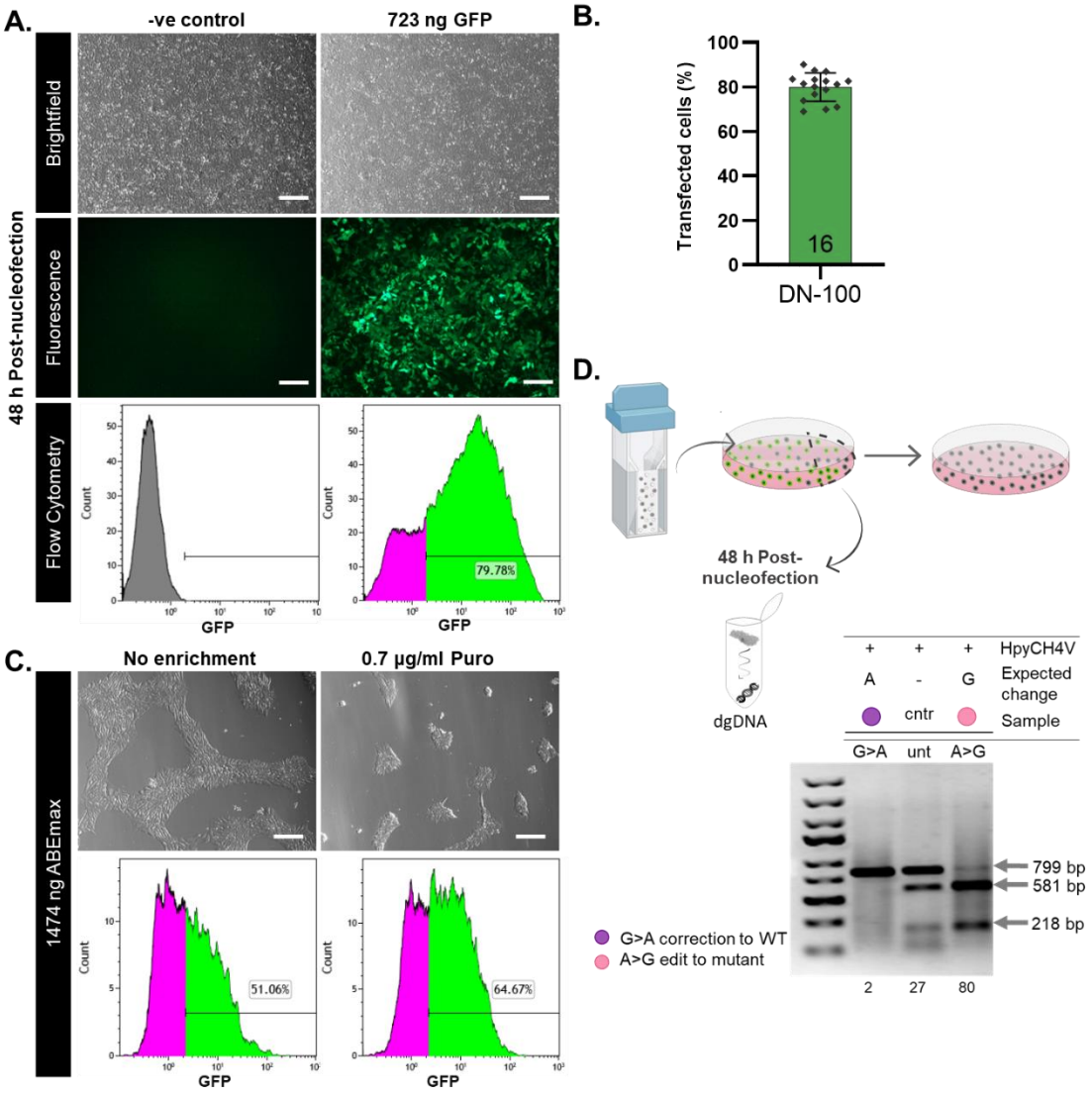


Figure 3.27 Optimised transfection parameters and transient puromycin enrichment

Transfection of IPF-patient-derived iPSCs checking cell viability and GFP expression 48 hours post-transfection using light and fluorescent microscopy, and flow cytometry. A) Panel shows results for GFP nucleofection using 723ng in comparison to the negative control. Panel in B) Quantification of transfection efficiency using 723ng GFP, DN-100 program and P3 buffer via flow cytometry. n=16. C) Panel shows results for ABEmax nucleofection using 1474ng with or without Puromycin enrichment (0.7ug/ml). D) Overview of the detection of edits at the population level by restriction digestion using the HpyCH4V enzyme. PCR products of populations successfully transfected with AncBE4max_P2A_GFP (purple) 799bp and ABEmax_P2A_GFP (pink) 218bp DNA bands. Untransfected cells were used as negative control. Scale bars= 100µm

Genotypic characterisation of targeted clones

To detect successful editing at the clonal level, puromycin-resistant clones were screened by PCR genotyping. A small portion of the colony was manually dissected to be used for genomic DNA extraction, and genotyping via PCR and HpyCH4V restriction digestion (**Figure 3.28A**). The other part of each of the dissected colonies were transferred to a well of 48-well plate for individual expansion. **Figure 3.28B** shows a representative 1% agarose gel of adenine base editing clonal screening using HpyCH4V enzyme assay to determine the number of clones correctly edited by base editing. As ABEmax generates mutant cells (A>G), the gel shows the absence of 799bp DNA bands corresponding to G>A in most clones, except for clones 5, 9 and 10, indicating an 72% efficiency (8/11 screened clones).

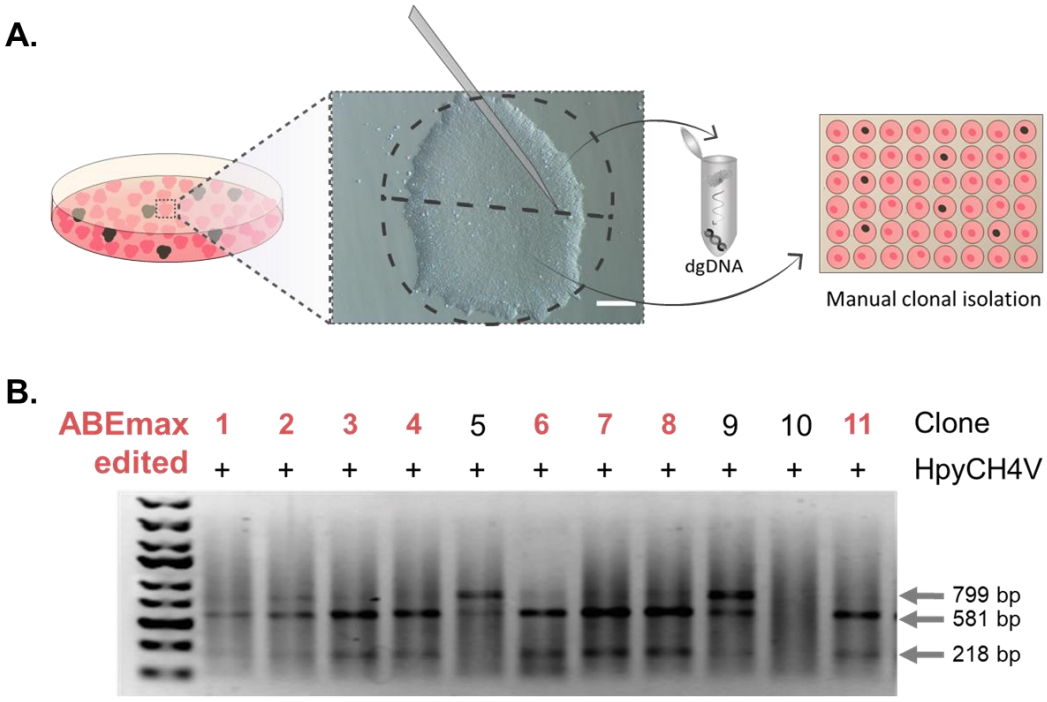


Figure 3.28 Base edited hiPSCs clonal isolation and screening

A) Manual clonal dissection and clonal screening to detect successfully targeted cells following puromycin treatment. A small fragment of each colony was collected for direct gDNA extraction and PCR genotyping, and the remaining fragment was placed in a well of a 24-well plate for further expansion. Scale bar= 100µm C) Representative clonal screening of cells transfected with pCMV_ABEmax_P2A_GFP. 8 out of the 11 shown clones were negative for A (799bp).

This PCR screening was repeated with cells modified using both base editor systems to validate their efficiency (data not shown). Sanger sequencing analyses were performed of the clones identified as positive from the PCR-restriction digestion screening to confirm whether the IPF-patient-derived hiPSCs had the desired

modification (**Figure 3.29**). Sequencing confirmed that the mutant cells generated with the ABEmax system contained only G's at the desired position (c.338A>G) and that the AncBE4max system successfully corrected the mutation, generating cells with only A's (c.338G>A). **Figure 3.30A** shows the karyotypes corresponding to the two generated cell lines SFTPC$^{WT/WT}$ (purple) and SFTPC$^{MUT/MUT}$ (pink).

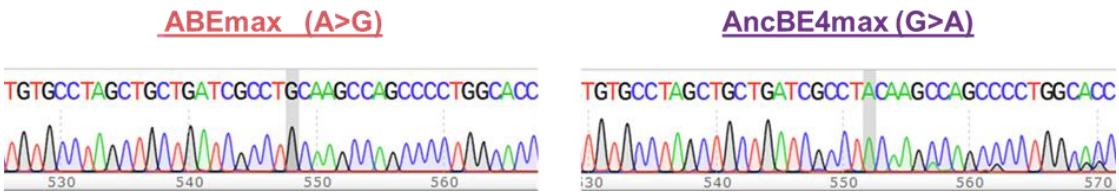


Figure 3.29 Validation of base editing process of IPF patient-derived hiPSCs

Representative Sanger sequencing confirming correct modification by base editors pCMV_ABEmax_P2A_GFP (A>G; pink) and pCMV_AncBE4max_P2A_GFP (G>A; purple).

The patient-derived iPSCs modified in this thesis retained pluripotency characteristics and were genomically stable after base editing. Both cell lines displayed a normal karyogram indicating genomic integrity upon editing. Furthermore, cells were analysed for the expression of pluripotency-associated transcription factors by flow cytometry and immunocytochemistry. **Figure 3.30B** shows the flow cytometry plots for SOX2 and OCT4 with a ~98% expression in both cell lines. Additionally, cells were seeded and expanded in Matrigel-coated plates to be tested for the presence of SOX2, OCT4 and NANOG in the monolayer of cells by immunocytochemistry. Representative images of the staining are shown in **Figure 3.30C** with positive expression of these pluripotency-associated markers in SFTPC$^{WT/WT}$ cells.

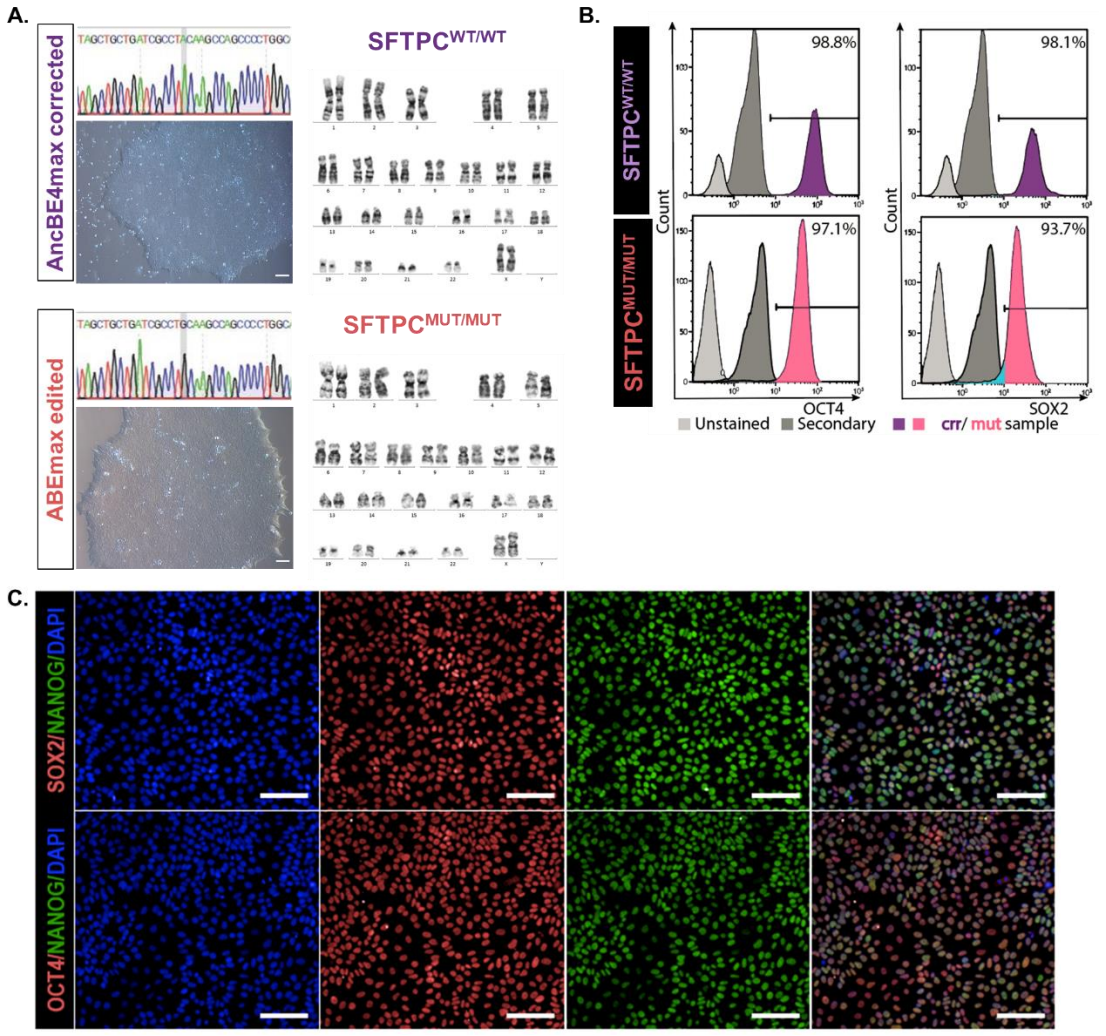


Figure 3.30 Characterisation of patient-specific corrected and homozygous SFTPC-mutant hiPSCs.

A) Quality control of gene-edited hiPSCs colonies showing chromatograms of the corrected and mutant cell lines, and their normal karyotypes. Scale bars= 100µm B) Flow cytometry of pluripotency markers OCT4 and SOX2 in edited cells. C) Representative immunocytochemistry images of SFTPC$^{WT/WT}$ hiPSCs showing homogeneous expression of pluripotency-associated markers SOX2 (red; top panel), OCT4 (red; bottom panel) and NANOG (green; top and bottom). Scale bars= 100µm

The successful generation of these hiPSC lines derived from individuals with IPF carrying a BRICHOS domain SFTPC heterozygous mutation, followed by the genetic modifications to obtain an isogenic control and a homozygous mutant cell line, represents a promising contribution to the study of IPF. These hiPSC lines can serve as a valuable platform to investigate the intricate genotype-phenotype interactions observed in IPF. Moreover, the direct differentiation of these hiPSCs into AT2 cells will allow the study of the entire spectrum of the SFTPC mutation in vitro. By utilising isogenic control lines, the genotype-dependent effects on IPF pathogenesis and

disease progression can be better discerned eliminating potential variations attributed to the genetic background in patient-specific hiPSCs. This more relevant human disease model provides a powerful tool for gaining insights into the molecular mechanisms underlying IPF.

Chapter 4 Development of an optimised protocol for the generation of hiPSC-derived AT2 cells

4.1 Introduction

The relevance of pulmonary hiPSCs-based disease models depends on how faithfully the in vitro differentiated cells can reproduce features of human pulmonary disease and the questions trying to be resolved with them. Therefore, in vitro differentiation protocols should successfully recapitulate the robust regulatory systems of lung morphogenesis to generate functionally mature cell types that provide an in vitro platform for studying human diseases. Based on the current knowledge of lung development, the protocols to derive lung epithelial cells from pluripotent stem cells have included a variety of conditions to mimic developmental milestones better and improve differentiation efficiencies [209]. Most protocols rely on specific medium compositions with different concentrations of growth factors, small molecules, and specific pathway activators and inhibitors. Furthermore, researchers have incorporated isolation steps of various cellular phenotypes via reporters or specific surface markers to their differentiation protocols. Likewise, the co-culture of the differentiating lung epithelium with stromal cells, their culture in 2D conditions on an air–liquid interface (ALI), or 3D Matrigel droplets have also been widely used [126]. However, the success of the available differentiation protocols has been hindered by the still incomplete understanding of the cellular and molecular mechanisms controlling lung development.

4.1.1 AT2 cells in vitro differentiation

The early attempts to generate lung epithelium from ESCs via directed differentiation were marked by inefficiency and unpredictability, employing incompletely defined media or depending on the inclusion of drug-resistance genes [136, 138, 140, 210]. Over time, the growth in knowledge regarding the development of the lung in vivo has paved the way for improved strategies in guiding the differentiation of hPSCs towards lung epithelial fate. These advanced approaches involve employing external signals to replicate the sequential cell signalling events that occur between endoderm and mesoderm, facilitating the definitive endoderm patterning and ultimately inducing the formation of NKX2.1+ lung progenitors with the potential to differentiate into mature functional lung epithelial cells. However, the conditions for directing the differentiation along a distal alveolar epithelial fate with homogeneity are yet to be refined and

optimised. Numerous groups have attempted to recapitulate the key milestones of embryonic lung development in vitro for the derivation of epithelial lung progenitors through the addition of sequential combinations of growth factors to ESC cultures [133–135]. They successfully differentiated ESCs into definitive endoderm via Nodal signalling activation, followed by the stage-specific inhibition of TGF-β and BMP signalling to pattern the endoderm into SOX2-expressing anterior foregut endoderm (AFE) [133]. This was followed by the addition of various combinations of Wnt activators, FGFs, BMPs, and retinoic acid (RA) to segregate the AFE into organ-specific domains, demonstrating the in vitro derivation of NKX2.1+ lung/thyroid competent progenitor cells [133–135].

Longmire et al. used an Nkx2.1:GFP reporter murine ESC line to isolate a relatively pure population of progenitor cells committed to the lung fate. This isolated population was then subjected to further differentiation using FGF2 and FGF10-supplemented DCI+K (DEXA, cAMP, IBMX and FGF7) media, leading to the maturation of both proximal and distal epithelial cells [134]. Building on the previous findings, Mou et al. successfully translated the directed differentiation conditions to generate NKX2.1 progenitors from hiPSCs, derived from patients with cystic fibrosis. Upon in vivo engraftment into immunodeficient mice, these progenitors demonstrated their ability to give rise to various lung cell types, including basal, club, goblet, ciliated and alveolar cells. Furthermore, researchers demonstrated that the reciprocal signalling between the lung epithelium and the mesenchyme during branching morphogenesis could be recapitulated in vitro. This was achieved by maintaining the NKX2.1+ lung progenitors under specific combinations of growth factors, including FGF, WNT, BMP4, RA, and SHH signalling [12, 143, 211]. The application of these conditions resulted in the emergence of alveolar type II cells, which exhibited characteristic features such as the presence of lamellar bodies and expression SFTPC and SFTPB.

One of the earliest studies on the applications of iPSCs for lung disease modelling was done by Wong and colleagues [212]. The study consisted of the derivation of lung epithelial cells from a cystic fibrosis (CF) hiPSC line with a mutation in the anion channel protein CFTR (cystic fibrosis transmembrane conductance regulator). They found that the 2D differentiated mutant hiPSCs, when exposed to C18, a small-molecule CF corrector compound, exhibited patches of cells expressing CFTR on their cell surface. This was an important proof of concept that CFTR-mutant hiPSCs could be used for modelling cystic fibrosis; however, the efficacy of their differentiation protocol was highly heterogeneous when using different cell lines [212]. Following on

this research, other groups also made use of hiPSC technology to generate CFTR-mutant cells and correct the mutation using CRISPR/Cas9, demonstrating the in vitro rescue of the channel function [213, 214].

Despite these promising advances, significant challenges remained in the field of in vitro lung development. One of the primary difficulties faced by researchers has been recreating the intricate lung developmental pathways and achieving precise proximal-distal patterning during the in vitro PSC differentiation process. This renders consistent and efficient lung-directed differentiation a challenging task. For instance, the induction of NKX2-1+ lung progenitors from both human ESCs and iPSCs using diverse protocols, has shown variable efficiencies, ranging from 40% to 80% [211, 212, 214]. Furthermore, although initial studies reported the induction of mature cell markers, like SFTPC, within their heterogeneous cultures, the expression of these markers appeared to be stochastic, leading to difficulties in obtaining a consistent population of alveolar cells [12, 134, 135]. This unpredictability and the heterogeneity of the cultures limited the ability to efficiently expand these presumable AT2 cells for downstream applications [215].

In an effort to improve the efficiency of differentiating PSCs into lung epithelial cells, researchers have adopted various strategies. These strategies include cell sorting using cell surface markers or engineered reporter cell lines to enrich the population of interest, which can be further cultured or used for experimental purposes. Another approach involves culturing the cells in 3D environments to induce the emergence of mature phenotypes or co-culturing them with other cell types. Advancements in cell sorting methods allowed the enrichment of more homogeneous populations of NKX2.1+ lung progenitor cells via surface markers (Carboxypeptidase M (CPM), CD47, CD26) or NKX2.1-GFP reporter hiPSC lines [157, 159, 216]. However, despite these breakthroughs, achieving the successful differentiation of these lung progenitors into mature functional epithelial cells, such as AT2 cells, was still a challenge given the extended periods of cell culture and the complex array of signalling pathways that needed to be accurately mimicked. Along with the introduction of sorting strategies, research groups employed prolonged in vitro cultures and xenografting into immunodeficient mice or co-culturing with other cell types, like foetal human lung fibroblasts, to increase the maturation of iPSCs-derived lung epithelial cells [157, 217]. Likewise, when NKX2.1+ lung progenitors were cultured in 3D conditions, using Matrigel or ECM components, their maturation improved, and the differentiated epithelial cells formed spheroids or organoids that were stable for up to several months

[157, 217–219]. The mature 3D structures contained cells corresponding to the distal and proximal airways, such as basal, ciliated, club and alveolar type I and II cells. The AT2 cells within these organoids exhibited a mature AT2 cell program, expressing surfactant proteins and phospholipids [157, 217–219].

Building on these previous findings, research groups have optimised and refined the directed differentiation protocols to yield higher numbers of AT2 cells for effective disease modelling. Researchers from the Mishima group used the NKX2.1 surrogate surface marker CPM, to purify the lung progenitors and differentiate them in 3D Matrigel cultures to obtain mature cells that could be propagated as alveolar spheroids with or without the presence of feeder cells [158, 159]. The generated AT2 cells were positive for SFTPA, ABCA3, SFTPC, and contained lamellar bodies. Similarly, the Kotton group used their previously generated BU3 hiPSC line engineered with the fluorescent reporters NKX2.1-GFP and SFTPC-tdTomato to facilitate the enrichment via cell sorting of differentiated lung progenitors to culture them in 3D in the presence of CHIR, KGF and the lung maturation factors. After two weeks in these conditions, the cells formed alveolospheres with presumable AT2 cells that could be enriched with the SFTPC reporter [145, 216, 218]. The generated expandable population of AT2 cells allowed them to model a genetic disease affecting the distal lung by using SFTPB-mutant (SFTPB121ins2) hiPSCs and demonstrated that CRISPR/Cas9-mediated correction rescued the phenotype [145]. Nonetheless , the NKX2.1 lung progenitors generated using this protocol were thoroughly scrutinised, and their transcriptome revealed that the cells maintain endodermal multipotency after lung specification, making them capable of differentiating into non-lung fates [19]. Furthermore, Sun et al. demonstrated that the sorting strategy using the SFTPC-tdTomato fluorescent reporter to enrich the AT2 population within the alveolospheres only identified the AT2 cells expressing the highest SFTPC levels, demonstrating the heterogeneity of the in vitro culture [220].

Additional studies employing a 3D culture strategy for the derivation of mature alveolar cells used hESCs to generate lung bud organoids containing goblet, AT1, and AT2 cells [219, 221]. Using CRISPR/Cas9, they introduced mutations in Hermansky-Pudlak syndrome (HPS) genes associated with HPS interstitial pneumonia in hESCs and differentiated them into branching lung organoids, which exhibited fibrotic changes [219]. Furthermore, genome-wide expression analysis revealed that IL11 was upregulated in the HPS mutant organoids. Making use of these data, Strikoudis et al. found that IL11 could induce fibrosis in healthy lung organoids, suggesting its potential

role as a therapeutic target [221]. Using the opposite approach, Korogi et al. generated hiPSCs from a patient carrying HPS2 mutations and successfully corrected it using CRISPR/Cas9 to be used for disease modelling [8]. They also compared the isolation efficiency of previously reported NKX2.1+ lung progenitor sorting methods, CPM and CD47/CD26 [157, 216]. They found that CPM-based sorting was able to isolate more NKX2-1+ cells than CD47high/CD26low-based methods. Furthermore, they incorporated the use of an antibody against the cell surface sodium-dependent phosphate transporter NaPi2b to isolate AT2 cells as it was more useful than the HTII-280 monoclonal antibody, usually used to isolate human adult AT2 cells, to maintain the hiPSC-derived AT2 cells [222].

Although there have been promising developments that enhance the in vitro differentiation of lung epithelial cells and AT2 cell generation, such as culture enrichment via cell sorting and 3D culture methods, the current state lacks an efficient and well-defined protocol for deriving and maintaining mature AT2 cells. To effectively model pulmonary diseases and conduct impactful research, it is crucial to focus on differentiation approaches that yield consistent cultures and render more mature, functional alveolar cells. Therefore, further efforts and advancements are necessary to address this challenge and pave the way for more accurate and reliable disease modelling using AT2 cells. By addressing this need, we can improve our understanding of lung diseases and potentially open new avenues for future therapeutic interventions.

4.1.2 Chapter aims and objectives

The previous chapter observed that the differentiation of hiPSCs derived from IPF patients toward an alveolar fate was not as efficient as desired. Notably, the findings indicated that a hybrid approach combining 2D and 3D cell culture strategies was necessary to effectively guide these cells towards a lung fate. Building upon these initial observations, it was determined that the SFPTC heterozygous mutant hiPSC line IPF1 was selected for gene editing to generate the SFTPC homozygous mutant and corrected cell lines to further refine the differentiation platform.

Therefore, the aims and objectives of this section are focused on optimising the differentiation platform for generating AT2 cells from gene-edited IPF mutant cell lines, and they are as follows:

- To improve the differentiation process at the definitive endoderm (DE) stage to enhance the yield of DE cells derived from IPF mutant cell lines. Flow cytometric analysis will be employed at this stage to confirm the high yields of DE cells obtained.
- To enhance the generation of lung progenitor cells by introducing an additional step involving anterior foregut ventralisation before proceeding with the induction and enrichment of lung progenitor cells using a surrogate surface marker for NKX2.1-positive cells.
- To adapt the differentiation platform into a 3D organoid system promoting the differentiation and maturation of the enriched NKX2.1 progenitor cells into an expandable population of mature AT2 cells.
- Following optimisation, the generated cells will be thoroughly characterised using bulk-mRNA sequencing. This analysis will investigate whether the optimised protocol successfully yields mature surfactant-producing AT2 cells, providing valuable insights into the quality of the generated cells.

4.2 Results

4.2.1 3D differentiation platform is suboptimal for IPF patient-derived hiPSCs.

Required enrichment of lung progenitors within the organoids.

The previous characterisation of the generated lung progenitors and AT2 cells using IPF-derived hiPSCs showed promising results; however, it was clear that further characterisation was needed, and that the differentiation protocol would likely need to be tailored to the new cell lines. Thus, before proceeding with the differentiation of the newly gene-edited hiPSCs, as their expansion got interrupted due to laboratory closure, the characterisation of the 3D organoids cultured in maturation media was reassumed. The poor quality of the organoid's immunocytochemistry images obtained using the Operetta microscope failed to delineate the expression of important cell markers before. Therefore, a new protocol tailored to 3D cultures and a different confocal microscope were used to characterise previously generated SFTPC$^{WT/MUT}$ organoids that were successfully thawed and had been kept in culture for 3 passages (**Figure 4.1A**) [182]. The cultured organoids were processed and stained to corroborate the presence of NKX2.1 and SFTPC cells, showing positivity for both markers. However, the positive cells were not homogenously distributed throughout the 3D structures (**Figure 4.1B**).

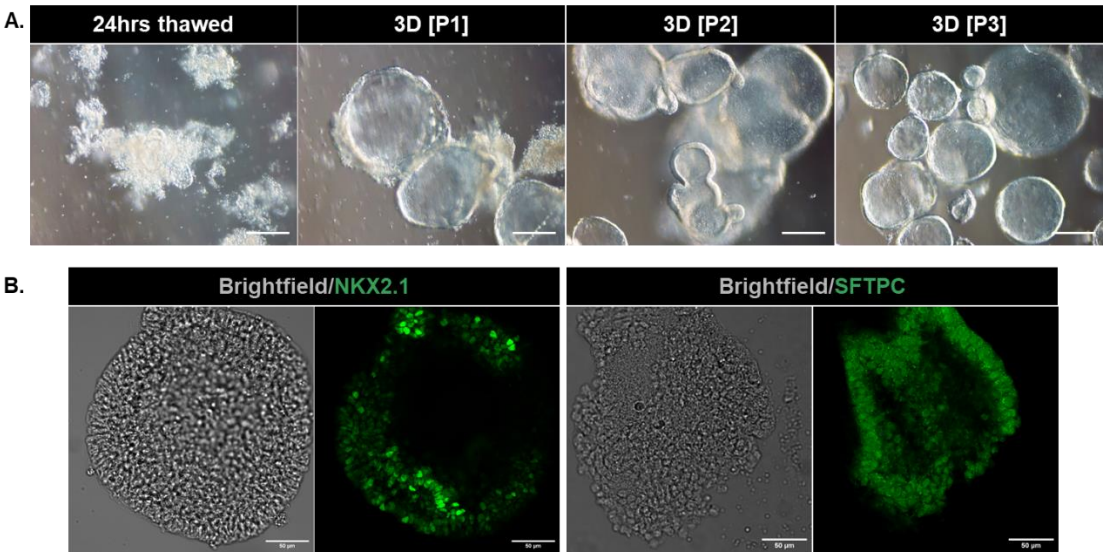


Figure 4.1 Thawed SFTPC$^{WT/MUT}$ organoids are positive for NKX2.1 and SFTPC.

A) Brightfield images of previously generated and frozen SFTPC$^{WT/MUT}$ hiPSC-derived organoids 24hrs after thawed and embedded in Matrigel droplets. Images also show the organoids as they were grown in culture and passaged 3 times ([P1], [P2] and [P3]) before proceeding with their characterisation. Scale bars, 200µm. B) Representative brightfield and corresponding

confocal fluorescence images of alveolar organoids on passage 3 showing positive staining of endogenous NKX2.1 (green) and SFTPC (green). Scale bars, 50 µm.

Ample evidence in the literature states how difficult is to maintain the AT2 cell phenotype in vitro. Other groups have reported the use of antibodies against the sodium-dependent phosphate transporter 2b (NaPi2b) and Carboxypeptidase M (CPM), cell membrane proteins that have been identified as surrogate surface markers for $SFTPC^+$ and $NKX2.1^+$ cells, respectively, to purify their cultures and obtain better yields of AT2 cells [8, 157]. To assess if this was a methodology that could be employed with our cultures, the organoids were also stained using commercially available antibodies against these surrogate surface markers as an indirect way of corroborating the presence of NKX2.1 and SFTPC cells. The results obtained showed organoids that co-expressed CPM on the surface of cells that were positively stained for nuclear NKX2.1 (**Figure 4.2A**). However, when examining the organoids for NaPi2b expression, the results were disappointing but not unexpected. Only a small proportion of cells within the organoids were positive for the SFTPC surrogate marker (**Figure 4.2B**).

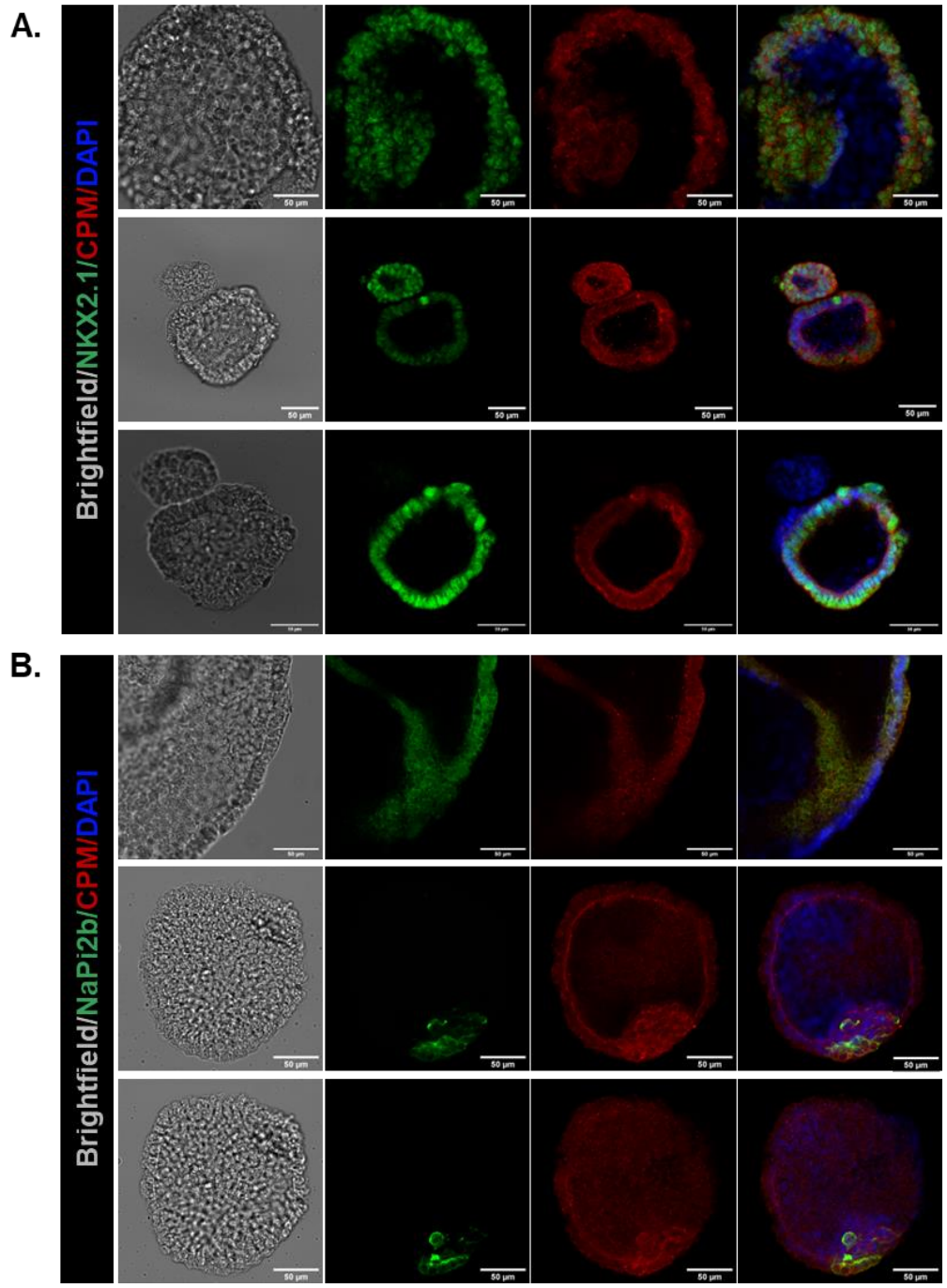


Figure 4.2 SFTPC$^{WT/MUT}$ organoids co-express surrogate surface markers for NKX2.1 and SFTPC.

Confocal fluorescence microscopy of SFTPC$^{WT/MUT}$ hiPSC-derived organoids in passage 3 showing A) three positively co-stained organoids with antibodies against nuclear NKX2.1 (green) and its surrogate cell surface maker, CPM (red). And B) shows three other organoids co-stained for NaPi2b (green), a surrogate cell surface marker for SFTPC, and CPM (red). Note the limited number of cells stained for NaPi2b in the two panels from the bottom. Grayscale brightfield images are shown to the left of each corresponding composite. Nuclei were stained with DAPI (blue). Scale bars, 50 µm.

Taken together, these results suggested that an enrichment strategy using these surface markers needed to be adopted in the differentiation protocol to isolate the progenitors and mature them to enhance the yields of $SFTPC^+$ cells obtained.

Suboptimal purification of AT2 cells growing in 3D conditions.

Thus far, the approach used to generate alveolar organoids was transferring clumps of monolayered cultures containing putative hiPSCs-derived $NKX2.1^+$ lung progenitors directly into 3D Matrigel droplets. By doing this, any contaminant lineages present in the monolayer with the lung progenitors were carried into the 3D culture, further potentiating their growth and potentially overtaking the $NKX2.1^+$ cells. To investigate how many of these lung progenitors were still in culture after being transferred to 3D, and whether they matured into surfactant-producing cells or not, an enrichment for CPM and NaPi2b positive cells was assayed using the $SFTPC^{WT/MUT}$ passage 6 organoids already in culture (**Figure 4.3**). For this, on the day of the sorting experiment, the organoids were enzymatically dissociated using first Dispase to disintegrate the embedding Matrigel and then trypsin to break up the organoids into a single-cell suspension. The obtained cells were then stained, as described in Chapter 2, using anti-CPM (Origene) and anti-NaPi2b (Cell Signaling) antibodies. However, the procedure was proven technically complicated, and a considerable number of cells were lost during the washing steps. Notwithstanding, the remaining cells were still run through the sorter. Strikingly, the number of cells recovered from the processed organoids was extremely low, with a recovery of 72 CPM positive and 16,661 negative cells and for the other surface marker, NaPi2b, only 14 positive cells were recovered and 17,107 negative cells.

As this was the first time the protocol was applied in the lab, it was believed that with some refinement to the multiple washing steps during the dissociation and staining protocols for FACS, the massive loss of cells could be prevented. After several attempts, the incorporation of 1% FBS/PBS as the washing reagent, instead of PBS on its own, and the use of 15ml falcon tubes in a benchtop centrifuge to spin down the cells, instead of 1.5ml snap cap tubes in a MiniSpin Plus (Eppendorf) centrifuge, were identified as two crucial changes to the protocol. Implementing these simple modifications, noticeably improved the number of cells obtained at the end of the staining protocol and ready to be sorted. However, when the sorting experiment for CPM and NaPi2b was repeated using the same organoids in older passages,7 and 9, the implemented changes did not reflect a higher recovery of positive cells, as shown in **Figure 4.3.**

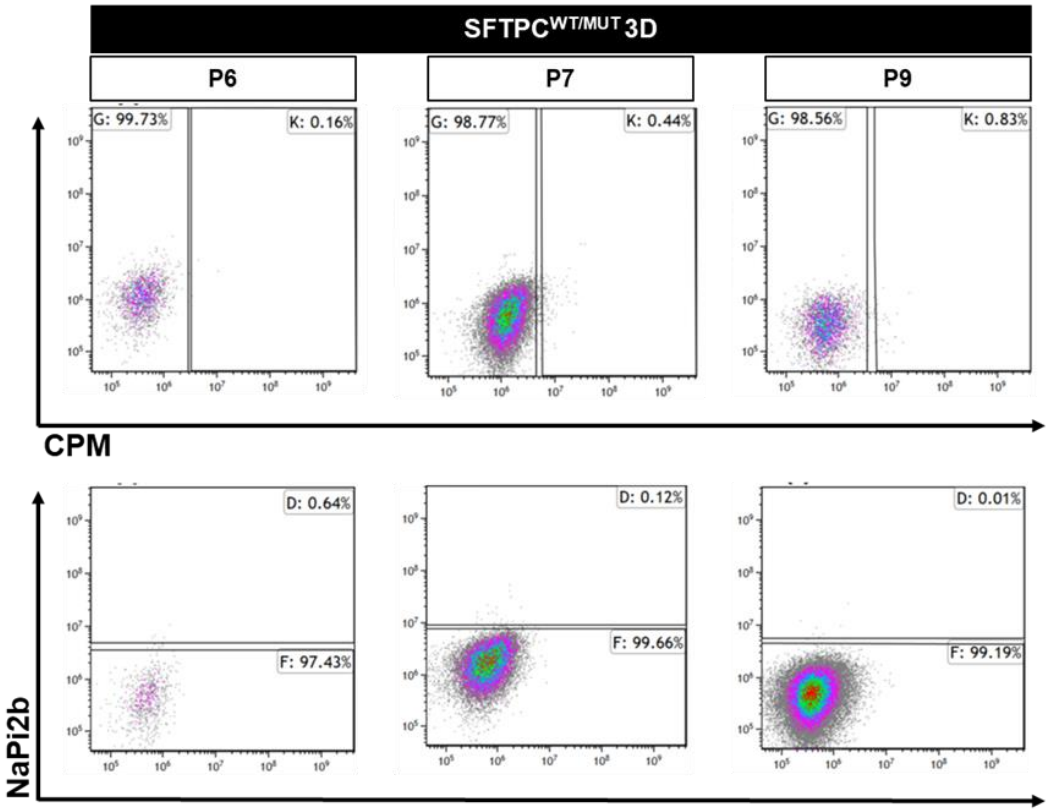


Figure 4.3 FACS failed to enrich cells using CPM and NaPi2b antibodies.

Flow cytometry of CPM$^+$ (top row, gate K) and NaPi2b$^+$ (bottom row, gate D) expression of SFTPC$^{WT/MUT}$ organoids in passages 6, 7 and 9 (P6, P7, and P9). The plots show an increase in the size of the sorted populations as the protocol was optimised towards P9, but the number of positive cells did not reach 1% for either of the markers on the 3 passages that were sorted. Gates were adjusted based on the unstained and secondary-only antibody controls.

Even if the thawing of the organoids appeared to have been successful based on their morphology and expression of key lung markers, it was important to corroborate that the freeze/thawing process did not affect the phenotype of the cells within the organoids processed for FACS. Therefore, this experiment needed to be repeated with freshly generated lung progenitors and alveolar organoids.

Yields of AT2 cells indicate that the differentiation protocol needed optimisation.

In an attempt to circumvent the potential issue of other lineages taking over the cultures during the 3D maturation conditions, once ready, the newly modified IPF-patient-derived cell lines were seeded and submitted to the differentiation protocol to assess their ability to generate distal lung progenitors by enriching for CPM and NaPi2b positive cells.

After 8 days of lung progenitors' induction in 2D culture conditions, the cells were either processed for cell sorting of CPM and NaPi2b positive cells or transferred to 3D

Matrigel droplets to continue with the maturation protocol. **Figure 4.4A** shows the FACS plots of the cells derived from SFTPC$^{WT/WT}$, SFTPC$^{WT/MUT}$ and SFTPC$^{MUT/MUT}$ lines that were sorted on lung day 8. The percentages of CPM^{+} cells were higher than the NaPi2b^{+} ones; however, the number of cells recovered from the sorter for both markers were between 1,000 to 5,000 cells. Despite the extremely low number of cells obtained, cells were reseeded in 3D Matrigel droplets to stimulate the emergence of organoids, but the low seeding densities did not allow the cultures to thrive.

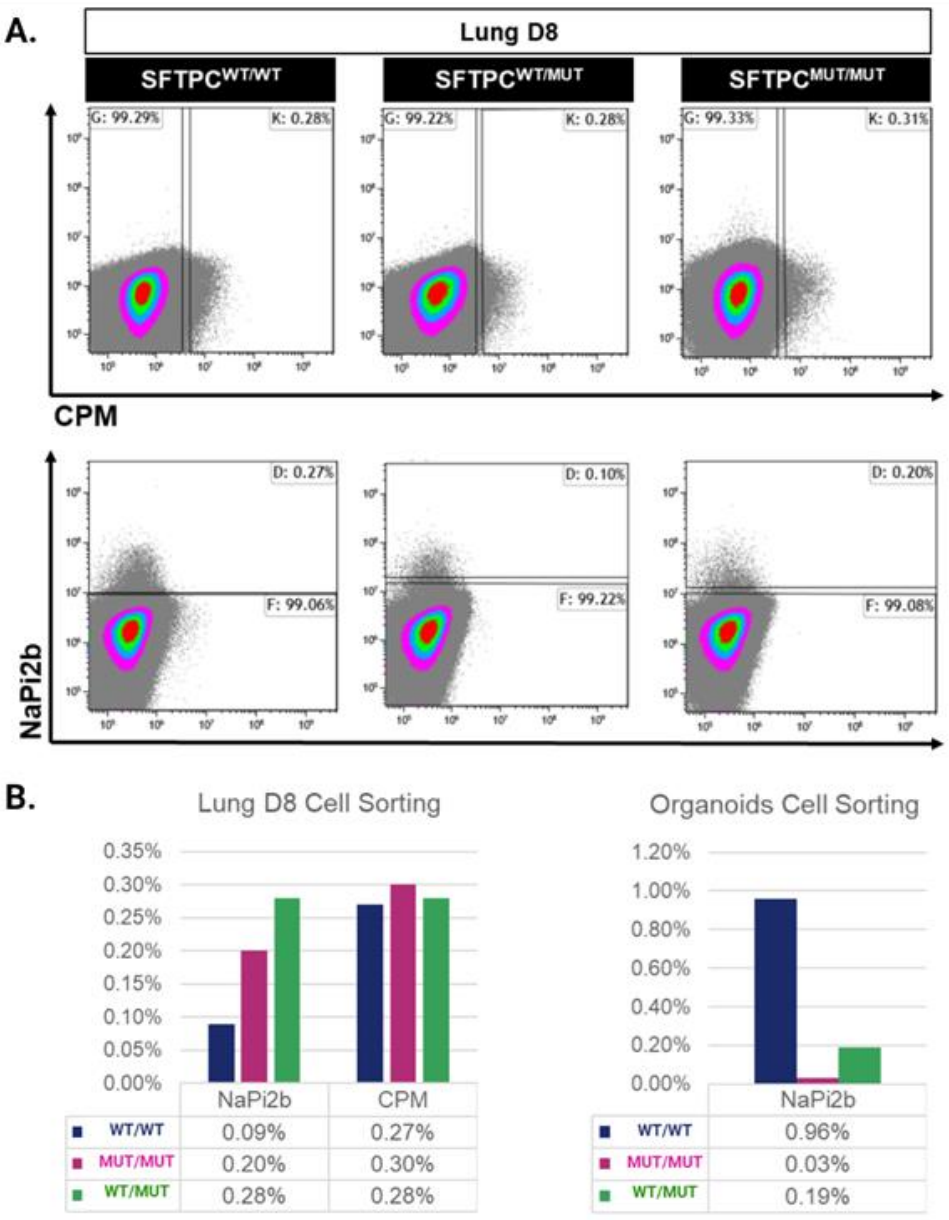


Figure 4.4 IPF-patient-derived cell lines differentiate into lung progenitors at low yields.
A) FACS plots of the gene-edited and unmodified cell lines at lung differentiation day 8, sorted for CPM (top row) and NaPi2b (bottom row). B) Graphs present the percentages of NaPi2b+

and CPM+ cells obtained in lung day 8 sorting (A) and compares them to the ones obtained from sorted cells derived from organoids at passage 3, n=1.

Interestingly, a few wells of organoids from the SFTPC$^{MUT/MUT}$ line survived, which were expanded for two more passages to have enough cells before processing them for further analysis. RNA was extracted from the organoids lysed on passage 4 to analyse the expression of lung progenitors' markers by qPCR. A comparison of the expression of the markers SOX9, NKX2.1, SFTPC and SFTPB was done between the CPM^{+}-derived organoids, lung day 12 monolayer and hiPSCs. It can be seen from the qPCR data in **Figure 4.5** that, even if not at the desired levels, the cells derived from the organoids had significantly higher expression levels of the markers of interest.

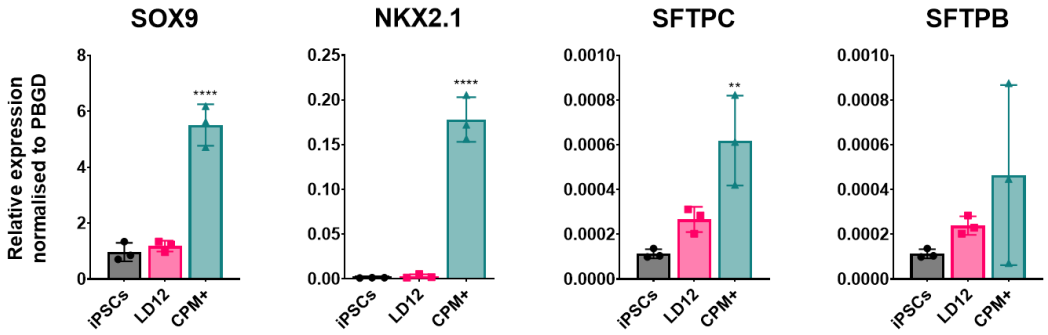


Figure 4.5 3D conditions enhance the expression of lung progenitor markers.

qPCR for lung progenitor markers SOX9, NKX2.1, SFTPC and SFTPB in hiPSCs, lung differentiation day 12 monolayer (LD12) and CPM^{+} SFTPC$^{MUT/MUT}$ cells-derived lung organoids at passage 4 (CPM^{+}). PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

In addition, the lung day-8 cells that were transferred directly to 3D culture, without undergoing any enrichment, were also processed for cell sorting after passaging them as clumps 3 times to simulate their maturation. For this experiment, cells were only stained for NaPi2b to assess if the culture conditions enhanced the yields of SFTPC^{+} cells. Unfortunately, according to the expression of the cell membrane co-transporter, the 3D maturation did not increase the percentages of SFTPC^{+} cells, and there were not enough cells to seed back into Matrigel droplets. A comparison between the cells sorted on lung day 8 and from passage 3 organoids can be observed in **Figure 4.4B.**

Following the disappointing finding of the low percentages of lung progenitors and AT2 cells obtained within our cultures, an assessment of the lung progenitors' stage of the differentiation was done. The recently reported differentiation culture platform by Jacob

A. et al. included a new media composition for generating lung progenitors [162]. Therefore, it was decided to compare the differentiation conditions used thus far (DIFF1) and the conditions described by Jacob et al. (DIFF2), as well as a combination of both, using their media composition with the growth factor's concentration of our condition (DIFF2+[1]). SFTPC$^{WT/WT}$ cells were seeded and differentiated up to lung day 8 using the three mentioned conditions, to be sorted using the CPM marker. The differences in the impact of the supplemented RPMI basal media used for the lung differentiation stage in DIFF1 or the supplemented IMDM/F12 in DIFF2 were not reflected in the production of CMP+ cells by lung day 8, as shown on the FACS plots in **Figure 4.6.**

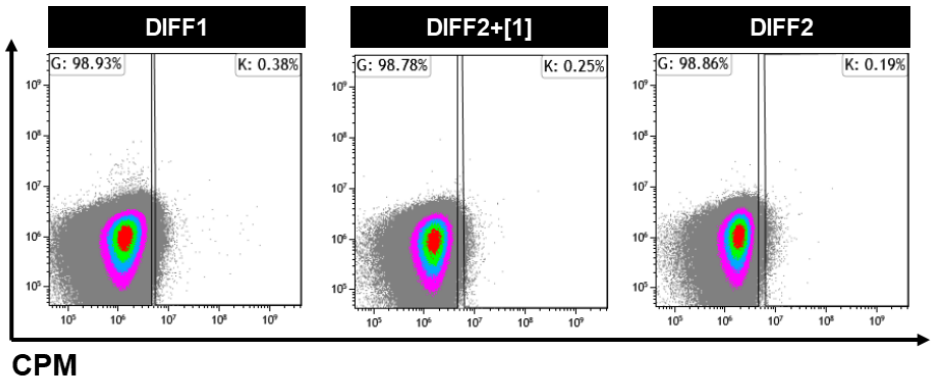


Figure 4.6 Distinct culture conditions do not affect CPM$^+$ yields.

Lung differentiation day 8 cells sorted for CPM. Cells were cultured under three different media compositions 1) DIFF1: basal DIFF1medium supplemented with 1µM CHIR99021, 250nm RA and 5ng/ml BMP4, 2) DIFF2+[1]: basal DIFF2 medium supplemented as 1), and 2) DIFF2: basal DIFF2 medium supplemented with 3µM CHIR99021, 100nm RA and 10ng/ml BMP4.

Thus far, the results obtained with the 3D-organoid platform showed the successful enrichment of scarce CPM$^+$ and NaPi2b$^+$ cells. However, the suboptimal differentiation efficiency of lung progenitors using the designed culture conditions made it clear that the platform needed optimisation. It was not feasible to continue with this exact methodology given the low CPM$^+$ cell numbers obtained from the sorter and the amount of time required to expand these scarce progenitors in 3D culture and mature them into precious AT2 cells. Furthermore, the fact that even after using three different media compositions, including one from a peer-reviewed protocol, failed to generate robust numbers of CPM$^+$ lung progenitors pointed towards the need for more thorough optimisation. Therefore, taking these results into account, the culture platform was re-evaluated from day 1 of the differentiation, starting from the earlier endoderm stage.

4.2.2 Optimisation of Definitive Endoderm differentiation step

Modulation of various signalling pathways to generate better yields of DE.

The specification of definitive endoderm (DE) from hiPSCs involves the temporal modulation of the activity of various signalling pathways, including Wnt, BMP, FGF, and nodal. These signalling pathways have been reported to play crucial roles in the development of the definitive endoderm germ layer during embryonic development [146, 149, 223, 224]. To identify the optimal combination of signalling activators for the differentiation of DE, different combinations of these molecules were screened. The goal was to find the combination that best supported the differentiation of DE using the IPF-patient-derived hiPSCs. For optimisation purposes, the differentiation was performed using the $SFTPC^{WT/WT}$ hiPSC lines BE31 and BE32.

Both cell lines were first seeded at a density of $1.5x10^4$ cells/cm^2 and maintained in HBE8 media for 24 hrs in the presence of TGFβ and FGF2. The following day, on differentiation day 0, cells were changed to DIFF1 media supplemented with the different experimental combinations: "WA" (Wnt [50ng/ml] + Activin [100ng/ml]), "WAB" (Wnt [50ng/ml] + Activin [100ng/ml] + BMP4 [10ng/ml]), or "WABF" (Wnt [50ng/ml] + Activin [100ng/ml] + BMP4 [10ng/ml] + FGF2 [16ng/ml]). After 3 days of culturing the hiPSCs with the specific growth factors and signalling molecules to induce DE cells, the Wnt signalling activator was removed from the media. The cells were then cultured for an additional 2 days in this new media to induce the formation of anterior foregut endoderm (AFE) cells (**Figure 4.7A**). Cells were harvested on days 3, 4 and 5 of the differentiation for RNA extraction to assess their gene expression through qPCR.

A comparison of the expression of endodermal markers was done between BE31 and BE32 cell lines cultured with "WA", "WAB", or "WABF" on the three different time points selected (**Figure 4.7B**). Both cell lines expressed the highest levels of SOX17, FOXA2, and GATA4 on the final day of the DE differentiation. However, the induction towards AFE was accompanied by a decrease in the levels of FOXA2 and GATA4 using the three culture conditions. Interestingly, even if the pluripotency marker NANOG was significantly downregulated as the differentiation progressed, BE32 showed higher levels than those of BE31, especially when using "WA" and "WAB". This indicates that some cells were not following the differentiation cues. BE32 also expressed higher levels of the posterior foregut endoderm marker CDX2, which would be expected to be decreasing to give rise to anterior foregut endoderm cells; therefore, BE31 was elected to move forward with the rest of the experiments. As WABF proved to induce

significantly higher levels of SOX17, FOXA2, and GATA4 at the AFE stage when using BE31, this condition was picked to continue the optimisation.

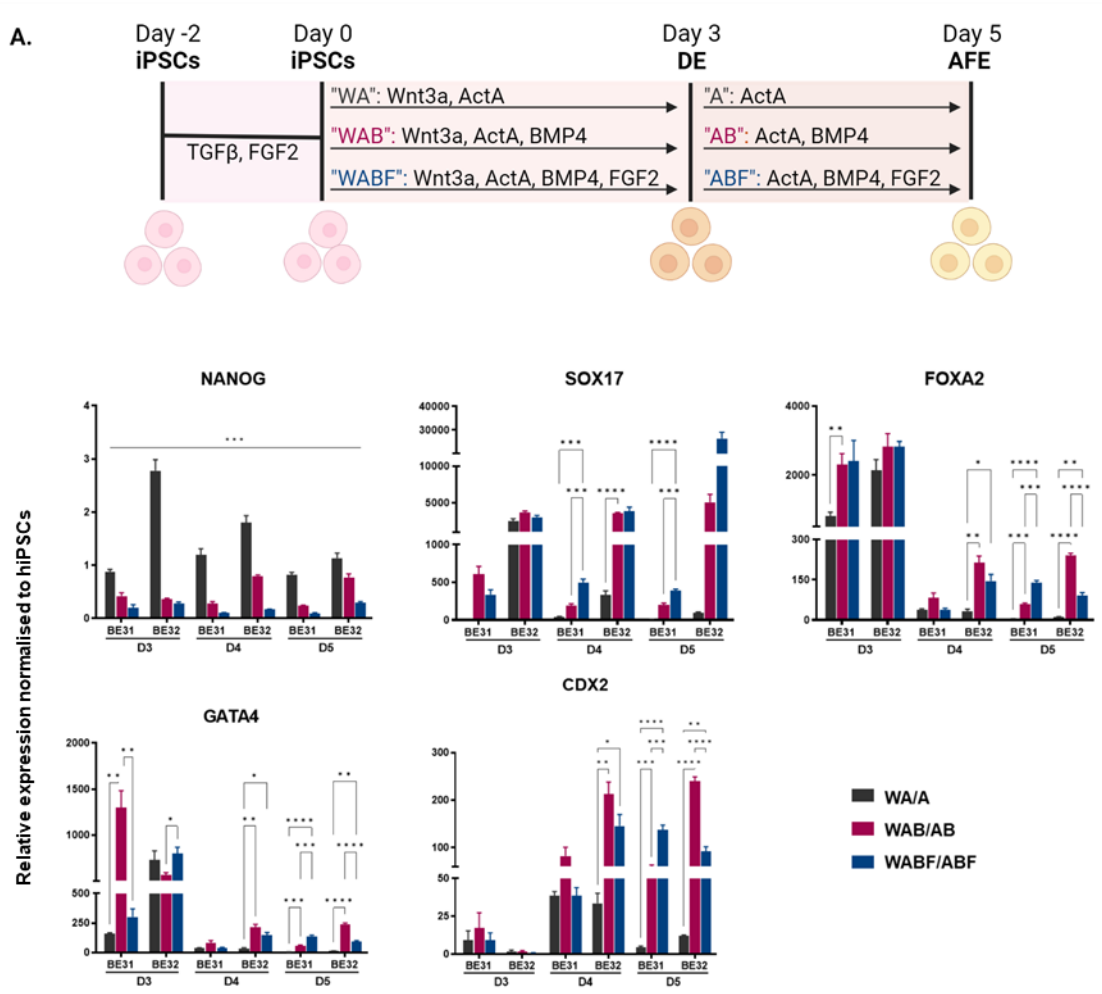


Figure 4.7 Definitive endoderm optimisation using Wnt, Act, BMP and FGF.

A) Schematic representation of the directed differentiation of hiPSCs into definitive endodermal cells, comparing three different culture conditions: “WA”, “WAB”, “WABF”. B) qPCR on days 3, 4 and 5 of the endoderm differentiation protocol using two SFTPC$^{WT/WT}$ hiPSC lines, BE31 and BE32. Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to undifferentiated hiPSCs. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1. All graphs show p-values <0.05, except SOX17, where every comparison was significant, so the graph only shows the ones with a p-value <0.001.

To evaluate if a different media composition could positively impact the differentiation, BE31 cells were cultured using the original DIFF1 media or the DIFF2 media described by the Kotton group, supplemented with "WABF" for the DE and AFE specification [162]. Gene expression analysis by qPCR indicated that both media supported the downregulation of NANOG, while endodermal markers SOX17, FOXA2 and GATA4 increased (**Figure 4.8**). However, no significant differences were identified between

media compositions, apart from for SOX17 on day 4 for cells growing in DIFF1 media. Interestingly, the hindgut marker CDX2 still got upregulated towards the end of AFE induction, though the fold changes relative to hiPSCs were lower when using DIFF2 media. Thus, the differentiation still needed to be adapted to augment the generation of cells with anterior foregut characteristics that could be primed to acquire the lung fate.

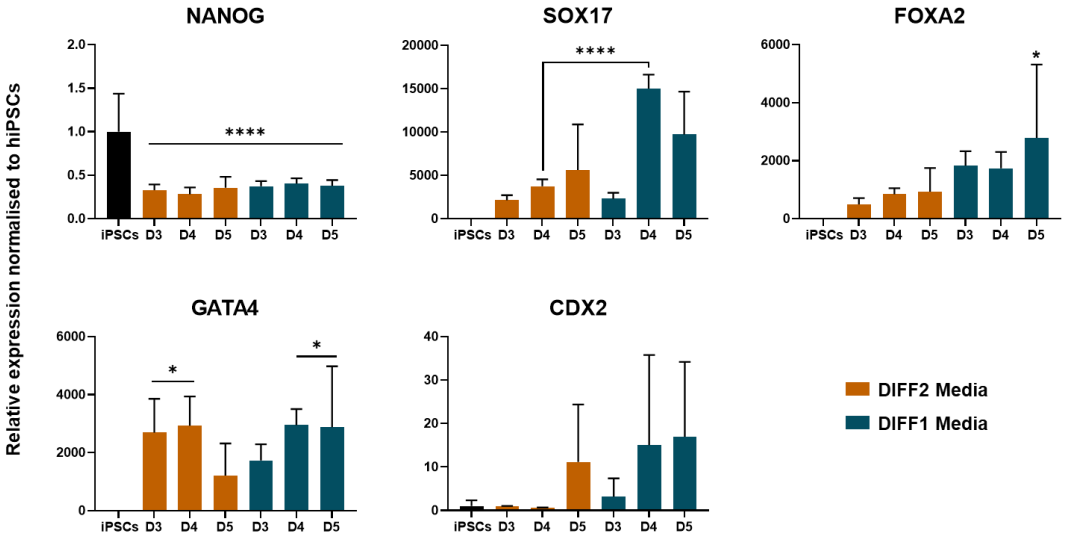


Figure 4.8 Definitive endoderm optimisation using two basal specification media.

qPCR for pluripotency (Nanog), endodermal (SOX17, FOXA2 and GATA4) and trophectoderm/foregut (CDX2) markers expression in endodermal differentiating cells on days 3, 4 and 5 cultured in DIFF1 or DIFF2 media. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using one-way ANOVA. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

In a study conducted by Green et al., they investigated the impact of inhibiting the BMP and TGF signalling pathways on the fate of foregut cells [133]. Their findings revealed that the addition of dorsomorphin (DS), a BMP inhibitor, and SB-43152 (SB), a TGFβ inhibitor, led to the generation of foregut cells with a propensity to differentiate into lung progenitors. Building upon these findings, the incorporation of this dual SMAD inhibition step into the protocol by utilising these two molecules for a duration of three days after the induction of definitive endoderm was aimed. This additional step was intended to promote the development of foregut cells that were more likely to differentiate into lung progenitors.

For this, the SFTPC$^{WT/WT}$ cell line BE31 was first differentiated using the basal media DIFF1 supplemented with "WABF" for 5 days, as described in **Figure 4.7A**.

Subsequently, the cells were changed to DIFF2 media supplemented with DS (2μM) and SB (10uM) for an additional 3 days. Cells were harvested at the end of the endoderm induction on day 5 and on every day of the ventralisation step, to perform qPCR analysis for NANOG, main endodermal markers and CDX2. It was reassuring to observe the significant downregulation of the pluripotency marker NANOG as the differentiation progressed, as well as CDX2. However, the expression of SOX17, FOXA2 and GATA4 also decreased with the inhibition of BMP/TGFβ signalling pathways (**Figure 4.9**).

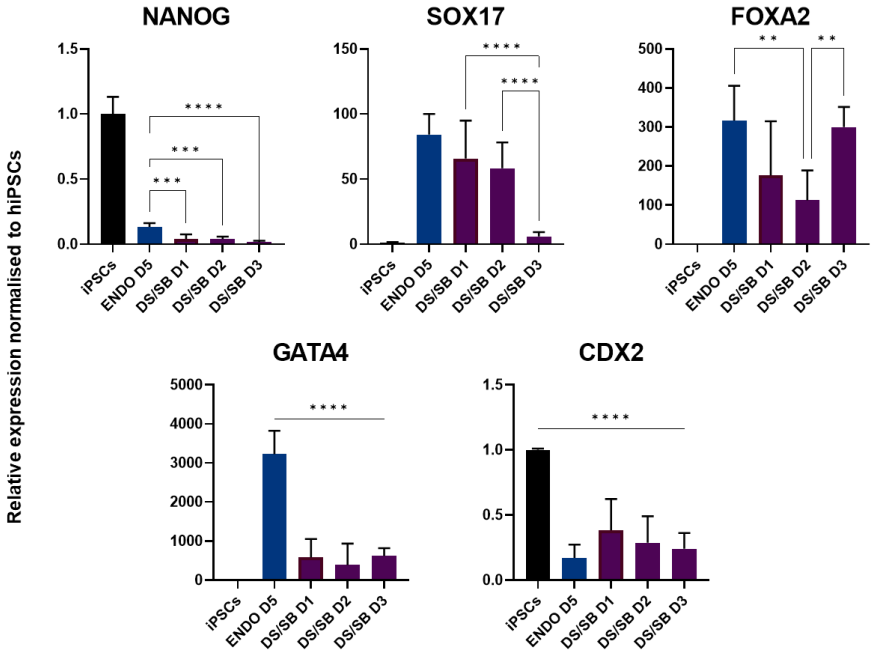


Figure 4.9 Dual SMAD inhibition does not improve the expression of endodermal markers.

qPCR showing the expression of endoderm-relevant markers on endoderm day 5 cells and on days 1, 2 and 3 with Dorsomorphin/ SB434542 (DIFF2 medium + DS/SB) treatment. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using one-way ANOVA. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

The objective of the AFE differentiation is to generate cells that have the potential to become NKX2+ lung progenitor cells. To assess whether the directed differentiation of foregut cells would effectively promote the generation of lung progenitor cells, it was decided to carry out the entire differentiation protocol and then perform flow cytometry to evaluate the impact on the production of CPM+ lung progenitor cells.

In this experiment, cells were initially differentiated using the basal media DIFF1 supplemented with "WABF" for 5 days, following the same approach as in the previous experiment. Afterwards, half of the cells underwent a 3-day ventralisation step using dual SMAD-inhibition, to be followed by the lung differentiation protocol, while the other half proceeded directly to the lung differentiation protocol. For the lung differentiation, cells were cultured using the culture conditions described in the previous section, which included DIFF1, DIFF2 (as described by Jacob et al., 2019) and a combination of both that used their media composition with the growth factors' concentrations of DIFF1 (DIFF2+[1]). At the end of the lung progenitors' protocol (lung day 9), cells were processed and stained against the cell surface marker CPM for cell sorting.

Figure 4.10 shows the FACS plots of the cells differentiated in the three different media compositions during the lung specification stage with or without the ventralisation step. The cells cultured in the presence of DS/SB with DIFF1 and DIFF2+[1] media had higher percentages of CPM^+ cells by lung day 9 (0.39% and 0.30%, respectively) than those of the cells cultured without DS/SB (0.04% for both). In the other hand, the cells cultured with DIFF2 media showed a higher percentage of CPM^+ cells when cultured without DS/SB (0.20% vs 0.14%). Regardless of the comparison, these results were not satisfactory. The percentages of positive cells were extremely low for all the conditions tested, and the number of cells retrieved from the sorter was insufficient for downstream analysis. Therefore, the differentiation approach used up to this point was far from optimal in obtaining the phenotype of interest with the IPF-patient-derived hiPSCs, indicating the need to make further adjustments and improvements to the protocol.

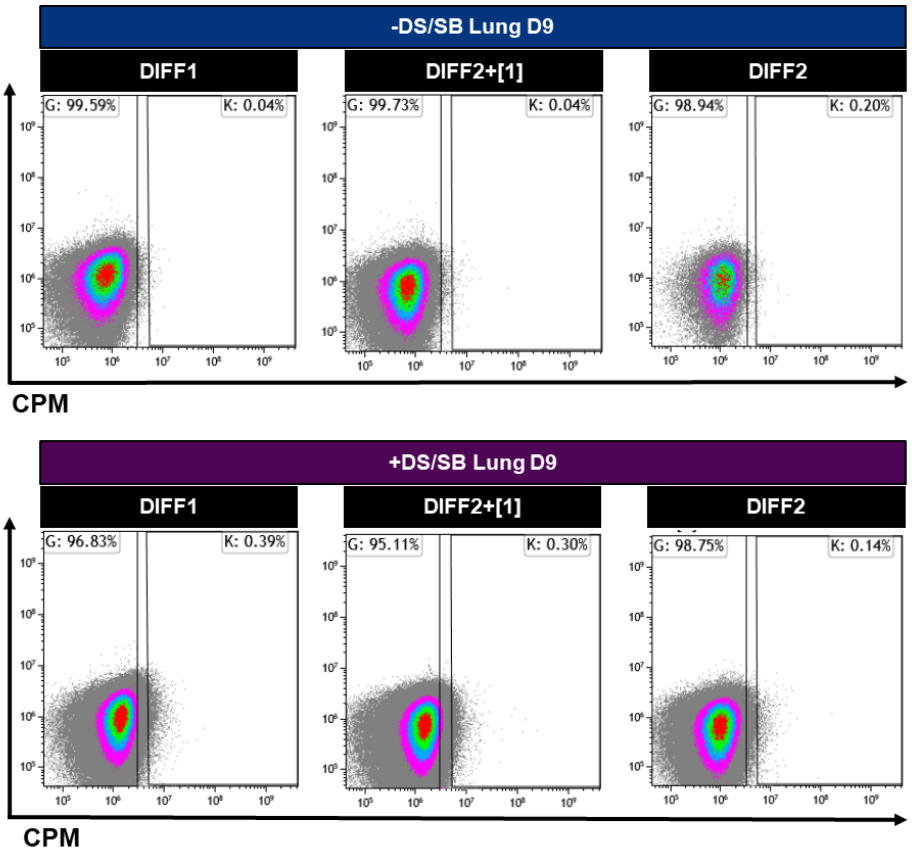


Figure 4.10 "WABF" endoderm and dual SMAD inhibition had no effect on the lung progenitors' stage of the protocol.

FACS plots comparing the cell sorting of lung day 9 CPM$^+$ progenitors cultured in three different media compositions. Cells were first differentiated using the "WABF" DE conditions for 5 days and continued with the lung differentiation (top panel) or with 3-day DS/SB treatment before the lung stage (bottom panel).

The efficiency of the first stage of the protocol, the definitive endoderm induction, was paramount to generating lung progenitor cells. Hence, it was decided to go back and finesse the first 5 days of the protocol. For this, the three conditions previously applied, "WA", "WAB", and "WABF", were used again, only with a modification on the FGF2 concentration (Wnt [50ng/ml] + Activin [100ng/ml] + BMP4 [10ng/ml] + FGF2 [100ng/ml]). This time the SFTPC$^{WT/WT}$ hiPSC line BE31 was seeded at 1x10^4 cells/cm^2 and differentiated for 5 days using supplemented DIFF1 basal media, as indicated in **Figure 4.7A**.

The morphological changes observed throughout the endoderm differentiation can be observed in **Figure 4.11A**. After one day of definitive endoderm induction (ENDO D1), the cells started growing into a monolayer of cobblestoned-shaped cells. By ENDO D3 the wells were fully confluent and the monolayers from the cells growing in the three conditions assayed presented areas of densely packed hiPSC-like colonies,

presumably undifferentiated cells. Towards the end of the 5-day protocol, the AFE induction caused abundant cell death observed as floating cells in the spent media and numerous gaps in the underlying monolayer. Expression of the pluripotency marker NANOG correlated with the presence of putative pluripotent colonies at day 3 of the differentiation in all the conditions, but especially in "WA" (**Figure 4.11B**). This condition also appeared to induce little to no expression of the markers of interest. As the differentiation progressed, the endodermal markers SOX17, GATA4 and CXCR4 were gradually upregulated, peaking on ENDO D5 for "WABF". Even if "WAB" generated good expression levels of these endodermal markers, the pattern was not ideal. Furthermore, "WAB" induced the highest levels of expression of the hindgut marker CDX2.

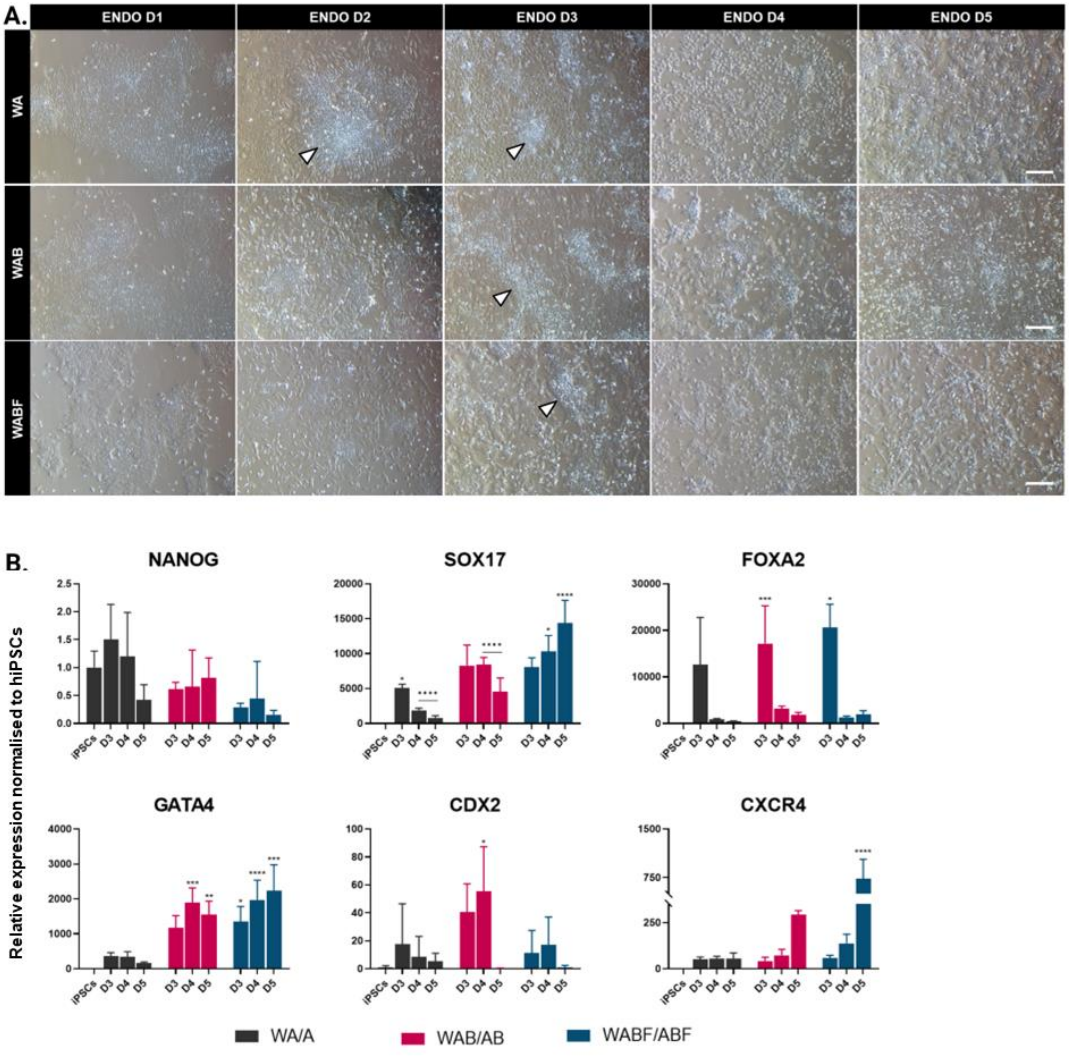


Figure 4.11 Lower seeding density and use of Wnt, Act, BMP and FGF for definitive endoderm optimisation.

A) Brightfield images of SFTPC$^{WT/WT}$ hiPSC seeded at 10,000 cells/cm2 and differentiated up to endoderm day 5 using "WA", "WAB", and "WABF" conditions. Note the presence of tight clusters of undifferentiated cells (arrows). *Scale bars, 200µm. B)* qPCR of cultured cells on days 3, 4 and 5 of the protocol for endoderm-relevant markers. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using one-way ANOVA. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

To further characterize the cells during the endoderm differentiation, flow cytometry for CXCR4 and c-KIT analysis was performed on the cells cultured in basal media DIFF1 supplemented with "WA," "WAB," or "WABF" on days 3, 4, and 5 of the protocol. According to the literature, a strong correlation exists between the robust generation of lung progenitors and a good expression of these endodermal markers [216, 225]. Jacob et al., suggest that although CXCR4 and c-KIT co-expression is not specific to definitive endoderm, their expression levels should be above ~90% to proceed with the differentiation protocol towards the generation of lung progenitors [162]. This criterion serves as an indicator of successful progression in the differentiation process.

Figure 4.12 presents the flow cytometry plots for the expression of CXCR4 and c-KIT markers, which serve as indicators of endodermal differentiation efficiency. Among the conditions tested, the least effective one for promoting endodermal differentiation was "WA". It showed the lowest expression for both markers on day 5 with 7% CXCR4 and 19% c-KIT. In contrast, the "WABF" condition demonstrated the best results across all three time points assessed. It exhibited a gradual increase in CXCR4 expression, reaching 41% on day 3, 52% on day 4, and 60% on day 5. Similarly, c-KIT expression showed levels of 53%, 46%, and 37% on the corresponding days. The "WAB" condition also showed a gradual increase in CXCR4 expression, peaking at 48% on day 5, accompanied by a c-KIT expression of 30%. However, despite these relatively high-efficiency percentages, according to the literature they were not sufficient to indicate successful differentiation progression for the IPF-patient-derived hiPSCs.

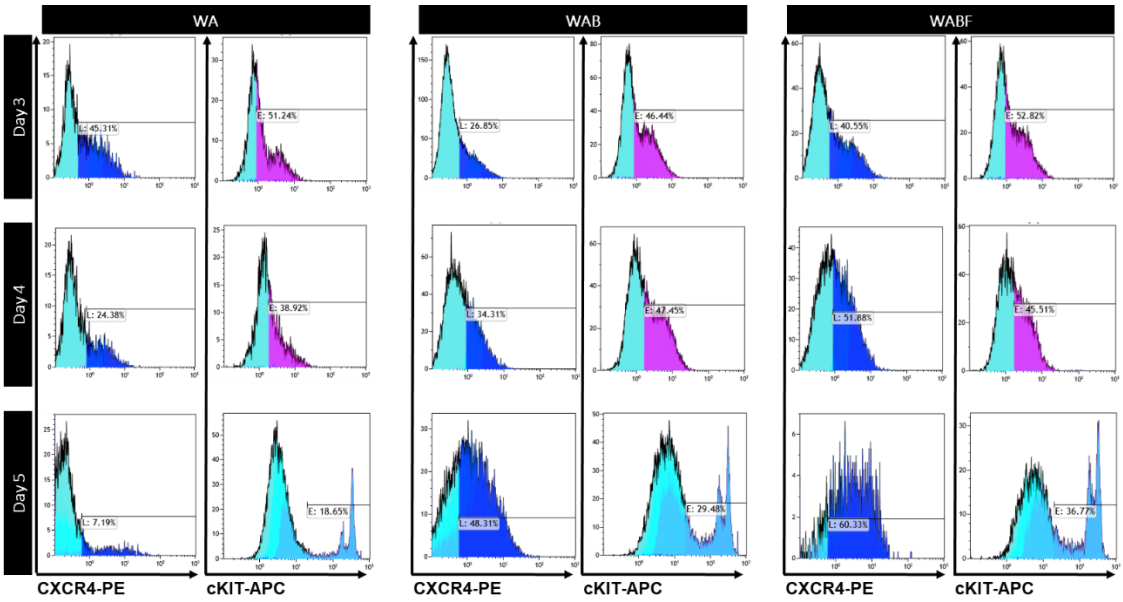


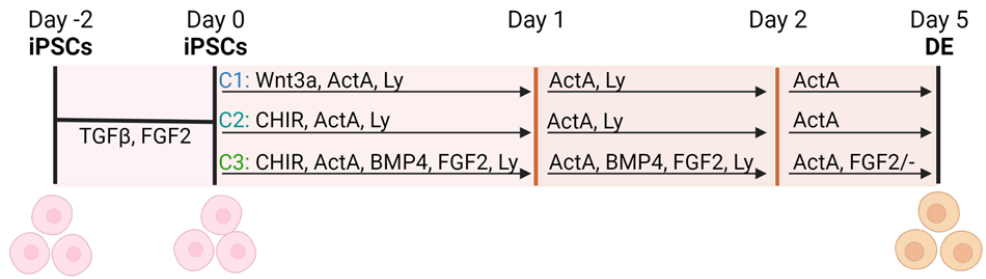
Figure 4.12 Use of FACS to assess definitive endoderm efficiency.

Definitive endoderm cells were differentiated using "WA", "WAB", and "WABF" and analysed for CXCR4 and c-KIT expression on days 3, 4 and 5 using flow cytometry. Histograms show positive expression gated against APC and PE conjugated isotype controls.

CHIR, ActA and LY94002 is the best condition to induce DE with IPF-patient-derived hiPSCs.

The presence of insulin has been reported to be refractory for the generation of definitive endoderm in vitro, resulting in limited differentiation. Siller et al. (2016) reported unsuccessful differentiation of hiPSCs lines characterized by tightly compacted cells that did not migrate out of the colonies [226]. This issue was resolved by removing insulin from the culture, which facilitated significant cell migration out of the colonies. Furthermore, McLean et al., (2006) suggested that the PI3K pathway, which is involved in insulin signal transduction, may hinder the differentiation of self-renewing human embryonic stem cells (hESCs) into definitive endoderm by antagonizing Activin/Nodal signalling activities [223]. They demonstrated that the use of LY294002, a potent inhibitor of PI3 kinase, successfully induced endodermal differentiation in hESCs. Based on these observations, it was decided to investigate whether the incorporation of LY294002 in the supplemented media used thus far could enhance the differentiation of DE cells. **Figure 4.13A** provides an overview of the timeline for the new DE differentiation procedure using the three investigated conditions, C1, C2, and C3. The specific signalling molecules and their concentrations used on each day of the 5-day protocol are described in **Figure 4.13B.**

A.



B.

	Condition 1	Condition 2	Condition 3
Day 0	Wnt3a (50ng/mL) + Act-A (100ng/mL) + Ly (10uM)	CHIR (3uM) + Act-A (100ng/mL) + Ly (10uM)	CHIR (3uM) + Act-A (100ng/mL) + bFGF (100ng/mL), BMP4 (10ng/mL) + Ly (10uM)
Day 1	Act-A (100ng/mL) + Ly (10uM)	Act-A (100ng/mL) + Ly (10uM)	Act-A (100ng/mL) + bFGF (100ng/mL) + BMP4 (10ng/mL) + Ly (10uM)
Day 2	Act-A (100ng/mL)	Act-A (100ng/mL)	Act-A (100ng/mL) + bFGF (100ng/mL)
Day 3	Act-A (100ng/mL)	Act-A (100ng/mL)	Act-A (100ng/mL)
Day 4	Act-A (100ng/mL)	Act-A (100ng/mL)	Act-A (100ng/mL)
Day 5	Change to Lung media	Change to Lung media	Change to Lung media

Figure 4.13 Incorporating PI3 kinase inhibitor in new conditions for definitive endoderm optimisation.

A) Schematic representation of the directed differentiation of hiPSCs into definitive endodermal cells using three new culture conditions. B) Table summarising the growth factors and small molecules with their respective concentrations used in conditions 1, 2 and 3 throughout the 5-day protocol.

BE31 cells were seeded at 1.25x10^4 cells/cm^2 and directly differentiated toward DE. Morphological changes were observed throughout the entire differentiation process, these are shown in **Figure 4.14**. There were significant differences in cell morphology observed from the beginning of the differentiation. Under condition 1, cells appeared sparse from day one of differentiation, and floating cells were frequently observed during media changes, resulting in gaps in the monolayer. By day 5, only a few cells remained. For condition 2, the presence of CHIR seemed to have a positive effect on the DE differentiation. The initially observed compacted colonies disappeared as the protocol progressed with the outgrowth of cells, and the amount of cell death was minimal. As for condition 3, the cells initially formed a monolayer by day 3, but clumps of tightly compacted cells reappeared by day 5.

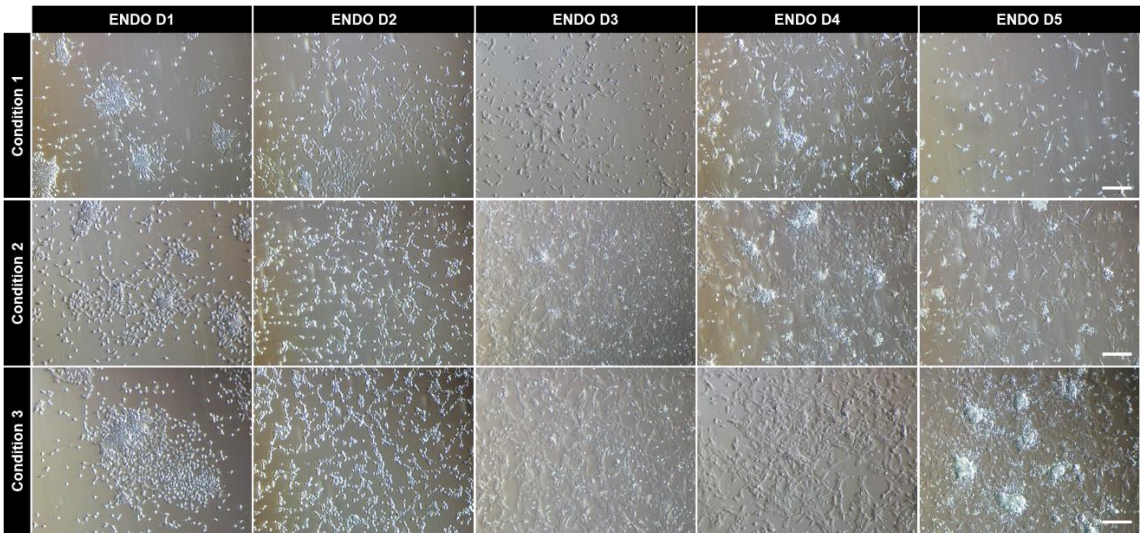


Figure 4.14 PI3 kinase inhibitor promoted differentiation of hiPSCs colonies.
Brightfield images of SFTPC$^{WT/WT}$ cells differentiated up to endoderm day 5 using LY294002 for the first 2 days, causing a marked change in cell morphology. Scale bars, 200µm

To characterise the cultures and evaluate the impact of LY294002 on the differentiation process, flow cytometry analysis was performed on harvested cells on days 3, 4, and 5 of the protocol. The expression of CXCR4 and c-KIT, as representative endodermal markers, was evaluated. **Figure 4.15A** illustrates the results, showing that all three new conditions supplemented with LY294002 significantly increased the expression levels of CXCR4 and c-KIT compared to the previous conditions. Among the three conditions, C2 exhibited the highest levels of CXCR4 and c-KIT expression. The percentages of positive cells for both markers increased from day 3 to day 5, with an average of 84% for CXCR4 and 73% for c-KIT in C2. For C1, the average percentages were 73% for CXCR4 and 62% for c-KIT, while for C3, the values were 59% for CXCR4 and 57% for c-KIT. In addition, to flow cytometry analysis, cells were harvested on day 4 of the differentiation to validate the findings of increased CXCR4 and c-KIT expression. Gene expression analysis was performed using qPCR to assess the expression of other endodermal markers, including SOX17, FOXA2, and GATA4 (**Figure 4.15B**). Consistent with the flow cytometry results, it was observed that C2 exhibited the highest expression levels of SOX17, FOXA2, and GATA4 compared to the undifferentiated control and conditions 1 and 3. This gene expression profile was also reflected in the expression levels of c-KIT and CXCR4, further confirming the effectiveness of C2 in promoting endodermal differentiation.

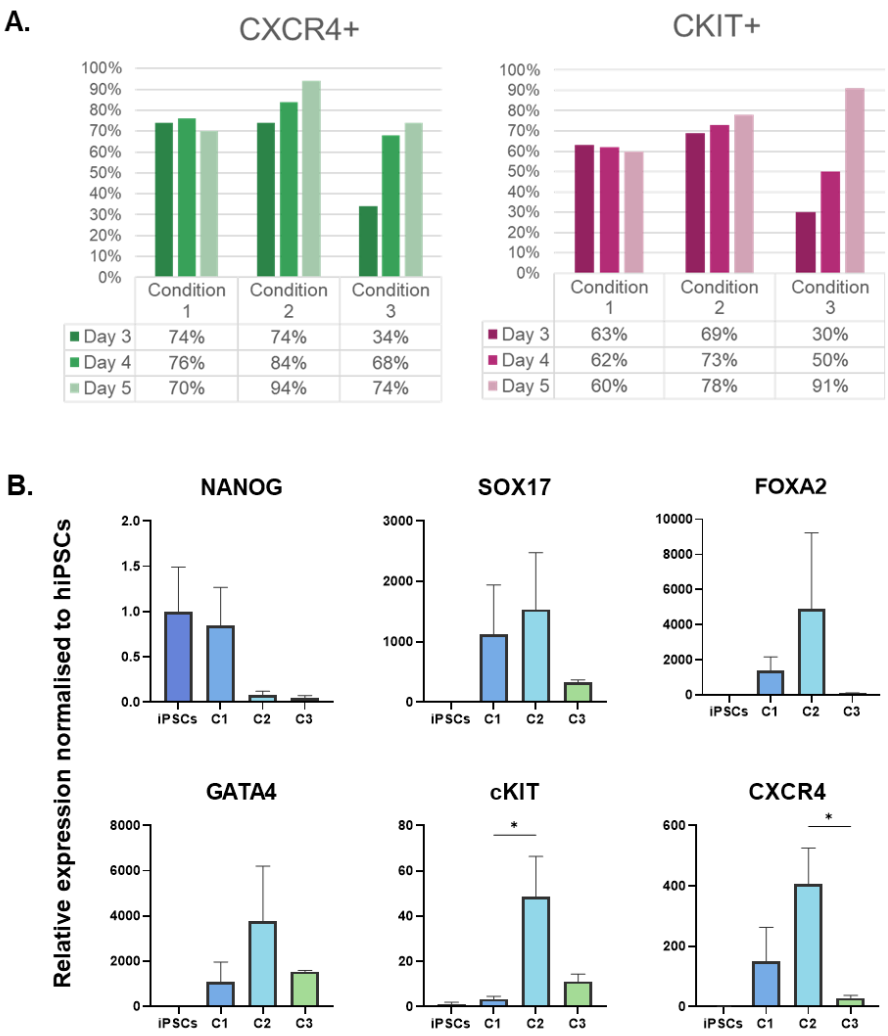


Figure 4.15 Condition 2 induces higher yields of endoderm markers expression.

A) Graphs show the percentages of CXCR4^{+} and c-KIT^{+} cells obtained via flow cytometry analysis on days 3, 4 and 5 of the definitive endoderm protocol. B) qPCR on day 4 of the differentiation protocol comparing expression of endoderm-relevant markers of cells cultured under conditions 1, 2 and 3. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using one-way ANOVA. *p≤0.05. n=1.

To ensure consistency in the differentiation process and avoid phenotypic drift, a specific day was chosen to stop the endoderm induction. Based on previous results showing high yields of positive CXCR4 and c-KIT endodermal cells, day 4 was selected as the endpoint for the differentiation. The experiment was repeated applying the same approach and the BE31 SFTPC$^{WT/WT}$ hiPSC cells to confirm the reproducibility of the optimised condition. Cells were plated and differentiated for various downstream applications, but samples were harvested on day 4 for flow cytometry analysis to assess the expression of the endodermal markers. The optimised condition (C2)

consistently generated cells with 97% CXCR4 positive cells and 83% c-KIT positive cells. **Figure 4.16** provides an overview of the robust definitive endoderm differentiation outcomes using C2 with the IPF-patient-derived cells.

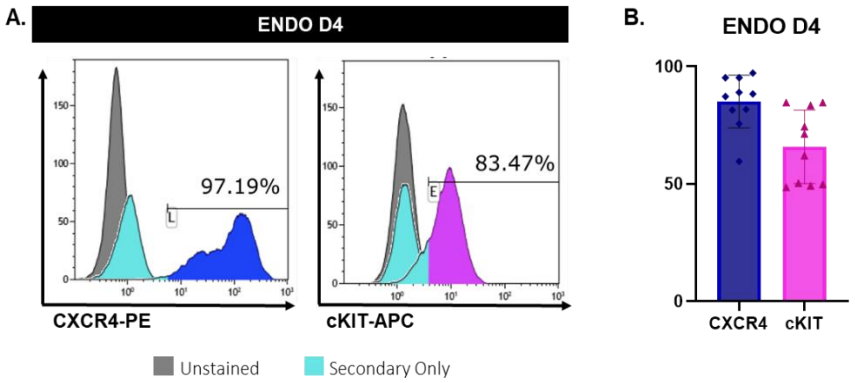


Figure 4.16 Condition 2 up to day 4 adopted as definitive endoderm differentiation protocol.

A) Representative flow cytometry of day 4 definitive endoderm cells analysed for CXCR4 and c-KIT. Histograms show expression above 80% for both markers, threshold established to continue with the lung differentiation B) Quantification by FACS of DE day 4 induction efficiency in SFTPC$^{WT/WT}$ cells, graph shows mean percentage of CXCR4^{+} and c-KIT^{+} cells in 10 independent experiments.

When analysing the morphological changes in the cultures during the endoderm induction, some gaps could still be detected on the monolayer as the differentiation progressed. To address this, the seeding density of cells was increased from 1.25 to 1.5x10^{4} cells/cm^{2}. This adjustment aimed to achieve a more uniform monolayer. Images depicting the establishment of the DE monolayer can be observed in **Figure 4.17A**. Notably, the presence of hiPSC-like colonies was absent throughout the entire protocol, indicating successful differentiation. Due to changes in the flow cytometry facility, the cytometer employed up to this point was no longer available, and the analysis protocol needed to be transferred to a new machine. Unlike the previous cytometer, this new machine allowed for easy co-staining analysis, which was not available before. The plots presented in **Figure 4.17B** demonstrate the nearly 100% expression of CXCR4 and c-KIT double-positivity in our cultures at day 4 of endoderm differentiation.

Finally, to gain further insights into the differentiation process, gene expression analysis was conducted as the cells progressed from day 1 to day 4 of the protocol by qPCR. The results, presented in **Figure 4.17C**, revealed high expression levels of endodermal markers throughout the entire differentiation process. Specifically, SOX17,

FOXA2, GATA4, and CXCR4 were significantly upregulated starting from the second day of differentiation and reaching their peak expression on day 4. It is worth noting that SOX17 exhibited a significant downregulation from day 3 onwards. Conversely, the expression levels of CDX2, a marker associated with posterior foregut endoderm, were significantly downregulated during the differentiation process.

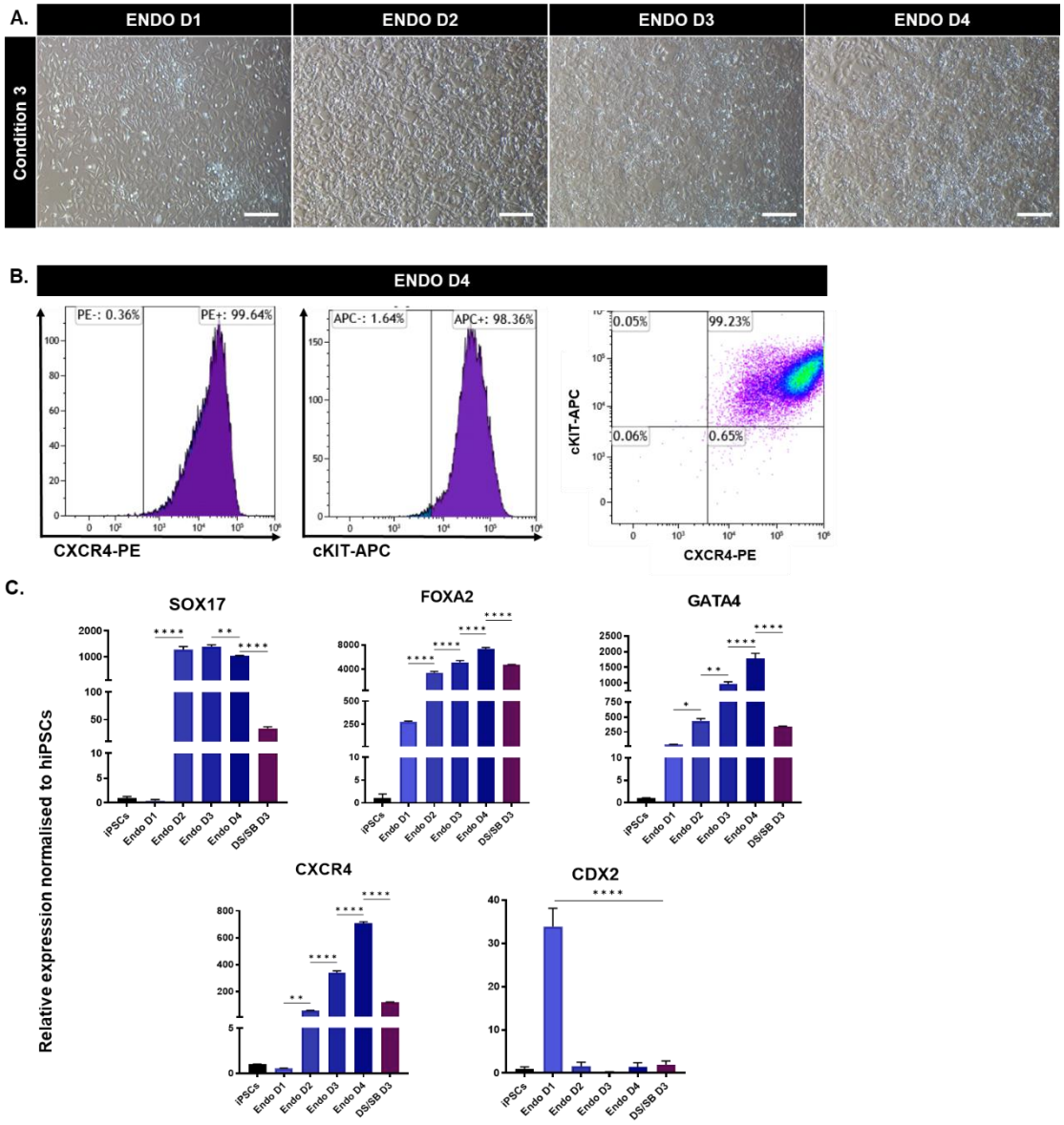


Figure 4.17 Optimised protocol generates reproducible yields of definitive endodermal cells.

A) Representative brightfield images of SFTPC$^{WT/WT}$ hiPSC differentiation to definitive endoderm using the 4-day protocol. Scale bars, 200µm. B) Representative FACS plots showing CXCR4 and c-KIT double staining of day 4 definitive endoderm cells. Dot plot shows >99% of the population falling in the double-positive gate. C) qPCR showing the expression of endodermal differentiating cells on days 1, 2, 3 and 4, and day 3 with Dorsomorphin/ SB434542 treatment. Data shows mean CT values ± standard error of the mean. Statistical

analysis for significance was performed in comparison to undifferentiated hiPSCs using one-way ANOVA. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Taken together, these observations highlight the morphological differences and suggest varying degrees of success in achieving definitive endoderm differentiation under different conditions. They also indicate that the changes in the protocol, including the incorporation of LY294002 in the culture media, effectively enhanced the expression of CXCR4 and c-KIT, evidence of the successful generation of endodermal cells. Based on the results from these experiments, the condition with the highest efficiencies for DE induction using the IPF-patient-derived cell line included 4 days under the influence of CHIR (3μM), LY94002 (10μM) and ActA (100ng/ml). Therefore, C2 was the elected combination of signalling molecules to move forward with the differentiation towards definitive endoderm.

Optimised protocol produces higher Definitive Endoderm (DE) cell yields than the old differentiation approach.

To compare the effectiveness of the newly established definitive endoderm differentiation protocol with the original protocol, the expression profiles of endodermal markers were analysed in $SFTPC^{WT/WT}$ cells differentiated using both conditions. The results demonstrated significant differences in the expression levels of key endodermal markers (**Figure 4.18**). SOX17, FOXA2, GATA4, and CXCR4 were significantly upregulated in the cells differentiated using the new protocol (C2). This indicates a more robust and efficient differentiation into the endodermal lineage. Furthermore, the pluripotency marker NANOG was significantly downregulated in the cells differentiated using the new protocol, confirming the successful transition towards the endoderm stage. Interestingly, the expression levels of CDX2 in the cells differentiated using the original protocol were lower compared to the cells differentiated using the new protocol. This observation suggests that some cells might be deviating from the intended endoderm fate and acquiring a posterior endodermal fate. In summary, these findings highlight the superiority of the new protocol in directing the differentiation process towards the desired endodermal lineage, while minimizing the potential for contaminating cell types.

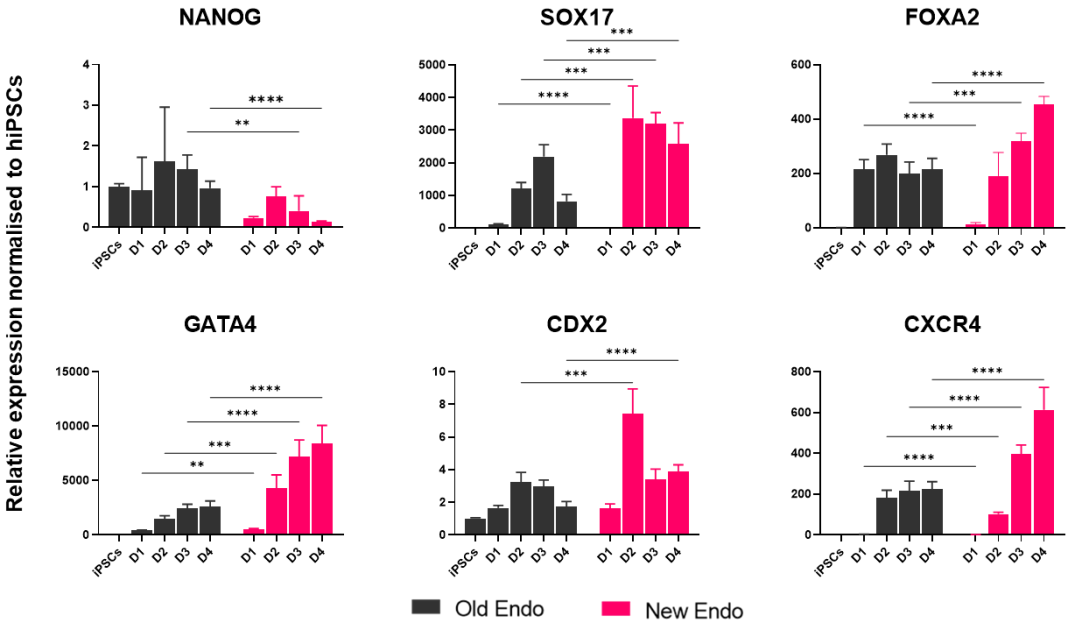


Figure 4.18 Optimised endoderm protocol is superior to the old protocol differentiating IPF-patient-derived hiPSCs.

qPCR showing the expression of endodermal differentiating SFTPC$^{WT/WT}$ cells on days 1, 2, 3 and 4. Data indicates that compared with the original endoderm differentiation protocol, the optimisation significantly increased the transcript levels of markers associated with definitive endoderm generation, such as SOX17, FOXA2 and GATA4. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using two-way ANOVA. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

In order to further emphasise the importance of optimising the endoderm stage of the protocol with the newly established cell lines, the healthy hiPSC cell line REBLPAT, previously used as a control, was subjected to both the original and new endoderm protocols for comparison. The original protocol had previously yielded satisfactory results in generating lung progenitors from REBLPAT cells. The aim was to assess the efficiency of the new endoderm protocol in generating comparable yields. REBLPAT hiPSCs were seeded at a density of 1.5x10^4 and differentiated using both protocols up to day 5 to ensure a fair comparison. **Figure 4.19A** illustrates the observed morphological differences between the treatment groups. By the second day of the protocol, cells cultured under both conditions had formed a homogeneous monolayer. However, when changing the media on day 3, most of the cells on the new protocol conditions were floating, leaving very few attached cells. The remaining cells struggled to survive until the end of the differentiation period. In contrast, the REBLPAT cells cultured using the original protocol thrived, maintaining a healthy and intact monolayer until day 5 of the differentiation.

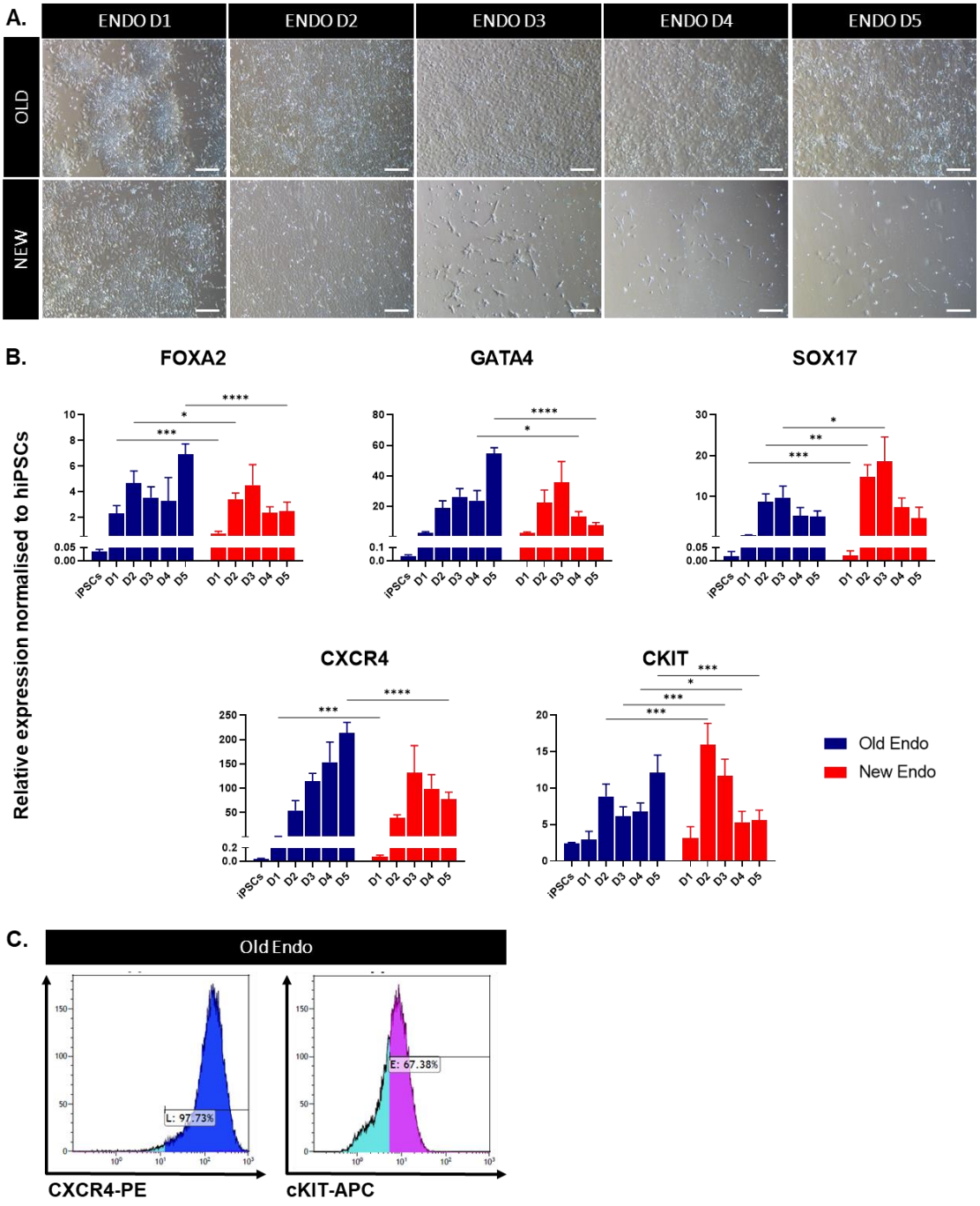


Figure 4.19 Old endoderm protocol is suitable for healthy hiPSC line REBLPAT.

A) Representative brightfield images of definitive endoderm differentiating cells using original 5-day protocol. Scale bars, 200µm B) Comparison of endodermal markers' gene expression by qPCR in REBLPAT cells differentiated for 5 days using the old and new DE differentiation protocols. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using two-way ANOVA. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1. C) Flow cytometry analysis of DE day 4 cells showing 97% CXCR4^{+} and 67% c-KIT^{+} cells.

Furthermore, gene expression analysis by qPCR was performed on cells harvested at each stage of the differentiation to compare the expression profiles between the two protocols (**Figure 4.19B**). Both conditions resulted in the upregulation of expected endoderm markers. However, the cells differentiated using the old protocol exhibited significantly higher levels of FOXA2, GATA4, and CXCR4 at every stage of the differentiation compared to the cells differentiated using the new protocol. The expression pattern of these upregulated markers gradually increased with the progression of differentiation. To be noted, the cells differentiated using the new protocol exhibited a peak expression of all the markers assessed by day 3 of the differentiation, indicating that with further optimisation, this condition could also be adapted to REBLPAT cells. Finally, when analysing the cells generated using the old protocol for the expression of CXCR4 and c-KIT via FACS at the end of the differentiation, the efficiency was surprisingly good. Flow cytometry plots in **Figure 4.19C** show how REBLPAT cells differentiated into 98% CXCR4 and 67% c-KIT-positive endodermal cells.

These results indicate that while the original protocol was indeed effective in promoting the generation of endodermal cells from REBLPAT hiPSCs, it did not yield desirable results with the newly derived SFTPC-mutant IPF-patient hiPSCs. Confirming that the development of the new endoderm protocol was crucial for achieving consistent and efficient endoderm differentiation outcomes in the context of the IPF hiPSCs.

4.2.3 Optimisation of Lung Progenitors differentiation step

Introduction of Anterior Foregut ventralisation step.

Having successfully determined the optimal conditions for definitive endoderm differentiation, the next crucial step was to ensure that the cells followed an anterior foregut endodermal fate. An attempt to establish these anterior foregut cells was previously made by subjecting the differentiating cells to a dual SMAD inhibition using DS and SB for three days after definitive endoderm induction. However, the results of this experiment were unsatisfactory as the DS/SB step was preceded by a suboptimal endoderm differentiation protocol. Therefore, the ventralisation step was reassessed by incorporating the optimised definitive endoderm protocol before it and applying either the lung progenitors' differentiation conditions DIFF1 or DIFF2 for 12 days. In addition to this, the experiment also included differentiating hiPSCs into endoderm cells followed by lung progenitors' induction (DIFF1) without subjecting them to the dual SMAD inhibition step. The DIFF1 conditions consisted of basal DIFF1 medium (RPMI,

B27 supplement, NEEA and P/S) supplemented with 1μM CHIR99021, 5 ng/mL BMP4 and 250nM RA., whereas the DIFF 2 conditions consisted of DIFF2 medium (IMDM, Ham's F12, GlutaMAX, B27 supplement, 7.5% BSA, N2 supplement, AA2P and 1-MTG) supplemented with 3μM CHIR99021, 10ng/mL BMP4 and 100nM RA.

SFTPC$^{WT/WT}$ hiPSCs were seeded at 1.5x10^4 cells/cm^2 and directly differentiated to DE cells using the previously optimised protocol. The differentiating cells were analysed with flow cytometry at the end of the DE induction to ensure a good expression of CXCR4 and c-KIT. Once ~90% co-expression of c-KIT/CXCR4 was confirmed, the cells were cultured with basal DIFF medium supplemented with 2μM DS and 10μM SB or differentiated directly into lung progenitors using DIFF1 conditions. Cells were harvested after 8, 10 and 12 days of culture in lung progenitors' media for further analysis. Using flow cytometry, the expression of the NKX2.1 surrogate surface marker, CPM, was investigated (Fujifilm antibody). The percentages of CPM$^+$ cells obtained under each condition are presented in **Figure 4.20**. Remarkably, the condition that yielded the highest percentage of CPM$^+$ cells, by a significant margin, was DIFF1 without the inclusion of the ventralisation step. Additionally, the differentiated lung progenitors on day 8 exhibited the highest value observed, with 81% of the cells being positive for the surface marker.

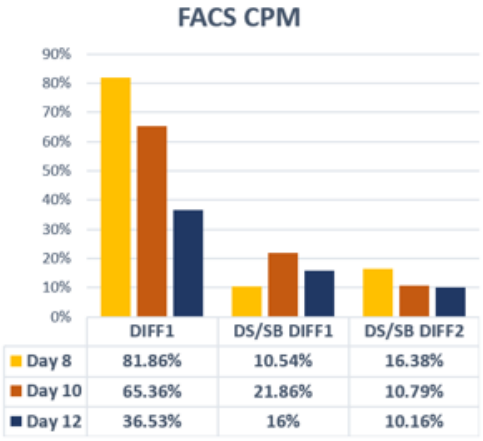


Figure 4.20 Lung progenitors differentiation stage optimisation.

Results of flow cytometry analysis for CPM expression in SFTPC$^{WT/WT}$ lung differentiated cells on days 8, 10 and 12. The graph shows the CPM percentages for lung progenitors grown on DIFF1 conditions with or without Dorsomorphin/SB-43152 (DS/SB) step and DIFF2 conditions with DS/SB step, n=1.

When the gene expression profile was analysed in the collected cells by qPCR, the results were not expected. The lung progenitor markers SOX9, SFTPC and NKX2.1 had the significantly lowest expression levels in the cells differentiated without the

DS/SB step. In fact, SFTPC was only upregulated in the cells cultured in DS/SB DIFF2 condition. Likewise, the AT2 cell markers were mainly upregulated in this later condition, and their levels were barely undetectable in the cells growing in DIFF1 only (**Figure 4.21A**). Contrastingly, the cells differentiated without the DS/SB step presented the highest significantly upregulated non-lung endodermal markers (**Figure 4.21B**). More specifically, the hepatic progenitor's markers TTR, AFP, Albumin and HNF4α, and the hindgut marker CDX2 were highly expressed in these cells, while their expression was significantly lower in the cells differentiated in the presence of the dual SMAD inhibition. Interestingly, the expression of PDX1, a pancreatic marker, was significantly upregulated in the cells growing with DS/SB DIFF2. However, the fold changes in gene expression were minimal, with 0.4 being the highest.

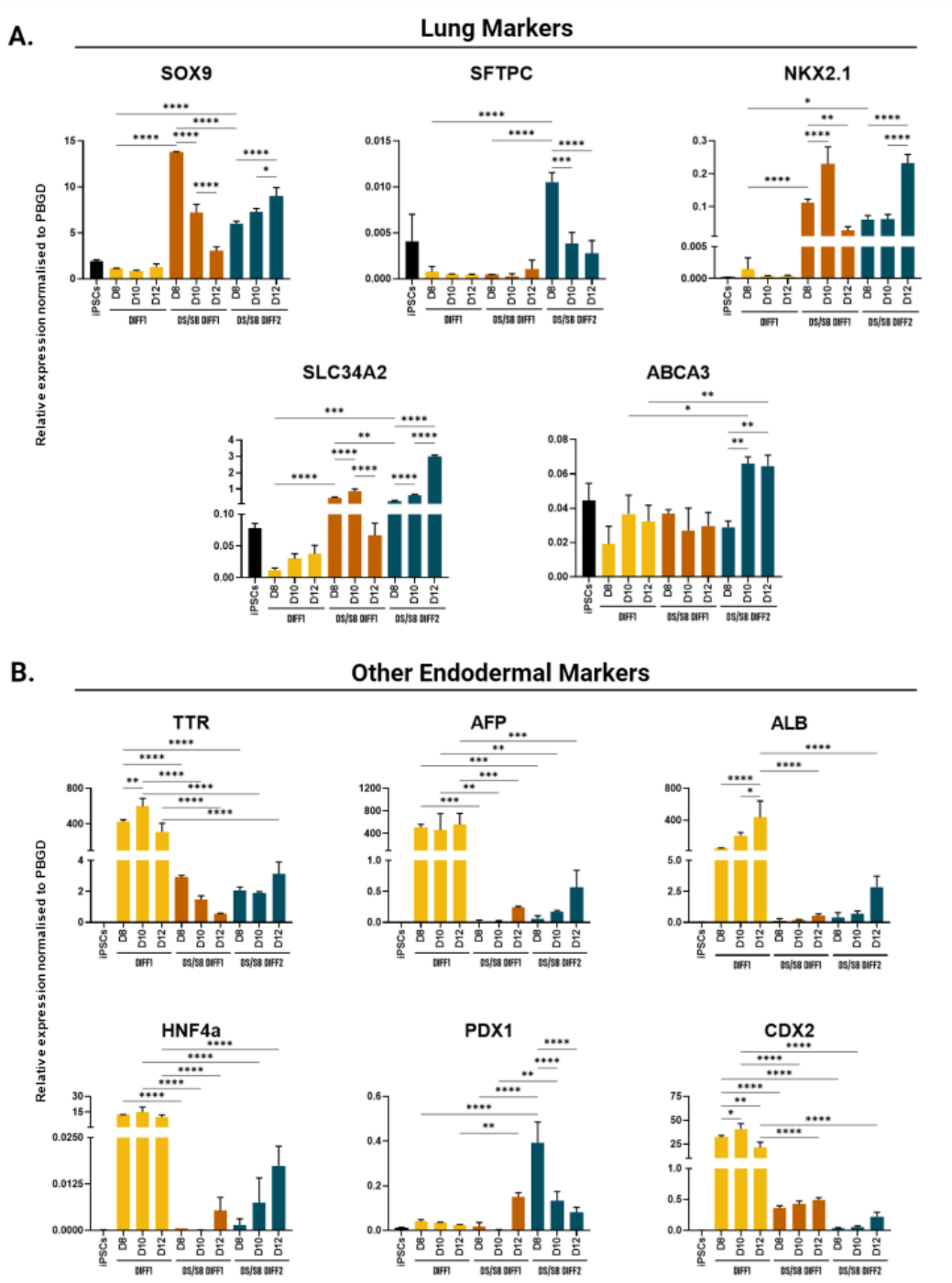


Figure 4.21 Expression of distal lung and other endodermal markers.

qPCR of cells harvested on days 8, 10 and 12 of the lung differentiation protocol using DIFF1, DS/SB + DIFF1 and DS/SB + DIFF2 conditions. A) Graphs showing the expression of distal lung markers and B) markers for other endodermal lineages, such as hepatic (AFP, ALB, HNF4a) and pancreatic (PDX1). PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

As experiments were being done simultaneously and based on the initial flow cytometry analysis results, where DIFF1 culture conditions on lung day 8 cells had the highest CPM percentages, these conditions were selected for flow cytometry cell sorting. The aim was to enrich the CPM^+ lung progenitors to transfer them into 3D culture conditions and induce their maturation into AT2 cells. The set-up of the previous experiment was replicated to enable a repeat of the flow cytometry analysis targeting the CPM^+ expressing cells. The only difference was that this time cells differentiated using DIFF1 condition were going to be sorted on long day 8 leaving no cells for analysis on days 10 and 12. In **Figure 4.22A**, the established sorting strategy is depicted, showing the gates for the cell population being analysed, followed by the selection of the single cells within the population and the exclusion of dead cells. The final sorting gate was determined against the unstained and secondary-only controls (not shown), indicating a positivity of 61% on lung day 8-cells differentiated using DIFF1 only. The cells recovered from the cell sorter (620,000 cells) were embedded in Matrigel droplets at a density of 300cells/µl and cultured with lung maturation media, as described in Chapter 2.

The bar graph in **Figure 4.22B** displays the CPM percentages obtained from the flow cytometry analysis of the cells cultured in the other two media compositions for lung day 8, 10 and 12. One more time, the expression of CPM^+ cells was higher in cells differentiated with DIFF1 compared to those differentiated in the presence of the dual SMAD inhibitors.

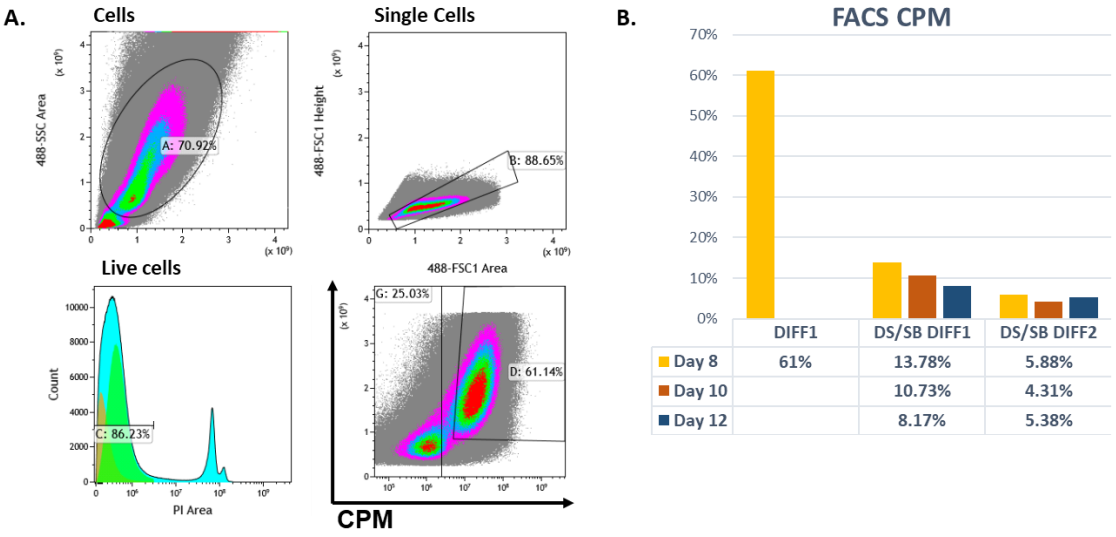


Figure 4.22 DIFF1 culture conditions on lung day 8 produced the highest yields of CPM$^+$ cells.

A) Gating strategy for CPM$^+$ cell sorting of SFTPC$^{WT/WT}$ lung differentiated cells on day 8 showing a positivity of 61%. B) Graph presents the percentages of CPM-expressing cells in the flow cytometry analysis of differentiated cells on lung days 8, 10 and 12 comparing the culture conditions: DIFF1, DS/SB + DIFF1 and DS/SB + DIFF2, n=1.

In this second experiment, the gene expression of lung and other endodermal markers was also evaluated via qPCR, and the results followed a similar pattern to the previous experiment. The lung progenitor markers SOX9, SFTPC, and NKX2.1 exhibited significantly higher expression levels in cells differentiated in the presence of the dual SMAD inhibitors compared to those cultured in DIFF1 alone. Moreover, the AT2 markers SLC34A2 and ABCA3 showed the lowest levels of expression in cells cultured without DS/SB (**Figure 4.23A**). Consistent with the previous experiment, the other endodermal markers analysed presented the highest expression in cells cultured with the DIFF1 condition, particularly TTR, with a 3,000-fold expression relative to the housekeeping gene. The expression of the other hepatic and hindgut cell markers was significantly lower in cells cultured with DS/SB. As in the first experiment, there was an exception for PDX1, which displayed high expression in DS/SB DIFF2, particularly on day 10 (**Figure 4.23B**).

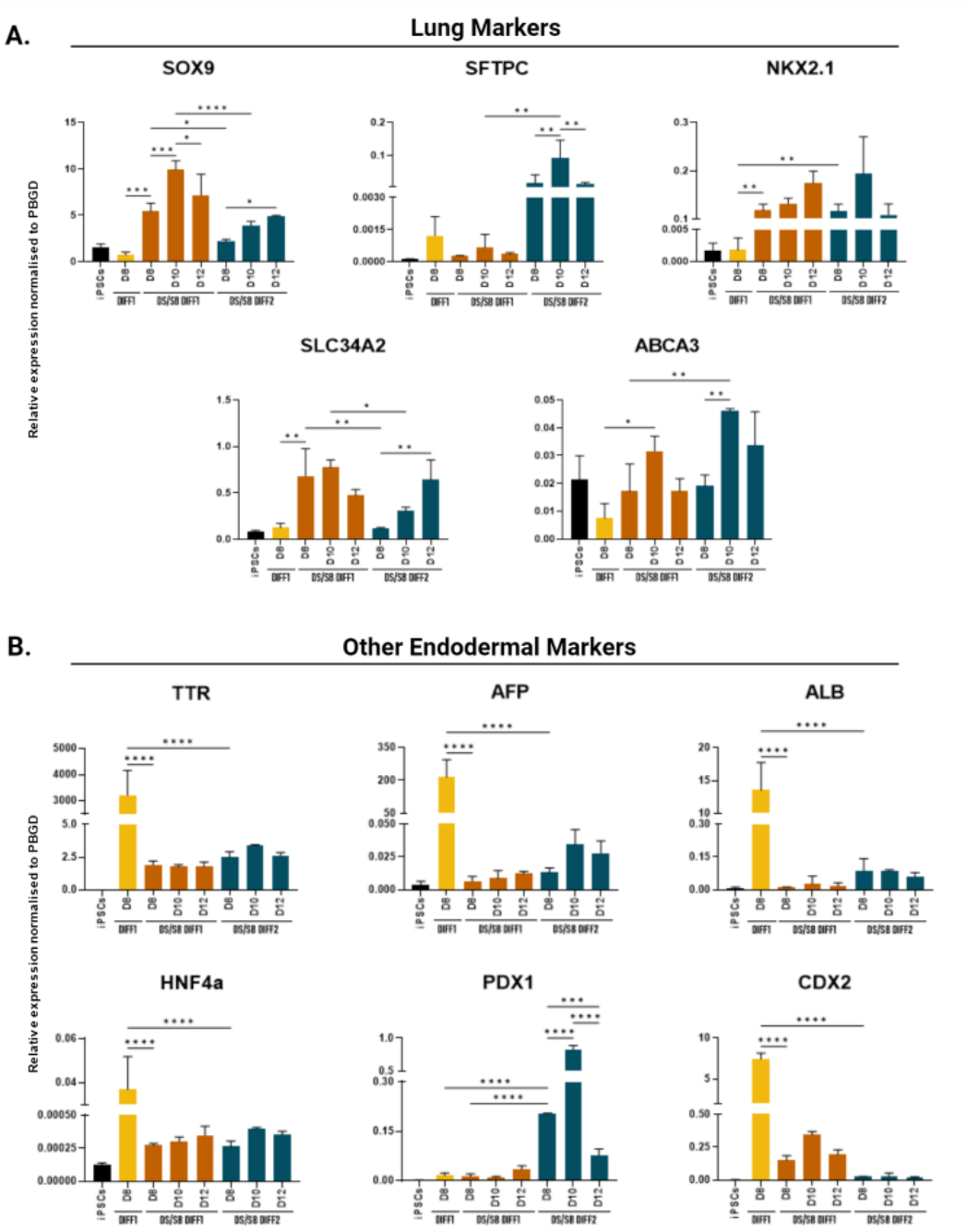


Figure 4.23 The expression pattern of the distal lung and other endodermal markers is replicated in an independent differentiation.

Comparison of gene expression of makers for A) distal lung and B) other endodermal lineages by qPCR in cells harvested on days 8, 10 and 12 of the lung differentiation protocol using DIFF1, DS/SB + DIFF1 and DS/SB + DIFF2 conditions. Cells cultured with DIFF1 condition were CPM-sorted on lung day 8, hence the absence of day 10 and 12 data. PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

To better characterise the CPM$^+$ cells sorted on lung day 8, the gene expression of markers of interest was analysed in comparison to a total lung adult sample (**Figure 4.24**). The expression of endodermal markers was conflicting, as SOX17 was significantly upregulated in the total lung marker whereas FOXA2 was upregulated in the CPM$^+$ cells. A similar conflicting pattern was observed in the expression of hepatic markers, where Albumin and HNF4α showed higher expression in the adult lung sample, while TTR and AFP were more highly expressed in the CPM$^+$ cells. Similarly, the pancreatic marker PDX1 and the hindgut marker CDX2 displayed significant upregulation in the CPM-sorted cells. Interestingly, the gene expression levels of CPM itself were significantly higher in the total lung sample compared to the CPM$^+$ cells.

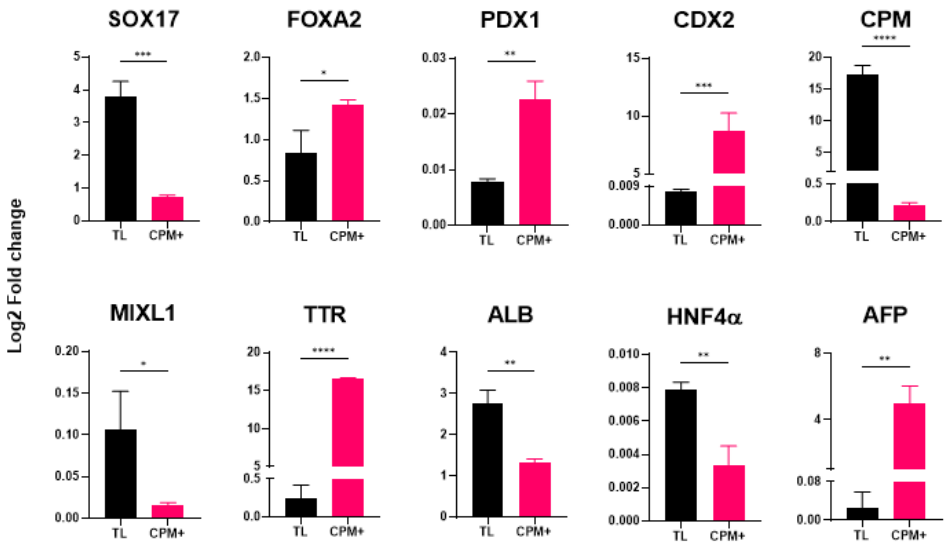


Figure 4.24 Use of a total lung sample to benchmark CPM$^+$ sorted cells.

qPCR data showing the comparison in the expression of various endodermal lineages markers in RNA from an adult lung (TL) and the differentiated IPF-derived hiPSCs sorted for CPM on lung day 8 (CPM+). Data shows mean CT values ± standard error of the mean. Unpaired student's t-test was performed as statistical analysis for significance *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Further characterisation of the sorted cells was performed, but this time the analysis was done using the cells growing in the organoids derived once the CPM-sorted cells were cultured in 3D conditions. The gene expression of the CPM-positive cells was compared to that of the CPM-negative cells obtained from the same sorting experiment. qPCR results in **Figure 4.25** show no statistically significant differences between sorted populations on the lung progenitor marker NKX2.1 expression levels and the AT2 markers SFTPC, SLC34A2 and ABCA3. Only SOX9 presented a significant difference, with a higher expression in the positive sorted cells. These findings suggested that the lung progenitor differentiation protocol resulted in very low

levels of NKX2.1 lung progenitors, and the sorting strategy employed was enriching for contaminant hepatic and pancreatic progenitors, along with a limited number of epithelial lung progenitors.

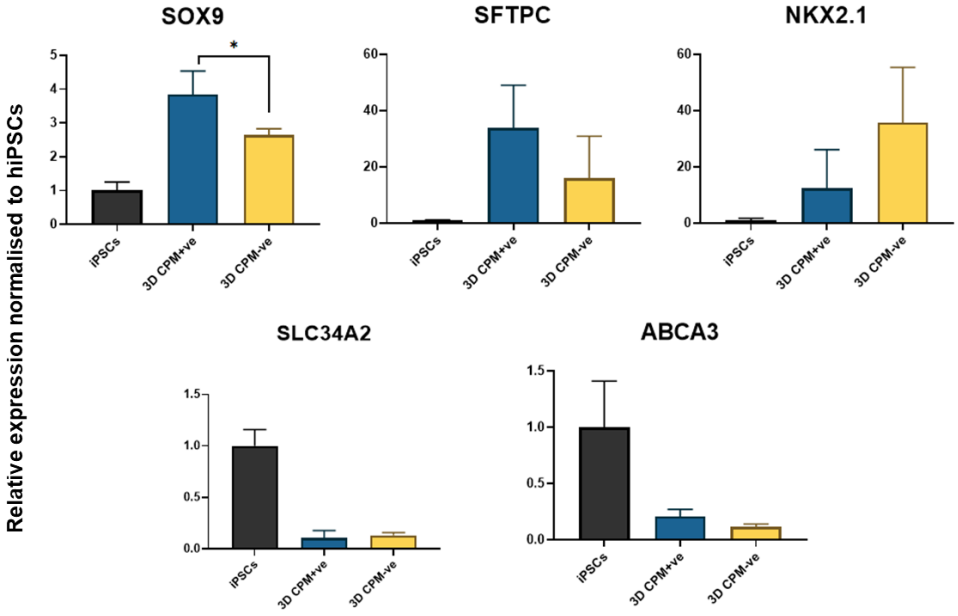


Figure 4.25 Cell enrichment did not improve the expression of lung markers.

qPCR of positive and negative CPM cells sorted on day 8 of the lung differentiation using DIFF1 condition, shows no significant difference in distal lung markers (ABCA3, NKX2.1, SFTPC and SLC34A2) except for SOX9. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using one-way ANOVA. *p≤0.05. n=1.

Several hiPSC in vitro differentiation protocols specify the differentiation of DE cells into SOX2^{+} anterior foregut endodermal cells through the inhibition of BMP and TGFβ signalling [133, 142, 217]. To achieve this, Noggin (BMP inhibitor) and SB143542 (TGFβ inhibitor) are often used. Furthermore, the temporal regulation of BMP signalling is crucial for the generation of lung progenitors. While its inhibition is necessary for the specification of anterior foregut endoderm cells, its presence is required for the ventralisation of these cells, leading to the induction of NKX2.1 expression and the emergence of lung progenitors [149]. In that note, prolonged inhibition of BMP signalling is associated with the posteriorisation of definitive endodermal cells, promoting the emergence of intestinal/pancreatic progenitors [165]. Therefore, it was hypothesized that adding a BMP4 inhibitor would reduce the emergence of NKX2.1+ lung progenitors in the cultures, with the decrease being proportional to the duration of its presence in the culture media.

To investigate this hypothesis, $SFTPC^{WT/WT}$ hiPSCs were seeded at a density of $1.5x10^4$ cells/cm^2 and underwent direct differentiation into DE using the optimized C2 condition. Following DE induction, the cells were further differentiated into lung progenitors using the DIFF1 condition. The DIFF1 basal media was supplemented with CHIR, RA, and BMP4, except for days 1, 2, 3, or 4 of the lung specification stage, during which BMP4 supplementation was replaced with Noggin at a concentration of 100ng/ml. After the specified days, the original DIFF1 condition was resumed for the remaining differentiation process. Cells cultured under the four experimental conditions were harvested on lung day 8 for flow cytometry and gene expression analysis.

Figure 4.26A displays the flow cytometry plots indicating the percentages of CPM^+ cells obtained under each experimental condition. The observed percentages of CPM^+ cells were not as high as those previously observed in cells maintained in the DIFF1 condition (17%, 24%, 29%, and 23% on BMP-inhibition days 1, 2, 3, or 4, respectively). Furthermore, no clear trend was identified that could be directly attributed to the duration of Noggin exposure. When analysing the gene expression of these samples using qPCR, only one gene, NKX2.1, exhibited significant differences. The biggest fold change in expression was observed in cells supplemented with Noggin for 3 days. However, this result did not correlate with the expression levels of the other lung markers assessed (**Figure 4.26B**). It is possible that the Noggin (Peprotech) used in the experiment was not biologically active, and the lack of BMP4 supplementation alone might have been responsible for the minimal changes in gene expression and CPM values observed. Consequently, the findings from this experiment were inconclusive, as the cells did not respond to the in vitro differentiation cues provided.

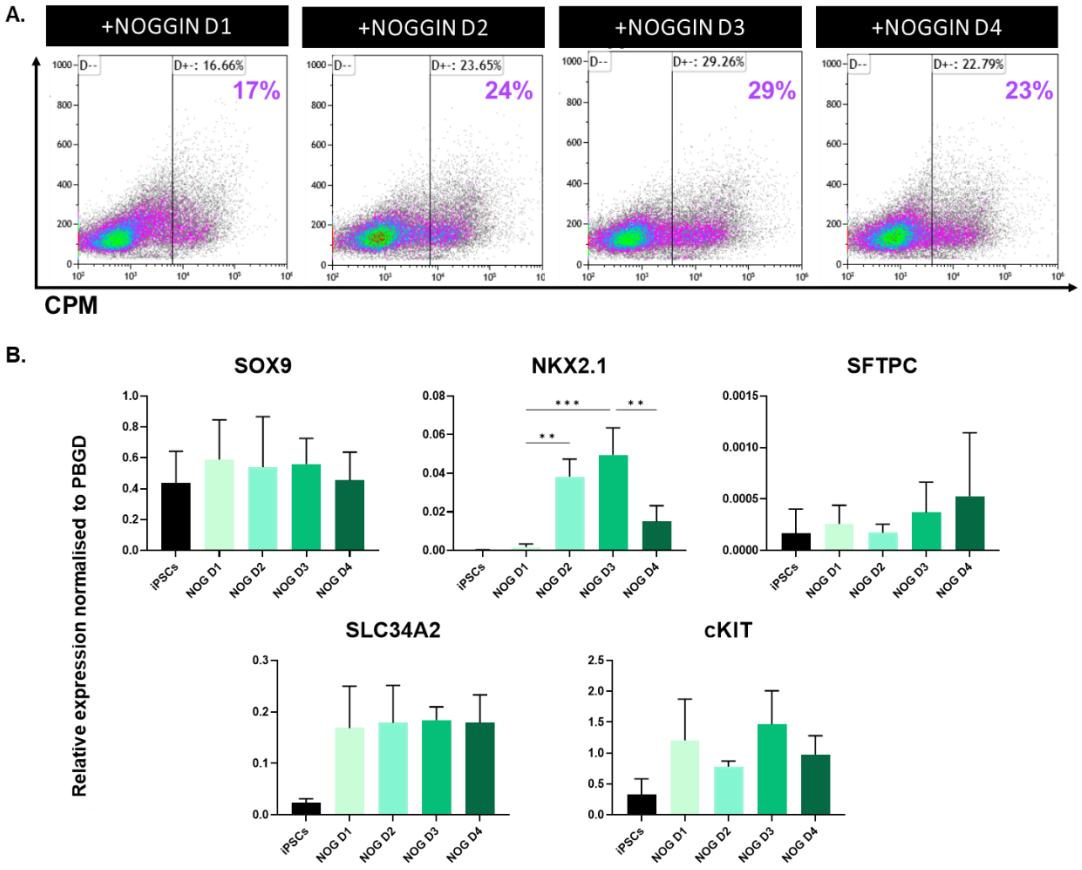


Figure 4.26 Assaying inhibition of BMP signalling at the lung progenitor's stage.

A) Flow cytometry analysis of CPM on lung day 8 progenitors differentiated using DIFF1 condition, adding NOGGING instead of BMP4 for 1, 2, 3 or 4 days. B) qPCR analysis of cells harvested on lung day 8, comparing the four BMP inhibition conditions. PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Since the previous differentiations were conducted exclusively with the BE31 cell line, which is the corrected SFTPC$^{WT/WT}$ hiPSC line, it was deemed necessary also to differentiate the mutated SFTPC$^{MUT/MUT}$ and parental SFTPC$^{WT/MUT}$ cell lines. This was done to obtain a full picture of the cell phenotype generated. The gene expression of SFTPC$^{WT/MUT}$ and SFTPC$^{MUT/MUT}$ hiPSCs differentiated using the optimised DE protocol and the DIFF1 condition to generate lung progenitors was analysed on endoderm day 4 and lung days 4, 6 and 8.

The expression profiles of SFTPC$^{WT/MUT}$ for the markers of interest can be observed in **Figure 4.27**. The analysis revealed that cells exhibited significant upregulation of crucial lung progenitor markers, like SOX9, towards the later stages of differentiation.

However, the fold expression changes observed for NKX2.1 and SFTPC were remarkably low. Conversely, the expression levels of the other endodermal markers that were assessed showed significant upregulation, particularly on lung day 8. The most substantial changes in expression were observed for TTR and AFP. Upon analysing the profiles of the differentiated SFTPC$^{MUT/MUT}$ cell line (**Figure 4.28**), the results were consistent with those of the SFTPC$^{WT/MUT}$ cell line regarding other endodermal markers. However, when examining the expression of lung-specific markers, the findings were even more discouraging for the SFTPC homozygous mutant cell line. There was no upregulation observed for NKX2.1, SOX9, or SFTPC at any stage of the differentiation process. This indicated a significant impairment in the ability of the SFTPC$^{MUT/MUT}$ cell line to express these key lung markers during differentiation using the DIFF1 condition.

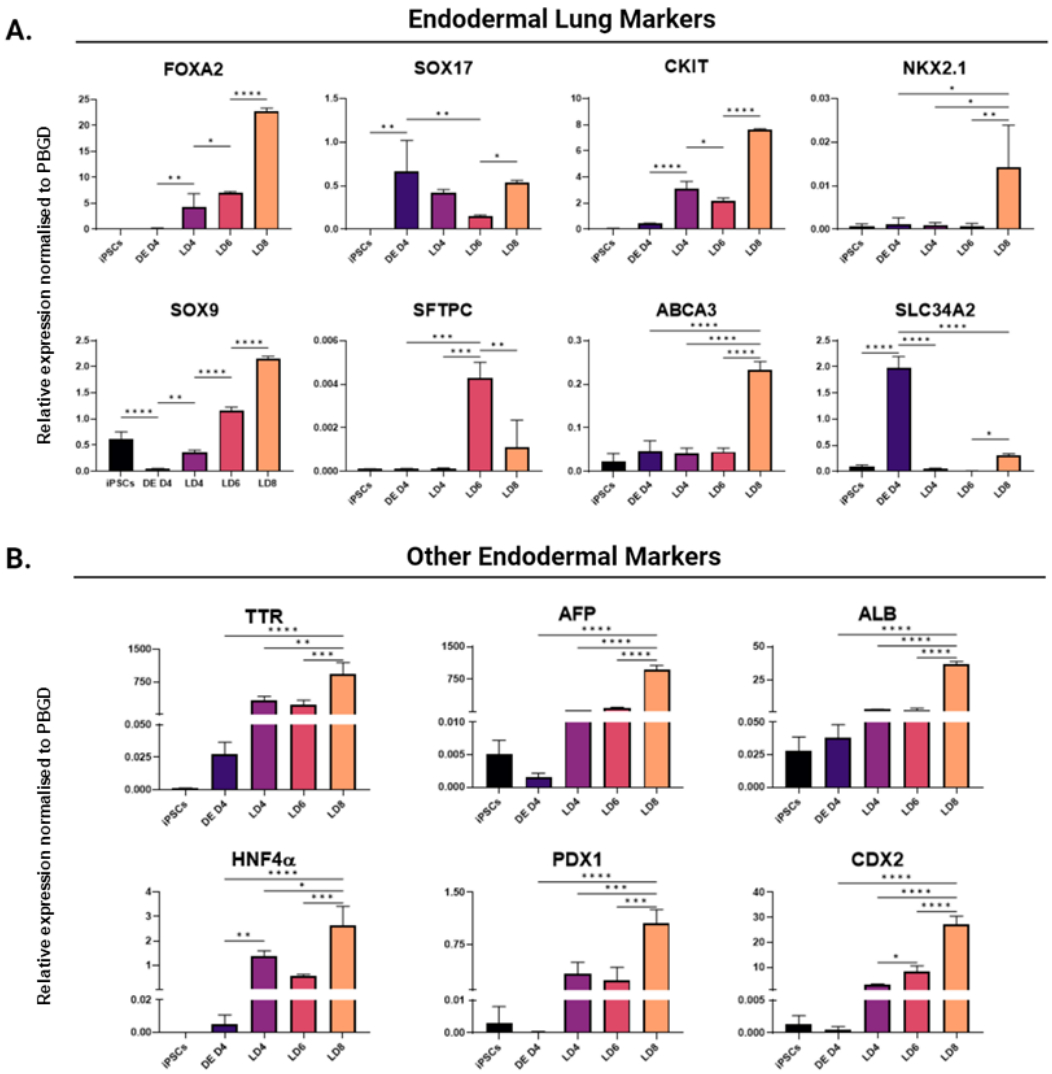


Figure 4.27 DIFF1 lung differentiation conditions were tested using SFTPC$^{WT/MUT}$ hiPSCs. qPCR data for A) lung progenitor markers and B) for other endodermal lineages. PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

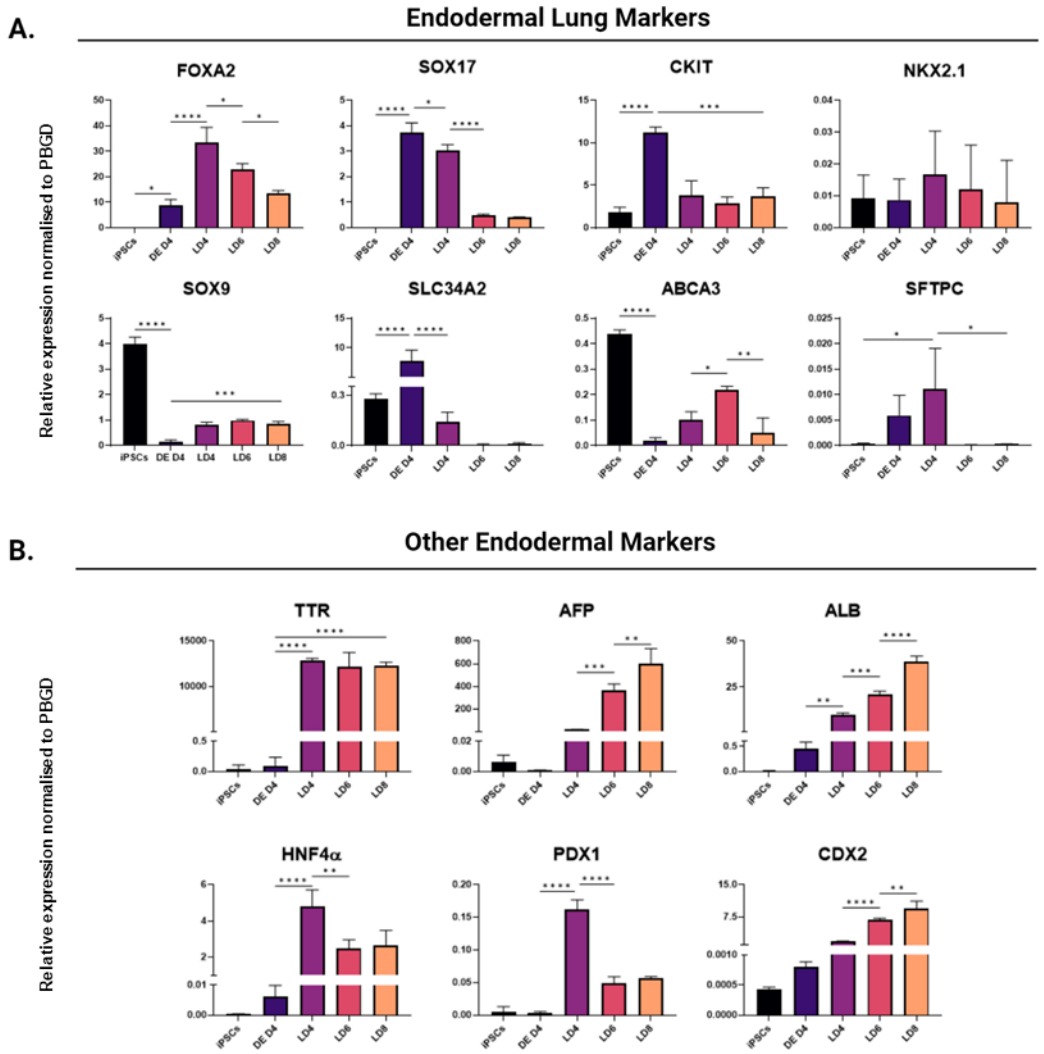


Figure 4.28 DIFF1 lung differentiation conditions were tested using SFTPC$^{MUT/MUT}$ hiPSCs. qPCR data for A) lung progenitor markers and B) for other endodermal lineages. PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Considering the collective findings, it became apparent that the DIFF1 condition applied for 8 days was not optimal for generating lung progenitors. Instead, it appeared to be promoting the generation of a significant number of hepatic progenitor cells, which were subsequently captured during the CPM enrichment process. As a result, it was

concluded that including a dual SMAD-inhibition step was necessary for ventralisation of the anterior foregut endoderm (AFE) population before initiating the induction of lung progenitors. This step was believed to be crucial to prime the epithelium to obtain a more refined population of lung progenitors and reducing contamination from other endodermal lineages.

Use of small molecules and recombinant growth factors for the directed differentiation of lung progenitors.

Strategies for the in vitro development of lung epithelium involve the exogenous addition of growth factors and inhibitors at specific times and concentrations to mimic the in vivo developmental cell signalling pathways. This has been achieved by sequentially specifying definitive endoderm, followed by anterior and ventral patterning, ultimately leading to the induction of NKX2.1+ lung progenitors, which further differentiate into lung epithelial cells [133–135, 155]. Efforts have been made to enhance the enrichment of this population of progenitor cells using cell surface markers such as CPM, CD47, and CD26 [157, 216]. Notably, CPM-based sorting methods have been shown to be more effective in isolating NKX2.1-positive cells compared to those utilizing CD47 [8].

To elucidate if the lung progenitors generated in this thesis could be better enriched using antibodies against CPM or CD47, presumable lung progenitors were analysed by flow cytometry on days 8, 9 and 10 of the lung differentiation protocol. SFTPC$^{WT/WT}$ cells were cultured using the three conditions previously described, DIFF1, DS/SB DIFF1 and DS/SB DIFF2. The flow cytometry plots in **Figure 4.29** illustrate the percentages obtained for CPM and CD47-positive cells from cells cultured under each condition. It was noteworthy to observe that the percentages of CD47-positive cells were consistently above 97% for all time points and conditions. This indicated that using CD47 as a surface marker for isolating NKX2.1-positive progenitors would not be effective since nearly all cells in the culture would be selected, defeating the purpose of enrichment. In contrast, the CPM-positive cell levels were lower than in previous experiments. Notably, on lung day 8, the DIFF1 condition did not yield the highest levels of CPM-positive cells. However, conditions that included the DS/SB step demonstrated the highest percentages of CPM-positive cells.

An intriguing observation was the identification of a rare but distinct population of highly positive CPM cells located at the far end of the X-axis on the flow cytometry plots. This population was more prominent in the plots of cells cultured under the DS/SB DIFF2

condition. Previous experiments (**Figure 4.20A** and **Figure 4.23A**) had demonstrated that this specific culture condition also led to the highest levels of SFTPC gene expression. This correlation suggested that this highly positive CPM population could represent the origin of NKX2.1 distal lung progenitors. Consequently, further investigation of this condition was deemed necessary.

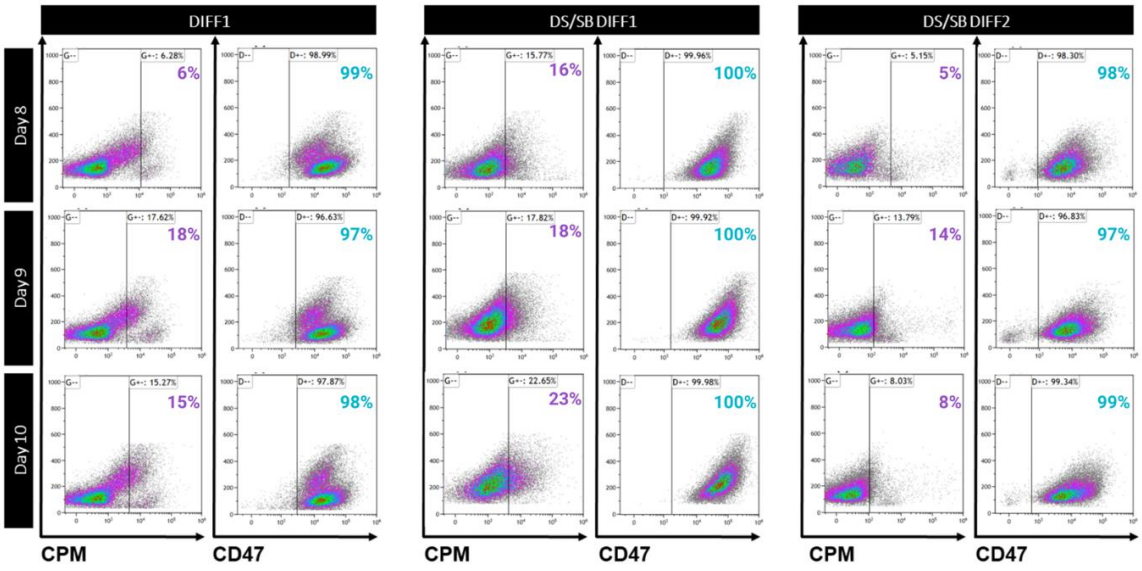


Figure 4.29 Adoption of DS/SB + DIFF2 condition for lung progenitors' differentiation.
Flow cytometry analysis for CPM and CD47 expression in SFTPC$^{WT/WT}$ lung differentiated cells on days 8, 9 and 10, comparing the culture conditions: DIFF1, DS/SB + DIFF1 and DS/SB + DIFF2.

To further characterize the CPM-positive high population, another differentiation experiment was conducted using SFTPC$^{WT/WT}$ hiPSCs. The cells were differentiated into lung progenitors using the DS/SB DIFF2 conditions and were sorted specifically on lung day 8, as the emergence of this distinct population was already evident. The CPM sorting strategy can be observed in **Figure 4.30A**, where different gates were defined to distinguish cells expressing different levels of the surface marker CPM. Disappointingly, it was found that only 0.11% of the processed cells on lung day 8 exhibited a high level of CPM positivity. However, an additional 11.5% of cells displayed moderate levels of CPM expression.

The different gated populations were sorted individually and subjected to gene expression analysis using qPCR (**Figure 4.30B**). They were compared to an undifferentiated control sample and the previously sorted CPM^{+} cells generated using the DIFF1 condition. A positive correlation was observed between the NKX2.1 gene expression and the highly positive CPM (CPM^{high}) cells. NKX2.1 was significantly upregulated in the CPM^{high} cells in comparison with the other sorted populations and

the undifferentiated control. Interestingly, the levels of NKX2.1 were higher in the CPM-negative population from this differentiation experiment than in the CPM-positive population from a previous differentiation performed using the DIFF1 condition. Additionally, the older CPM-positive sample exhibited significantly higher expression levels of the hepatic endodermal markers TTR, AFP, and ALB compared to all the other samples. The gene expression levels of TTR were also significantly upregulated in the CPM^{high} population but not those of AFP and ALB.

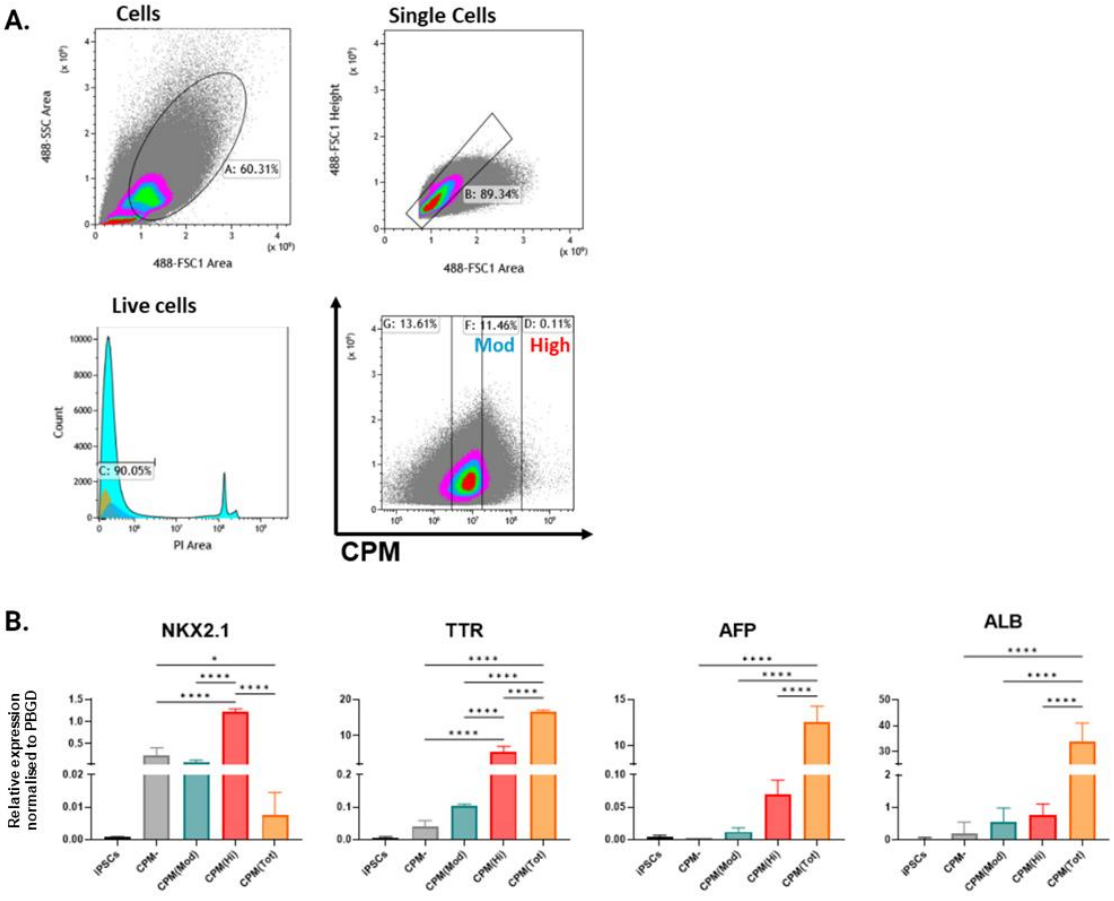


Figure 4.30 Change of cell sorting gating strategy to enrich highly positive CPM lung progenitors.

A) Sorting of CPM^{+} SFTPC$^{WT/WT}$ cells on lung day 8. The gating strategy was modified to distinguish and enrich the negative CPM cells, moderately positive (Mod) and highly expressing ones (High). B) qPCR data comparing the gene expression of the lung (NKX2.1) and hepatic markers (TTR, AFP and ALB) in the CPM populations sorted in A (CPM^{-}, mod, and Hi), and the previously isolated 61% CPM^{+} cells (total- Tot). PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

These findings were promising as they indicated that sorting for the highly positive CPM population alone could help eliminate contaminating cells from the cultures.

However, the extremely low recovery of CPM^{high} cells needed to be addressed and improved.

The published protocol, from which the DS/SB DIFF2 condition was adapted [162], recommended performing a cell splitting step at the end of the endoderm induction stage. In the study from Jacob et al., the cells were replated at an optimized ratio per cell line before proceeding with the ventralisation step. In an effort to improve the percentages of CPM^{high} cells obtained in our cultures, it was decided to implement this splitting step prior to culturing the cells with DS/SB. Following the completion of the endoderm differentiation, a couple of wells of the cultured plates were processed to analyse the expression of CXCR4 and c-KIT. If the majority (>80%) of the endoderm cells exhibited double positivity for these markers, indicating their readiness for lung differentiation, the remaining cells on the plate were dissociated to be re-plated Different seeding ratios were tested to identify the one that yielded the highest number of CPM^{high} cells.

The endodermal cells were split and seeded at 1:1, 1:1.5, 1:2 and 1:3 ratios, and subsequently cultured in DS/SB DIFF2 conditions. Extra care needed to be taken during the cell splitting process, as the protocol advised against complete dissociation of the monolayer. Instead, it recommended to keep the cells in clumps to improve their viability and recovery after being seeded in DS/SB media. After 3 days of DS/SB treatment and 8 days of lung induction, the cells were processed for flow cytometry analysis. **Figure 4.31A** illustrates the flow cytometry plots, which were gated to include either the entire CPM-positive population or only the CPMhigh cells. The clear separation of the highly expressing cells from the rest of the population is evident in the plots. Notably, the highest percentages for both gating strategies were observed in the cells replated at a ratio of 1:3 on endoderm day 4. These populations were also analysed by qPCR on endoderm D4 (before the split) and lung days 4 and 6 (**Figure 4.31B**). The results indicated that the cells split at lower ratios presented the highest gene expression levels for NKX2.1 and CPM. However, these samples also presented the highest expression levels for the hepatic markers TTR, AFP and Albumin. Based on these results, a seeding ratio of 1:3 was selected to be implemented in the differentiation protocol.

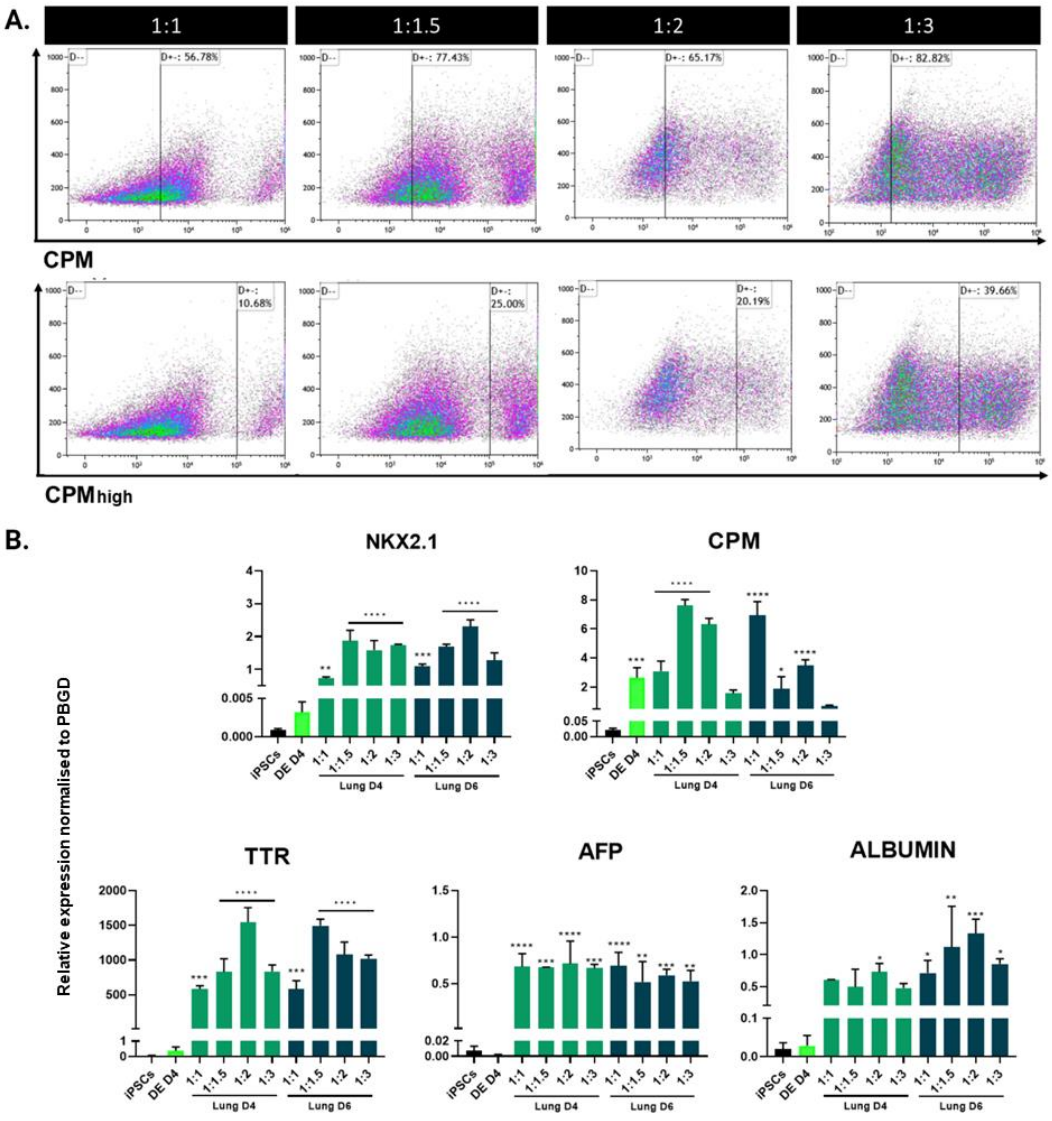


Figure 4.31 Dissociating and replating endodermal cells enhances lung progenitors' differentiation.

A) Flow cytometry analysis for CPM of lung day 8 cells dissociated on endoderm day 4 and reseeded at 1:1, 1:15, 1:2 and 1:3 ratios to continue with DS/SB+DIFF2. FACS plots on the top row show the gating of total CPM^{+} cells, whereas the bottom row shows the gating of CPM^{high} cells. B) qPCR of the endoderm day 4 (DE D4) and the reseeded cells harvested at lung days 4 and 6. PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

The experiment was repeated to account for potential user-dependent variations associated with seeding cells as ratios, especially when clump-based seeding was involved instead of specific seeding densities. The objective was to confirm whether the 1:3 seeding ratio was indeed the most effective. Flow cytometry analysis was

conducted once again on lung day 8, and the cells were gated to determine the percentage of CPM^{high} cells (**Figure 4.32**). The results reaffirmed that the 1:3 seeding ratio yielded the highest generation of CPM^{high} cells.

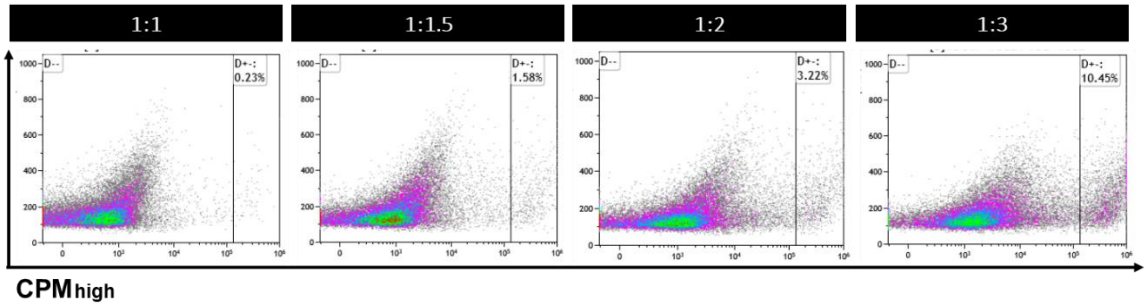


Figure 4.32 Replating DE D4 cells at a 1:3 ratio generates the highest yield of CPM^{high} cells.

Flow cytometry analysis plots of lung progenitors on day 8 show distinct populations of CPM^{high} cells, with the highest percentage (10.7%) from the cells replated at a 1:3 ratio.

Based on these findings, the replating ratio of 1:3 on endoderm day 4 was selected to be implemented in the differentiation protocol to proceed with the dual SMAD inhibition step.

Increased expression of lung progenitors in 2D culture.

The optimised differentiation protocol for generating lung progenitors in monolayer cultures was applied to SFTPC$^{WT/WT}$ hiPSCs in order to enrich the CPM^{high} cell population identified in the previous flow cytometry analysis. In our initial attempt to sort this population, the cells were not split at the end of the endoderm differentiation, resulting in a low yield of only 0.11% CPM^{high} cells (Figure 4.30 Change of cell sorting gating strategy to enrich highly positive CPM lung progenitors.**Figure 4.30A**). However, after implementing the 1:3 seeding ratio prior to ventralising the cells, significant improvements were observed. **Figure 4.33A** depicts the flow cytometry sorting gates from two independent experiments. The percentages of CPM^{high} cells obtained were 2.53% and 23.18%. The variation in percentages is likely attributed to the dexterity acquired with practice in clump-based cell splitting. The CPM^{high} cells retrieved from the sorter were either transferred to 3D Matrigel droplets in maturation culture conditions or processed for RNA extraction for further analysis.

The enriched CPM^{high} lung progenitors were subjected to gene expression analysis via qPCR, comparing them to CPM-moderately positive cells and CPM-negative cells. The results shown in **Figure 4.33B** demonstrated that the enrichment of CPM^{high} cells was associated with an elevation in the transcript levels of the lung progenitor marker

NKX2.1. This observation correlated with the gene expression profile of CPM. Additionally, the enrichment of CPM^{high} cells also led to an enhancement in the expression of markers specific to AT2 cells, namely SFTPC and ABCA3. These findings indicate that the enrichment process increased the abundance of lung progenitors and promoted the expression of markers associated with mature AT2 cells.

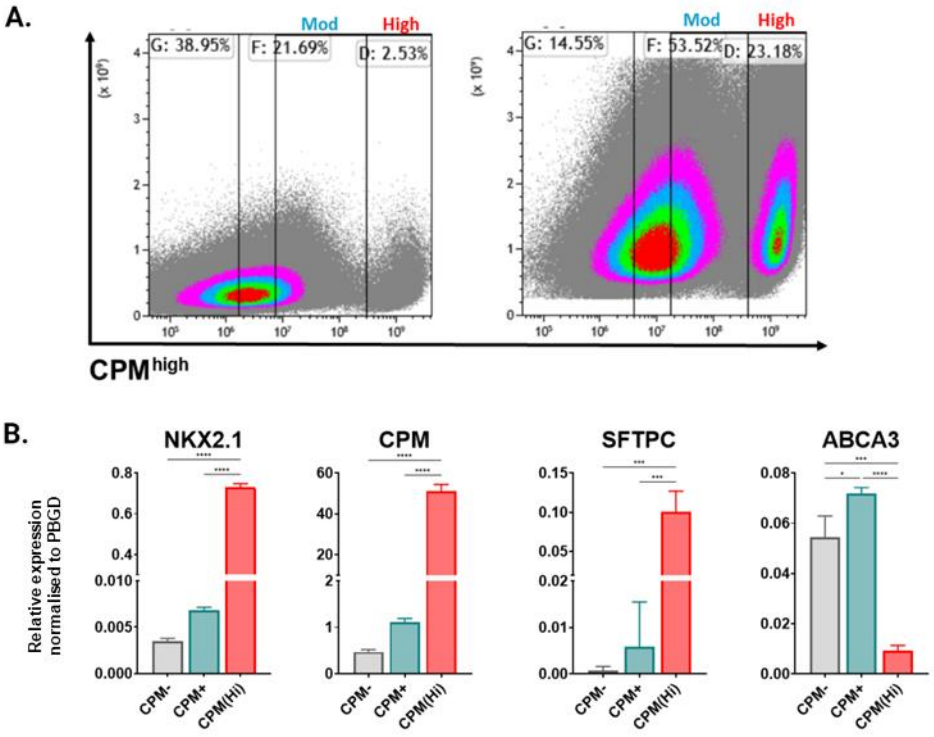


Figure 4.33 CPM^{high} sorted cells express the required distal lung progenitors' signature.
A) Sorting lung day 8 CPM^{high} cells using SFTPC$^{WT/WT}$ hiPSCs in two independent experiments. B) qPCR for gene expression of distal lung markers comparing CPM negative, moderately positive, and highly positive cells. PBGD (Porphobilinogen Deaminase): internal reference gene. Results are shown as log2 fold change between normalised sample sets ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

The cell phenotype of interest was further confirmed through immunocytochemistry analysis. SFTPC$^{WT/WT}$ hiPSCs differentiated using the optimized protocol were stained using antibodies against lung progenitor markers NKX2.1 and SFTPC, along with their respective cell membrane surrogate markers CPM and NaPi2b. In addition, to ensure the absence of contaminant cell types in the cultures following the differentiation of lung progenitors, the cells were also stained for hepatic markers TTR, AFP, and Albumin. The immunofluorescent images in **Figure 4.34** demonstrate the positive punctate cytoplasmic SFTPC staining along with its corresponding surface marker

NaPi2b. Similarly, nuclear NKX2.1-positive staining and cell membrane CPM staining were found to co-localize in the cultures. Importantly, negative staining was observed for Albumin and AFP, indicating the absence of hepatic cell contamination. It is worth noting that positive staining for TTR in the NaPi2B-stained images was expected, as it aligns with the consistent gene expression of TTR observed in the qPCR analysis of the different cell populations.

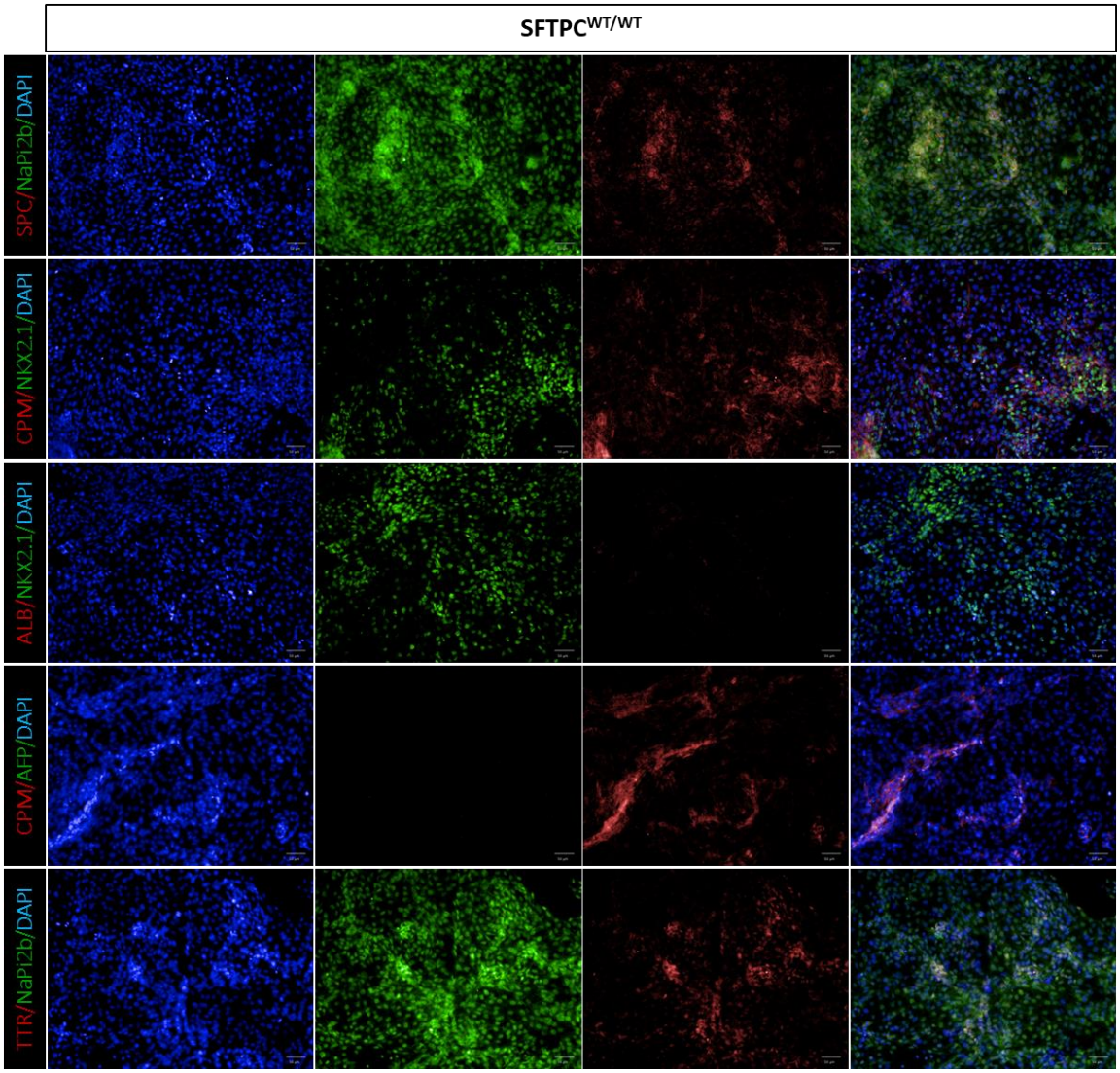


Figure 4.34 SFTPC$^{WT/WT}$ hiPSCs-derived lung progenitors co-express NKX2.1 and SFTPC with surrogate markers.

Representative immunofluorescence microscopy with antibodies against SPC, CPM, ALBUMIN, TTR (red) and NaPi2b, NKX2-1 and AFP (green) in day 8 of the lung progenitors differentiation protocol. Nuclei were stained with DAPI (blue). Scale bars, 50 µm.

Immunocytochemistry analysis was conducted on differentiated SFTPC$^{WT/MUT}$ and SFTPC$^{MUT/MUT}$ cells using the newly optimized protocol. The images in **Figure 4.35** provide confirmation of the positive co-staining of nuclear NKX2.1 and cytoplasmic

SFTPC, along with their respective surrogate markers CPM and NaPi2b, which stain the cell membrane. It is noteworthy that the intensity of SFTPC staining in these cells was not as pronounced as the staining observed in the wild-type cell line.

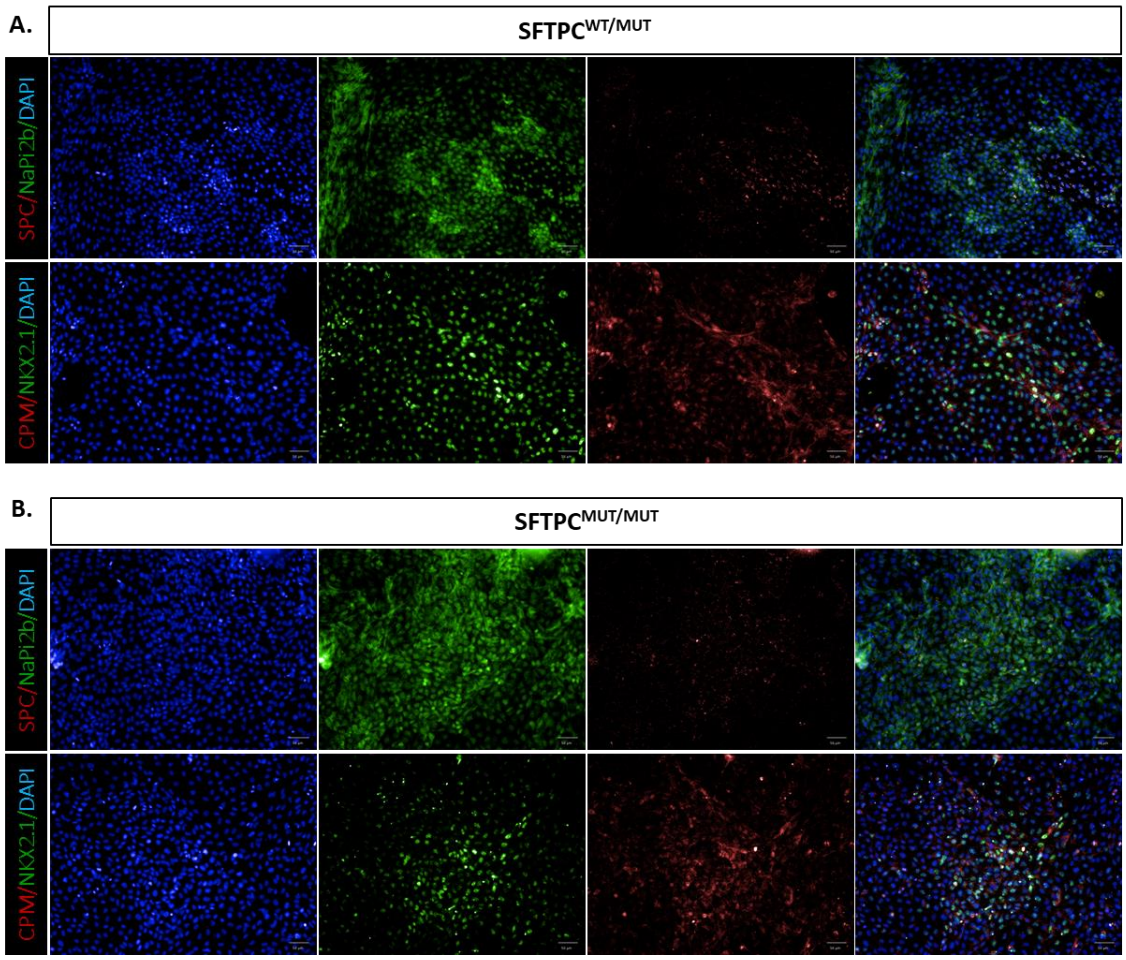


Figure 4.35 SFTPC$^{WT/MUT}$ and SFTPC$^{MUT/MUT}$ hiPSCs-derived lung progenitors co-express NKX2.1 and SFTPC with surrogate markers.

Representative immunofluorescence microscopy of A) SFTPC$^{WT/MUT}$ and B) SFTPC$^{MUT/MUT}$ day 8 lung progenitors co-stained with antibodies against SPC or CPM (red) and NaPi2b or NKX2-1 (green). Nuclei were stained with DAPI (blue). Scale bars, 50µm.

As with the definitive endoderm protocol, for comparison purposes, the healthy hiPSC cell line REBELPAT was cultured using both the original and the newly optimized protocol for lung progenitors' differentiation. Differentiated cells were harvested throughout the protocol for analysis using qPCR and flow cytometry (**Figure 4.36**). Flow cytometry analysis of lung day 8 cells revealed no presence of a high CPM population in either condition. However, the percentages of CPM-positive cells were significantly higher when using the original differentiation protocol, with nearly 50% of the cells staining positive for the surface marker. On the other hand, the newly

optimized protocol resulted in higher gene expression levels of the lung markers SOX9 and CPM through qPCR analysis.

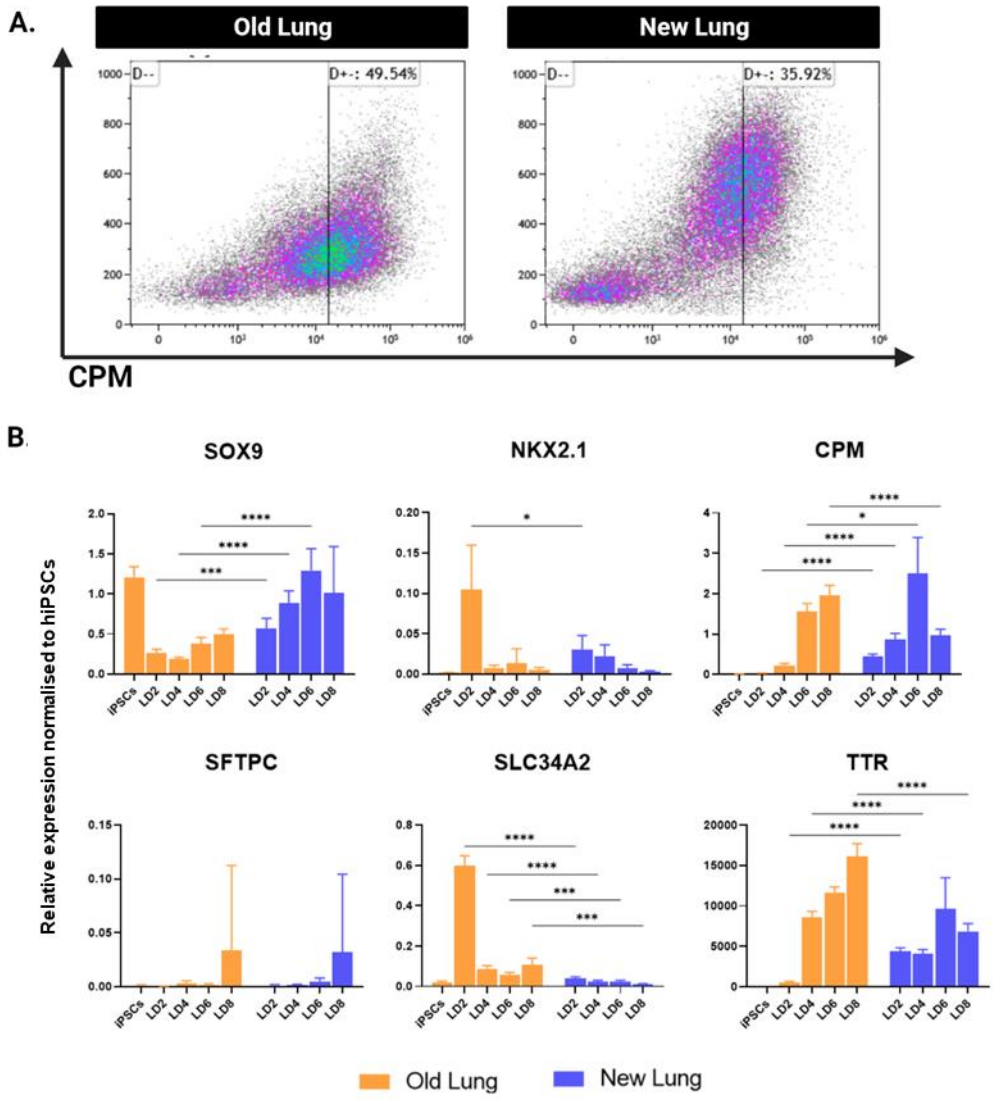


Figure 4.36 Healthy hiPSC line generates higher yields of CPM^{high} cells using original lung differentiation protocol.

A) Flow cytometry analysis of REBPLPAT lung day 8 cells showing 50% CPM^{high} cells using the old lung protocol and 36% using the new protocol. C) Comparison of lung markers' gene expression by qPCR in cells harvested on days 2, 4, 6 and 8 of the lung progenitors differentiation using the old and new protocols. Data shows mean CT values ± standard error of the mean. Statistical analysis for significance was performed in comparison to undifferentiated hiPSCs using one-way ANOVA. *p≤0.05. n=1.

Together, these results confirm the thorough optimisation of the lung progenitors' differentiation protocol was paramount to generate NKX2.1^{+} cells derived from the IPF-patient hiPSC lines generated in this thesis.

4.2.4 AT2 cells purification, maturation, and expansion

Transferring NKX2.1 cells to Matrigel droplets enhances the expression of distal lung markers specific to alveolar cells.

Our previous experiments showed that transferring our monolayered cultures containing lung progenitors into 3D Matrigel droplets improved expression of distal lung markers like SFTPC. However, it was also noted that other unwanted lineages were transferred along with the desired NKX2.1$^+$ cells. To circumvent this issue the approach was to enrich the lung progenitors using an anti-CPM antibody before inducing their maturation in 3D to increase the yields of SFTPC$^+$ cells. Organoids were generated from CPM$^+$ and CPM$^-$ cells sorted after 8 days of lung induction in 2D. However, it was discovered that even though CPM-positive (61% positivity obtained with DIFF1 conditions), the first cells sorted were not solely lung progenitors. This was due to the use of a suboptimal differentiation for the IPF-patient-derived cells. Nevertheless, when analysing the gene expression of the organoids derived from CPM$^+$ and CPM$^-$ populations, the SFTPC and SOX9 levels were higher in organoids from the positive population than in the negative population (**Figure 4.25**). This indicated that the 3D conditions specifically enhanced the production of SFTPC in CPM$^+$ cells.

NaPi2B FACS is suitable for isolating pure populations of SFTPC$^+$ cells.

In addition to the qPCR analysis, some of the organoids generated from CPM$^+$ cells differentiated using the unoptimised DIFF1 conditions were also processed for FACS using an anti-NaPi2b antibody (Cell Signaling) to enrich for the SFTPC$^+$ cells (**Figure 4.37A**). After 10 days of exposure to the 3D maturation conditions, NaPi2B$^+$ cells' yields were extremely low. Of the million cells stained on the first passage [P0], only 1.3% were gated NaPi2b positive and the cells recovered from the sorter were not enough for downstream analysis. Two days after splitting them for the first time, CHIR was removed from the maturation media for a 5-day, followed by an add-back period of 5 days. This step was implemented to slow down the proliferation of the alveolar cells and promote their maturation into SFTPC-producing cells [19]. Following the CHIR withdrawal and add-back, the rapid-proliferating organoids were processed again [P1] for NaPi2b sorting. Surprisingly, the percentage of cells that were positively stained for the surface marker had increased to 89.3%. This drastic change in percentages was attributed to the unsuitability of the antibody to be used for cell sorting and a different anti-NaPi2b antibody was ordered (NSJ Bioreagents).

The organoids were maintained in culture until the cells could be sorted using the new antibody. In passage 6 [P6], the experiment was repeated, and all the cells were stained to enrich for the presumable AT2 cells. Contrary to expectations, the percentage of NaPi2b+ cells dropped to 0.85%, allowing the recovery of only around 6,000 cells. This time the recovered cells were not seeded back into Matrigel droplets but instead were seeded into Matrigel-coated plates so they could be fixed and stained for SFTPC to evaluate the enrichment of AT2 cells. A comparison between the sorted NaPi2b negative and positive cells stained with an anti-SFTPC antibody is presented in **Figure 4.37B**. Both populations exhibited cells that were positively stained for SFTPC; however, the majority of cells in the NaPi2b^{+} group appeared to be strongly positive for SFTPC. This observation was confirmed through image analysis, where it was found that 96% of cells in the NaPi2B^{+} sample were positive for the exclusive AT2 cell marker, in contrast to 40% in the NaPi2b^{-} sample. Furthermore, the intensity of the SFTPC staining was noticeably higher in the NaPi2b positive cells, as demonstrated in **Figure 4.37C**. Together, these results confirmed that the introduction of a NaPi2b enrichment step to purify the SFTPC^{+} population within the organoids was a suitable approach to introduce to the differentiation protocol.

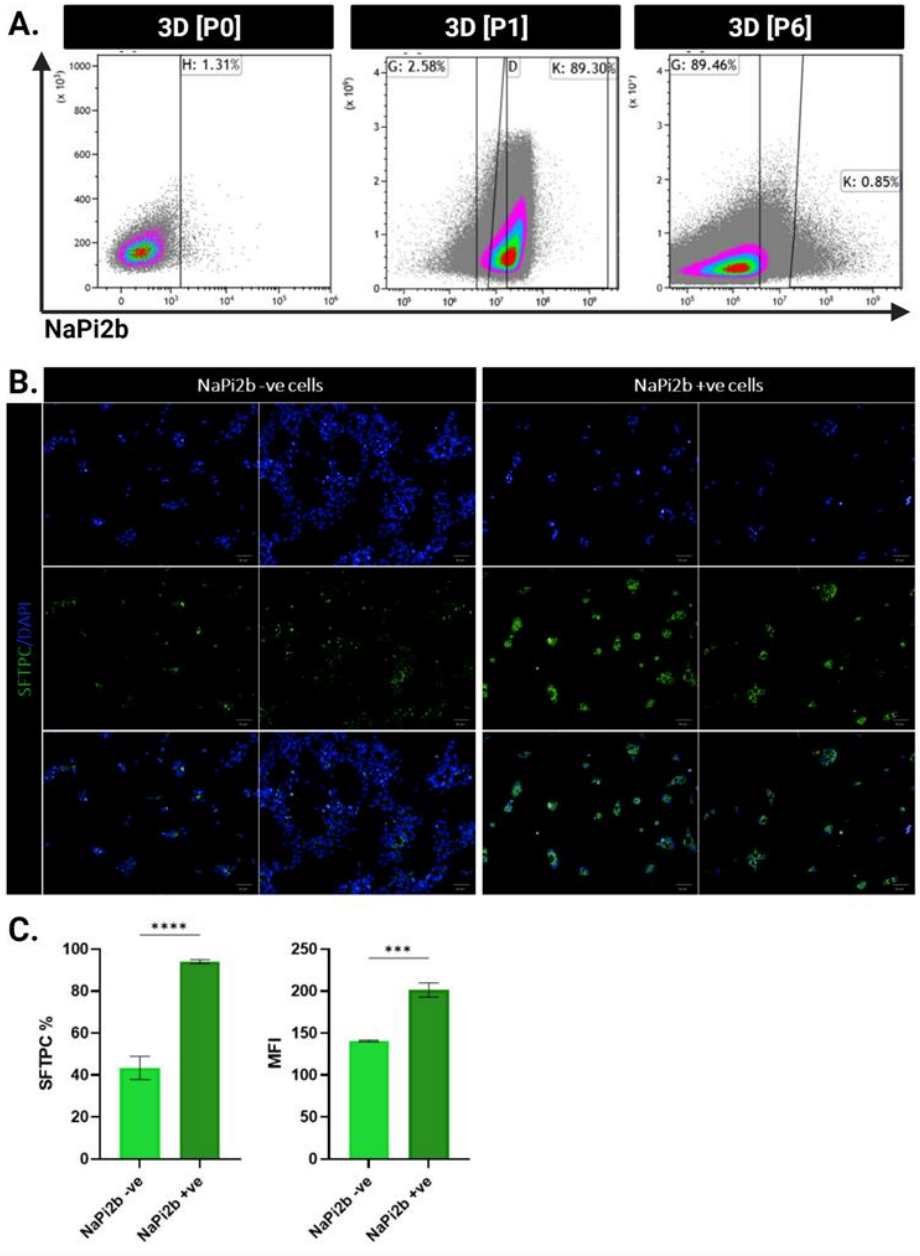


Figure 4.37 Enrichment of SFTPC⁺ cells using antibody against NaPi2b.

A) NaPi2b cell sorting of organoids generated from CPM⁺ single cells differentiated using the DIFF1 conditions. Cells were enriched on the first split [P0], seeded back into 3D Matrigel droplets to generate organoids and enriched again on the second [P1] and seventh [P7] splits. The dot plots show the established NaPi2b⁺ sorting gates at 1.3%, 89.3%, and 0. 85%, for P0, 1 and 6, respectively. B) Immunocytochemistry images showing sorted NaPi2b negative (left panel) and positive (right panel) cells derived from organoids maintained in culture for 6 passages. Sorted cells were collected, seeded on 2D Matrigel-coated plates, and stained for SFPTC (green). Nuclei were stained with DAPI (blue). Scale bars 50µm. C) Proportion of SFTPC⁺ cells in NaPi2b negative (-ve) and positive (+ve) sorted cells on the left and the Median Fluorescence Intensity (MFI) emitted by these two populations on the right. Graphs represent the average of three image fields analysed ± standard error of the mean , statistical analysis for significance was performed with a two-tailed unpaired Student's t-test. ****p≤0.0001.

Recognizing the potential of the two enrichment steps implemented thus far in improving the culture purity, the next objective was to enhance the percentages of AT2 cells obtained at the end, enabling their application for downstream analysis and experimental purposes. However, previous experiments using the NaPi2b antibody from a specific source yielded conflicting results. To address any potential issues related to antibody source variability, it was decided to employ the anti-NaPi2b MX35 antibody, which was used in the original study that pioneered the isolation of AT2 cells in vitro [8]. This antibody, initially provided by the Ludwig Institute for Cancer Research, NY, was not commercially available [227].

The previous NaPi2b sorting results were obtained from an experiment using lung progenitors differentiated with the original protocol; therefore, this needed to be repeated with the optimised DS/SB DIFF 2 conditions. However, before proceeding with sorting AT2 cells growing in 3D culture conditions, it was decided to sort for the AT2 cells already present at the end of the differentiation of lung progenitors. $SFTPC^{WT/WT}$ cells were plated at $1.5x10^4$ cells/cm^2, differentiated using the optimised endoderm C2, evaluated for the CXCR4/c-KIT co-staining on endoderm D4, obtaining 97% double-positive cells. Cells were then dissociated and seeded at a 1:3 ratio to continue with the DS/SB step, followed by the lung differentiation using DIFF2 culture conditions. On lung day 8, the cells were dissociated and stained using the MX35 antibody for FACS. A new analysis protocol was established in the sorter to suit the new antibody (**Figure 4.38**). Not surprisingly, the percentage of AT2-positive cells present in the monolayered cultures was only 2.95%. Notwithstanding, those cells had not been exposed to the maturation conditions known to help them thrive.

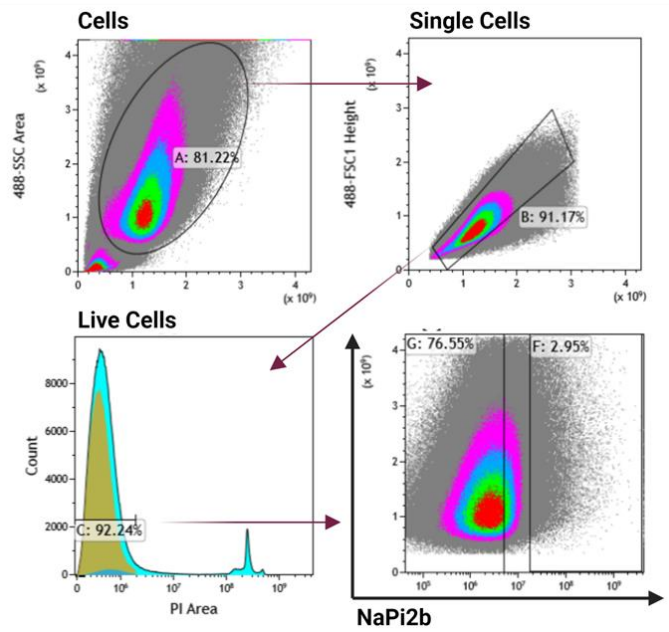


Figure 4.38 Sorting strategy to enrich for NaPi2b cells on day 8 of the lung differentiation protocol using DIFF2 conditions.

Representative flow cytometry of LD8 cells and the sorting gates used to exclude debris (Cells), exclude doublets (Single cells), include live cells (PI area), and select for NaPi2b^{+} cells. Gate for NaPi2b positivity (2.95%) was established against the unstained and secondary-only controls.

Maturation cocktail was essential for the long-term expansion of AT2 cells.

The utilization of a maturation media, supplemented with various combinations of glucocorticoids, cAMP, IBMX, and growth factors has been a widely adopted strategy for enhancing the expression of lung epithelial-specific genes, such as surfactant proteins [134, 228]. This specialized media helps maintain optimal levels of NKX2.1 and plays a vital role in supporting the differentiation and maturation of AT2 cells. Moreover, the incorporation of 3D cultures into the differentiation offers an added advantage by generating more complex and representative in vitro models, which in turn facilitates the production of functionally mature AT2 cells.

Based on previous studies where the isolation of human alveolar epithelial progenitors was achieved by sorting for EpCAM and HTII-280 co-expressing cells it was deemed appropriate to incorporate the use of an anti-EpCAM antibody alongside the new anti-NaPi2b antibody [36, 38]. This combined approach aimed to enhance the purity of sorted AT2 cells derived from our IPF-patient-derived hiPSCs organoid cultures. On LD8, differentiated SFTPC$^{WT/WT}$ cells were subjected to purification to isolate the CPM^{high} population from the cultures. This process resulted in a positivity of 23% and the recovery of approximately 5x10^{5} cells. These purified cells were then transferred into 3D droplets, with a density of 400 cells per µl of Matrigel, in the presence of DIFF2

medium supplemented with 3μM CHIR99021, 10ng/mL FGF10, 50nM Dexamethasone, 0.1mM cAMP, 0.1mM IBMX and 10ng/mL FGF7 10μM Y-27632 and 10% CloneR2. The single cells were allowed to expand and mature, forming presumable alveolar organoids, before the first round of NaPi2b/EpCAM sorting (**Figure 4.39A**). To maintain the population of SFTPC-expressing AT2s between passages, the enriched NaPi2b+/EpCAM+ cells were seeded back into 3D Matrigel droplets with a density of 200 cells/μl for maintenance and expansion. At the initial passage [P1], 8% of the cells were identified as NaPi2b^{+}/EpCAM^{+}. Subsequently, for the second sorting [P4], the percentage increased to 12%, for the third sorting [P7], it rose to 34%, and for the final sorting [P8], it reached 42% (**Figure 4.39B**).

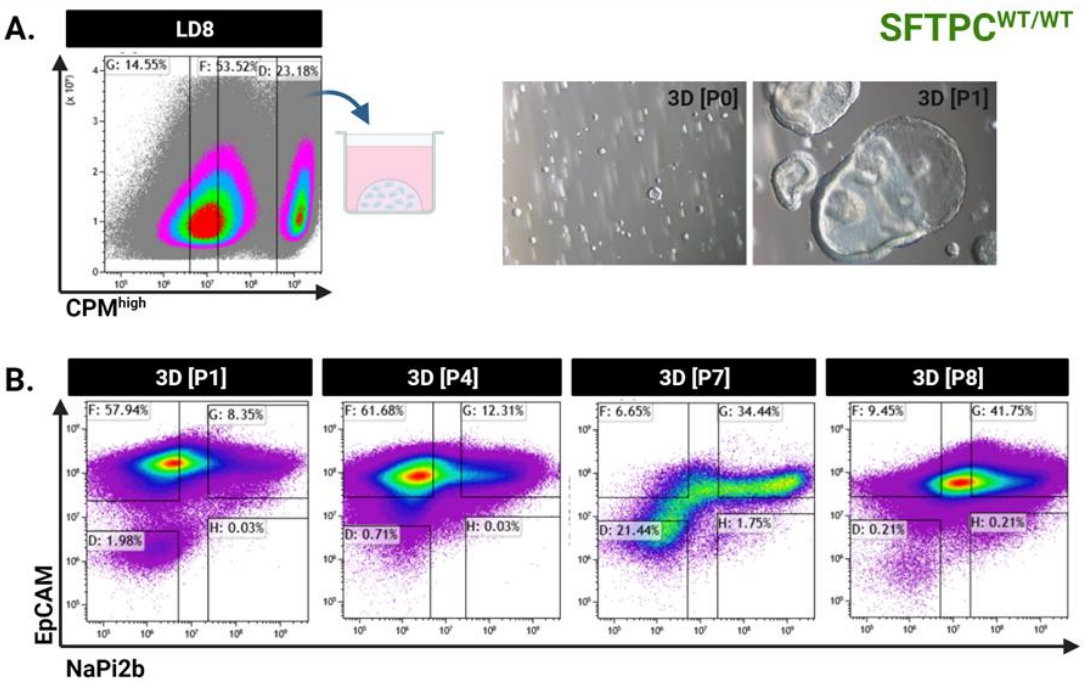


Figure 4.39 3D-maturation conditions are advantageous to the culture of NaPi2b^{+} cells.
A) SFTPC$^{WT/WT}$ hiPSCs differentiated and purified for NKX2.1 lung progenitors using antibody against CPM. The flow plot on the left displays the CPM^{high} population formed by LD8 cells constituting 23.1% of the cells. Diagram illustrates seeded cells in 3D-Matrigel droplets with maturation media. Brightfield images on the right show compacted spherical structures formed by the CPM cells on [P0], and how they grew into bigger cystic structures upon organoid passaging [P1]. B) FACS gates of CPM^{+} organoid-derived single-cells that were sorted for double positive EpCAM/NaPi2b on passages 1 (8%), 4 (12%), 7 (34%), and 8 (42%) as they were replated and expanded in 3D culture.

The SFTPC$^{WT/WT}$ [P4] alveolar organoids from above, which had been cultured in 3D for approximately 40 days, were dissociated for regular passaging without any enrichment steps. A portion of the presumed AT2 cells was set aside for

immunostaining. The cells were briefly seeded on a Matrigel-coated culture plate before fixation for staining and imaging, as shown in **Figure 4.40**.

Interestingly, all cells within the organoids exhibited positive staining for the epithelial cell marker ECAD and for SFTPC while the distribution of the surface marker NaPi2b positive cells was limited. Additionally, even after an extended period in 3D culture, the cells continued to express the surrogate surface marker for NKX2.1, CPM. This observation was supported by quantitative analysis of the immunofluorescent images, revealing that 98% of the cells were positive for SFTPC, while only 47% showed positive staining for NaPi2b. Notably, near the totality of the NKX2.1^{+} cells were also positive for CPM, further validating the expression of this marker in the cultured cells.

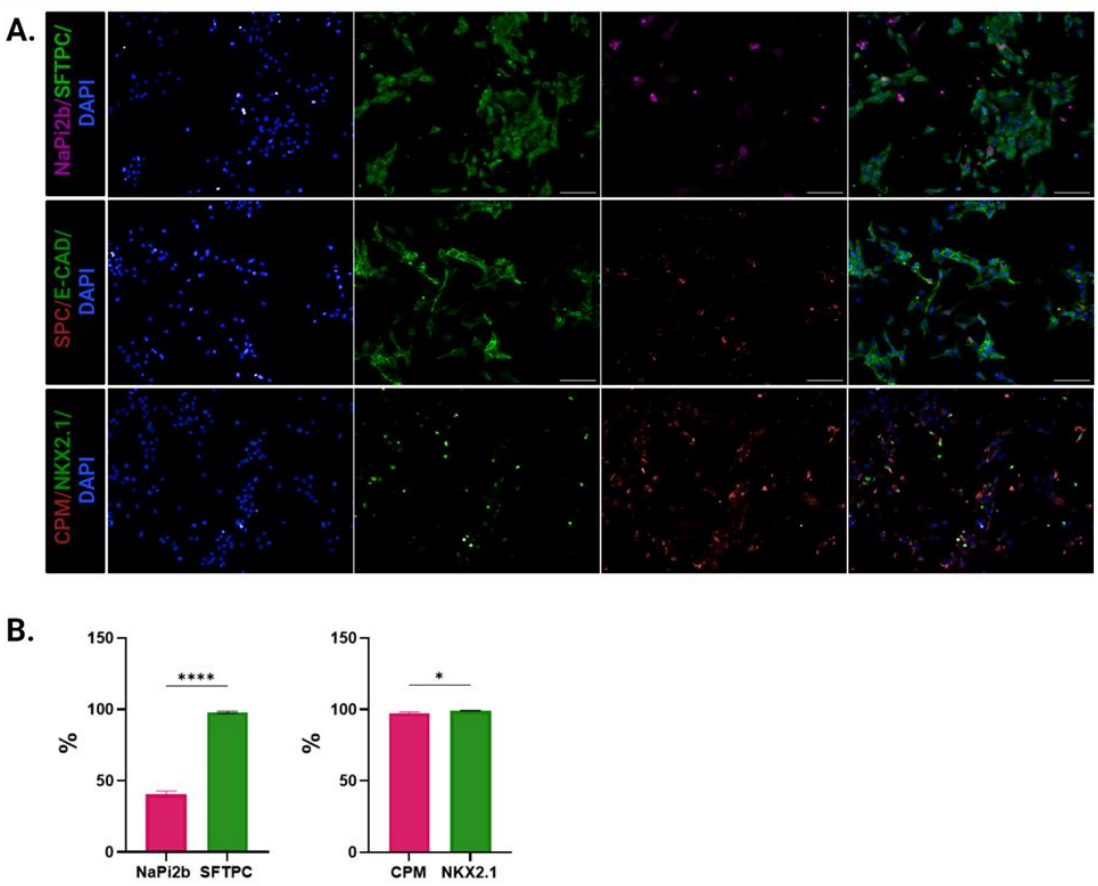


Figure 4.40 Incorporation of CPM and NaPi2b enrichment steps aid maintaining AT2 cells phenotype.

A) Representative immunofluorescence microscopy of dissociated SFTPC$^{WT/WT}$ passage 4 organoids co-stained with antibodies against: top row= NaPi2b (red) and SFTPC (green), middle row= SPC (red) and ECAD (green), and bottom row= CPM (red) and NKX2.1 (green). Nuclei were stained with DAPI (blue). Scale bars, 50 µm. B) Graph on the left represents the percentage of NaPi2b^{+} and SFTPC^{+} cells per total DAPI (representative images in top row from A) and the graph on the right, the percentage of CPM^{+} and NKX2.1^{+} cells per total DAPI (representative images in bottom row from A). Statistical analysis for significance was

performed with a two-tailed unpaired Student's t-test. ****p≤0.0001.*p≤0.05 and ****p≤0.0001.

The observed decrease in the expression of NaPi2b when the cells were seeded in 2D culture could explain the need for frequent enrichment steps. This decline in expression does not necessarily imply a loss of the AT2 cell phenotype but may indicate that not all AT2 cells express the surrogate surface marker. It has been observed that hiPSC-derived AT2 cells in vitro exhibit a heterogeneous and dynamic nature, characterized by varying proliferative potentials and maturation states. Sun et al., (2021), demonstrated this heterogeneity by separating AT2 cells based on their SFTPC expression levels [220]. Using their SFTPCtdTomato and ABCA3:GFP reporters, they revealed that the GFP single-positive cells exhibited higher levels of progenitor markers and lower levels of SFTPC expression. On the other hand, the tdTomato-positive population selectively represented cells with higher levels of SFTPC expression.

To assess the heterogeneity of our organoid cultures and determine if a population of AT2 cells was positive for SFTPC but not detected during NaPi2b enrichment, AT2 organoids were processed as for normal NaPi2b cell sorting. Subsequently, the cells were fixed and stained with intracellular SFTPC for analysis using the ImageStream flow cytometer (**Figure 4.41**). The results obtained from the ImageStream analysis revealed that almost all of the processed cells showed positive staining for SFTPC (97%). In contrast, only a small fraction (8%) of this SFTPC-positive population also exhibited NaPi2b positivity. While NaPi2b enrichment enables the purification of AT2 cells with high SFTPC expression, it discards the remaining alveolar cells in the organoid culture that also express SFTPC, likely representing a more immature cell population. These findings are consistent with the observations reported by Sun et al. and highlight the importance of considering the heterogeneous nature of the cell population and the potential limitations of relying on purification steps.

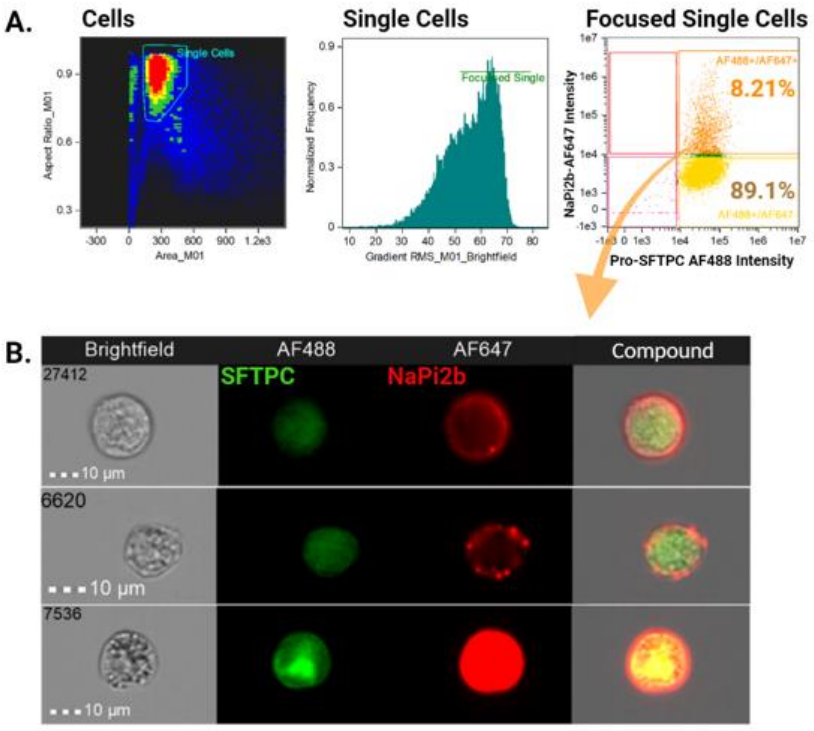


Figure 4.41 Enrichment of NaPi2b^{+} cells favour a small proportion of SFTPC^{+} cells.

A) Image Stream flow cytometry plots showing the gating strategy to select for single and focused single cells, then to select the co-stained for SFTPC and NaPi2b. B) Shows three representative composites of single cells co-stained for SFTPC (green) and its surrogate cell surface marker, NaPi2b (red). Grayscale brightfield images of each cell are shown to the left of the composite. Scale bars, 10µm.

Gene expression analysis was performed using qPCR to further characterise the enriched AT2 cells obtained from the 3D organoid cultures (**Figure 4.42**). A comparison was made between NaPi2b cells that were maintained in 3D culture prior to sorting and the sorted NaPi2b positive and negative cell populations. The gene expression analysis results supported the enrichment strategy's efficacy using NaPi2b as a surrogate surface marker for SFTPC-expressing cells. The NaPi2b positive cells demonstrated significantly higher levels of NKX2.1 and SFTPC gene expression, indicating their AT2 cell identity. Importantly, the enrichment process using NaPi2b also significantly reduced the expression of contaminant hindgut lineages, which were consistently observed in our previous cultures and expressed markers such as CDX2, VILLIN, and LGR5. These findings highlight the effectiveness of the NaPi2b-based enrichment method in obtaining a purer population of AT2 cells for further analysis in this thesis.

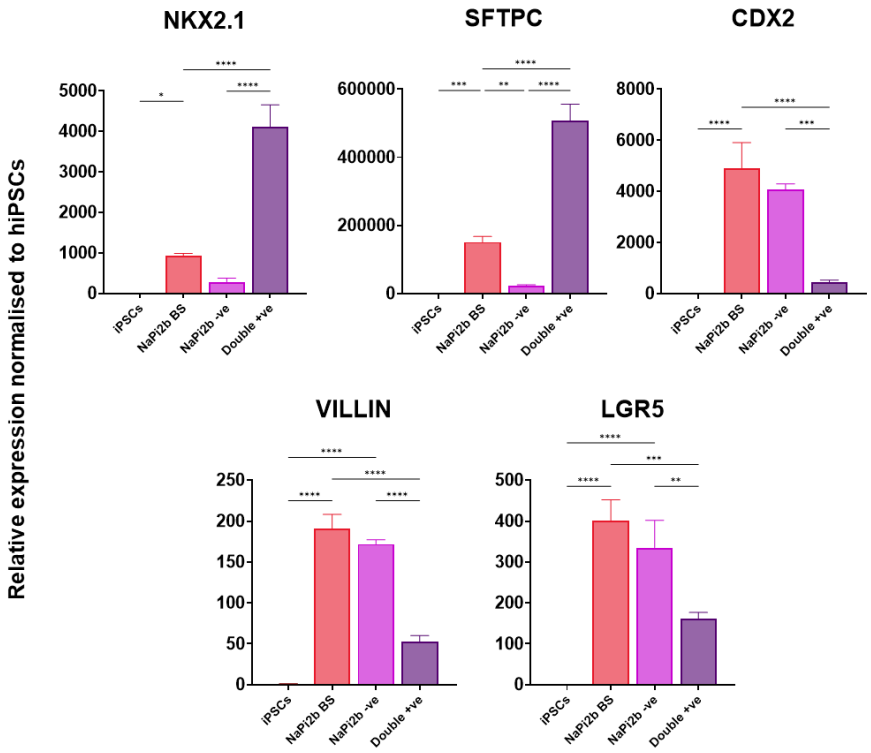


Figure 4.42 Thrive of other cell lineages is significantly reduced by AT2 cells enrichment strategy.

Gene expression of AT2 (NKX2.1, SFTPC) and intestinal (CDX2, VILLIN, LGR5) cell markers by qPCR in single cells derived from alveolar organoids sorted for NaPI2b. NaPI2b BS (Before Sorting), NaPI2b -ve (Negative), Doble +ve (Double positive). Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to undifferentiated hiPSCs. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Having established that the NaPi2b negative cells within the organoids were also positive for SFTPC, it was important to investigate whether these cells could acquire NaPi2b expression with extended maturation in 3D culture conditions. To address this, cells from both EPCAM+/NaPi2b+ and EPCAM+/NaPi2b- populations were sorted and reseeded into 3D droplets, then cultured under maturation conditions for two passages (**Figure 4.43**). This experiment was conducted using SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ hiPSC-derived AT2 organoids. Notably, significant morphological differences were observed between the two groups, with the NaPi2b- organoids appearing more compact and exhibiting lower confluence compared to those derived from NaPi2b+ cells, despite being seeded at the same density.

Furthermore, EpCAM/NaPi2b sorting results in both cultures yielded surprising findings. In the case of SFTPC$^{WT/WT,}$ the percentage of double-positive cells derived

from the NaPi2b+ population was 42%, whereas it was 47% for cells derived from the NaPi2b- population. However, the results for SFTPC$^{MUT/MUT}$ were less remarkable, with 49% double-positive cells obtained from the NaPi2b+ population and only 2% from the NaPi2b- population. Nevertheless, regardless of the percentages, these results demonstrate that the maturation culture conditions can promote the emergence of highly expressing SFTPC-AT2 cells from a previously sorted NaPi2b- population.

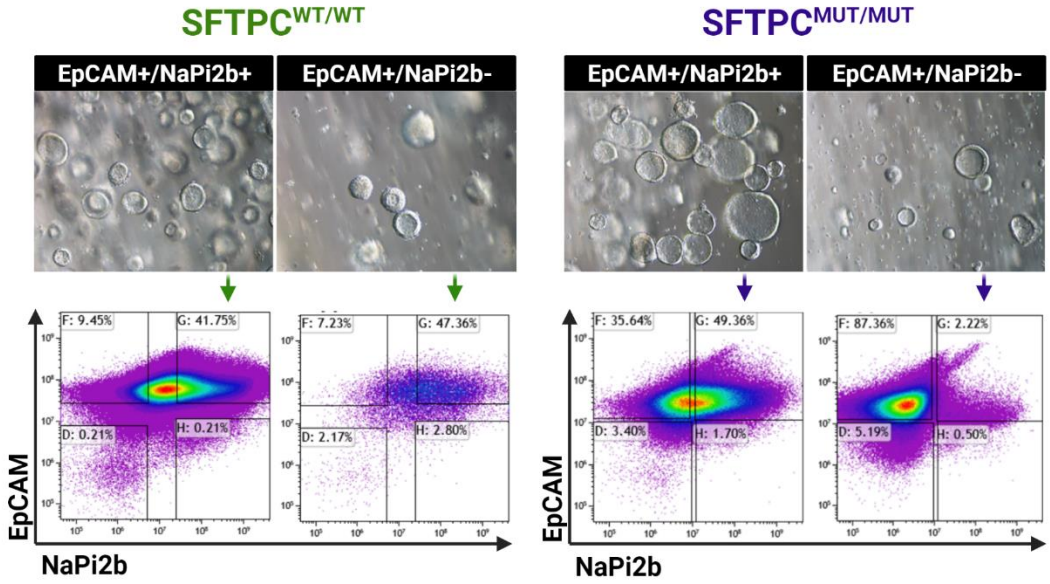


Figure 4.43 NaPi2b negative population can give rise to NaPi2b positive cells.
Brightfield images of SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ cells sorted for EpCAM/NaPi2b double positive and EpCAM positive/NaPi2b negative cells. The flow cytometry plots at the bottom show the percentages of EpCAM/NaPi2b cells derived from re-sorting these cells after expansion in 3D. Notice the emergence of NaPi2b positive cells derived from EpCAM positive/NaPi2b negative population.

Throughout this thesis, the entire differentiation protocol was specifically optimized for the SFTPC$^{WT/WT}$ cell line. Despite the fact that both the SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ cell lines were derived from the parental IPF-patient-derived hiPSC line (SFTPC$^{WT/MUT}$), variations in their differentiation efficiencies were expected. To assess the extent of these variations, all three hiPSC lines were differentiated toward AT2 cells using the same protocol and conditions. **Figure 4.44**, **Figure 4.45** and **Figure 4.46** depict the cell sorting gates applied to each cell line to enrich for the CPM^{high} lung progenitors on day 8 of the lung differentiation process. The percentages of CPM^{high} cells obtained were 1.53% for SFTPC$^{WT/WT}$, 0.44% for SFTPC$^{MUT/MUT}$, and 0.53% for SFTPC$^{WT/MUT}$. Although the values were not particularly high for any of the cell lines, the SFTPC$^{WT/WT}$ line exhibited the highest percentage of CPM^{high} cells. Morphologically, no discernible

differences were observed among the three cell lines when the sorted single cells were transferred to 3D-Matrigel droplets and allowed to proliferate and form organoids.

When the Matrigel droplets reached confluency, the organoids were dissociated, and the single cells were subjected to double staining for EpCAM/NaPi2b sorting. Surprisingly, the expected increase in the percentage of double-positive cells was not observed, as the cells were maintained and passaged in 3D culture. In the SFTPC$^{WT/WT}$ cell line, the percentage of EpCAM^{+}/NaPi2b^{+} cells decreased from 4.52% to 1.3%. The homozygous mutant cell line initially showed 4.4% EpCAM^{+}/NaPi2b^{+} cells, which then decreased to 3.67% before rebounding to 49.36%. Unfortunately, for the SFTPC$^{WT/MUT}$ cell line, only 1.75% of EpCAM^{+}/NaPi2b^{+} cells were recovered before the cultures became accidentally contaminated, preventing further analysis.

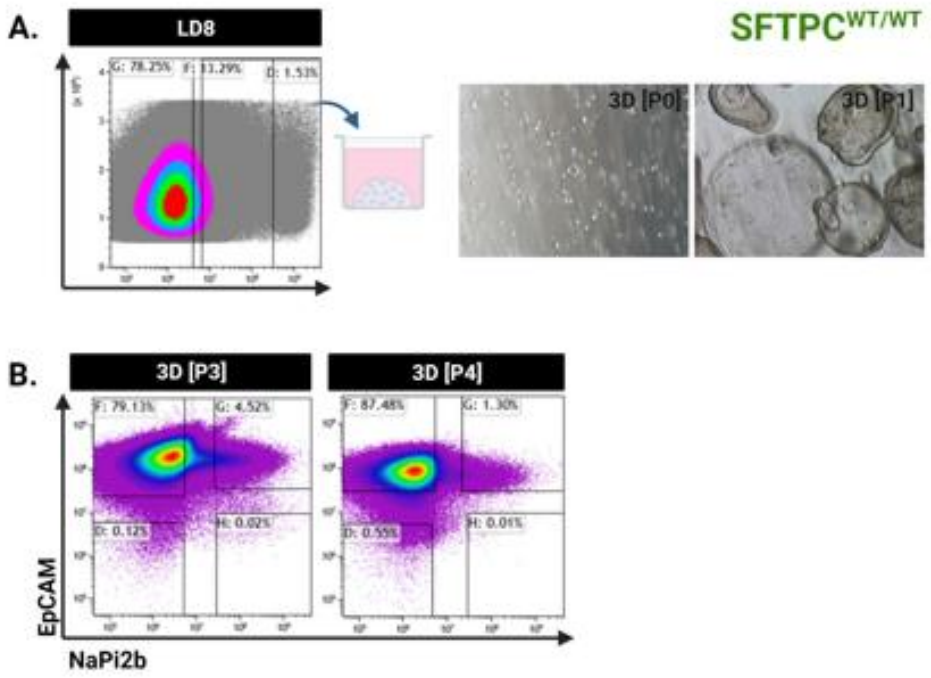


Figure 4.44 AT2cells can be derived from the IPF-patient-derived SFTPC$^{WT/WT}$ hiPSCs.
A) Flow cytometry plot of CPM^{high} cells sorted on day 8 of the lung progenitors' differentiation. Enriched cells seeded in 3D-Matrigel droplets are shown in the brightfield images on the right at passages 0 [P0], and 1 [P1]. B) CPM^{high} cells sorted in A were expanded in 3D culture and enriched for double positive EpCAM/NaPi2b cells on passages 3 (5%) and 4 (1.3%).

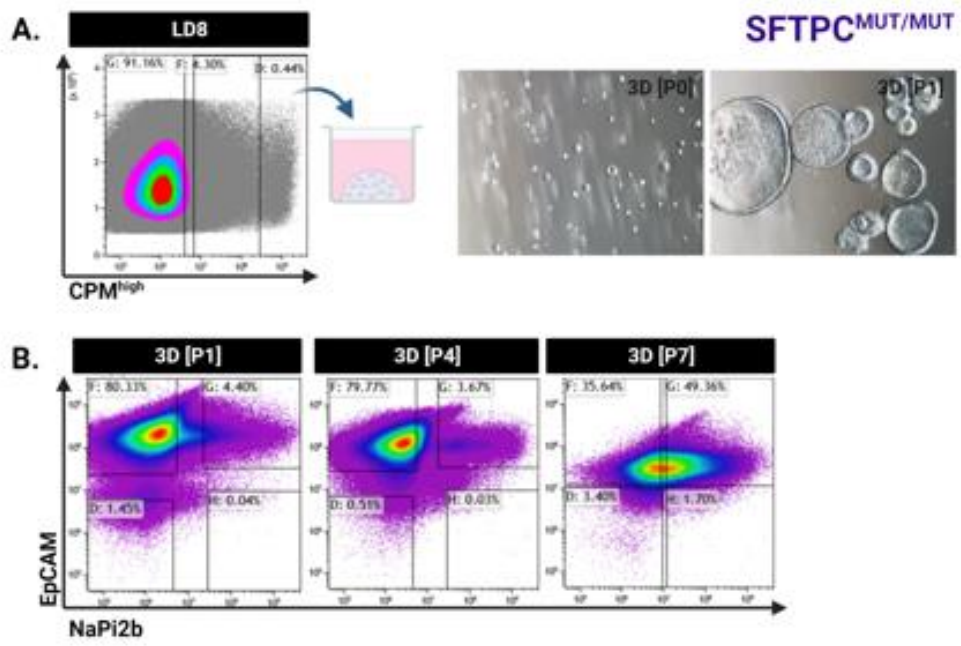


Figure 4.45 AT2cells can be derived from the IPF-patient-derived SFTPC$^{MUT/MUT}$ hiPSCs.

A) Flow cytometry plot of CPM^{high} cells sorted on day 8 of the lung progenitors' differentiation. Enriched cells seeded in 3D-Matrigel droplets are shown in the brightfield images on the right at passages 0 [P0], and 1 [P1]. B) CPM^{high} cells sorted in A were expanded in 3D culture and enriched for double positive EpCAM/NaPi2b cells on passages 1 (4%), 5 (4%) and 7 (50%).

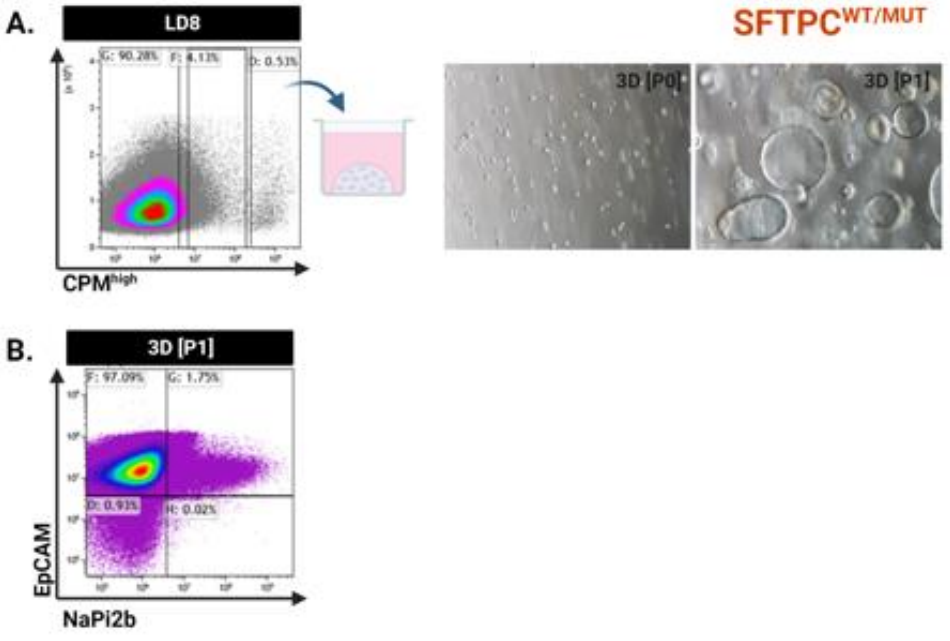


Figure 4.46 AT2cells can be derived from the IPF-patient-derived SFTPC$^{WT/MUT}$ hiPSCs.

A) Flow cytometry plot of CPM^{high} cells sorted on day 8 of the lung progenitors' differentiation. Enriched cells seeded in 3D-Matrigel droplets are shown in the brightfield images on the right at passages 0 [P0], and 1 [P1]. B) CPM^{high} cells sorted in A were expanded in 3D culture and

enriched for double positive EpCAM/NaPi2b cells after one passage, obtaining 2% positive cells.

To corroborate the phenotype obtained with the homozygous mutant cell line, SFTPC$^{MUT/MUT}$ [P5] alveolar organoids were dissociated for regular passaging in 3D culture without sorting for NaPi2b cells, and a subset of AT2 cells was seeded in 2D for immunostaining. Once the cells attached to the culture plate, they were fixed and subjected to immunocytochemistry staining for imaging. The panel in **Figure 4.47** demonstrates the positive co-staining of SFTPC with its surrogate surface marker NaPi2b. This surfactant protein was also co stained with the epithelial marker E-cadherin. Similar to SFTPC$^{WT/WT}$, all of the cells that stained positive for NKX2.1 were also positive for CPM.

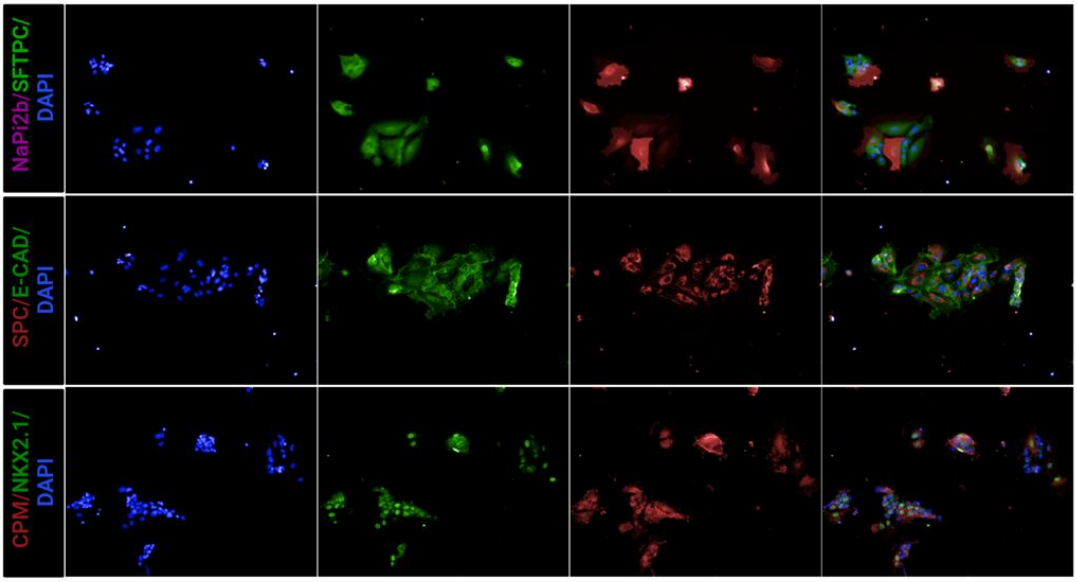


Figure 4.47 Enrichment steps also help maintain SFTPC mutant-AT2 cells.

A) Representative immunofluorescence microscopy of dissociated SFTPC$^{MUT/MUT}$ passage 5 organoids co-stained with antibodies against: top row= NaPi2b (red) and SFTPC (green), middle row= SPC (red) and ECAD (green), and bottom row= CPM (red) and NKX2.1 (green). Nuclei were stained with DAPI (blue). Scale bars, 50 µm.

The efficiencies obtained in this simultaneous experiment for sorted CPM^{high} and EpCAM/NaPi2b populations using the patient-derived hiPSCs were lower than envisioned; however, the results proved that the differentiation platform also work with the homozygous and heterozygous mutant cell lines. In future works, minor optimisation to tailor the platform to these two cell lines and equalise the efficiency of the three IPF-patient derived hiPSCs could be beneficial.

Taken together, these promising findings validate the thoroughly optimised differentiation platform, incorporating the enrichment steps and 3D-maturation conditions. This approach yielded an expandable population of AT2 cells, providing a valuable tool for disease modelling. Therefore, this protocol was employed for subsequent characterisation and disease modelling experiments in this thesis.

4.2.5 Optimised protocol generates robust AT2 cells.

After meticulous optimisation of each stage of the differentiation process, a refined protocol was established specifically for IPF-patient-derived hiPSCs. The finalised protocol consisted of a hybrid approach, blending monolayered and 3D culture conditions. In brief, hiPSCs were seeded, and 24hrs later, the differentiation started with a four-day definitive endoderm induction using DIFF1 basal media supplemented with 3μM CHIR, 100ng/ml ActA and 10μM LY94002. Upon confirmation of key endodermal markers, CXCR4 and c-KIT, via flow cytometry analysis, the cultures underwent dissociation, and cells were reseeded at a 1:3 ratio in DIFF2 basal media supplemented with 2μM Dorsomorphin and 10μM SB434542. Following the anteriorisation step, the lung progenitor induction was performed using DIFF 2 basal media supplemented with 3μM CHIR99021, 10ng/mL BMP4 and 100nM RA. After eight days, the NKX2.1^{+} lung progenitors were isolated from the cultures using an anti-CPM antibody and subsequently seeded in 3D-Matrigel droplets in maturation media (DIFF2 basal media supplemented with 3μM CHIR99021, 10ng/mL FGF10, 50nM Dexamethasone, 0.1mM cAMP, 0.1mM IBMX and 10ng/mL FGF7) with 10μM Y-27632 and 10% CloneR2. As the single cells proliferated and formed abundant organoids, the droplets were allowed to reach confluency. To enrich the desired SFTPC-expressing AT2 cells, the dissociated organoids were sorted using an anti-NaPi2b antibody.

To better understand the transcriptomic profiling of the AT2 cells produced with the optimised differentiation protocol using hiPSCs derived from an IPF patient, mRNA sequencing was performed to identify differentially expressed genes among samples at crucial stages of the differentiation process. Additionally, a comparison was made between the expression profiles of the AT2 cells generated in this thesis, those of AT2 cells generated by another group, and those of adult lung and foetal samples to provide a broader context for the analysis [19, 229]. To facilitate this analysis, the widely used hiPSCs line SFTPC$^{WT/WT}$ and the isogenic mutant hiPSCs lines SFTPC$^{WT/MUT}$ and SFTPC$^{MUT/MUT}$ were simultaneously differentiated into AT2 cells. The samples sent for sequencing included the undifferentiated hiPSCs, definitive endoderm day 4 cells,

unsorted lung day 8 lung progenitors and CPM/NaPi2 sorted populations (**Figure 4.48A**).

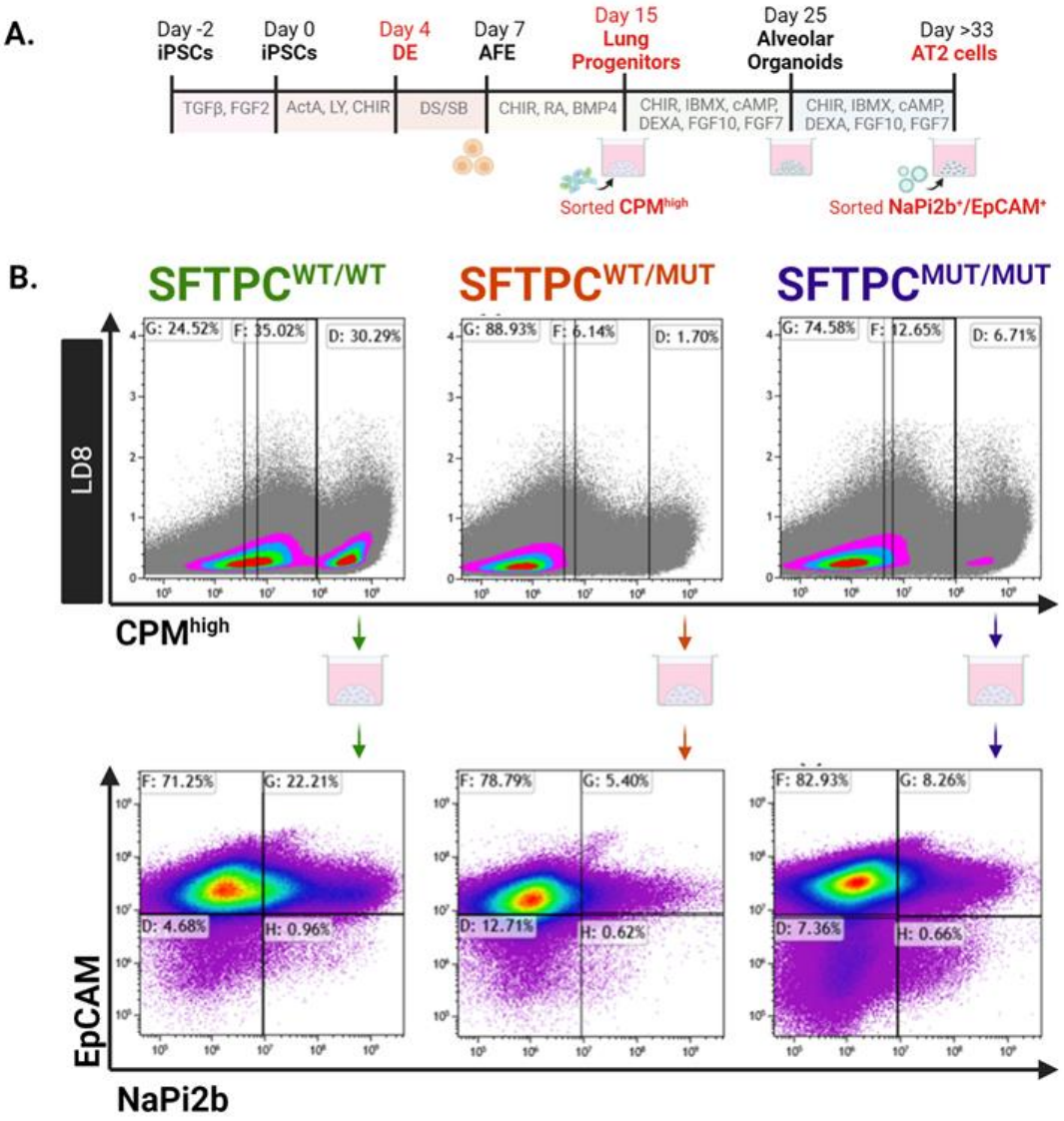


Figure 4.48 Directed differentiation protocol sequentially recapitulates embryonic lung development to produce AT2 cells.

A) Schematic representation of the in vitro differentiation of human iPSCs into AT2s, highlighting in red the key stages at which cells were collected for RNA-seq analysis. B) Flow cytometry plots of the sorted samples sent for RNA-seq. Plots on the top row show the percentages of CPM^{high} lung progenitors obtained on day 8, and the bottom plots show the percentages of EpCAM/NaPi2b double positive cells obtained after culturing the progenitors in 3D conditions.

Notably, the differentiation efficiencies achieved with each of the three cell lines varied significantly. When differentiated to lung progenitors in 2D and sorted for CPM, the SFTPC$^{WT/WT}$ line yielded 30% CPM^{high} cells, while the SFTPC$^{WT/MUT}$ and SFTPC$^{MUT/MUT}$ lines produced only 2% and 7% CPM^{high} cells, respectively. A proportion of the sorted

cells was used for RNAseq purposes, and the rest was seeded back into culture to continue with the differentiation protocol. Once transferred to 3D droplets for their maturation into AT2 cells, the percentages of $EpCAM^{+}/NaPi2b^{+}$ cells were 22% for $SFTPC^{WT/WT}$, 5% from $SFTPC^{WT/MUT}$, and 8% from $SFTPC^{MUT/MUT}$. These observed differences in the yields produced by each cell line were anticipated since the differentiation protocol was optimised specifically for the wild-type cell line (**Figure 4.48B**).

RNA Sequencing

Transcriptome analysis was performed using RNA samples isolated from undifferentiated $SFTPC^{WT/WT}$ hiPSCs and cells collected at each time point indicated in red in **Figure 4.48A**. This analysis included both positive and negative sorted populations. To assess the variation among the different samples throughout the AT2 differentiation protocol using $SFTPC^{WT/WT}$ hiPSCs, principal component analysis (PCA) was conducted. The results showed that the most significant variation in gene expression was observed between clusters comprising undifferentiated hiPSCs and Definitive Endoderm (DED4) samples and the remaining lung progenitors and AT2 differentiated samples. Notably, the lung progenitors' samples, including pre-sorted and enriched, formed a distinct cluster separated from the AT2 NaPi2b positive and negative samples, as illustrated in **Figure 4.49A**. The next PCA assessed the primary separation of gene expression patterns between NaPi2b positive (NaPi2b_pos) and negative (NaPi2b_neg) samples derived from SFTPC wild-type and mutant cell lines. The NaPi2b_neg samples from the three cell lines were found to cluster separately from the positive samples, displaying less variation among the homozygous cell lines. Conversely, the positive samples exhibited greater variance among themselves without forming distinct clusters. This suggests that the highly expressing SFTPC AT2 cells derived from the three cell lines are less closely related to each other than the rest of the AT2s. An additional PCA examined the correlation between primary foetal samples at different gestation weeks (16, 17.5, 20, and 21), human adult AT2 cells and total lung tissue, and the putative AT2 cells differentiated from hiPSCs (NaPi2b_Pos) [19, 229]. Surprisingly, the isolated adult AT2 cells displayed the highest variance and did not cluster with any of the other included samples. On the other hand, the other adult samples (total lung) showed a strong association with the NaPi2b_Pos cells. Furthermore, the four foetal samples did not form a distinct cluster but were relatively closer to each other than to the remaining samples. These findings suggest that the

differentiated AT2 cells resemble the adult lung more closely than the foetal lung.

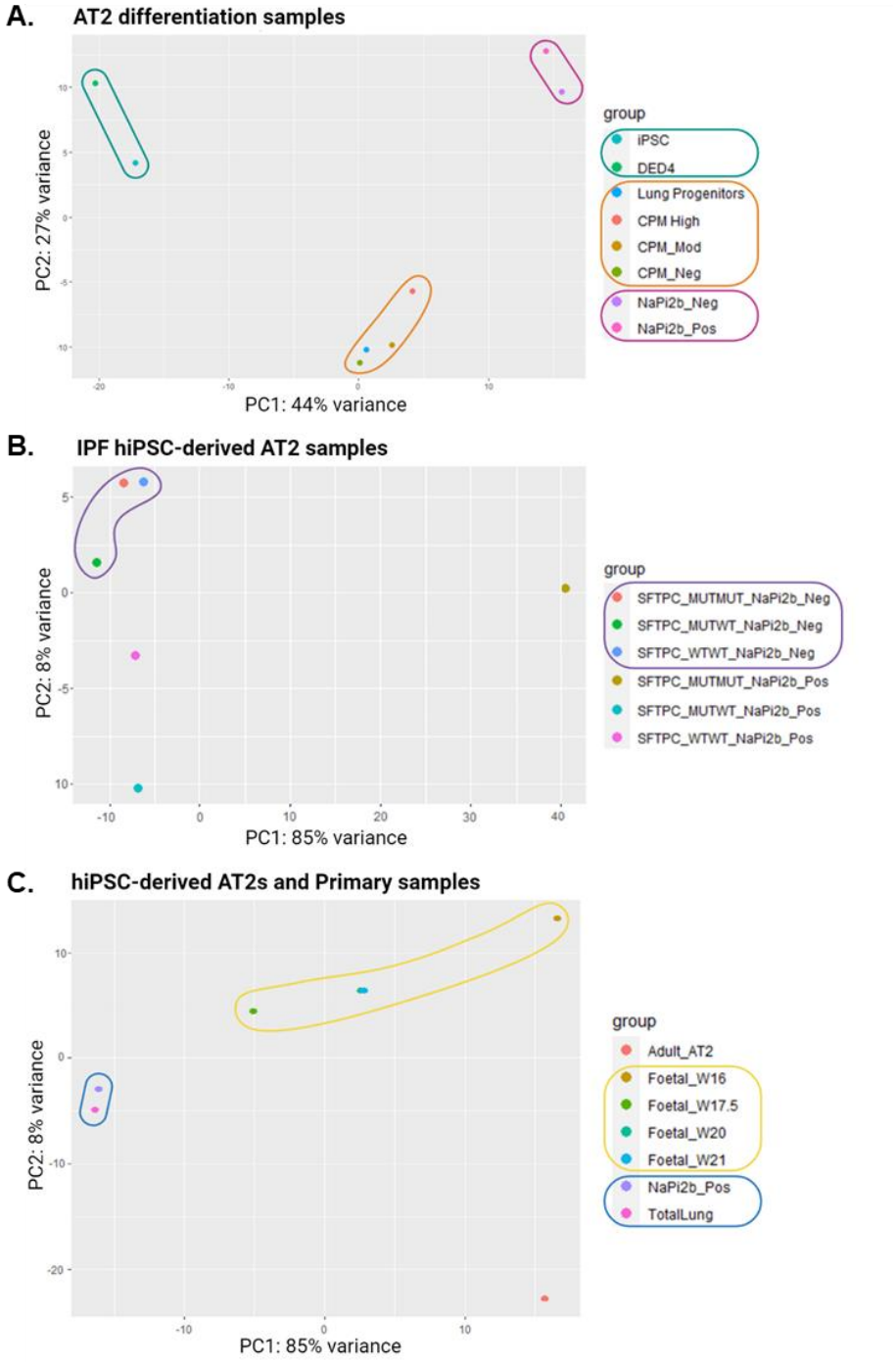


Figure 4.49 Transcriptomic profiling of directed differentiated hiPSC-derived AT2 cells.

Principal component analysis (PCA) plots show the primary separation of samples by gene expression. A) Gene expression across key stages of the directed differentiation protocol using SFTPC$^{WT/WT}$ hiPSCs (DED4- definitive endoderm, Lung progenitors- pre-sorted cells on day 15, CPM_High, Mod, Neg- sorted cells on day 15, NaPi2b_neg, pos- sorted cells on day >33). B) Gene expression across NaPi2b positive and negative cells generated from SFTPC$^{MUT/MUT}$, SFTPC$^{MUT/WT}$ and SFTPC$^{WT/WT}$ cell lines. C) Gene expression across hiPSC-derived AT2s

(NaPi2b_Pos), foetal lung tissue at different weeks of gestation (W16, W17.5, W20, W21), adult AT2s (Adult_AT2) and adult Total lung tissue.

To better evaluate the changes that occur during the differentiation process and if they correlate to the ones observed during development, the gene expression patterns between samples from key time points of the differentiation were compared. The transcriptomic data set was tested for differentially expressed genes (DEGs) between samples using the *limma* package from Bioconductor. DEGs were filtered based on criteria including a normalised Log2 Fold Change of mRNA expression >0.5 and a false discovery rate (FDR) <0.05. The first comparison was between day 4 definitive endoderm (DE D4) and lung progenitors' samples. The heatmaps in **Figure 4.50** show the clear upregulation of definitive endoderm markers SOX17, GATA6, GATA4 and FOXA2 in the DE D4 sample, whereas the main DEGs upregulated in the Lung Progenitors sample were SOX9, SOX2, SHH, PDGFC, NKX2.1, FGF10 and BMP4. These transcriptional signatures align with the differentiation of epithelial lung progenitors resulting from the epithelial-mesenchymal crosstalk during the branching morphogenesis of the lung primordium [101, 230, 231]. When compared to the unenriched lung progenitors, the transcriptome profile of the enriched CPM^{high} cells revealed a distinct set of upregulated transcription factors including NKX2.1, EPCAM, NAPSA, LPCAT1 and LAMP3. The upregulation of NAPSA and LPCAT1, which are involved in the processing and metabolism of surfactant components, as well as LAMP3, a marker associated with lamellar bodies, suggest that the enriched CPM^{high} cells represent a relatively more mature population of lung progenitors associated with AT2 cell fate [103, 232].

To interrogate the differences between CPM^{high} and $EpCAM^{+}/NaPi2b^{+}$ sorted cells ($NaPi2b^{+}$), the analysis focused on identifying transcripts associated with the differentiation and maturation of AT2 cells. Among the top DEGs between these samples, markers known to be present in primary adult differentiating AT2 cells were highly upregulated in the $NaPi2b^{+}$ population. These included surfactant protein genes SFTPB and SFTPC, as well as SLC34A2, NAPSA, CEBPD, and CLDN18 [233]. Furthermore, the $NaPi2b^{+}$ population exhibited significant upregulation of additional DEGs associated with AT2 cell maturation, such as SFTPA2, SFTPA1, PGC, LYZ, and CXCL5. On the other hand, DEGs including MYCN, NKD2, and BAMBI showed downregulation with maturation. These observations further emphasise the marked differences between the $NKX2.1^{+}$ lung progenitors and the differentiated AT2 cells expressing mature markers. The upregulation of specific genes associated with

surfactant proteins and maturation markers further supports the notion that the NaPi2b^{+} population represents a differentiated state of surfactant producing AT2 cells.

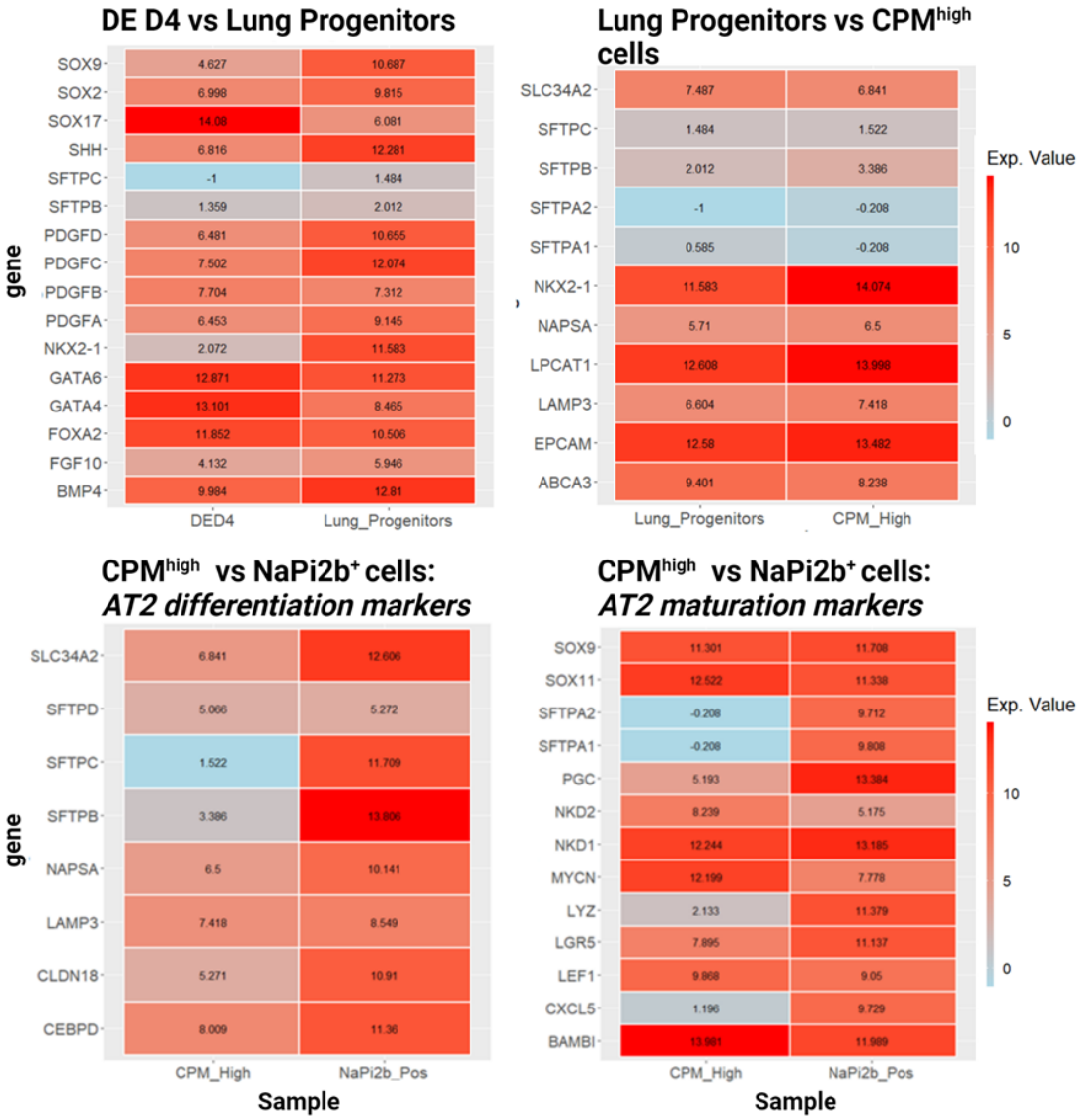


Figure 4.50 Differentially expressed genes between all sample groups throughout the differentiation.

Heatmaps comparing relative gene expression of differentially expressed genes in cell populations at different stages of the differentiation (DE D4- definitive endoderm, Lung progenitors- pre-sorted cells on day 15, CPM^{high}- sorted cells on day 15, NaPi2b^{+} sorted cells on day >33) (FDR<0.05).

To investigate whether enriching for CPM^{high} cells favoured other lung epithelial lineages the DEGs between lung progenitors' samples and CPM^{high} cells were studied for transcripts associated to AT1, basal, multiciliated, secretory and goblet cell markers. Heatmaps in **Figure 4.51** illustrate the comparisons of these gene expression profiles.

Among the top DEGs between lung progenitors and CPM^{high} cells, the AT1 cell-associated genes more highly expressed in both populations were TSPAN13 and COL4A1. The genes, together with AQP5, AGER and HOPX were upregulated in the CPM^{high} sample; however, no major differences in expression were identified. On the other hand, the AT1 markers, PDPN and CAV1 were slightly upregulated in the lung progenitors' sample. When analysing DEGs related to basal cells in both samples, the well-known basal markers TP63 and KRT5 were found to be downregulated in the CPM^{high} population. However, when considering markers associated with a recently described lung basal resting cell, such as KRT19, IGFBP2, CYR61a, and BCAM, they were upregulated in the enriched CPM^{high} population [234]. To be noted, the transcript levels of these markers were also high in the lung progenitors' population, indicating that they are not exclusive to the CPM^{high} population. Contrastingly, among these markers, PERP and KRT17 were found to be downregulated in the CPM^{high} sample. In terms of multiciliated cell markers, both populations expressed similar levels for most markers. FOXJ1 and TUBB3 showed upregulation in the CPM^{high} sample and CD24, which was the most highly expressed in both samples. Regarding secretory cell markers, SCGB1A1 and SCGB3A2, were expressed at low levels in both the CPM^{high} and lung progenitor samples, with slight upregulation observed in the CPM^{high} population. Similarly, the goblet cell markers MUC5AC and MUC5B displayed very low expression levels in both populations but were slightly upregulated in the CPM^{high} sample.

These findings suggest that the enrichment of CPM^{high} cells did not strongly bias the differentiation towards specific lung epithelial lineages, such as AT1, multiciliated, secretory, or goblet cells. While markers associated with these cell fates were expressed in both the CPM^{high} and lung progenitor populations, the overall expression patterns did not reveal significant differences to favour these specific lineages. It is important to note that the culture conditions for lung progenitors may have favoured the development of a lung basal resting phenotype, and this phenotype was not affected by the CPM enrichment process. Nevertheless, the observed expression patterns indicate that the enrichment of CPM^{high} cells primarily selects a population of maturing lung epithelial cells rather than biasing them towards specific lineages. Furthermore, the enriched CPM^{high} cell population seems to represent a group of maturing lung epithelial cells that, given the appropriate cues and signals, have the

potential to differentiate into diverse mature and functional lung cell populations.

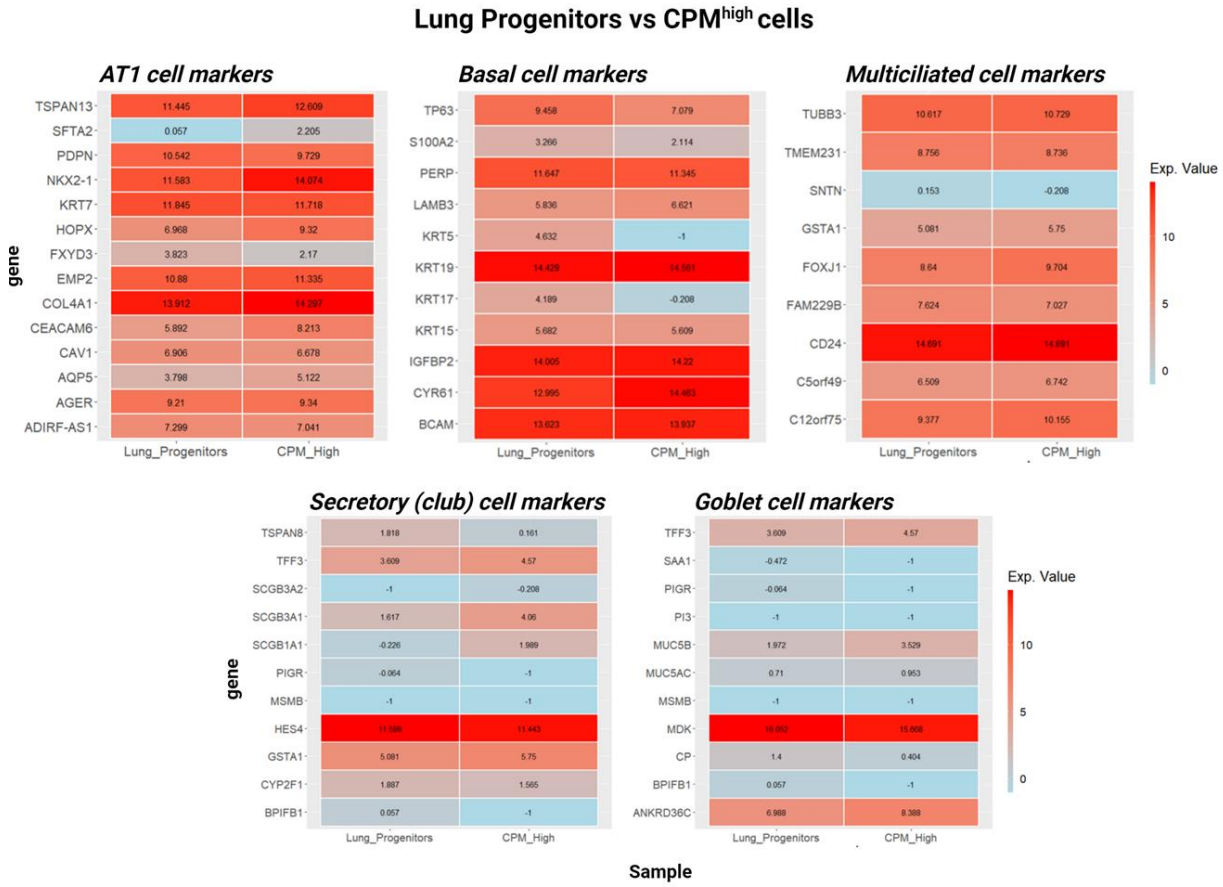


Figure 4.51 CPM^{high} sorting strategy discriminates AT1 and proximal airway cell populations.

Heatmaps show representative airway genes and their relative expression in pre-sorted lung progenitors versus CPM^{high} sorted cells at day 15 of the differentiation protocol.

The NaPi2b negative (NaPi2b_Neg) and positive (NaPi2b_Pos) sorted populations were further analysed to investigate their proliferative states. The DEGs were examined, focusing on top markers associated with AT2 cell proliferation, AT2 transitional states and the expression of WNT signalling markers. Heatmaps comparing the relative gene expression of these markers are shown in **Figure 4.52**. Among the top DEGs associated with AT2 cell proliferation, the majority (seven out of nine) were found to be downregulated in the NaPi2b_Pos population. Conversely, the most upregulated gene in the NaPi2b_Neg population was EIF1AX, a translation initiation factor. This aligns with the upregulation of KRAS and EGFR transcripts in the NaPi2b_Neg sample, as these genes are known to regulate the self-renewal aspect of the stem cell program in AT2 cells [111]. Similarly, three out of nine DEGs related to AT2 transitional cells were upregulated in the NaPi2b_Pos sample, including KRT7,

RNASE1, and SFTA1P. Additionally, genes involved in the WNT signalling pathway, such as WNT isoforms, NKD1, and LEF1, showed upregulation in the NaPi2b_Pos sample. This finding is somewhat conflicting, as WNT signalling pathways are known to promote AT2 cell proliferation and self-renewal [145, 161]. These results suggest that the enrichment strategy employed in this thesis captures a population of AT2 cells with a spectrum of proliferative states. It has been reported that the maturation of AT2 cells is inversely related to their proliferation state [235]. Therefore, the population of AT2 cells generated in this study exhibits potential for expansion due to their proliferative capacity, while the more mature cells are desirable for disease modelling purposes.

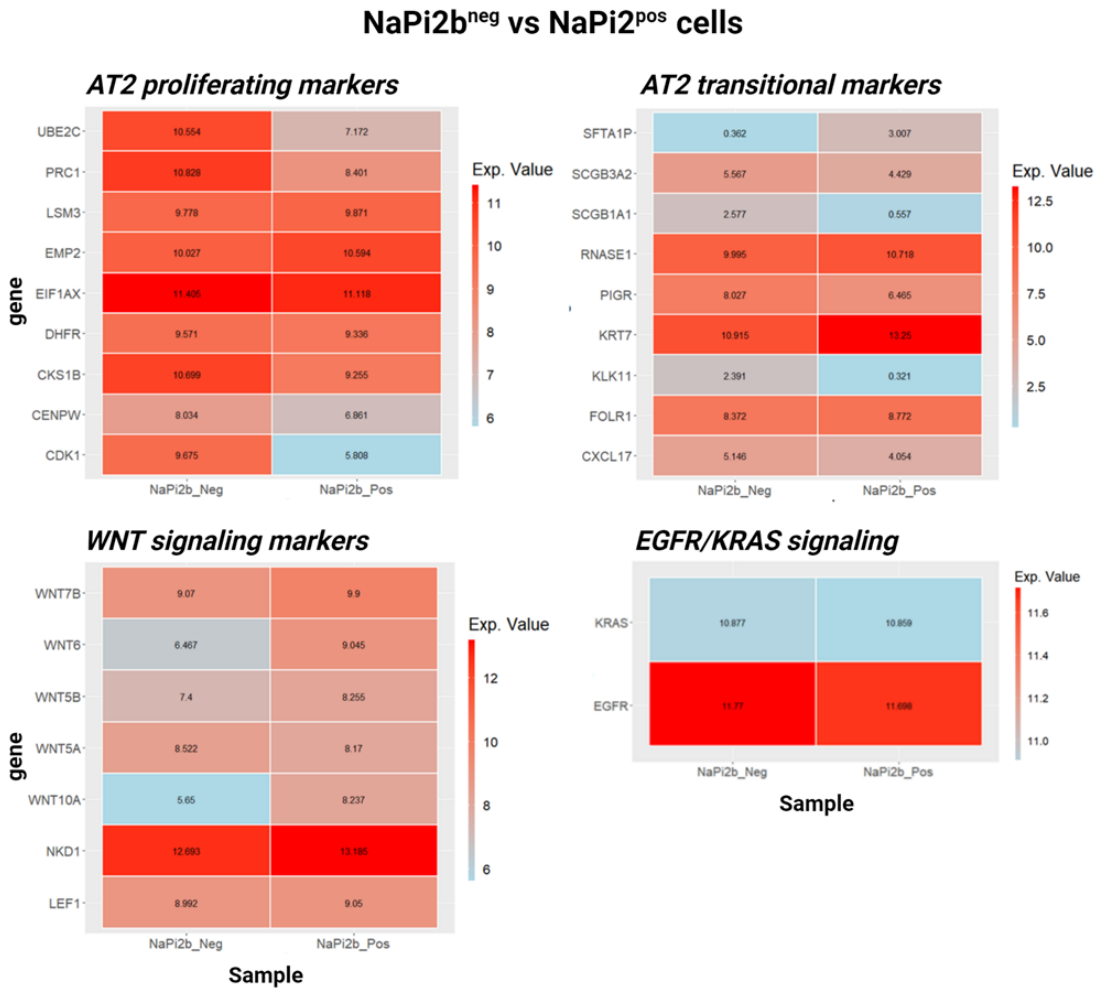


Figure 4.52 hiPSC-derived AT2s express markers of proliferation and self-renewal.

Heatmaps comparing relative gene expression of proliferative and transitional AT2 markers as well as WNT and EGFR/KRAS signalling markers differentially expressed in sorted NaPi2b negative and NaPi2b positive cells.

The global transcriptomics of the enriched AT2 cells, both the negative and positive populations, using the anti-NaPi2b antibody in this thesis, were compared and benchmarked against the published data from Alysandratos et al. (2021) [236]. The reference dataset used in the comparison consisted of AT2 cells that were sorted based on their SFTPCtdTomato reporter.

The heatmap in **Figure 4.53A** shows representative AT2 cell markers and their relative gene expression in NaPi2b sorted and $SFTPC^{tdTomato}$ sorted cells. The most significant changes between the $NaPi2b^{neg}$ and $NaPi2b^{pos}$ sorted cells was the upregulation of SLC34A2, SFTPC, SFTPB, SFTPA2, SFTPA1, NAPSA, LAMP3 and ABCA3 in the $NaPi2b^{pos}$ population. However, when comparing this positive population to the $SFTPC^{tdTomato\_Neg}$ population, the upregulation of these genes was not as evident, apart from for LAMP3 and SFTPA1. Interestingly, the transcript levels of SFTPB were the same for these two populations. Not surprisingly, when comparing the $SFTPC^{tdTomato}$ sorted populations, SFTPC was the most evident DEG, followed by SFTPA2, SFTPA1 and NAPSA. Nevertheless, the differences in transcript levels for the rest of the DEGs were very subtle. Lastly, when comparing both positive populations, the $SFTPC^{tdTomato\_Pos}$ sample had higher expression levels of SFTPC, SFTPB, SFTPA2 and LPCAT1; however, the difference was minimal. On the other hand, SFTPA1 and LAMP3, markers associated with AT2 maturation, were evidently upregulated in the NaPi2bPos sample. Furthermore, the AT2 proliferating marker, EMP2, was significantly upregulated in the $SFTPC^{tdTomato}$ sorted populations (**Figure 4.53B**). As for the rest of the AT2 proliferating markers, even if at lower transcript levels, they were all upregulated in the NaPi2b sorted populations in comparison to the SFTPC reporter ones. These findings provide additional evidence supporting the presence of mature and proliferating cells within the AT2 cells' population generated by the thoroughly optimised differentiation protocol.

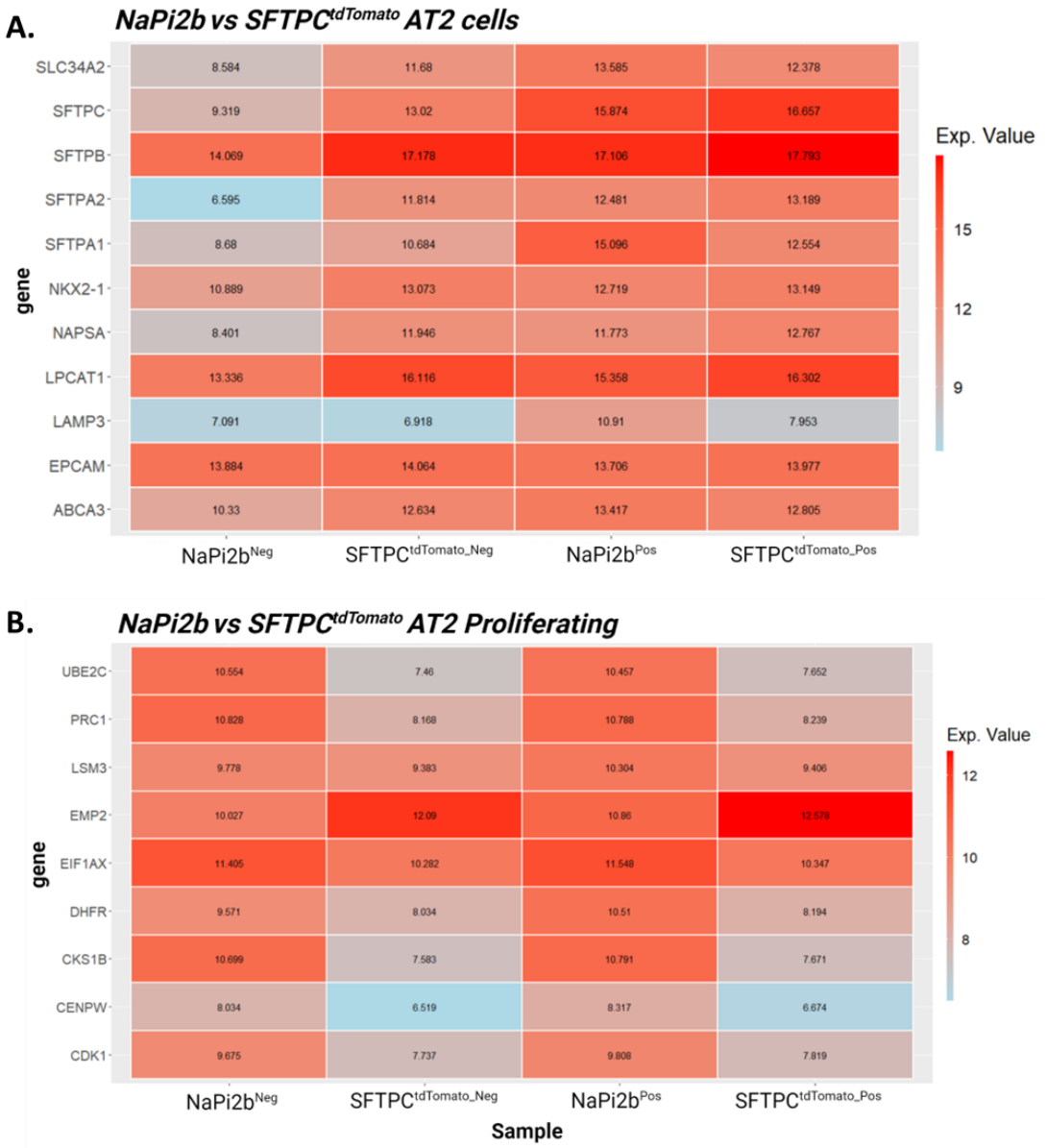


Figure 4.53 NaPi2b^{Pos} AT2 cells express a similar transcriptomic profile to SFTPC^{tdTomato} AT2 cells.

Heatmaps of representative A) AT2 cell and B) proliferating AT2 cell genes differentially expressed in positive and negative cells sorted using the surface marker NaPi2b or the SFTPC^{tdTomato} reporter.

To further validate the transcriptomic profile of the NaPi2b^{Pos} cells, an additional comparison was done by benchmarking this population and the SFTPC^{tdTomato_Pos} population from above against published data from adult and foetal primary samples from Hurley et al, 2020 [233]. The heatmap in **Figure 4.54** illustrates the similarity in transcript levels for key AT2 markers such as SFTPC, SFTPB, SFTPA1, and ABCA3 between the hiPSC-derived AT2 cells and adult AT2 cells. When examining the foetal samples, the closest resemblance to the transcriptomic profile of the sorted AT2 cells

was observed in the latest gestational sample available (week 21), while lower expression levels of SFTPA2 and SFTPA1 were noted in the foetal lung samples across all assessed gestational weeks. Additionally, the total lung RNA sample previously used as a benchmark for qPCR experiments was included and exhibited comparable transcript levels to those of in vitro differentiated AT2 cells. Notably, SLC34A2 and SFTPB were identified as the two DEGs with relatively consistent expression across all analysed samples.

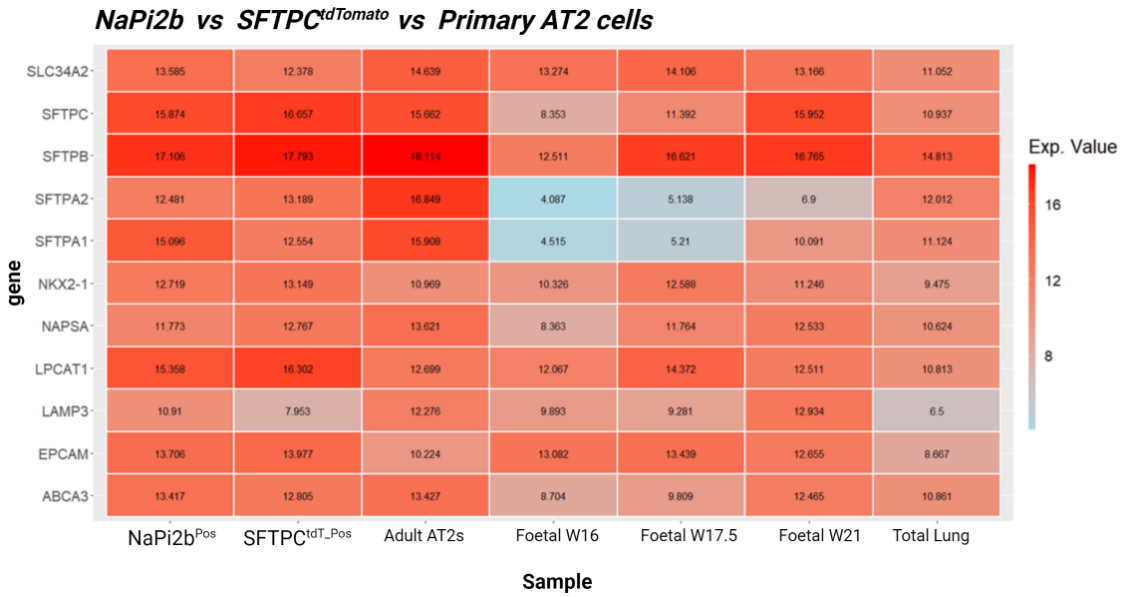


Figure 4.54 Human adult and foetal AT2 cells provide a benchmark for hiPSC-derived AT2 cells characterisation.

Heatmap comparing relative gene expression of differentially expressed known AT2 genes in sorted SFTPC^{tdTomato}-positive and NaPi2b positive populations versus primary samples (adult and foetal (W16, W17.5, W20, W21).

When examining the DEG analysis related to AT1 genes in the NaPi2b and SFTPC^{tdTomato} sorted populations, the heatmap in **Figure 4.55A** showed that out of the 13 genes included, 12 were upregulated in the negative and positive cells sorted using the SFTPC^{tdTomato} reporter. In contrast, only FXYD3 showed upregulation in both NaPi2b sorted populations. Notably, the NaPi2b^{Pos} sample exhibited significantly higher expression levels of AQP5 compared to the other samples. Furthermore, when including the foetal and primary adult samples in the comparison, the SFTPC^{tdTomato_Pos} sample exhibited the highest levels of expression for TSPAN13, HOPX, and EMP2 compared to the other samples (**Figure 4.55B**). The levels of KRT7 and COL4A1 in the SFTPC^{tdTomato_Pos} sample were only surpassed by those observed in the foetal W17.5 sample. These findings suggest that the SFTPC^{tdTomato} reporter has the potential

to enrich alveolar cells expressing high levels of transcripts associated with the AT1 cell fate.

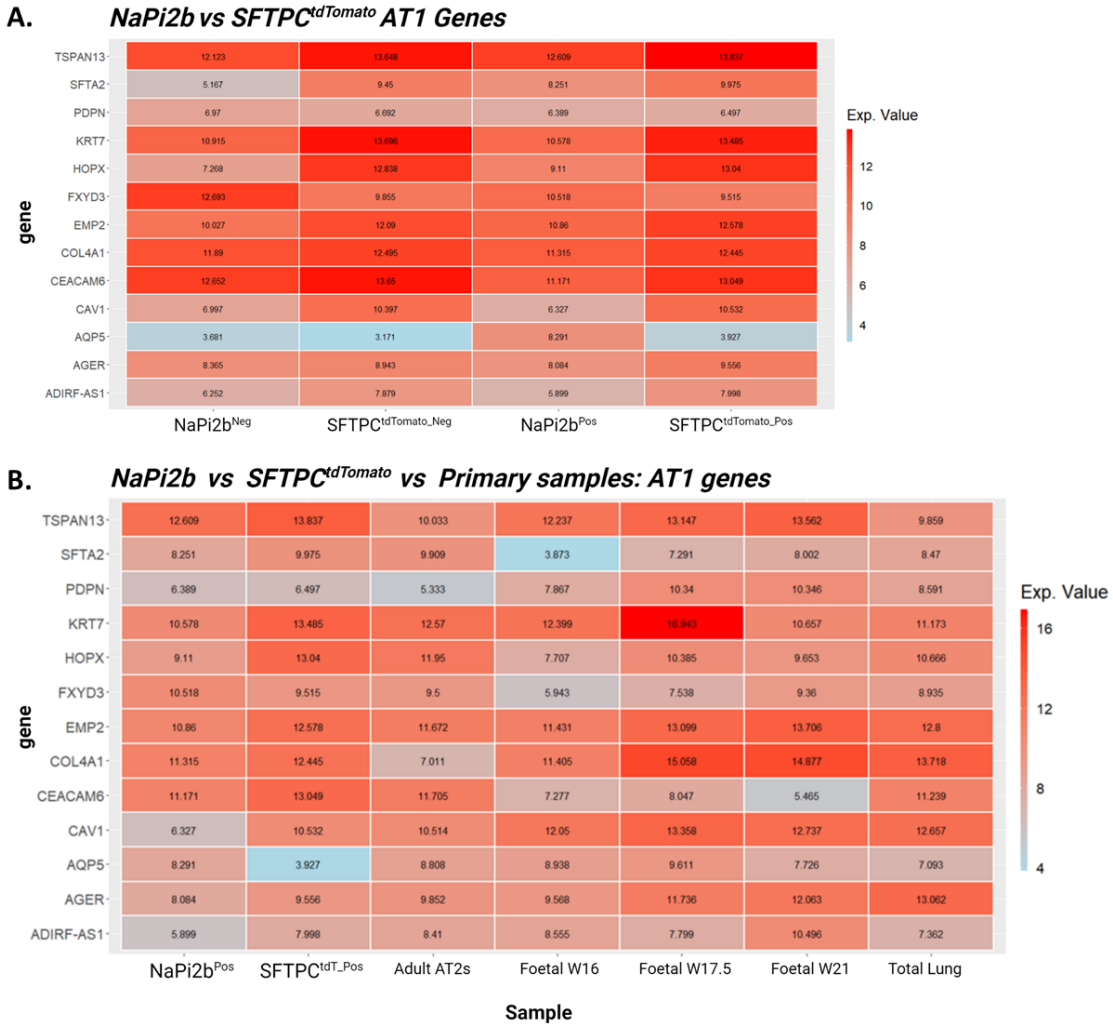


Figure 4.55 Expression of AT1 genes is lower in NaPi2b^{Pos} sorted samples than in the SFTPC^{tdTomato} ones.

A) RNAseq gene expression heatmap of representative AT1 cell genes in NaPi2b or SFTPCtdTomato sorted cells. B) Heatmap of AT1 cell genes differentially expressed in in sorted SFTPC reporter positive and NaPi2b positive populations versus primary samples (adult and foetal (W16, W17.5, W20, W21).

The transcriptomic profile of the differentiated AT2 cells from SFTPC$^{WT/WT}$ or SFTPC$^{MUT/MUT}$ were compared to investigate if they had a phenotype prone to an inflammatory state or expressed markers related to the AT2-AT1 transitional state widely reported in association to IPF [237–240]. The heatmaps in **Figure 4.56** show the main DEGs found in these two categories.

When investigating the representative inflammatory DEGs and their relative expression in both samples, the SFTPC$^{MUT/MUT}$ NaPi2b^{Pos} presented a series of upregulated markers, including TGFB1, IL6R, IL1B, IL17REL, IL17C, CTGF, which have been described as released by AT2 cells in pulmonary fibrosis [10, 172, 241]. Conversely, the SFTPC$^{WT/WT}$ NaPi2b^{Pos} sample exhibited upregulated DEGs such as TGFB2, PDGFRL, PDGFD, PDGFC and a series of interleukin genes, IL17RD, IL17RA, and IL11, associated to the activation of fibroblasts by AT2 cells [242, 243]. Moving into the analysis of the AT2-AT1 transitional markers, most of the DEGs were upregulated in the SFTPC$^{WT/WT}$ NaPi2b^{Pos} sample, except for CLDN4, which exhibited the highest levels of expression among all the markers in the SFTPC$^{MUT/MUT}$ NaPi2b^{Pos} sample. Transitional AT2-AT1 cells with an enriched CLDN4 signature have been identified in cell populations in IPF tissues [239]. Furthermore, YAP1 and STAT3 transcripts were highly expressed in both samples but upregulated in the SFTPC$^{WT/WT}$ cell line. Interestingly, the SFTPC$^{WT/WT}$ NaPi2b^{Pos} sample also showed upregulation of KRT17, a marker that has been reported as fibrosis-specific in cell populations of IPF patient tissues [244, 245]. These observations are not enough to conclude that the AT2 cells derived from the SFTPC$^{MUT/MUT}$ or SFTPC$^{WT/WT}$ have a basal pro-inflammatory or transitional state; however, these findings do not exclude the possibility that the SFTPC$^{MUT/MUT}$ cell line may exhibit an increased response upon stimulation.

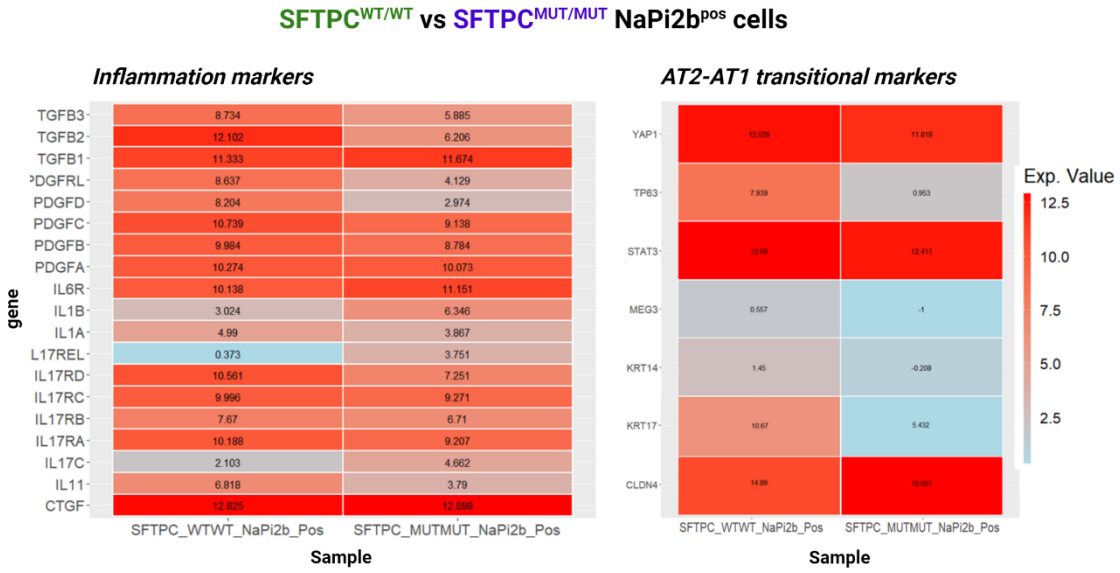


Figure 4.56 Differentially expressed inflammatory and AT2 transitional markers between SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ AT2 cells.

The heatmap on the left compares the relative gene expression of the top 19 inflammatory markers differentially expressed in wild-type and mutant AT2 cells, and the heatmap on the right represents the relative expression of AT2 to AT1 transitional markers.

Comparing the expression patterns of the enriched AT2 cells with previously published data provided a comprehensive assessment of their similarities and differences. This comparison served as validation for the reliability of the differentiation platform developed in this thesis, demonstrating that the AT2 cells generated here closely resemble those generated using a well-established differentiation protocol [162]. These results highlight the robustness and consistency of the differentiation process and further support the potential of the generated AT2 cells for a wide range of applications and research studies.

4.2.6 Further optimisation potential

Modulation of lung progenitors culture conditions enhanced yields of NKX2.1⁺ cells

The differentiation protocol used in this thesis to generate the AT2 cells for disease modelling purposes has been thoroughly described. Nevertheless, there is an ongoing need to develop better in vitro models that closely recapitulate the in vivo alveolar environment to understand lung pathogenesis and discover potential targets for clinical interventions. Therefore, constant efforts are being made in the lab to optimise the differentiation platform further. A significant breakthrough was using a new basal differentiation medium, DIFF3 medium, to culture the differentiating patient-derived hiPSCs from the anteriorisation step onward. This medium consisted of 70% IMDM (Iscove's Modified Dulbecco's Medium) and 23% Ham's F12, with the remaining 7% composed of PVA, ITS-X, lipids, aMTG, AAP, GlutaMAX, NEAA, and antibiotics.

Initially, a slight modification to the endoderm part of the protocol was assessed, where LY294002 was added only on the first day of the protocol instead of for two days as previously optimised. **Figure 4.57A** shows the morphological differences identified between both conditions. On the second day, the optimised condition had more cell death, and bigger gaps could be observed on the monolayer. However, for DE D3, the cells in both conditions had formed a densely packed monolayer with larger cell bodies. Cells on DE D4 were processed and analysed for the co-expression of CXCR4/c-KIT by flow cytometry. **Figure 4.57B** shows a 96.6% double CXCR4/c-Kit positive cell population generated with the modified condition, and an emerging population can be observed in the dot plot, whereas the optimised condition generated 98.4% double positive cells.

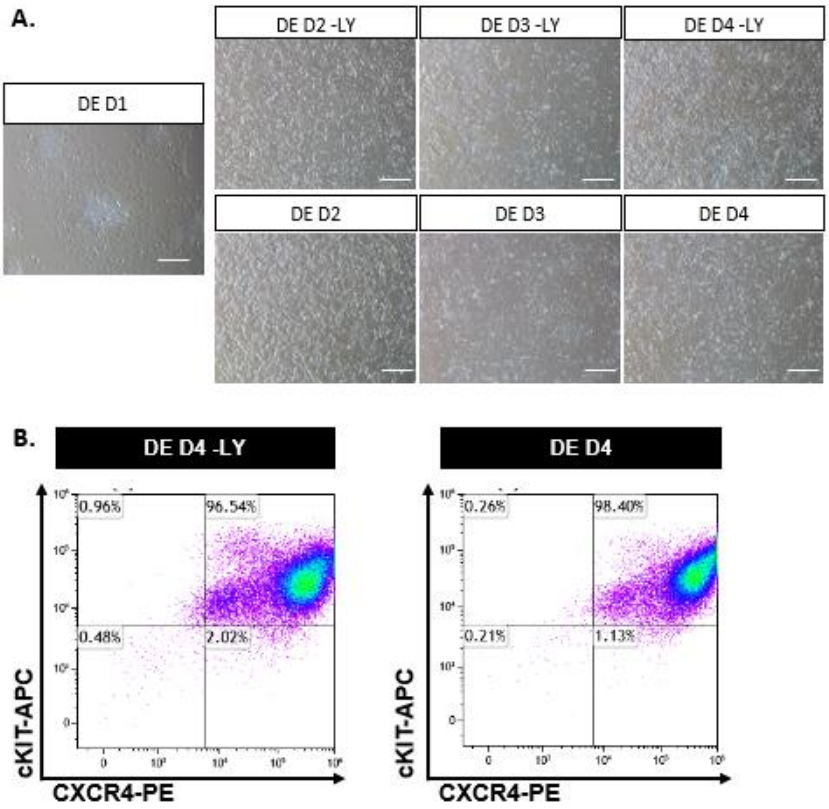


Figure 4.57 One-day PI3 kinase inhibition did not show a better DE differentiation efficiency.

A) Brightfield images of SFTPC$^{WT/WT}$ cells differentiated up to definitive endoderm (DE) day 4 using LY294002 only for one day (-LY, top row) or for 2 days (bottom row). Scale bars, 200µm B) Flow cytometry of day 4 DE cells from A. analysed for CXCR4-PE and c-KIT-APC. Dot plots show a co-expression of 97% for -LY and 98% for normal protocol.

DE D4-LY cells were used to continue with the DIFF2 DS/SB and lung differentiations to assess if the modified endoderm condition could positively impact the end stage of lung differentiation. **Figure 4.58A** shows brightfield images demonstrating normal morphological changes as the cells progressed throughout the differentiation. After plating the cells at a 1:3 ratio for anteriorisation induction, the seeded clumps proliferated, and cells started to expand out. By the third day of lung induction, the cells had formed a compacted cobblestoned monolayer. The cells grew tighter together, and swirl-like structures were observed in some areas by day 7. On the last day of lung progenitor differentiation (LD8), cells were stained to assess their expression of SFTPC. **Figure 4.59** displays immunocytochemistry images of the LD8 monolayer, where the previously described swirls corresponded to areas of intensely stained SFTPC-positive cells, while the rest of the monolayer also exhibited positive staining for SFTPC. The results indicated that the earlier removal of LY294002 from the culture media at the endoderm stage did not impact the endoderm differentiation.

The introduction of the new DIFF3 media was initially performed from the anteriorisation stage onward. Cells were differentiated towards endoderm following the optimised protocol and split on DE D4 to continue the anteriorisation stage. DIFF3 culture media was supplemented with DS/SB only or with the addition of 100nM RA and 50ng FGF10, 5μM IWR-1, 3μM CHIR, or 100μM RA. Unfortunately, none of the cultured conditions progressed beyond this experiment's first day of the anteriorisation stage, including the DIFF2 DS/SB control (**Figure 4.58B**). In addition to the DIFF3 media, the temporal inhibition of the Wnt signalling pathway was also assessed during the same differentiation experiment. For this, the Wnt inhibitor IWR-1 was added at a concentration of 5μM during the first 24, 48, and 72 hours of the lung induction stage. However, as shown in **Figure 4.58C**, the absence of Wnt signalling strongly affected cell survival during the differentiation, resulting in the death of almost every cell in the culture after a 72-hour inhibition. The low survival of the cells with every condition assessed might have been due to an issue with the splitting stage on DE D4, making it difficult to determine the specific effects of the conditions, and thus the experiments needed to be repeated.

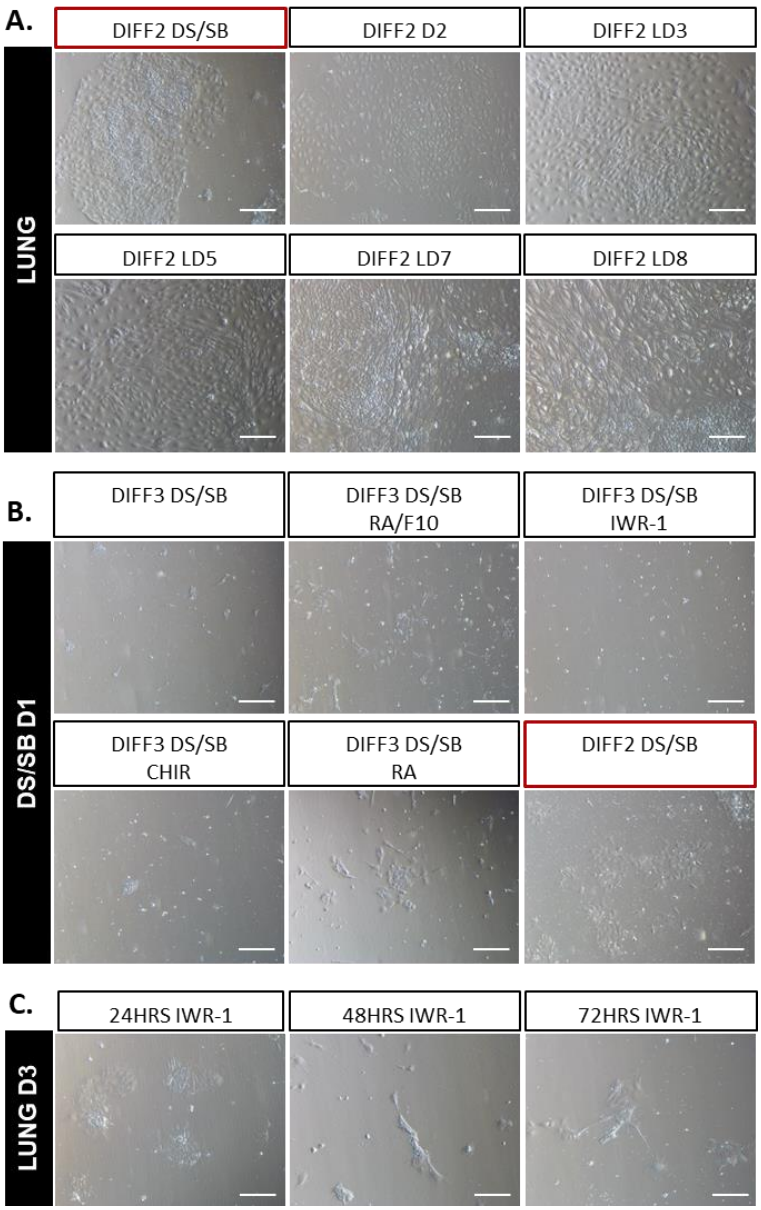


Figure 4.58 Distinct optimisation approaches.

Representative brightfield images of A) lung differentiating cells derived from DE D4 -LY and DIFF2 DS/SB conditions on day 2, 3, 5, 7 and 8 of the lung protocol. B) Anteriorisation day 1 (DS/SB D1) cells generated from DE D4 normal endoderm. Cells were cultured using DIFF2 DS/SB condition as a control (red square) and DIFF3 DS/SB conditions supplemented with RA, F10, IWR-1 or CHIR. C) Lung day 3 cells derived from DE D4 normal endoderm and DIFF2 DS/SB conditions with the inhibition of Wnt (IWR-1, 5µM) for the first 24, 48, and 72 hours of the lung stage.

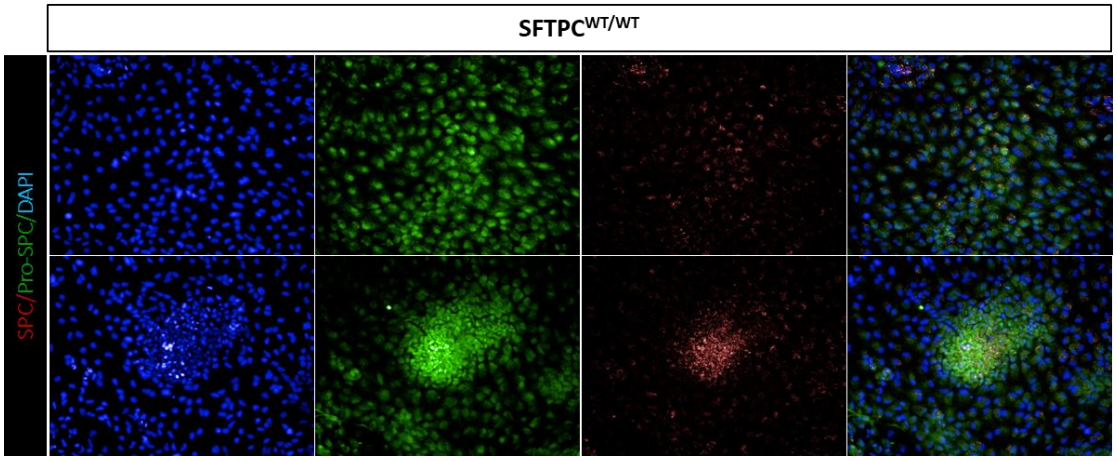


Figure 4.59 SFTPC$^{WT/WT}$ 2D lung progenitors.

Immunocytochemistry images of SFTPC$^{WT/WT}$ day 8 lung progenitors growing in a 2D monolayer were stained for pro-SPC (green) and SPC (red). The cells were differentiated using LY294002 only for one day during the definitive endoderm induction stage. Nuclei were stained with DAPI (blue).

SFTPC$^{WT/WT}$ hiPSCs were again differentiated using the previously optimised endoderm differentiation protocol (**Figure 4.60**). The cells exhibited the expected morphological changes with minimal cell death, and the assessment of CXCR4/c-KIT co-expression on DE D4 resulted in 99.2% double-positive cells. These cells were subsequently used to study the effect of the Wnt inhibitor, IWR-1. In this case, a lower concentration of IWR-1 was used to avoid potential toxicity. After the anteriorisation stage using DIFF2 media, the cells were subjected to lung induction using the normal lung differentiation protocol, adding 1µM IWR-1 for 24, 48, 72, 96, 120, or 144 hours.

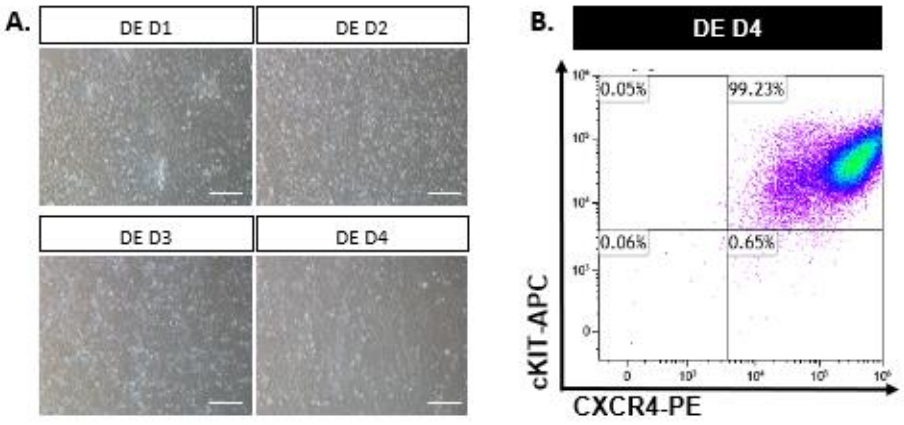


Figure 4.60 Optimised definitive endoderm day 4 cells.

Definitive endoderm (DE) cells up to day 4 of the differentiation using DS/SB DIFF2 conditions with SFTPC$^{WT/WT}$ hiPSCs shown in the brightfield images (scale bars, 200µm) in A) and the respective flow cytometry analysis in B) for the co-expression of CXCR4-PE and c-KIT-APC (99%) on day 4.

The brightfield images in **Figure 4.61** demonstrate the effects of the Wnt inhibitor on the differentiation of lung progenitors. Massive cell death was observed starting from 48 hours of inhibition onward, with remaining elongated and balled-up cells. These morphological changes correlate with the results obtained from the flow cytometry analysis on lung day 8, analysing the generation of CPM^{high} cells. Only the top two conditions (IWR-1 24 and 48 hours) generated a visible CPM^{high} population on the flow cytometry plots, with significant differences in percentages of 13.22% and 1.28%, respectively. The remaining conditions showed CPM^{high} positive cells at levels below 0.5%.

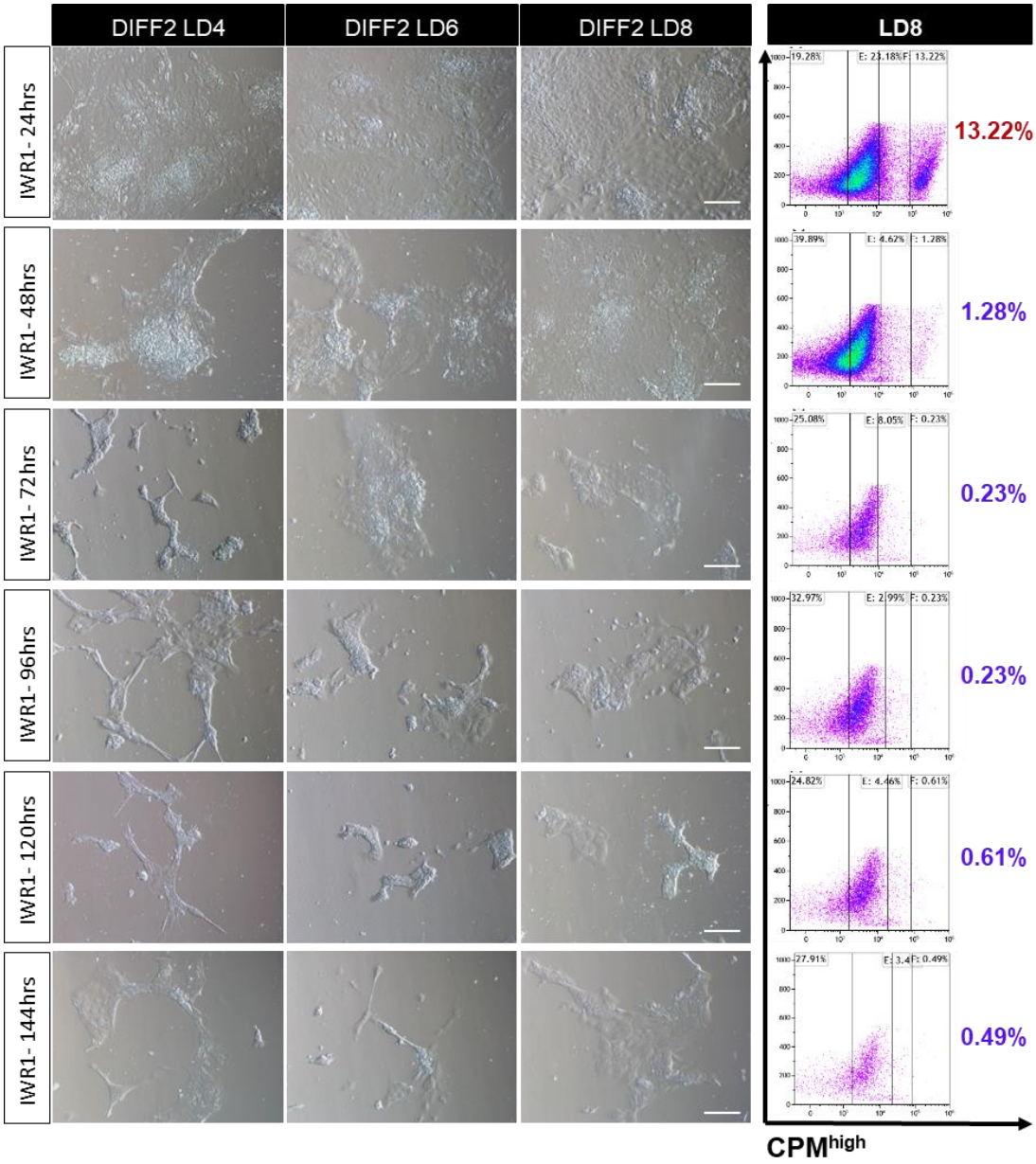


Figure 4.61 Wnt inhibition seems to be toxic for the lung differentiating progenitors.
Representative brightfield images of differentiating cells on days 4, 6 and 8 of the lung stage of the protocol. After the DS/SB step in DIFF2 culture media, DIFF2 media was supplemented with the Wnt inhibitor, IWR1 for 24, 48, 72, 96, 120, or 144 hours. Scale bars, 200µm. On the right, the flow cytometry analysis of lung day 8 cells shows the expression of CPM^{high}. Cells cultured with IWR1 only for 24hrs had the highest percentage of CPM^{high} cells (13.22%)

Finally, the same DE D4 cells were used to reassess the effects of DIFF3 media and the different concentrations of signalling molecules. In this experiment, the cells were transferred to DIFF3 media from the anteriorisation stage, either with DS/SB alone or supplemented with RA, FGF10, CHIR, and IWR-1 at the previously mentioned concentrations. The cells continued with the normal lung differentiation growth factors (3µM CHIR, 10ng/mL BMP4, and 100nM RA) in DIFF3 media. This time, the differentiation was successful for every experimental and control conditions except for the cells supplemented with the Wnt inhibitor (**Figure 4.62**). Interestingly, by lung day 2, all the conditions growing in DIFF3 media had already formed a confluent monolayer, while the control differentiation still had islets of expanding cells. Furthermore, every DIFF3 media condition induced the formation of rounded 3D-like "buttons" of cells scattered throughout the monolayers. These structures were more pronounced in the DS/SB-only condition and the one supplemented with CHIR. The cultures were then analysed for CPM^{high}-expressing cells. Surprisingly, these promising results indicated that the condition generating the highest yield of CPM^{high} positive cells was DIFF3 DS/SB without any additional supplementation, resulting in a percentage of 68.3% compared to the 23.12% from the thoroughly optimised protocol developed throughout this thesis.

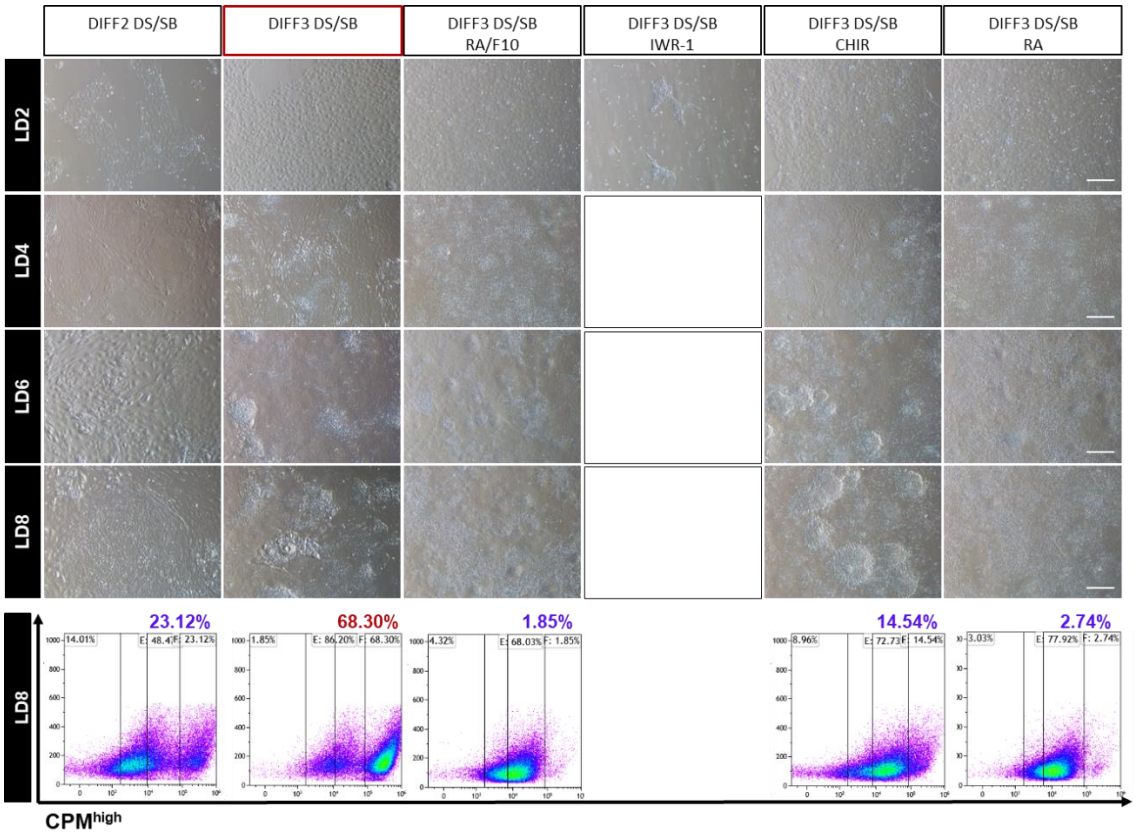


Figure 4.62 DIFF3 condition enabled a more efficient differentiation CPM^{high} lung progenitors.

Representative brightfield images of lung differentiating cells on days 2, 4, 6 and 8 of the protocol growing in DIFF2 or DIFF3 lung progenitors' media. The anteriorisation step of the protocol was performed using DIFF2 DS/SB condition as a control and compared to DIFF3 DS/SB condition on its own or supplemented with RA, F10, IWR1 or CHIR. Scale bars, 200µm. The flow cytometry plots at the bottom show the analysis of lung day 8 cells for the expression of CPM^{high} obtained with each condition. Cells cultured with DIFF3 conditions, without supplementing the media, had the highest percentage of CPM^{high} cells (68.3%).

Further characterisation of the lung progenitors generated using the DIFF3 media for the anteriorisation step and the lung stage of the protocol was performed through bulk RNA sequencing. On lung day 8, the cells were processed and sorted based on CPM gating for moderately positive (Mod) and highly positive (High) populations. The sorted cells, along with samples generated using the original, optimised protocol, were subjected to sequencing. **Figure 4.63** presents a heatmap displaying the differential gene expression of known lung progenitors' genes between the samples generated in the Hannan lab using both differentiation conditions and data from a Kotton lab sample generated using their published differentiation protocol [162].

It is worth noting that among all the analysed samples, the DIFF3_CPM^{high} cells generated using the DIFF3 conditions exhibited the highest expression levels of the

ultimate lung progenitor marker NKX2.1, with a value of 14.07 compared to the second highest expression of 13.1 in Kotton's lab sample (LP CD47^{pos}). This trend was also observed for its surrogate surface marker CPM, with values of 15.1 and 14.98, respectively. Similarly, the distal lung marker SOX9 showed higher expression in the DIFF3_CPM^{high} sample. Interestingly, the gene expression of SFTPC was significantly lower in the CPM-sorted sample compared to the CD47-sorted sample. However, it should be noted that at the lung progenitor stage, high levels of SFTPC were not expected. Overall, these results indicated that the DIFF3 media surpassed the efficiency of the optimized DIFF2 conditions. Additionally, this preliminary data showed that the DIFF3 media might be superior to the peer-reviewed conditions developed by the Kotton lab that are widely used in the field.

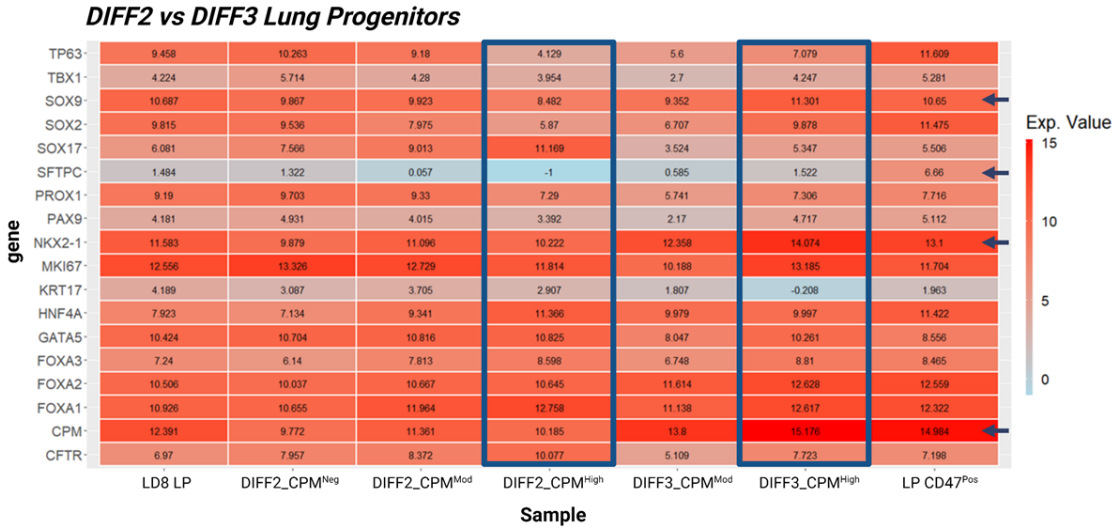


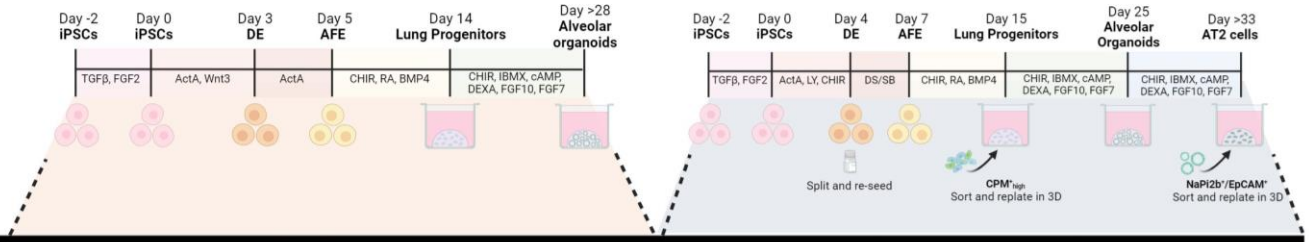
Figure 4.63 Heatmap of representative genes in various lung progenitors' samples.

RNAseq gene expression heatmap of representative lung progenitors' genes differentially expressed between pre-sorted and sorted lung progenitors' samples on lung day 8 (LD8 LP- pre-sorted LD8 lung progenitors, DIFF2_CPM$^{Neg/Mod/High}$- cells differentiated in DIFF2 medium sorted for CPM on lung day8, DIFF3_CPM$^{Mod/High}$- cells differentiated in DIFF3 medium sorted for CPM on lung day8, LP CD47Pos- cells sorted for CD47 on lung day 8 [data obtained Hurley et al., 2020 [233]).

The significance of this discovery must be considered, as the identification of a culture condition that can generate nearly three times more CPM^{high} cells carries immense potential for advancing alveolar organoid generation and enhancing the yields of AT2 cells in subsequent stages. Furthermore, this finding highlights the possibility of further optimising the maturation stage of the protocol to achieve even more remarkable outcomes. However, the timing of this breakthrough was not ideal, given that most disease modelling experiments had already utilised the optimised DIFF2 conditions.

Overall, this chapter presents the results of the optimised protocol development derived from a previously unsuccessful lung differentiation protocol for generating AT2 cells from hiPSCs derived from IPF patients. The tailored conditions matched the differentiation requirements of the SFTPC mutant cell lines developed in this thesis. The optimised protocol consists of several crucial stages in cellular-directed differentiation and organoid generation and processing. Initially, hiPSCs are guided to become definitive endoderm (DE) in basal DIFF1 media, and the differentiation efficiency is evaluated based on CXCR4 and c-KIT co-expression. Subsequently, DE cells with a minimum of 80% yield efficiency are clump-passaged and induced to form anterior foregut endoderm (AFE) by transitioning to basal DIFF2 media. AFE cells are then directed to become lung progenitor cells and later purified using Flow Cytometry (FACS) with an anti-CPM antibody. Furthermore, 3D organoids are generated from the enriched lung progenitors by embedding them in Matrigel droplets, followed by their maturation in a specific cocktail. As the emerging AT2 cells mature, they can be further enriched using an anti-NaPi2b antibody and maintained in 3D conditions for their expansion for experimental purposes, such as the infection modelling. A summary of the specific combinations and concentrations of growth factors and small molecules used in the optimised protocol, along with a comparison to the original protocol, is detailed in **Table 4.1**.

Table 4.1 Summary of lung differentiation protocols employed in this thesis.



Differentiation Stage	Original protocol			Optimised protocol		
	Media	Component	Final concentration	Media	Component	Final concentration
iPSCs	Home-brew E8 medium	FGF2	100 ng/ml	Home-brew E8 medium	FGF2	100 ng/ml
		TGF-β1	2 ng/ml		TGF-β1	2 ng/ml
Definitive endoderm (DE)	Basal DIFF-1 medium	Activin A	100 ng/ml	Basal DIFF-1 medium	Activin A	100 ng/ml
		Wnt 3a	50 ng/ml		CHIR99021	3 µM
					LY294002	10 µM
Anterior foregut endoderm (AFE)	Basal DIFF-1 medium	Activin A	50 ng/ml	*Split DE cells at 1:3 ratio*		
				Basal DIFF-2 medium	SB431542	10 µM
					Dorsomorphin	2 Mm
Lung progenitors	Basal DIFF-1 medium	CHIR99021	1uM	Basal DIFF-2 medium	CHIR99021	3 µM
		RA	250nm		RA	100 nM
		BMP4	5ng/ml		BMP4	10ng/mL
				CPM+high cell enrichment		
Alveolar organoids	***Transfer cells to 3D culture***					
	Basal DIFF-1 medium	CHIR99021	3 µM	Basal DIFF-2 medium	CHIR99021	3 µM
		FGF10	10 ng/mL		FGF10	10 ng/mL
		DEXA	50 nM		DEXA	50 nM
		CAMP	0.1 mM		CAMP	0.1 mM
		IBMX	0.1 mM		IBMX	0.1 mM
		FGF7	10 ng/mL		FGF7	10 ng/mL
				NaPi2b+ cell enrichment		
	Suitable for healthy hiPSC lines REBLPAT and LOPCK up to lung progenitor's stage			Suitable for IPF-patient-derived hiPSC lines IPF1 SFTPC$^{WT/WT}$, SFTPC$^{WT/MUT}$, SFTPC$^{MUT/MUT}$		

Chapter 5 Infection modelling using hiPSC-derived AT2 cells

5.1 Introduction

Even if the pathogenesis of IPF is not yet fully understood, it has been putatively accepted that the condition results from the integration of damaging environmental stimuli and genetically predisposed lung epithelium. This interaction compromises AT2 cell function and reduces its capacity to regenerate the damaged alveolar epithelial tissue [72, 74, 75]. Among the possible environmental triggers giving rise to alveolar epithelial cell injury, including smoking, occupational exposure, micro-aspiration, air pollution, exposure to infecting agents is the most commonly associated to the pathogenesis and progression of IPF. In genetically predisposed AT2 cells, such gene-environment interactions lead to AT2 cells' apoptosis or adoption of a senescent phenotype with the release of a host of factors driving inflammatory and fibrotic responses [246]. The ER stress present in this dysfunctional AT2 cells turns the cells more susceptible to second external stimuli, leading to the release and exaggerated activation of pro-fibrotic mediators such as TGFβ and PDGF, resulting in extracellular matrix synthesis and accumulation [3, 247].

Clinical and culture-independent studies indicate there is a strong association between the lung microbiome and IPF prognosis, with both viruses and bacteria having the capacity to modulate the host response to injury in IPF [84]. In fact, it has been demonstrated that bacterial burden rather than specific microbes is a predictor of the rate of functional decline and risk of death in IPF patients [248]. A study to estimate the time-progression in IPF identified the presence of Streptococcus, Staphylococcus and respiratory viruses was strongly associated with disease progression [249]. Furthermore, there is evidence that supports the mechanistic role of airborne infecting agents in the pathogenesis and progression of IPF [250].

5.1.1 Acute exacerbations

Acute exacerbations (AE) are a life-threatening complication in patients with IPF and have been defined as episodes of clinically significant respiratory deterioration with an unidentifiable cause [251]. They are characterised by the appearance of diffuse alveolar damage, leading to the presentation of worsening dyspnoea, new onset cough, chest pain, fever, and fatigue. Patients may exhibit signs of respiratory distress

such as tachypnoea and low oxygen saturation levels, accompanied by bilateral infiltrates on chest imaging, consistent with underlying IPF. Unfortunately, these episodes are associated with a mortality rate as high as 85% and a 50% reduction in IPF patients' life expectancy [251, 252]. The median survival following an AE in IPF is approximately three to four months and independently of the treatment, 40% of IPF deaths are related to an AE event [253]. The negative outcome of patients during AE are due to the respiratory dysbiosis generated by the dysregulation of the pulmonary ecosystem associated with a dysregulated host immune response. These events are triggered by an inflammatory insult, such as infections, that initiate a downstream response that accelerates the dysbiosis-inflammation cycle [254].

There is a growing number of evidence that demonstrates that infectious pathogens, particularly viruses may be responsible for a substantial number of AE cases in IPF, as viruses have been identified in up to 40% of IPF patients suffering an AE [84]. Viral sequences have also been detected in the nasopharyngeal swab of patients with stable IPF and those presenting with AE. Among these, HHV and influenza type A accounted for the majority of the viral burden. Additionally, IPF patients with AE showed upregulation of inflammatory cytokines like IL-6, IL-17, IL-9 and IFN-gamma [255]. Furthermore, animal models have also demonstrated that viral infections can exacerbate established fibrosis [256, 257]. These data suggest virus infections coupled with an elevated anti-virus immune response might be an important trigger for AE in IPF. A better understanding of the complex pathophysiology of this IPF complication is needed to advance the development of new therapeutic agents resulting in an improved prognosis for this severe condition.

5.1.2 Influenza A virus subtype H1N1 infection

Influenza type A is the dominant strain causing seasonal influenza epidemics each year. The antigenic strain variations caused by adaptive mutations in the hemagglutinin esterase (HA) gene in influenza A limit the success of human herd immunity [258]. Studies have reported that, among other viruses, influenza virus can cause long-term damage to the lung epithelium becoming a risk factor to develop pulmonary fibrosis [91].

Localization of influenza A H1N1 antigen in association with diffuse alveolar damage was the main parenchymal abnormality found during the 2009 influenza A pandemic [259]. Other histologic features identified were AT2 cell hyperplasia and fibrosis [260]. In response to the infecting insult, the alveolar epithelium triggers a series of

inflammatory repair pathways that may lead to the increased activation of fibroblast, playing an important role in the development of fibrosis in H1N1 patients. Among these pathways is the activation of the potent pro-fibrotic cytokine TGFβ, activated by AT2 cells that express the cell surface integrin αvβ6 [73][6]. The release of the mature TGFβ in response to the infection induces the proliferation of fibroblast and the production of extracellular matrix proteins [247].

Determination of viral infection in epithelial cells

An essential characteristic of viruses is their ability to infect and replicate within a cell. The cycle of the replicating virus is associated with several biochemical and morphological changes within the host cells that often lead to cell death [261]. In order to perform reproducible experiments with viruses it is important to determine the viral titre, which is the number of viral particles in a given volume. This allows a consistent number of viruses to be used in experiments. Furthermore, determining the number of active and infective virions of the virus strain to be used, known as infectivity, is also crucial for experiments [262].

There are numerous methods to measure the viral titre and infectivity, most depending on observing the morphological changes inflicted on the host cells, referred to as the virus cytopathic effect (CPE) [262]. These methods, among others, include the haemagglutination assay, plaque assay, viral Tox-Glo™ assay, and virus staining. The interpretation and analysis of the results obtained from these assays allow the identification of the viral titre and/or infectivity, each with its advantages and disadvantages.

Haemagglutination assay

The haemagglutinin (HA) protein is an antigenic glycoprotein on the surface of influenza virus particles. It is responsible for the membrane fusion between the virus and the host cells of the respiratory tract. HA protein is responsible for the first stage of virus infection via binding to sialic acid residues on the host cell surface. The influenza virus can also bind to the sialic acid residues on red blood cells, causing them to aggregate [263]. Based on this, the haemagglutination assay, a standard method to measure viral titres, can detect the presence of influenza virus in a sample. Chicken red blood cells (RBCs) are usually used in this titration assay as they are nucleated and sediment faster than mammalian red blood cells, which lack a nucleus [264].

When a sample containing influenza virus is added to RBCs, the present HA binds to the cells and causes them to agglutinate, promoting a lattice of cells all over the well,

inhibiting their sedimentation; this is called haemagglutination. If no virus is present, the RBCs will form visible red sediment at the bottom of the well in the form of a sphere ('button'), as shown in **Figure 5.1**.

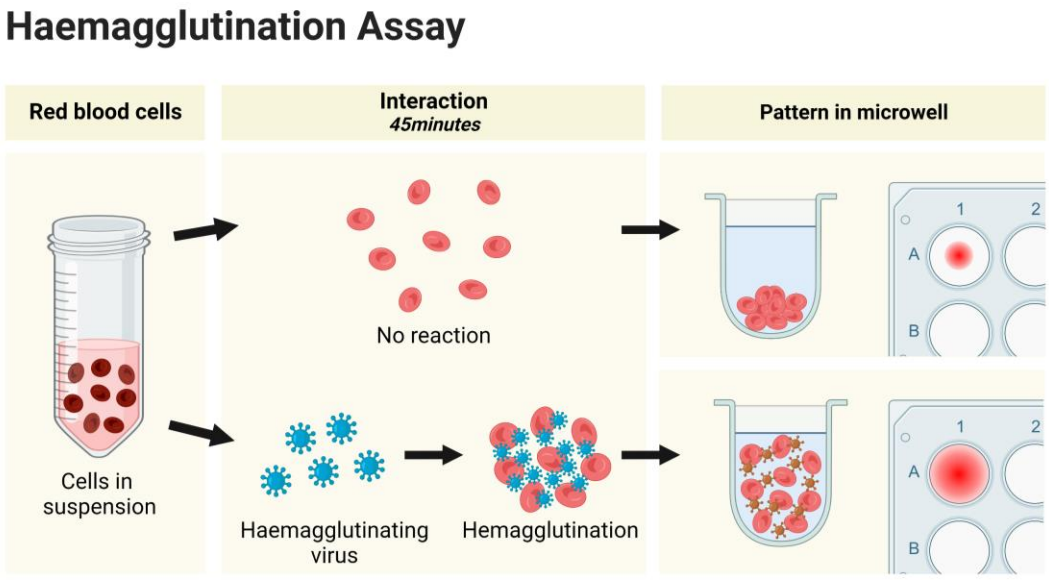


Figure 5.1 Schematic representation of the haemagglutination assay.

Red blood cells (RBCs) in suspension fall to the bottom of a V-shaped well when not agglutinated by the presence of the virus. When virus is added to the well, it binds RBCs and forms a matrix deposited across the whole well by linking RBCs together, also known as agglutination, preventing RBCs from settling in the diluent. Adapted from Killian, 2014 [265]. Created with BioRender.com.

Thus, doubling serial dilutions of a viral sample can be mixed with an equal volume of chicken RBCs of known concentration to assess their interaction. For agglutination to occur, the virus should be in sufficient concentration to bind several RBCs simultaneously, leading to the formation of a lattice. If the virus is insufficient, the RBCs will fall to the bottom of the well, and a well-defined pellet will appear; this way, the lattice or button formation can be used to determine the viral titre in the sample. Therefore, the haemagglutination titre of a virus suspension is defined as the reciprocal of the highest dilution which caused complete agglutination before a pellet appeared, and it is expressed as the number of HA units per unit volume (HAU) for a given concentration of RBCs (**Figure 5.2**)[266].

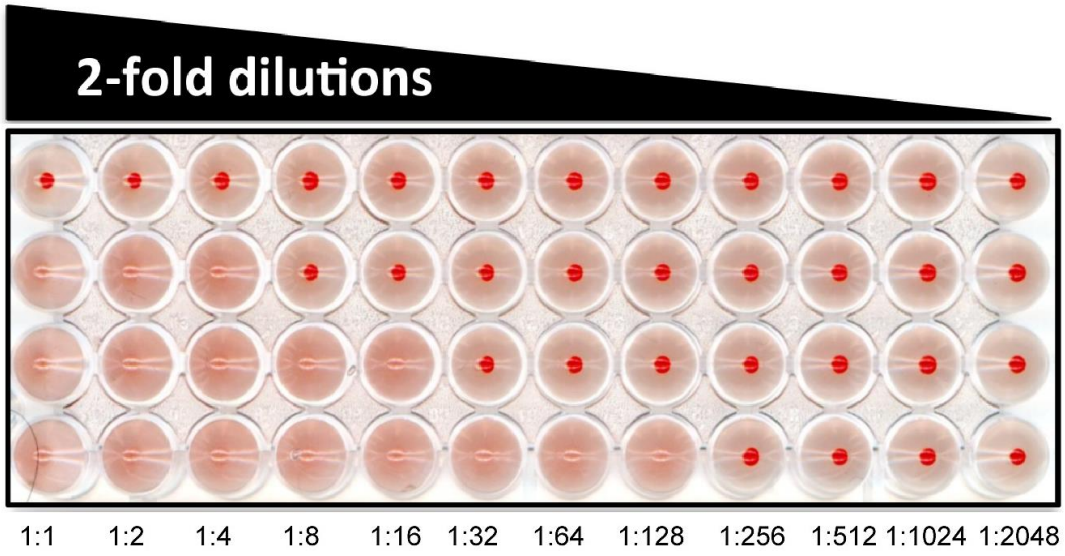


Figure 5.2 Interpretation of haemagglutination assay results.

Microtiter plate with positive and negative haemagglutination results. The HAU titre represents the last dilution before a pellet appears. Results for each row: 1) shows no haemagglutination; 2) positive haemagglutination 4 HAU; 3) positive haemagglutination 16 HAU; 4) positive haemagglutination 128 HAU.

The HA assay is a simple and cost-effective technique to determine the relative amount of influenza virus particles in infected samples. It is, to some extent, a quantitative approach, as each haemagglutination unit is equal to approximately 5–6 logs of virus [264]. However, this assay does not assess virus infectivity or the presence of a viable virus. Degraded or inactivated viral particles that are no longer infectious can also be detected by the HA assay as remanent haemagglutinin protein can still cause red blood cells to agglutinate [267].

Plaque assay

The plaque assay relies on the principle that lytic viruses infect cells, replicate and finally lyse the initial host cells as the progeny virions escape only to infect the surrounding cells. The replication-lysis-infection cycle will continue as the newly produced virions infect new cells, further propagating the infection [268]. After the initial infection and application of an immobilising overlay to prevent virus spread and restrict its growth, the lack of viable cells left results in increasingly distinct and visible plaques, even to the naked eye. As the number of plaques is proportionally related to the concentration of viable virus, they can be used to determine the viral titre in the sample used (**Figure 5.3**).

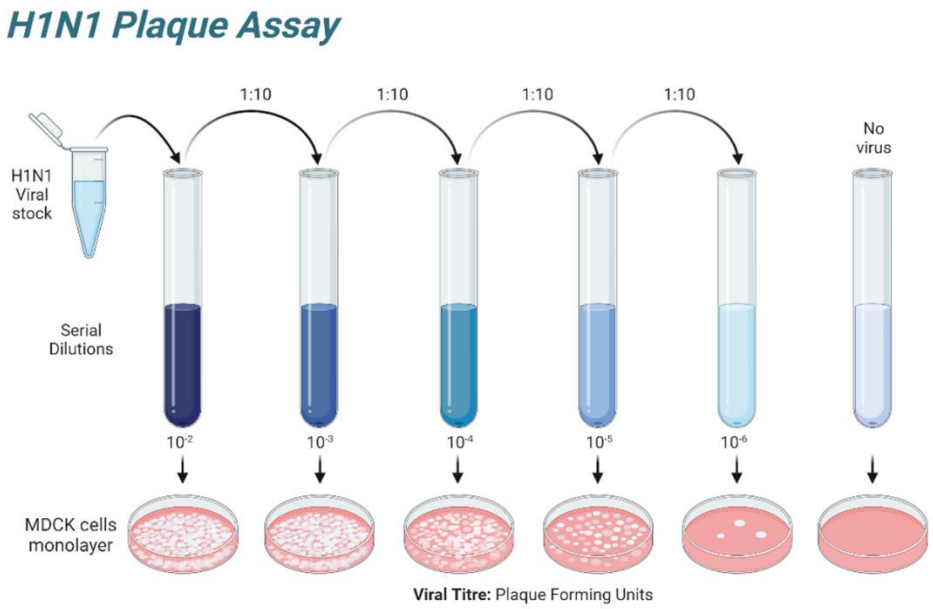


Figure 5.3 Schematic representation of the H1N1 plaque assay.

Serial 10-fold dilutions of the influenza virus A H1N1 stocks are prepared in MEM media with TPCK-trypsin. A confluent monolayer of MDCK cells is inoculated using 3 wells per dilution being tested. After a short incubation, cells are overlayed with agarose and stained once cytopathic effect is observed. Virus titre is determined by counting the number of plaques formed on each well and multiplying it by the dilution factor. Plaques counted from the 3 replicates are averaged. Image created using a modified Biorender template.

Viral Tox-Glo assay™

The Viral Tox-Glo™ assay consists of the identification of cytopathic effect (CPE) induced by viral infection, and it does it by measuring cellular adenosine triphosphate (ATP) using firefly luciferase. As ATP is the main energy storage molecule and is involved in a variety of cellular enzymatic reactions, it is considered one of the most sensitive indicators of cell viability [269]. When cells are lethally damaged, and the integrity of their membrane is lost, the ability to synthesise ATP is impaired, and endogenous ATP levels decrease drastically. Since ATP degrades rapidly after cell death and its concentration is related to the number of cells, measuring it determines the number of live cells left [269]. The assay uses luciferin as a substrate and the enzyme luciferase to catalyse the reaction to oxyluciferin, generating photons of light [270]. Because this reaction is ATP-dependent, only live cells would produce a signal, and the result is a linear relationship between the intensity of the luminescent signal and ATP concentration/viable cell number (**Figure 5.4**). The measurement of ATP depletion upon the infection can then be correlated with the viral burden and corresponding tissue culture infective dose (TCID) [271].

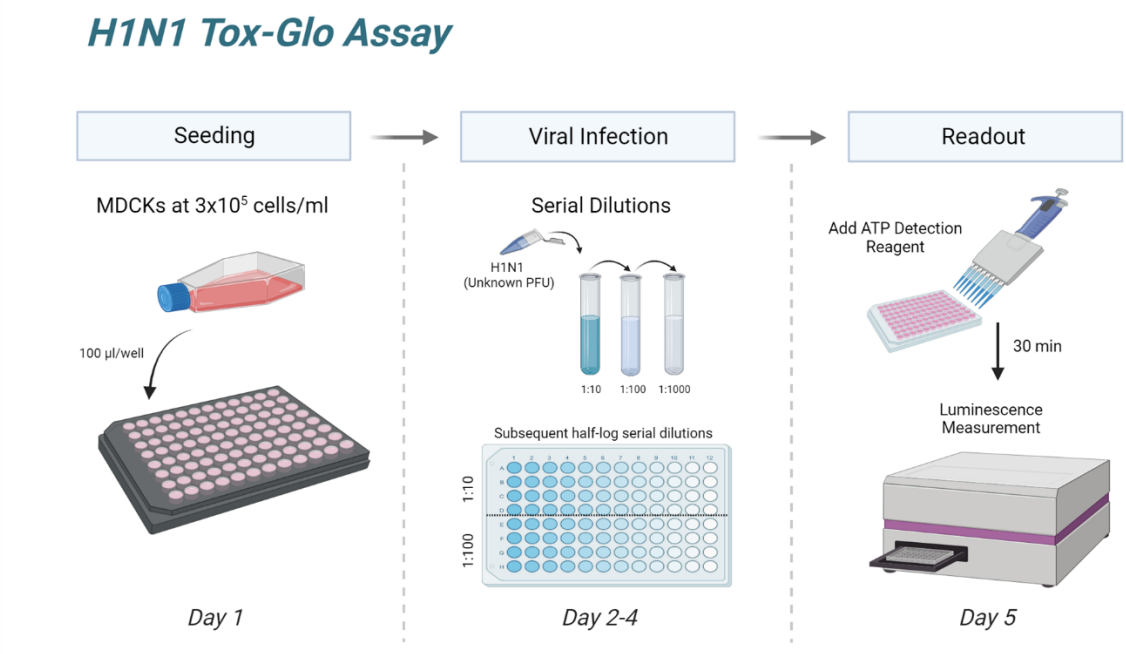


Figure 5.4 Schematic representation of the ATP-depletion assay.

MDCK cells seeded in an opaque 96-well plate are incubated for 24 hrs before adding half-log viral serial dilutions per column. After a 3-day incubation, 100µl of ATP detection reagent are added to each well. Following 30 minutes, the luminescence generated with the ATP of the remaining viable cells is measured with a luminometer. The reciprocal of the dilution that caused a 50% decline in the ATP levels compared to blank wells is the $TCID_{50}$. Created with Biorender.

Co-culture assay

The principle of the TMLC- iHBECs co-culture assay is to determine the αvβ6-mediated TGFβ activation [272]. This activating process does not lead to the secretion of active TGFβ into the extracellular space, thus, the traditional antibody-based methods used to measure the active form of TGFβ cannot be used. Instead, a system where the reporter cell line TMLC is cultured in direct contact with αvβ6-integrin expressing cells is needed. The assay consists of co-culturing iHBECs with the reporter TMLC cell line, which expresses firefly luciferase driven by the TGFβ-inducible portion of the plasminogen activator inhibitor-1 (PAI-1) promoter [173]. iHBECs are a continuously dividing cell line that retain many of the properties of primary epithelial cells, including the expression of αvβ6 integrins in vitro [172].

When iHBECs are infected with influenza A virus, the double-stranded viral RNA from the replicating virus translocates to the nucleus activating the transcription of pro-inflammatory mediators, including TGFβ. The produced latent TGFβ pro-peptides are secreted into the extracellular space, where they remain inactive until the mature TGFβ is released from its latent complex [273]. The viral particles entering the cell are also

processed and detected by the TLR3 in endosomes [274, 275]. The activation of TLR3 in iHBECs triggers actin cytoskeletal forces that are transmitted through the β subunit of the αvβ6 integrin located on the plasma membrane. Upon latent-TGFβ detection, the conformational change on αvβ6 from extended-closed to extended-open releases the mature TGFβ, emitting a signal . TMLCs do not express the αvβ6 integrin, but they present an abundant expression of TGFβ receptors [92]. This way, TMLCs drive the expression of firefly luciferase in response to the PAI-1 promoter stimulation by TGFβ activated on the neighbouring iHBECs (**Figure 5.5**). The luciferase activity can be then measured using a luciferase assay reporter kit assessing the luminescence expressed as relative luminescence units (RLU) in a luminometer. The nature of the assay allows the measurement of dose-dependent increases in the concentration of TGFβ. Thus, measuring TGFβ activation is an indirect way of measuring viral activity.

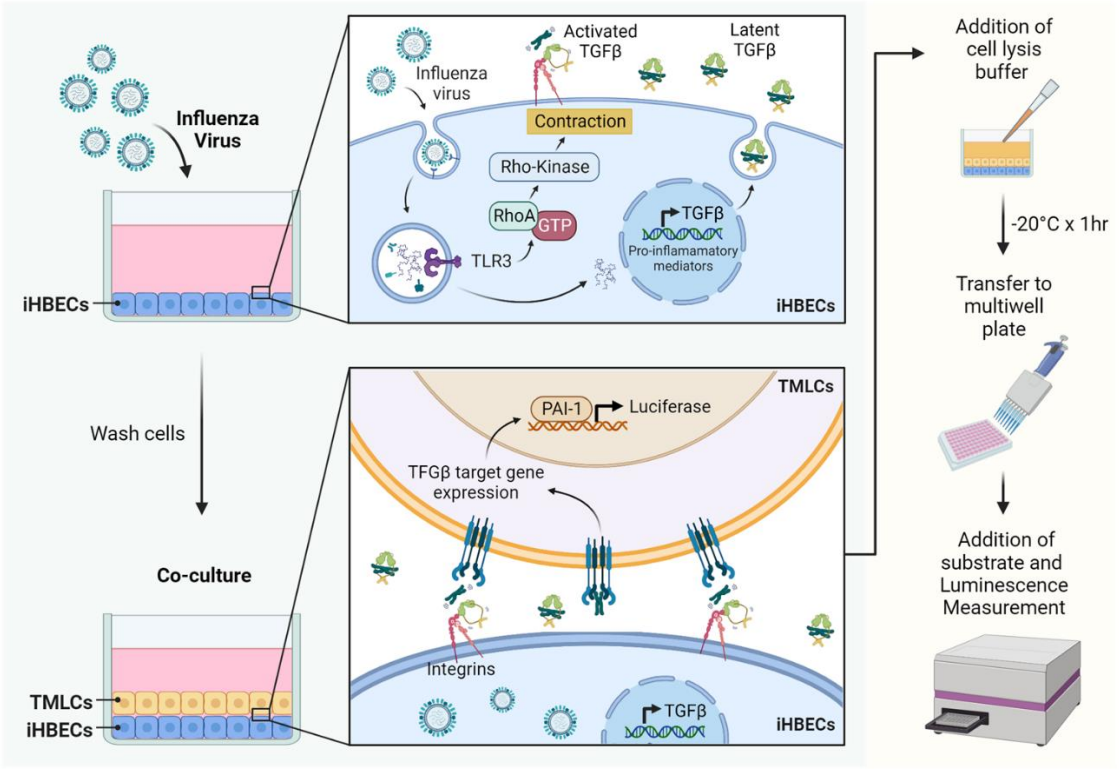


Figure 5.5 Schematic representation of the TMLC- iHBECs Co-culture Assay to measure the in vitro activation of TGFβ.

iHBECs are culture to confluency in a 96-well plate, growth-arrested, and inoculated with influenza virus for 60 minutes. After the incubation, cells are thoroughly washed and TMLCs are added directly on top and incubated for 16 hours. This avoids the direct exposure of the reporter cell line to the virus. Lysis buffer is then added to each well and the plate is frozen to aid cell lysis. The lysates are then transferred to a 96-well luma plate and placed in the luminometer, which adds 100µl of luciferase substrate per well to then read the emitted luminescence. The measured relative luminescence units can then be plotted after subtracting

the data from the unstimulated TMLC cells to generate a standard curve. Created with Biorender.

Regardless of the methodology, a reliable approach to replicate and isolate the virus, as well as a robust assay to determine the concentration and infectivity of the expanded virus are needed to implement viral inoculations for in vivo or in vitro infection modelling.

5.1.3 AT2 cells response to infection

AT2 cells are continuously exposed to mechanical, metabolic and environmental factors and their main function is to maintain alveolar homeostasis. Failure to fulfil this task can result in the accumulation of misfolded proteins, aberrant signalling cascades activation, inflammation, cell death and eventually parenchymal lung disease. It is now well established that dysfunctional AT2 cells play a pivotal role in the impaired repair response in lung fibrosis after damaging stimuli [112, 276–278]. Genetically compromised AT2 cells' exposure to infecting agents might exacerbate the activation of intrinsic pathways leading to inflammation and recruitment of immune system cells, worsening the lung scarring. However, the exact mechanisms leading to the aberrant wound healing seen in IPF patients after infections are not clear.

In vivo response

The murine model of lung injury has been incredibly useful in studying alveolar destruction and regeneration upon respiratory infections. Previous research has shown that latent virus infection in mice could augment the fibrotic response when exposed to a second insult like bleomycin with an increase in collagen deposition when compared to that of uninfected controls [279]. Other groups also identified that influenza viral infection leads to collagen deposition in mice lungs with large fibrotic areas, thickened alveolar walls and collapsed alveoli [93]. H1N1 infection also leads to an increased expression of TGFβ and SMAD-2 by AT2 cells and macrophages, indicating that the observed fibrosis might have been due to the viral activation of TGF-β/SMAD pathway [280]. Furthermore, when infecting aged and young mice, it was found that viral-induced lung injury was exacerbated in the aged group [95]. Older mice had a higher expression of TGFβ and a significant increase in apoptotic responses. Apoptosis of AT2 cells was associated with an increase in ER stress markers in ageing mice, making them more susceptible to injury, activation of profibrotic pathways and cell death [94]. Infection modelling using mice has also led to the discovery of lung progenitors, such as KRT5+ lung epithelial cells that were found to be capable of regenerating the

damaged alveolar epithelium after injury [281, 282]. Likewise, influenza-induced lung injury resulting in tissue damage with different zones of injury helped identifying a subtype of AT2 cells that are Wnt-responsive and act as facultative progenitors in the distal lung [36].

In vitro response

Despite the advances in the field due to the knowledge gained using mouse models, there are significant differences between mouse and human lungs that fail to recapitulate the pathologic changes of human IPF making it difficult to study the initial stages of the disease [6].

This highlights the need for more reliable models to widen our knowledge of the human lung homeostasis and disease. Therefore, the in vitro use of primary human tissues and technologies such as hiPSC-derived lung epithelium are a promising tool to help elucidate human-specific mechanisms in alveolar repair and host cell factors in respiratory infections.

Several approaches have been made to model in vitro infections of AT2 cells with airborne infecting agents. Research conducted in organoid models for COVID-19 infections has indicated that AT2 cells, after being infected by the virus, activate aberrant pathways including interferon signalling and apoptotic pathways [132, 283] . Similarly, a study of influenza A virus tropism and replication kinetics demonstrated that human airway organoids provided results comparable to the ones obtained from human ex-vivo bronchus cultures infecting ciliated and goblet cells, demonstrating that organoids could be used as a physiologically relevant experimental model [284]. Furthermore, airway organoids derived from human lung tissue containing airway epithelial cells ciliated, goblet, club, and basal cells that shared morphological and functional characteristics with the human airway epithelium, allowed modelling of viral infections and offered the possibility to study neutrophil–epithelium interactions in vitro [285, 286].

Making use of this organoid technology, a group generated alveolosphere cultures derived from healthy adult human primary alveolar AT2 cells, that gave rise to self-renewing AT2 cells, as well as presumable AT1 cells [128, 287]. The AT2 cells in such organoids proliferated upon IL-1 and TNFα stimulation in vitro, mimicking the in vivo findings of Katsura et al, in influenza virus-infected murine lungs [287]. These AT2 cultured cells also expressed viral receptors making them permissive to SARS-CoV2-infection. When infected with the virus, cells activated IFN and inflammatory pathways

to exert antiviral defence mechanisms. SARS-CoV-2 infection also induced loss of surfactant proteins SFTPB and SFTPC, as well as cell death [130]. Using a similar approach but with primary human distal lung organoids containing pure populations of AT2s, which expressed influenza receptors, were infected with an H1N1 strain PR8 modified to express GFP upon viral replication [131]. Nearly 80% of the AT2 cells in the organoids expressed GFP 12 hours post-infection and the viral copies identified per ml of supernatant peaked at 96 hours post-infection. These AT2 organoids also expressed SARS-CoV-2 receptor ACE2 and processing protease TMPRSS2 mRNAs. Therefore, the group adapted an apical-out suspension culture polarisation method that caused the epithelial organoids to reorganise, so the apical membrane faced the spheroid exterior, this was to facilitate access of SARS-CoV-2 to ACE-expressing luminal cells. Only 10% of the AT2 organoids were directly infected by SARS-CoV-2 and showed production of infectious virions, the remaining organoids were devoid of infection [131]. Another SARS-CoV-2 infection study applying a 3D culture model of AT2 cells derived from primary tissue proved the feasibility of in vitro infection modelling. The organoid-cultured AT2 cells showed rapid viral replication after SARS-CoV-2 infection with expression of type I and III interferons [132].

The studies mentioned above use primary AT2 cells to derive their organoids, however, modelling respiratory virus infection has also been done with organoids derived from hiPSCs [283, 288]. Porotto et al, generated organoids that provided a tissue environment that was reminiscent of the branching airways and alveolar structures, helping to maintain the authentic viral genome of parainfluenza, respiratory syncytial virus and measles virus [288]. The changes observed upon infection simulated those observed in human infant lungs, like the detachment and shedding of infected cells into the lumens of the organoids. This suggested the model could be useful to study pathogenesis and if genetically engineered, a more relevant model for specific viral infections, demonstrating the value of organoid technology for virus research [288].

Mutant cell lines

The use of patient-derived hiPSCs renders a potent tool for the study of specific diseases. An example of this is a study showing that patient-derived hiPSC-derived lung epithelial cells carrying a heterozygous null mutation in IRF7, gene encoding for the interferon regulatory factor 7, produced reduced amounts of type I IFN and displayed increased viral replication in response to infection with influenza virus [289]. This correlated with the observations in the patient, who suffered a life-threatening influenza infection and who's immune cells produced negligible amounts of type I and

III interferons [289]. Similarly, another group generated hiPSCs from patients carrying a complete Toll-like receptor 3 (TLR3) deficiency presenting with severe influenza pneumonitis [290]. TLR3 acts a sensor for extracellular double-stranded RNA and is a potent inducer of interferons in response to viruses. The generated patient-derived hiPSCs lung epithelial were inoculated with influenza virus. Their findings agreed with the clinical observations, the proportion of viral particles was higher in cells derived from TLR3-/- hiPSCs than for those derived from healthy controls [290]. These studies are examples of how a suitable in vitro models could aid with the understanding of host-pathogen interactions and the development of in vitro interventions to modify the progression of the disease, that could be later translated to the clinic.

Primary tissue-derived alveolar organoids, differentiation of hiPSCs or a combination of these technologies as a model for disease in vitro fail to fully replicate the native environment of the human lung and features like the complex virus-host interactions and the host immune response. Furthermore, these models fail to completely recapitulate the gene expression profiles and functionality of those of their mature counterparts in the human adult lung. However, this technology provides a better replica of the human alveoli than previously developed models, offering a strong platform for the study and understanding of viral replication, pathogenesis and potential targets to develop effective clinical interventions for prevention and treatment in vitro.

5.1.4 Chapter aims and objectives

The thorough optimisation performed in chapter 4 laid the foundation for the objectives outlined in this chapter. The robust differentiation platform developed as a result of the optimisation will now be used to achieve the objectives related to establishing an in vitro infection model for hiPSC-derived AT2 cells. This model will enable a deeper understanding of IPF pathogenesis, focusing on studying the effects of mutant SFTPC and infections in the context of IPF. The following key objectives are outlined:

- To validate the in vitro infection model for AT2 cells as proof of concept. This will be demonstrated by exposing healthy hiPSC-derived AT2 cells to Pseudomonas aeruginosa LPS. The chosen healthy cell line, REBLPAT, allows for AT2 cell generation in 2D culture, eliminating the need for a 3D maturation step facilitating exposure to the infecting agent.
- To adapt the infection methods with hiPSC lines possessing an IPF genetic background, which require a 3D Matrigel droplet culture for mature AT2 cell generation.
- To quantify Influenza A virus strain H1N1 A/Puerto Rico/8/34 viral particles for the successful infection of AT2 cells generated in the previous chapter. The viral titre and infective dose needed for the in vitro disease modelling will be determined using cell-based assays.
- To infect the IPF patient-derived $SFTPC^{WT/MUT}$, $SFTPC^{MUT/MUT}$, and $SFTPC^{WT/WT}$ hiPSCs with the quantified H1N1 virus.
- To conduct a global transcriptomic analysis to compare the profiles of infected and uninfected cells. This will shed light into the physiological characteristics and responses of mutant and wild-type AT2 cells to infections. This data can potentially aid in identifying disease-triggering factors, serving as potential therapeutic targets or diagnostic biomarkers.

5.2 Results

5.2.1 In vitro infection of AT2 cells with Pseudomonas aeruginosa LPS

This section reports, as proof of principle of AT2 cells infection, the in vitro exposure of hiPSC-derived AT2s to Pseudomonas aeruginosa LPS.

First, healthy hiPSC-derived AT2 cells were treated with LPS in an air-liquid interface (ALI). The healthy cell line REBLPAT was used due to its capacity to generate AT2 cells in 2D culture without the need for a 3D maturation step. This way, the infection could be performed without digesting the Matrigel droplets to extract the cells or worrying about viral penetration. The second infection method used was done using hiPSC with an IPF genetic background. These cell lines need to be embedded in 3D Matrigel droplets in order to generate mature AT2 cells. Thus, it was essential to assess if 3D-grown alveolar organoids containing AT2 cells could be infected with LPS without digesting the extracellular matrix surrounding them.

Infection of healthy AT2 cells on an air-liquid interface

The healthy hiPSC line, REBLPAT, was differentiated into lung progenitors in 2D culture conditions for 12 days as per the original protocol previously described. For the maturation stage, the cells were transferred in maturation media to Matrigel-coated hanging cell culture inserts with a pore size of 0.4µm. The following day, the media from the inserts was removed, exposing the cells to the air on the apical surface and the maturation culture media on the basolateral side. These ALI culture conditions allowed the seeded lung progenitors to be in an environment that simulated the airway epithelium physiological conditions (**Figure 5.6A**). The platform promoted the expression of SFTPC and NKX2.1, shown by qPCR and ICC results (**Figure 5.6B, C**). Cells were also stained for the distal lung epithelium markers SOX9 and GATA6, showing co-expression with NKX2.1. This data demonstrated that the ALI culture efficiently promoted the maturation of the healthy differentiated AT2 cells.

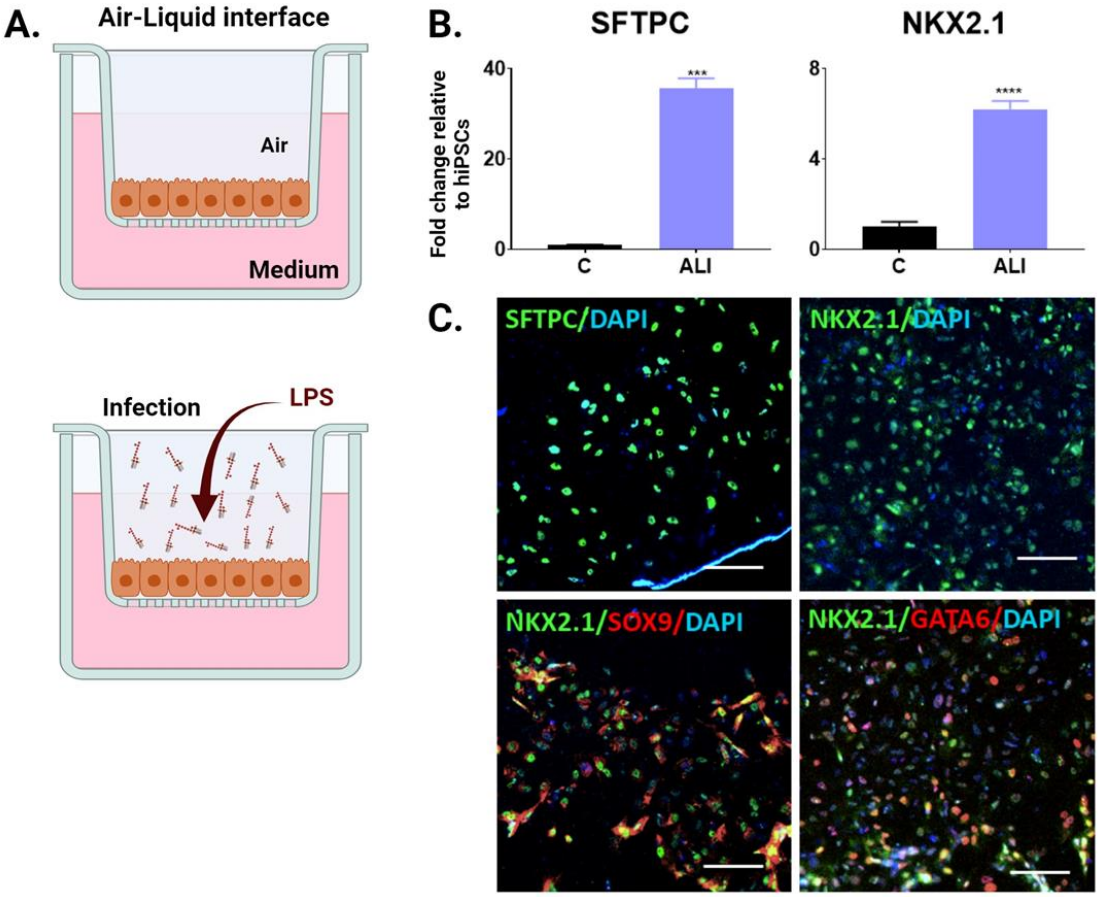


Figure 5.6 Air-liquid interface supports AT2 cell maturation.

A) Schematic representation of the culture platform employed for lung cell maturation and infection. B) qPCR data demonstrated that ALI significantly increased the expression of mature lung cell markers NKX2.1 and SFTPC. C, undifferentiated hiPSC control; ALI, Air-Liquid interface. Data shows mean CT values ± standard error of the mean. Unpaired student's t-test statistical analysis for significance was performed, *** p≤0.001 and ****p≤0.0001. n=1. C) Immunocytochemistry imaging confirmed these results, SFTPC and NKX2.1 (green). Additionally, cells were positively stained for distal lung markers, GATA6 and SOX9 (red). Scale bars 100µm.

In order to test the ability of the REBL-PAT-derived AT2 cells on ALI culture to model respiratory infection, cells were co-cultured with 1µl/ml of Pseudomonas aeruginosa LPS for 48 hours and samples were harvested at 24 and 48 hrs post-exposure. qPCR analysis showed that LPS-induced injury modified the functionality of the AT2 cells, significantly decreasing the expression of SFTPC by >20-fold (**Figure 5.7**). qPCR data showed that TGFβ and NFκB genes significantly decreased after 24 hours. Contrastingly, TNFα and the inflammatory interleukin IL18 were significantly upregulated, along with the anti-inflammatory IL10. Data also showed that, compared to the untreated control, LPS exposure induced a fibrotic response as observed in the

upregulation of mRNA levels of Collagen Type I Alpha 1 Chain (COL1A1), COL1A1 (COL3A1), Fibronectin (FN) and Laminin Subunit Alpha 1 (LAMA1) towards the latest timepoint of 48 hours. Interestingly, these four extracellular matrix components' expressions were significantly down-regulated by the 24-hour infection time point.

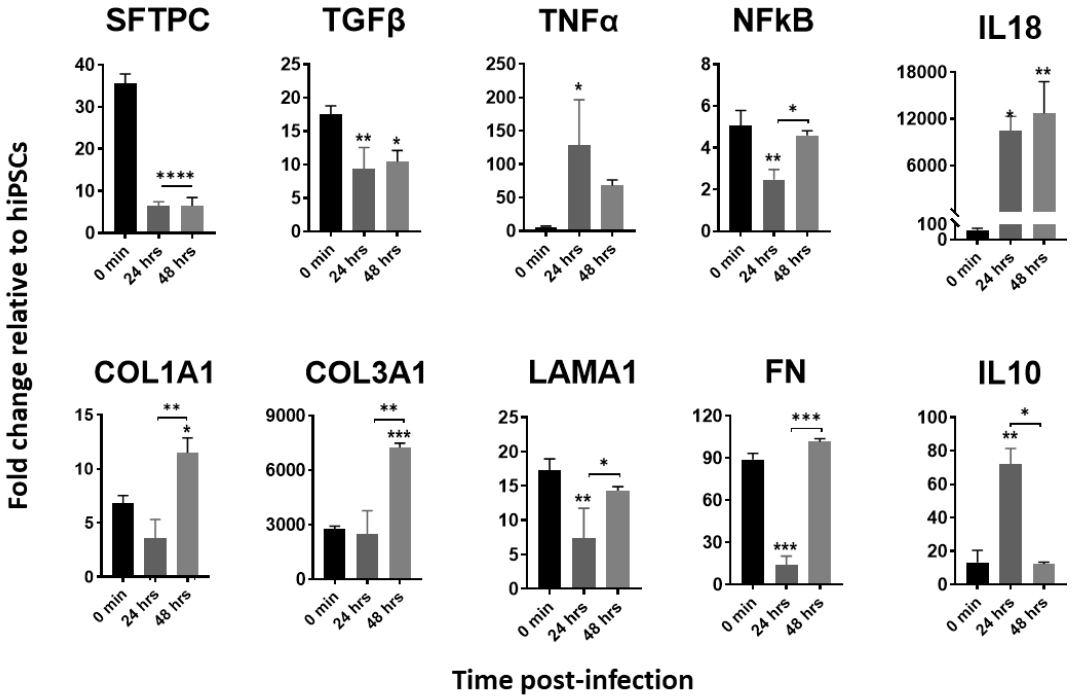


Figure 5.7 Healthy AT2 cells response upon LPS infection.

qPCR shows changes in the expression of SFTPC and inflammatory markers (top row) and fibrotic markers (bottom row) at 24- and 48-hours post-Pseudomonas aeruginosa LPS infection. Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to untreated sample. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

The host-defence response activated by Pseudomonas aeruginosa LPS and its receptor, Toll-like receptor 4 (TLR4), was further assessed using a membrane-based antibody array. AT2 cells were infected with Pseudomonas aeruginosa LPS for 120 minutes, and samples were harvested at 0, 30, 60 and 120 minutes of exposure. The densitometry analysis of the membranes revealed an overall significant increase in the phosphorylation levels of TOR, similar to what was observed for GSK3β, the 30-minute infection induced a significant increase in the phosphorylation levels of ERK1 and ERK2. However, the trend changed dramatically after this time point, with a significant decrease towards the 120-minute exposure. Compared to the untreated control, the 120-minute LPS exposure significantly decreased P53 phosphorylation levels (**Figure 5.8**).

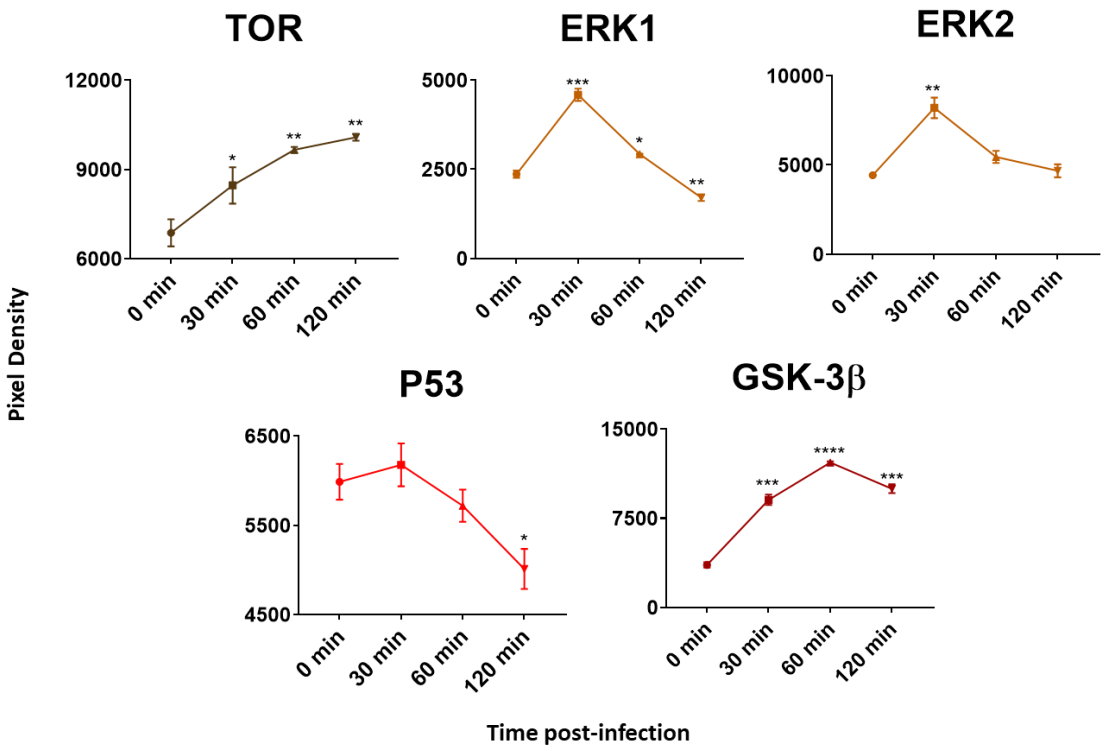


Figure 5.8 LPS activates TLR4 signalling pathways in AT2 cells.

Graphs of AT2 cells' relative phosphorylation levels, based on pixel density, for protein kinases downstream TLR4 after treatment with 1µl of LPS for 30, 60, and 120 minutes before lysis, n=1. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001.

Infection of mutant AT2 organoids

Two SFPTC$^{WT/MUT}$ cell lines (IPF1 and IPF5) were differentiated into lung progenitors for nine days using the original differentiation protocol. Afterwards, the cells were embedded in 3D Matrigel droplets to induce their differentiation and maturation into alveolar organoids containing AT2 cells. The cells were maintained as organoids for approximately 50 days, passaging them every 7-10 days with no morphological differences detected between cell lines (**Figure 5.9**). Organoids cultured in Matrigel droplets were treated with 1ul/ml of Pseudomonas aeruginosa LPS for 2 hours. Samples were harvested 0-, 30-, 60-, and 120-minutes post LPS exposure to assess the cellular response to the infecting agent.

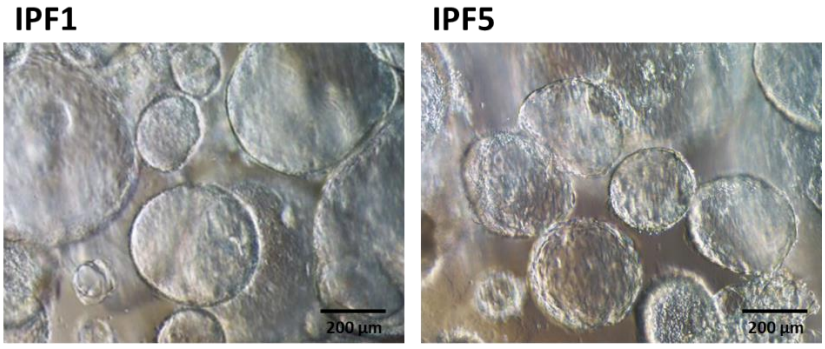


Figure 5.9 IPF-mutant hiPSC-derived alveolar organoids in culture.

Brightfield images of alveolar organoids differentiated from two different hiPSC IPF-mutant cell lines at day ~50 of culture in 3D Matrigel droplets before infection with Pseudomonas aeruginosa LPS infection. Scale bar= 200µm

The infection with LPS induced a significant loss in expression towards the 120-min time point of the lung marker NKX2.1 in IPF1 cell line; contrastingly, no significant changes in expression were observed in IPF5 as demonstrated by qPCR results in **Figure 5.10**. The same was observed in genes related to inflammatory pathways, such as TGFβ, TNFα and the Interferon$^{\gamma}$ inducing factor, IL18. The anti-inflammatory cytokine IL10 followed a similar trend, but no significant changes were identified. Interestingly, the stress rapid-acting transcription factor NF^{k}B revealed a rising trend in both cell lines; still, not statistically significant. The rest of the inflammatory cytokines in **Figure 5.10**, IL8, IL1B, and IL23, were also downregulated in IPF1 towards the end of the LPS exposure. However, IPF5 showed a surprisingly significant increase of IL8 and IL23 at the 60min-infection time point.

Similarly, when analysing the gene expression of fibrotic markers, no significant differences were identified for COL3A1 and FN for either cell line and the expression trends conflicted between IPF1 and IP5. On the other hand, COL1A1 significantly increased after 120 min of LPS exposure in IPF1. Moreover, LAMA1, a crucial extracellular matrix element that is highly expressed in the lungs of patients with IPF and has been identified as a genetic modifier of TGFβ1–induced pulmonary fibrosis in mice, showed a significant downregulation at the 60-minute time point when exposed to IPF5 (**Figure 5.11**)[291].

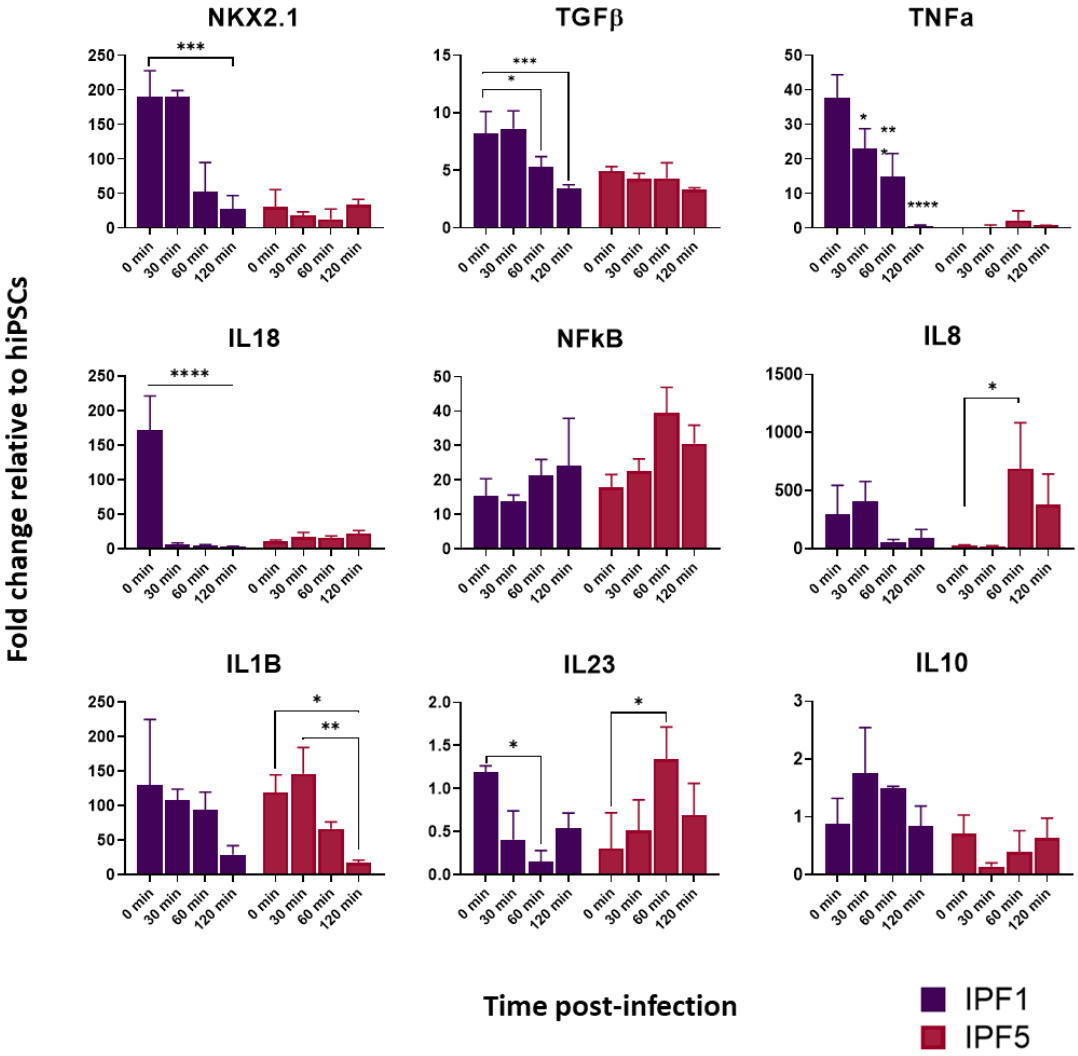


Figure 5.10 Effects of alveolar organoids exposure to Pseudomonas aeruginosa LPS.

Gene expression profile of IPF1 and IPF5-derived alveolar organoids infected with 1ul/ml of LPS while embedded in Matrigel droplets. Inflammation markers were assessed at 0-, 30-, 60- and 120-minutes post-infection using qPCR. Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to untreated samples. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

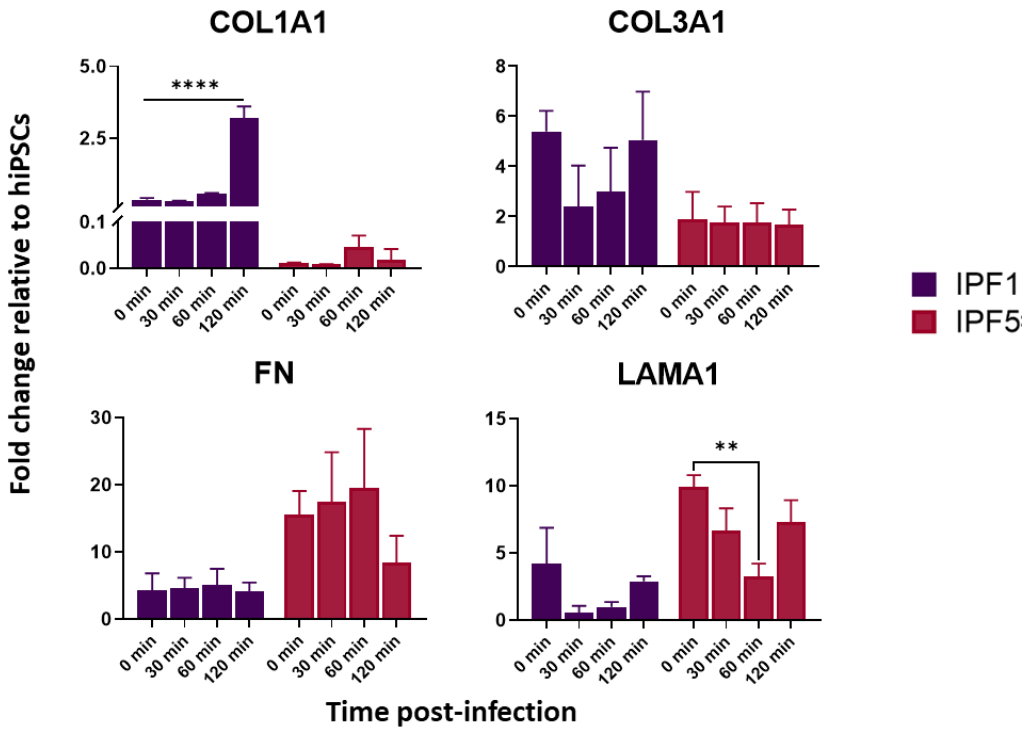


Figure 5.11 Alveolar organoids 2-hour LPS treatment induces the expression of collagen 1.

qPCR of COL1A1, COL3A1, FN and LAMA1 at 0, 30, 60 and 120 minutes after LPS exposure of organoids embedded in 3D Matrigel droplets. Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to untreated samples. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001. n=1.

Stressing stimuli, such as infectious agents, can cause cells to accumulate unfolded or misfolded proteins in the endoplasmic reticulum, leading to ER stress. In response to this, the cell activates the unfolded protein response (UPR) as a protective mechanism to restore ER homeostasis by improving protein folding, processing, and degradation. To assess if exposure to LPS triggered UPR on the infected SFPTC$^{WT/MUT}$ cell lines, a series of genes essential to the UPR pathway were measured (**Figure 5.12**).

The ER stress-sensing chaperon molecule binding immunoglobulin protein (BiP) or GRP78 showed a statistically significant downregulation towards the 120-min LPS infection in IPF1. This downregulation was also significant in IPF1 expression of the UPR signalling activators inositol requiring enzyme 1 (IRE1) and activating transcription factor 6 (ATF6). Activation of IRE1 leads to the downstream splicing of the target UPR regulator X-box binding protein 1 (XBP1). While no significant differences were found in the IPF1 levels of total XBP1 (tXBP1), the unspliced form (usXBP1) was significantly downregulated at the 120-min time point, which inversely correlated with a significant increase in the spliced form of XPB1 (sXBP1). The

increased levels of sXBP1 towards the end of the infection might be due to the activity of IRE1 at the earlier time point. The splicing of XBP1 promotes the upregulation of UPR target genes like the ER degradation-enhancing mannosidase-like protein (EDEM). Levels of EDEM were significantly downregulated in IPF1 after 120min of LPS exposure; contrastingly, significant upregulation was observed at the same time point for IPF5. EDEM, together with BiP, were the only UPR markers showing a significant increase in mRNA levels after infecting IPF5 with LPS. Even though IPF5 show relatively higher levels in most UPR markers compared to those expressed by IPF1, the trend remained relatively stable within time points without significant differences apart from the two previously mentioned, discarding a more dominant response. The overall gene expressions of IPF1 and IPF5 UPR markers suggest that no ER stress pathways were significantly activated upon LPS exposure for 120 minutes.

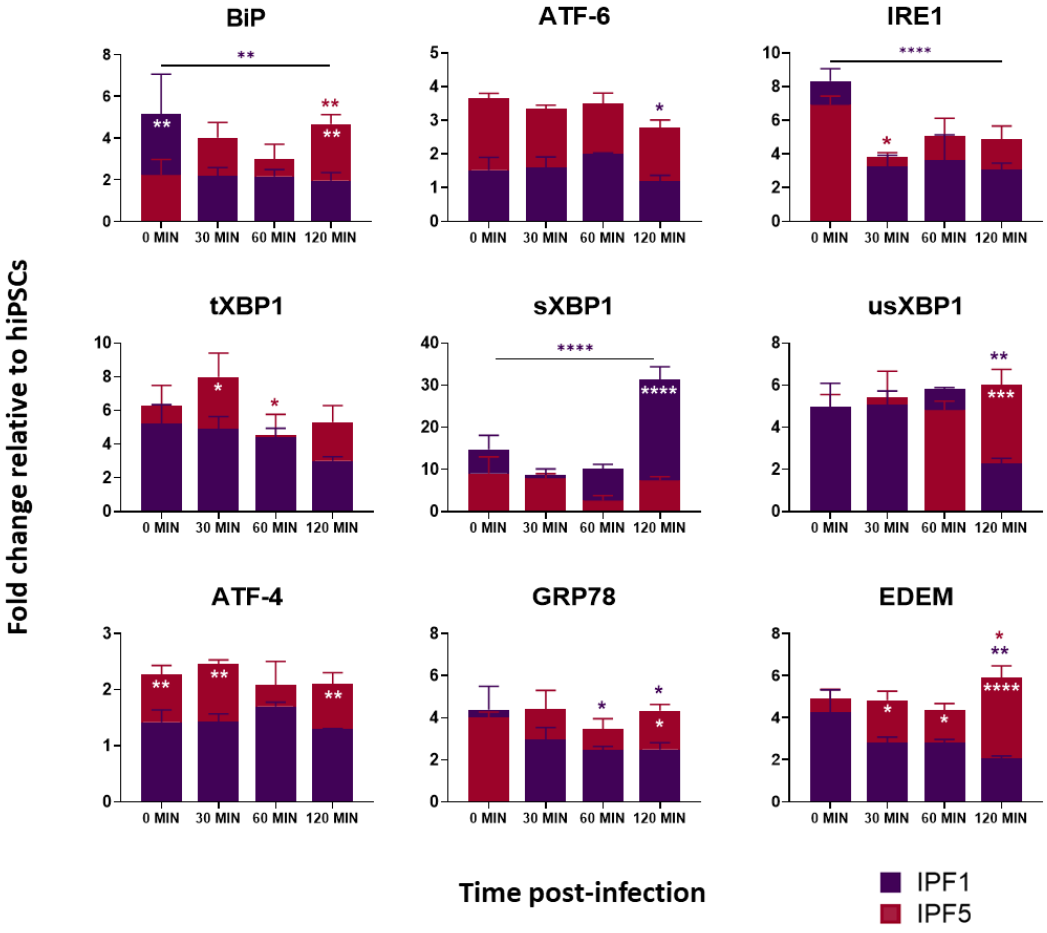


Figure 5.12 LPS treatment for 2 hours did not induce an unfolded protein response on alveolar organoids.

Transcript profile of unfolded protein response regulatory genes including stress-sensing molecule binding immunoglobulin protein (BiP), activating transcription factor 6 (ATF6),

inositol-requiring 1 (IRE1), total, spliced, and unspliced X-box binding protein-1 (total, tXBP1; spliced, sXBP1; unspliced, usXBP1), ATF4, glucose-regulated protein 78 (GRP78) and ER degradation-enhancing mannosidase-like protein (EDEM). Data shows the mean CT values ± standard error of the mean. One-way ANOVA was performed for statistical analysis for significance in comparison to untreated samples. White asterisks indicate statistically significant differences between IPF5 and IPF1, n=1. *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001.

In summary, these conflicting data demonstrate that AT2 cells generated an inflammatory response when exposed to LPS in a 2D platform; however, when the cells were treated in an organoid structure, the cellular response was vague. A possible reason for this lack of response is the impaired penetration of the LPS to the core of the organoids cultured within the droplets in a short period. Although increasing the exposure time may induce a more significant response, it does not ensure that all the cells within the organoid culture will be exposed to the infecting agent to the same degree. Therefore, 2D culture was adopted as the preferred infection method for future experiments, circumventing the need to optimise an infection protocol where the number and size of the organoids infected could be controlled to generate reproducible and reliable results.

5.2.2 Influenza A (H1N1) virus propagation and titration

As stated previously, Influenza A (H1N1) virus is one of the most prevalent infectious agents found in the lungs of IPF patients, especially in the context of acute exacerbations. Therefore, infecting the AT2 cells generated with the optimised protocol described in Chapter 4 with the H1N1 virus was paramount for this thesis.

Replication of virus in MDCKs

For the reproducibility of viral in vitro infections, it is essential to know the concentration of virus particles present in the initial virus suspension and such quantitation can be done through diverse methods. To generate the initial virus suspension, influenza virus strain H1N1 (A/PR/8/34) was replicated using MDCK cells. The cells were cultured in MEM serum-free medium supplemented with 1 µg/ml TPCK-trypsin and infected with 5 HAU/µl influenza (Public Health England). In vivo, proteolytic cleavage of the HA protein of the replicating virions by serum proteases is required to continue with the virus infection cycle [292]. This allows the fusion of the viral and host cell membranes and further internalisation of the viral RNA in the cell. In vitro, this was accomplished by adding exogenous trypsin to the cell culture in serum-free conditions to avoid the inactivation of the enzyme. After three days, the supernatant containing the replicated virus was harvested, aliquoted and kept at -80°C for later quantitation.

HA assay optimisation

The viral stocks obtained from the influenza A (H1N1) virus replication using MDCK cells were titrated using the HA assay. The initial step of the experiment involved determining the optimal concentration of chicken RBCs required for button formation in the absence of the virus. This was accomplished by evaluating four different conditions. Firstly, two dilutions of 0.5% RBCs were tested - one using the RBCs stock directly without washing, and the other with a washing step to remove any remaining Alsever's solution (an anticoagulant). Additionally, two different dilutions of 1% and 2% RBCs were examined, both of which underwent the washing process. After preparing the four different solutions, the experimental procedure dictated the addition of 50µl of two-fold viral serial dilutions to each well.

Figure 5.13 shows the HA assay plates using these four conditions. When using washed and non-washed 0.5% RBCs, no "buttons" were observed even at the highest viral dilutions and no-virus control. This indicated that a higher concentration of RBCs was needed for the assay. However, the same results were obtained when doubling the RBCs concentration to 1%. When using 2% RBCs, the no-virus control showed no haemagglutination, and a well-defined mass of cells settled at the bottom of the wells. This showed that the number of RBCs was enough for the cells to form visible sediment on the plates. Therefore 2% RBCs was appropriate to be used for the HA assay. Using this concentration to determine the titre of the virus, haemagglutination could be observed up to 1:128 dilution. Since the assay started with 50µl of virus stock, the viral titre in haemagglutination units (HAU) was 128 HAU per 50µl or 2.56 HAU per µl.

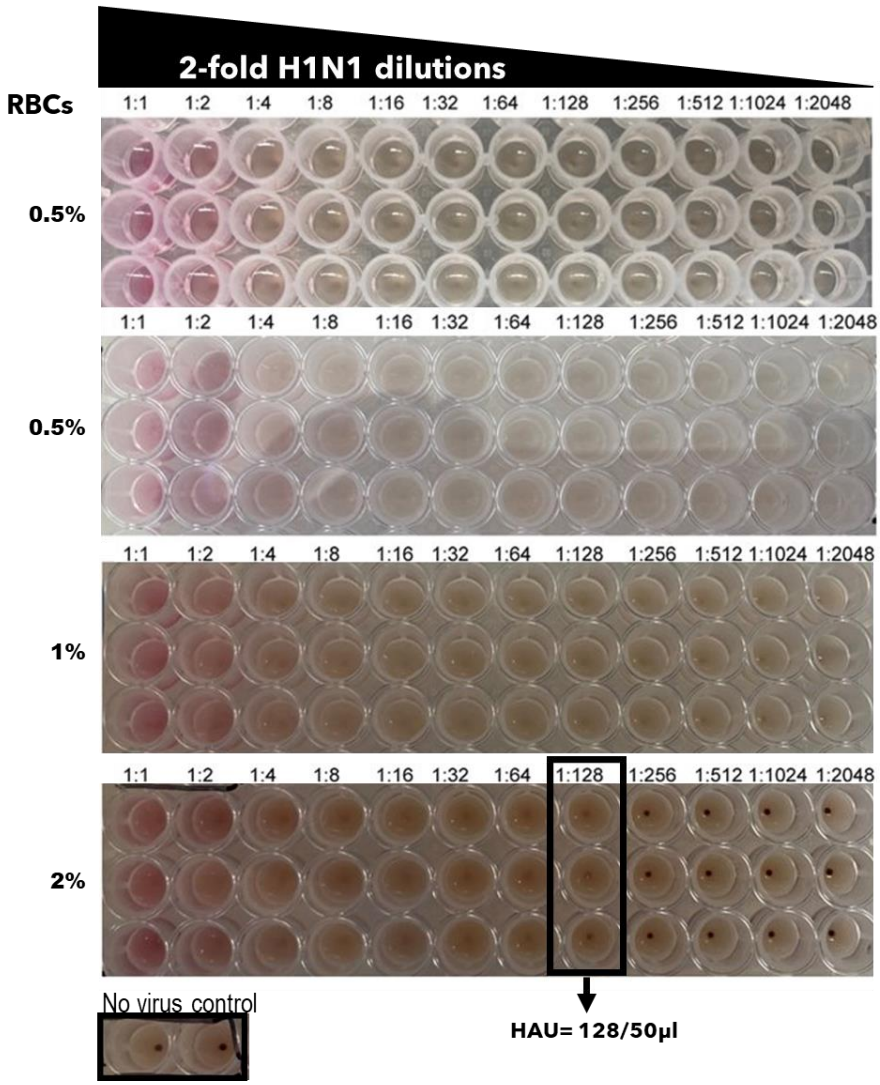


Figure 5.13 Influenza A (H1N1) virus haemagglutination assay optimisation.

Doubling viral dilutions were mixed with different concentrations of chicken RBCs in triplicates. RBCs concentrations are designated on the plates' left side, and the virus dilution factor is designated across the top. The top three plates with RBCs concentrations of 0.5% and 1% show failed attempts of the haemagglutination assay as no lattices or "buttons" are observed. The bottom plate with 2% RBCs shows positive haemagglutination 128 HAU/50ul, and no signs of incomplete haemagglutination are observed.

Plaque assay optimisation

After confirming the presence of high titre virus in the initial suspension, a plaque assay was carried out as it is known as one of the most accurate and sensitive forms of assaying virus infectivity [268].

Thawed influenza virus (A/PR/8/34/H1N1) stock was inoculated in 10-fold dilutions in triplicates, on MDCK cells as described in chapter 2. MDCK cells were seeded at a density of $9x10^4$ and $4.5x10^5$ cells/well, to determine the optimal density to proceed

with the plaque assay. Wells seeded with the lowest cell density were too sparse and did not form a monolayer until 3 days after seeding. **Figure 5.14** shows cells seeded at the higher density 24 and 48hrs post-seeding. Images demonstrate that cells had grown into a confluent monolayer by the 48hrs post seeding, as recommended by the protocol, and were ready to be infected. The cells were treated with the virus dilution for a period of 2hrs. Following this, a solid overlay was added to limit the CPE effects caused by the virus to the initial site of infection. Only three days after the treatment, CPE were visible in the high concentration viral dilution of 1:2. However, no CPE or any morphological changes were observed at the higher 10^{-3} dilution.

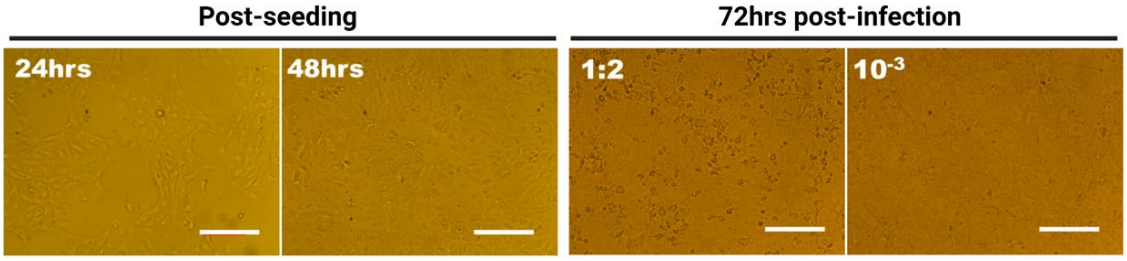


Figure 5.14 MDCK cells seeding for Influenza A (H1N1) virus plaque assay.
Brightfield images of MDCK cells seeded into 6-well plate format at $4.5x10^5$ cells/well and infected with influenza virus at 1:2 and 1^{-3} dilution. Scale bar= 200µm.

The appearance of CPE on the experimental dilutions was crucial to proceed with the next step of the assay, which consisted of staining the wells to visualise the formation of plaques. After 6 days, no plaques were observable by the naked eye, thus the wells were stained with Neutral Red (**Figure 5.15**). Each experiment set included a negative control, 1:2 and 1:10 dilutions as highly concentrated controls, apart from the experimental 10-fold dilutions. Macroscopically, no clearly defined plaques were observable with the aid of the lightbox on any of the experiments, despite of the staining. The negative control and the 1:10 dilution wells had a very similar hazy aspect to that of the rest of the wells, making the identification of plaques impossible. In contrast, the well with the highly concentrated 1:2 virus dilution appeared to be translucent, with no plaques at all.

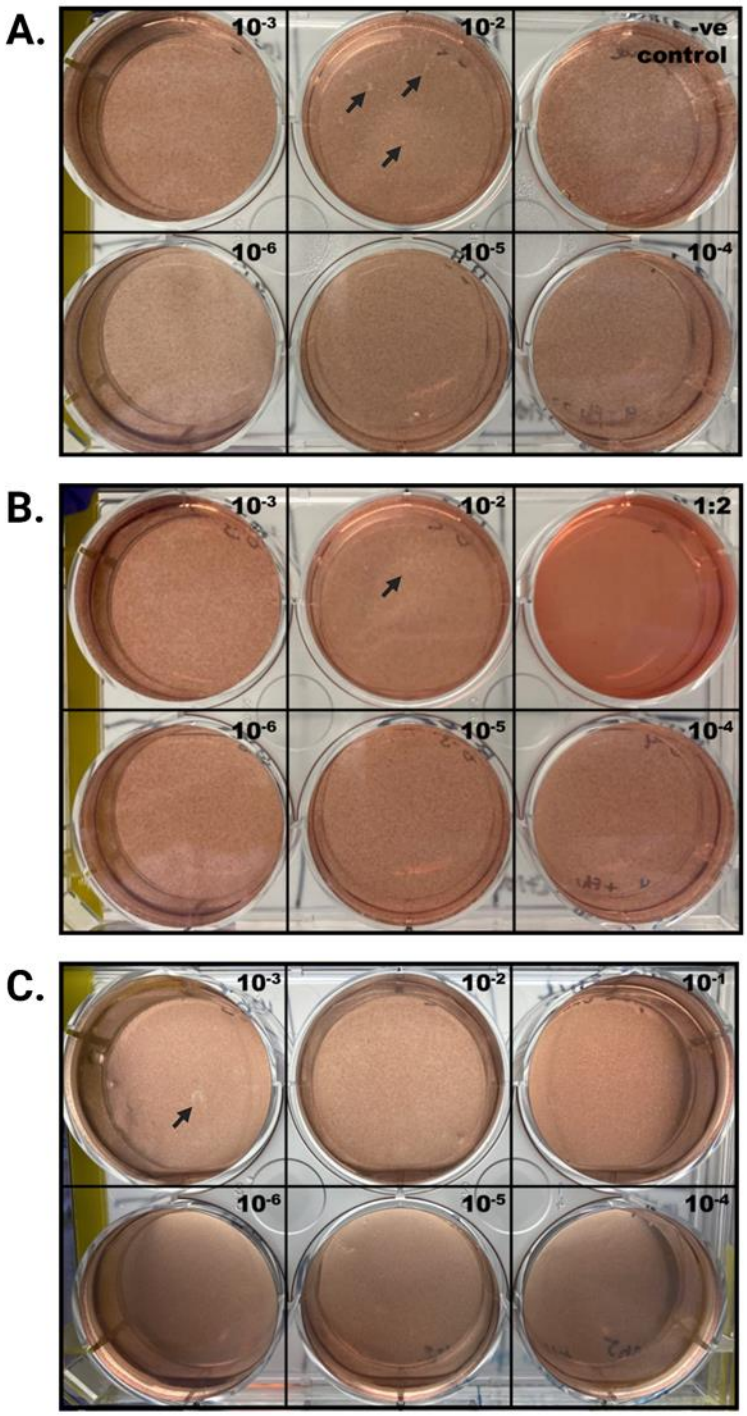


Figure 5.15 Neutral Red staining of influenza A(H1N1) virus plaque assay.

Representative images of 3 different plaque assay experiments using a lightbox for the visualisation of Neutral Red stained wells. Plates show a range of influenza virus dilutions going from 10^{-2} to $10^{-6.}$ Top right corner of the plates shows A) negative control, B) 1:2 dilution and C) 1:10 dilution. Arrows indicate faint plaques with diffuse borders.

This was further confirmed when looking at the wells under the microscope (**Figure 5.16**). The totality of the cells on the negative control had captured the Neutral Red staining, whereas the positive 1:2 control had killed all the cells in culture, as no absorption of stain was observed. The 10^{-2} dilution was enough to infect and kill the cells; however, the pattern produced by the dead cells was remarkably diffuse and no plaques were distinguishable. On the 10^{-3} dilution, some of the areas looked promising, with defined borders between zones of live and dead cells, still, the rest of the well had a diffuse pattern of cell death. The higher dilutions were not enough to cause visual CPE.

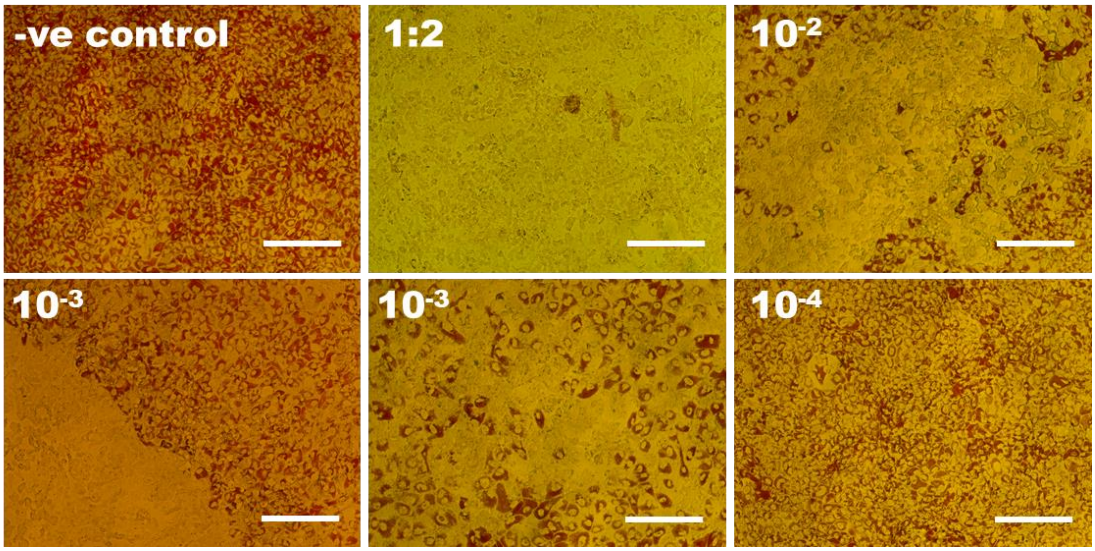


Figure 5.16 Influenza A(H1N1) virus plaque formation.

Brightfield images of MDCK cells stained with Neutral Red 6 days after inoculation with influenza virus to visualise plaques not visible to the naked eye. Scale bar= 200µm.

Together, these results confirmed this method was not suitable for the downstream quantification of the virus. Further optimisation would be needed to finesse the appearance of defined plaques. While the diffused plaques were not quantifiable to determine the viral titre, it was evident that the influenza virus at higher concentrations had killed all or most of the cells within the inoculated well. This was enough evidence to confirm the presence of viable infective virus in the initial viral suspension. Therefore, adopting a different technique to quantify the infective influenza particles was necessary.

Viral Tox Glo assay

To determine the tissue culture infective dose (TCID) of the initial influenza A virus suspension, ATP depletion was measured using the viral Tox-Glo assay, so it could be correlated with the viral burden.

MDCK cells were seeded in an opaque-walled 96-well plate at $3x10^4$ cells per well [293]. After becoming confluent, 24 hours later, the cells were inoculated with virus stock dilutions of 1:10 or 1:100 and subsequent half-log serial dilutions, as per the protocol. Cells were incubated in the presence of the virus until CPE were observed. The infected MDCK cells showed minimal morphological differences and no CPE compared to the untreated control, upon 24-hour infection (**Figure 5.17**). Contrastingly, 48 hours post infection, the cells started to detach, and floating debris was observed in the media, whereas the untreated control showed no signs of cell death. These observations only became more evident towards the 72-hour infection timepoint.

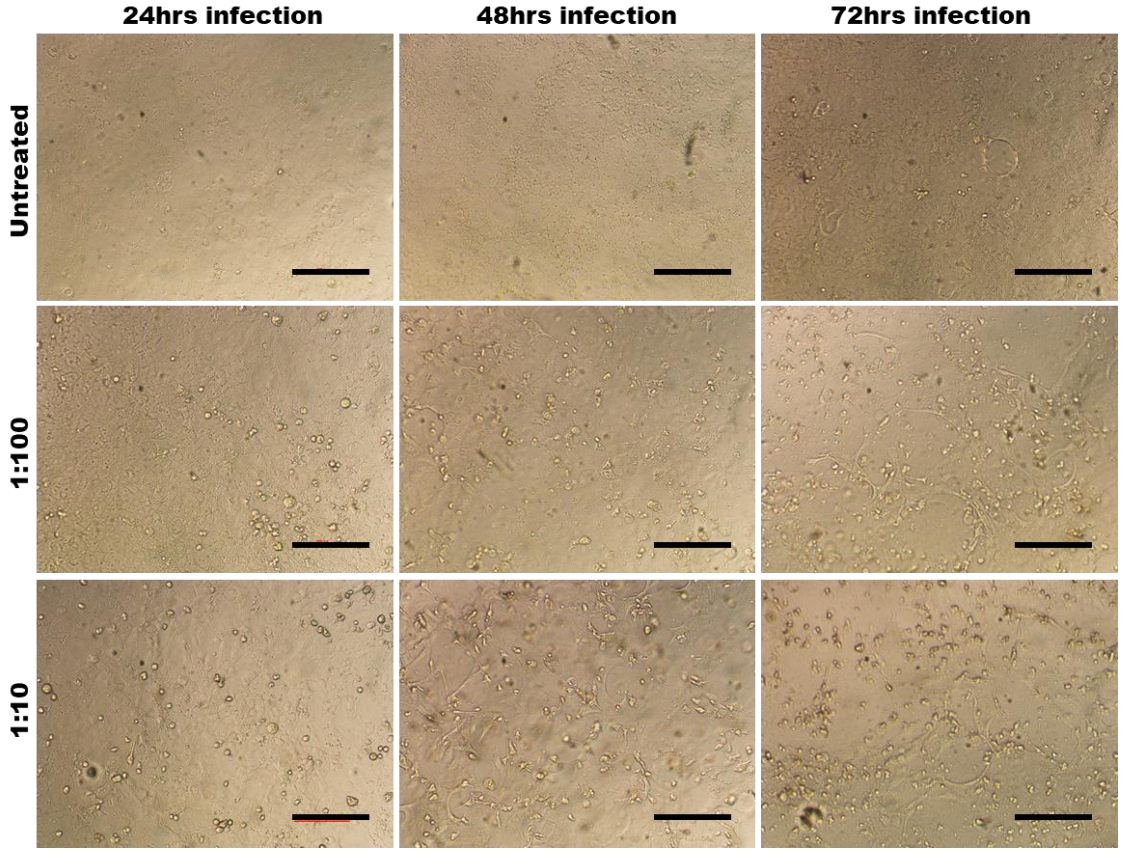


Figure 5.17 Morphological changes on MDCK cells at different influenza inoculum concentrations.

Brightfield images of MDCKs 24, 48 and 72 hours after treatment with influenza virus at 1:100 (middle row), 1:10 (bottom row) dilutions and untreated control (top row). Scale bar= 200µm.

After identifying CPE in most wells, cells were incubated with the ATP detection reagent for 30 minutes, followed by the bioluminescence measurement using a luminometer. The luminescence data, minus the average of the no-cell control wells obtained from the assay, was plotted against the influenza viral dilution factor used per well to calculate the infectivity of the virus. GraphPad Prism fitted these values by a dose-response, special (asymmetrical) algorithm. Therefore, the $TCID_{50}$ value identified corresponded to the reciprocal of the dilution that produced a 50% decline in ATP levels compared to untreated controls. Applying the equation of the sigmoidal curve, it was determined that the amount of virus required to induce CPE in 50% of the virus-inoculated cells ($LogIC_{50}$ or $TCID_{50}$) was 9,178-fold dilution, and the purple dotted line represents it in the graph of **Figure 5.18**.

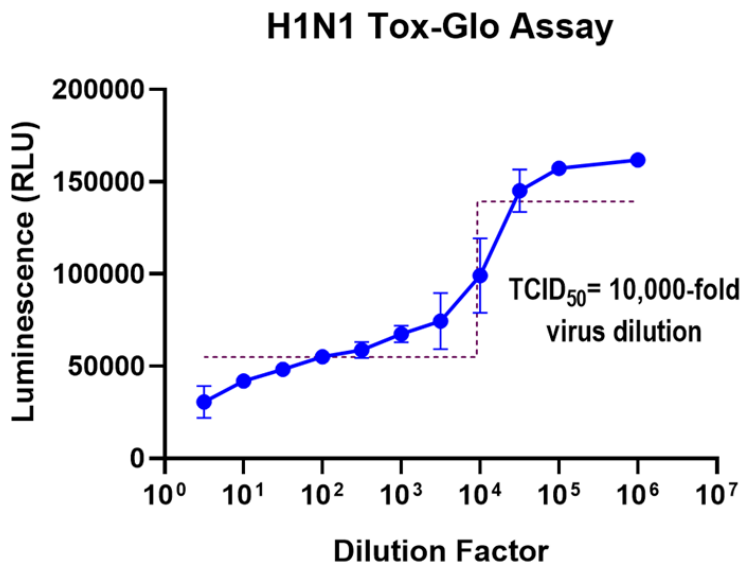


Figure 5.18 Calculation of influenza virus $TCID_{50}$ value using Tox-Glo Assay.

Dose-response curve obtained from the 72-hour inoculation of MDCK cells with half-log serial dilutions of influenza virus. An ATP detection reagent was added to the wells, and the generated luminescence was measured. The raw luminescence (y-axis) was plotted versus the virus dilution factor (x-axis), and the LogIC50 value revealed the virus concentration that gives the half-maximal effect, infecting 50% of the cells in the culture.

Using the obtained $TCID_{50}$ of 10^4-fold dilution (rounding up the sigmoidal curve equation's result) and including the initial 1:10 viral dilution used for the ToxGlo Assay in the calculation, the final $TCID_{50}$ for the viral stock was determined to be a 10^5-fold dilution. Multiplying this value by 0.7 the PFU present in the 46µl of viral dilution used in the assay was calculated as 70,000 PFU per 46µl or 1,521,739 PFU per mL ($PFU = 0.7xTCID50$). The PFU obtained could then be used to calculate the MOI for subsequent experiments by dividing the PFU value by the number of cells to be

infected ($MOI = \frac{PFU}{total\ amount\ of\ cells}$). In the case of hiPSCs-derived AT2 cells, the infections were performed by seeding 5x10^5 cells and using 300µl of initial viral stock. The PFU in 300µl was determined to be 456,521, resulting in an MOI of 0.91 (MOI=~1) when divided by the number of cells.

TMLC- iHBECs Co-culture Assay

In order to indirectly measure the influenza virus infection of epithelial cells, the TMLC co-culture bioassay was performed to quantify the dose-dependent increases in TGFβ concentration in response to the injury. Before proceeding to the inoculation of the cells with influenza A (H1N1) virus, the assay was first conducted using TGFβ and TGFβ agonists (LPA and Poly I:C). For this, iHBECs, an immortalised epithelial cell line that expresses αvβ6 integrins in vitro, was seeded at a density of 1x10^4 cells per well. The cells were grown to confluency for 72 hours. Subsequently, TMLC cells, a highly specific TGFβ reporter cell line, were seeded on top of the iHBECs and left to adhere for an hour. Once the co-culture was established, cells were inoculated for 16 hours with TGFβ, LPA and Poly I:C, as described in Chapter 2. The next day, the luciferase substrate was added to the wells to detect the TGFβ activation of the TMLCs reporter via the emitted luminescence. This was measured using a luminometer, and the TGFβ-dependent luciferase activity values were plotted by subtracting the data from the unstimulated TMLC controls to produce their standard curves (**Figure 5.19**).

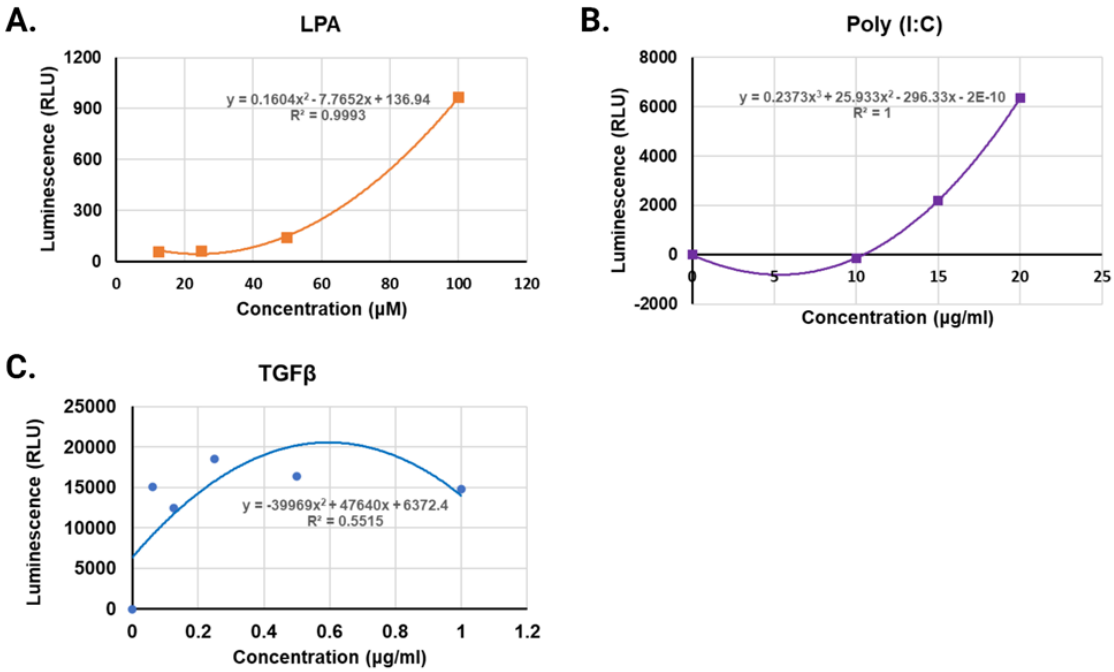


Figure 5.19 Quantification of TGFβ activation on TMLC-iHBEC co-culture assay.

Standard curves determine the dose-dependent increase in TFGB concentration (ug/ml) plotting the analytical luminescence (RLU) signal emitted by TMLCs upon stimulation with known concentrations of A) LPA, B) Poly (I:C), and C) TGFβ. R-square values are indicated on each graph.

The results obtained using LPA and Poly I:C demonstrated that they induced dose-dependent increases in TGFβ concentration, as shown by their standard curves with an R-squared above 0.98 (**Figure 5.19A-B**). However, when adding TGFβ, the standard curve generated was not precise, with an R-squared value under the ideal >0.95 (**Figure 5.19C**). For the assay to be applicable, an established TGFβ standard curve must be used to plot the luminescence values obtained after the influenza virus infection. This is to determine the concentration of TGFβ produced by the virus infected AT2 cells. Without a suitable reference, the absorbance results obtained after the viral infections would not be interpolated on the graph, and no TGFβ concentration would be determined. In practical terms, the system could be used to co-culture TMLCs and hiPSC-derived AT2 cells. Nevertheless, the stimulation with TGFβ required additional optimisation to produce a reliable TGFβ standard curve, but given the limitations on time, this path was not pursued. Therefore, the TMLC co-culture assay for detecting TGFβ activity was not fit as an indirect measurement of viral infection in AT2 cells in this thesis.

5.2.3 Optimisation of AT2 cells monolayer infection

The experimental format destined to work in vitro with influenza A (H1N1) throughout this thesis has been 2D cell culture which relies on growing the cells in monolayers. Moreover, the results obtained with the LPS infection in a 3D format were questionable. Thus, being able to infect the generated AT2 cells in 2D was an important consideration. Therefore, it was crucial to investigate if the AT2 cells grown and maintained in 3D culture conditions could be transferred to 2D culture without losing their characteristic phenotype.

Alveolar organoids growing in 3D-Matrigel droplets were dissociated into single cells and prepared for EpCAM and NaPi2b, the surrogate marker for $SFTPC^+$ cells, FACS sorting. The cells coming from the double positive gate were sorted and seeded in a 48-well plate at a density of $1x10^4$ cells/well (**Figure 5.20A**). After 24hrs in 2D culture, the cells were fixed and stained as described in Chapter 2, to ensure that changing the culture conditions did not alter the expression of the exclusive marker of AT2 cells, SFTPC. Immunocytochemistry results revealed that most of the Napi2b sorted cells retained the epithelial marker ECAD as well as their SFTPC expression as shown by the pro-SFTPC and SPC staining. However, while some of the cells were strongly positive for NaPi2b staining, the culture did lose expression of this surface marker after being grown in 2D culture (**Figure 5.20B**). Taken together, these results confirmed that the optimised differentiation protocol from Chapter 4 was effective at generating putative AT2 cells using isogenic hiPSC-IPF lines, and if transferred to 2D-culture conditions, they retain the expression of AT2 cell markers. These cells can then be transferred to monolayer cultures to proceed with the influenza virus infection experiments and provide a more straightforward, consistent infection modelling.

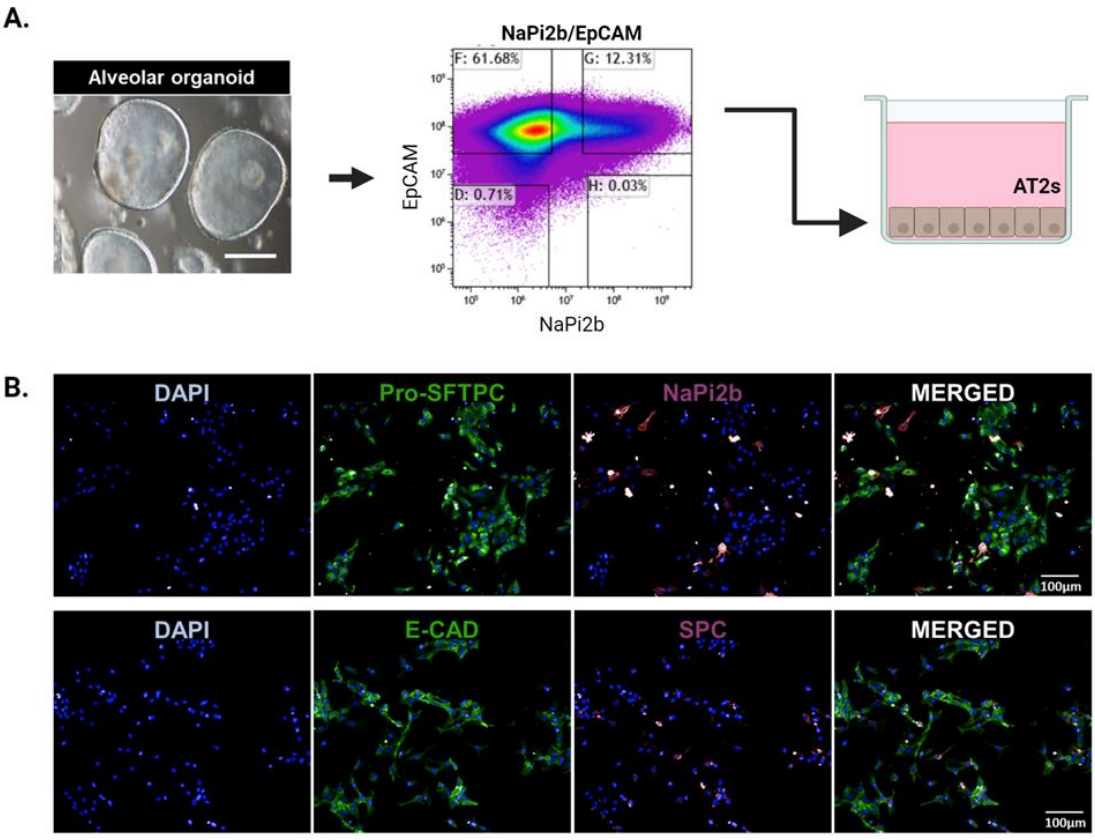


Figure 5.20 3D-generated AT2 cells retain phenotype after being transferred to 2D culture conditions.

A) Alveolar organoids containing SFTPC$^{WT/WT}$ AT2 cells were FACS sorted for NaPi2b^{+}/EpCAM^{+} cells, and the obtained single cells were seeded in a 48-well plate in 2D culture. B) Immunostaining images of AT2 cells 24hrs after seeding. The top row shows cells co-stained for Pro-SFTPC (green) and NaPi2b (red), the bottom row shows co-staining of ECAD (green) and SPC (red), nuclei stained with DAPI (blue). Scale bar= 100µl.

After verifying that AT2 cells maintained the expression of SFTPC after the transition to 2D culture, it was then necessary to determine the optimal conditions to perform the in vitro infections using monolayer cultures. The key parameter to establish was the AT2 cells seeding density. When seeding the cells to confirm the expression of SFTPC, it was noted that cells were too sparse, and after 24 hours, they were only ~50% confluent. Hence a higher seeding density was assessed. Dissociated single SFTPC$^{WT/WT}$ AT2 cells were seeded in a 96well plate at a density of 1x10^{5}cells/cm^{2} in maturation media with Rho-Associated Kinase Inhibitor (ROCKi). Cells were incubated for 24hrs for attachment before infecting them. Subsequently, cells were inoculated for 2 hours with 100µl of two different influenza virus dilutions, 1:50 and 1:100. The virus dilution was then replaced with maturation media without ROCKi. **Figure 5.21** shows the cells before and 24, and 48 hrs after the infection with the two dilutions used. Just

before the inoculation, it was observed that even if the cells were attached to the wells, they were not expanding and were too sparse to form a monolayer. 24hrs after, the control and the inoculated cells were balling up and detaching from the well. By the 48hrs post-infection, all cells were dead and floating on the culture media. Cells and supernatant were harvested at 24 and 48hrs time points; however, the RNA obtained was insufficient for reverse transcription and further gene expression analysis. The findings from this infection experiment suggest that the seeding density needed to be further increased and the time of infection adjusted to allow the cells to expand before inoculating them with the virus.

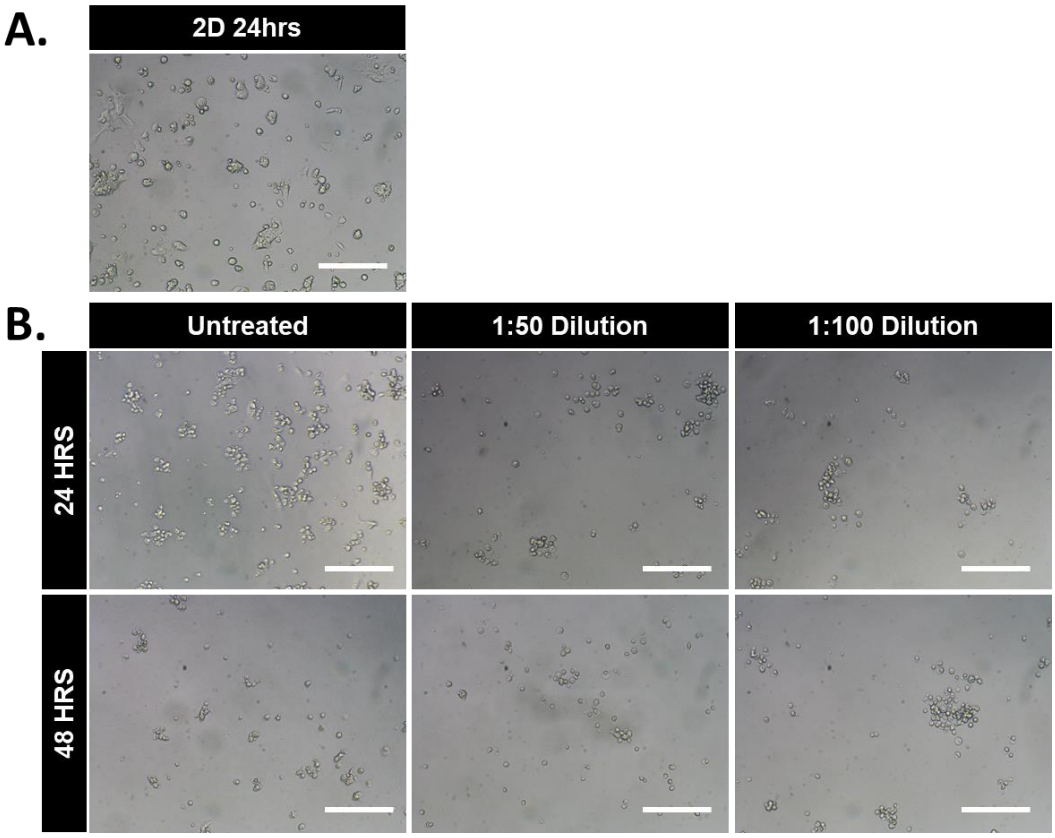


Figure 5.21 Initial seeding density assessment of wild-type AT2 cells for infection purposes.

A) Brightfield image of SFTPC$^{WT/WT}$ AT2 24hrs after seeded in 2D at 30,000 cells/well. B) AT2 cells imaged 24 and 48hrs after inoculated with 1:50 and 1:100 H1N1 viral dilution and compared to untreated cells. Scale bar= 200µm.

Before optimising the AT2 cells seeding density for infection purposes, it was decided to change from 96-well plates to 12-well plates in order to obtain enough RNA for downstream analysis after the infections. For this, SFTPC$^{WT/WT}$ AT2 cells were seeded at 1.5x10^5 cells/cm^2 in a 12-well plate. However, instead of using ROCKi, the

maturation media was supplemented with CloneR2 when seeding the cells in 2D culture conditions. CloneR2 was beneficial when seeding single AT2 cells coming from the cell sorter into 3D-Matrigel droplets for organoid generation. When seeding the cells in 2D, it was noted that 24 hours was insufficient for the AT2 cells to expand into a monolayer. Instead, within this time frame, the cells only managed to attach themselves to the surface of the culture plate, without much proliferation. Hence, the cells were left for 48hrs before proceeding with the influenza virus inoculation (**Figure 5.22A**).

This time, the untreated cells grew into a complete monolayer by 48hrs, indicating that the seeding parameters were adequate for the experiment. For the purpose of obtaining a 48hrs post-infection sample to analyse a more prolonged period infection without having killed all the cells, a comparison of MOI 1 and MOI 3 was made. When infecting the cells using MOI 1, floating debris was observed from 24hrs post-infection with zones of disrupted cell monolayer underneath. 24hrs after, most of the cells were floating in the media. Contrastingly, when MOI 3 was used, only a few cells were still attached to the plate 24hrs post-infection, and only debris was observed by the end of the experiment (**Figure 5.22B**). Samples were collected 24 and 48 hrs post-infection for MOI 1 and only 48hrs post-infection for MOI 3. Cells infected with MOI 1 for 48hrs had higher mRNA transcripts of inflammation-related genes, including TNFα, Nuclear Factor Kappa B (NFkB), IL6 and IL-1B, which were significantly upregulated (**Figure 5.22C**). Interestingly, TGFβ and IL23 were significantly downregulated at both time points using MOI 1 or MO1 3. Using a higher viral concentration with MOI 3 did not increase the inflammation signature of the infected AT2 cells.

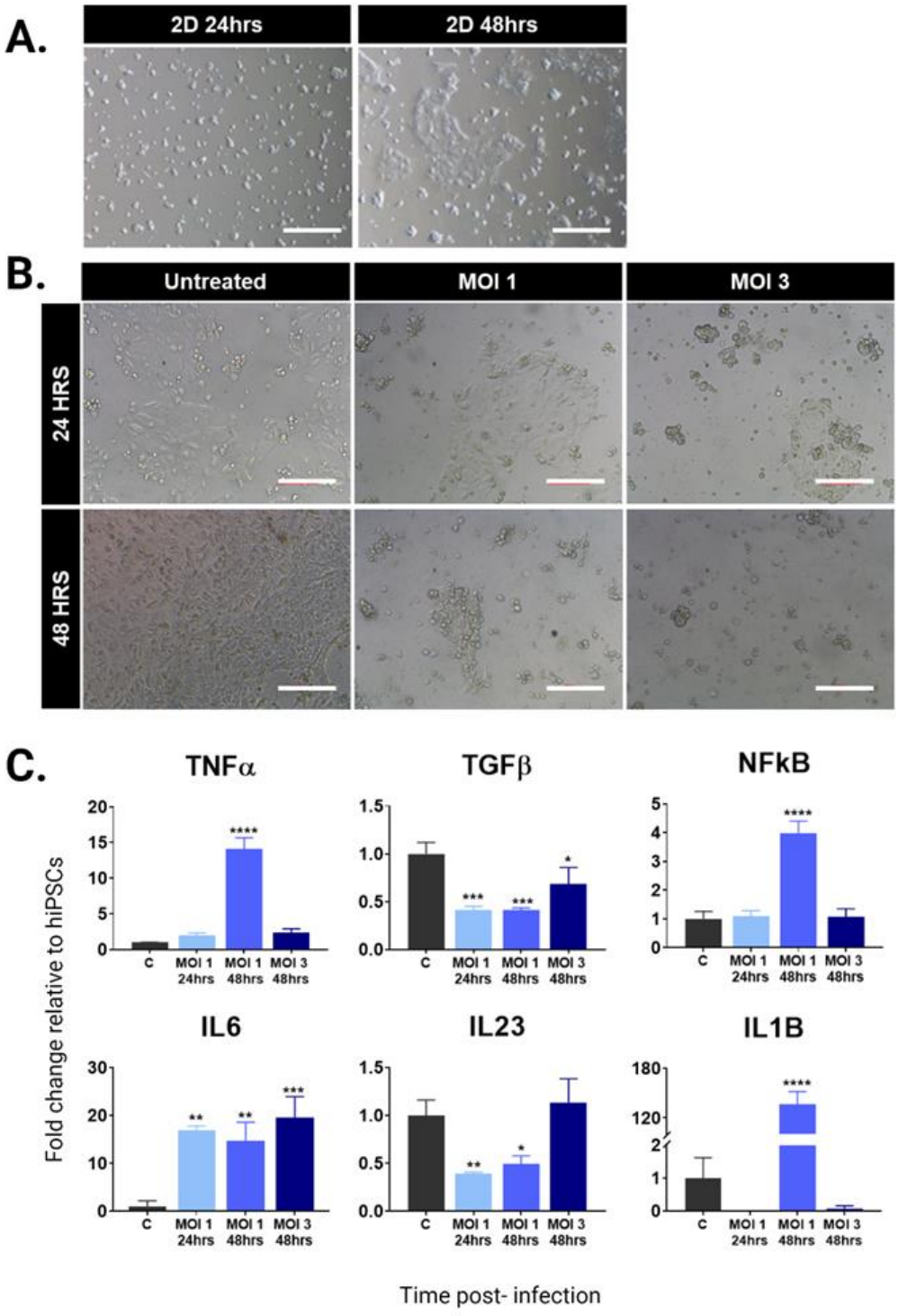


Figure 5.22 Testing infection parameters using wild-type AT2 cells.

Brightfield images of SFTPC$^{WT/WT}$ AT2 cells A) 24 and 48hrs after cultured in 2D conditions B) Uninfected AT2 cells compared to the virus infected cells using MOI 1 and MOI 3 at two different timepoints. Scale bare= 200µm C) Gene expression of inflammation markers post-infection with H1N1 MOI 1 or MOI 3 was assessed using qPCR. Data shows the mean CT values ± standard error of the mean. One-way-ANOVA statistical analysis for significance was

performed in comparison to the untreated control, corresponding to "C". *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001.

SFTPC$^{WT/MUT}$ AT2 cells were also seeded at 1.5x10^{5}cells/cm^2 to elucidate if the mutant AT2 cells were also permissive to influenza A H1N1 infection. This time cells were only inoculated using an MOI 1 as it proved to be enough to infect AT2 cells and induce changes in cell gene expression. The heterozygous mutant cells showed excellent attachment after 24hrs of seeding them in 2D culture conditions and spread out into more extensive areas of cells after 48hrs (**Figure 5.23A**). A complete monolayer could be observed on the untreated wells even at the 24hrs time point, which only grew into a tighter monolayer by the end of the experiment. The AT2 cells infected with MOI 1 were still in a monolayer with a few areas of lifted cells 24 hrs post-infection. By the 48hrs timepoint, cell debris was observed in the media, and the cells underneath were no longer a monolayer, but attached cells were still observed (**Figure 5.23B**). RNA samples were harvested at 24 and 48hrs time points. When analysed, qPCR showed similar results to the ones obtained with SFTPC$^{WT/WT}$. A significant upregulation of TNFα, NFkB and IL6 was observed on the 48hrs timepoint, including IL23, which was significantly downregulated on the wild type. Likewise, IL1B mRNA transcripts were significantly upregulated 24hrs post-infection and decreased significantly after 24hrs, opposite to what was observed in the wild-type cells (**Figure 5.23C**). The expression of TGFβ shows the same trend, with a significant downregulation 24hrs post-infection. The results from this experiment show that 1.5x10^{5}cells/cm^2 is a suitable seeding density to infect the cells 48hrs after seeding and that MOI 1 is sufficient to induce a viral response from the AT2 cells. Indicating that AT2 cells are permissive to influenza virus H1N1 infection and replication.

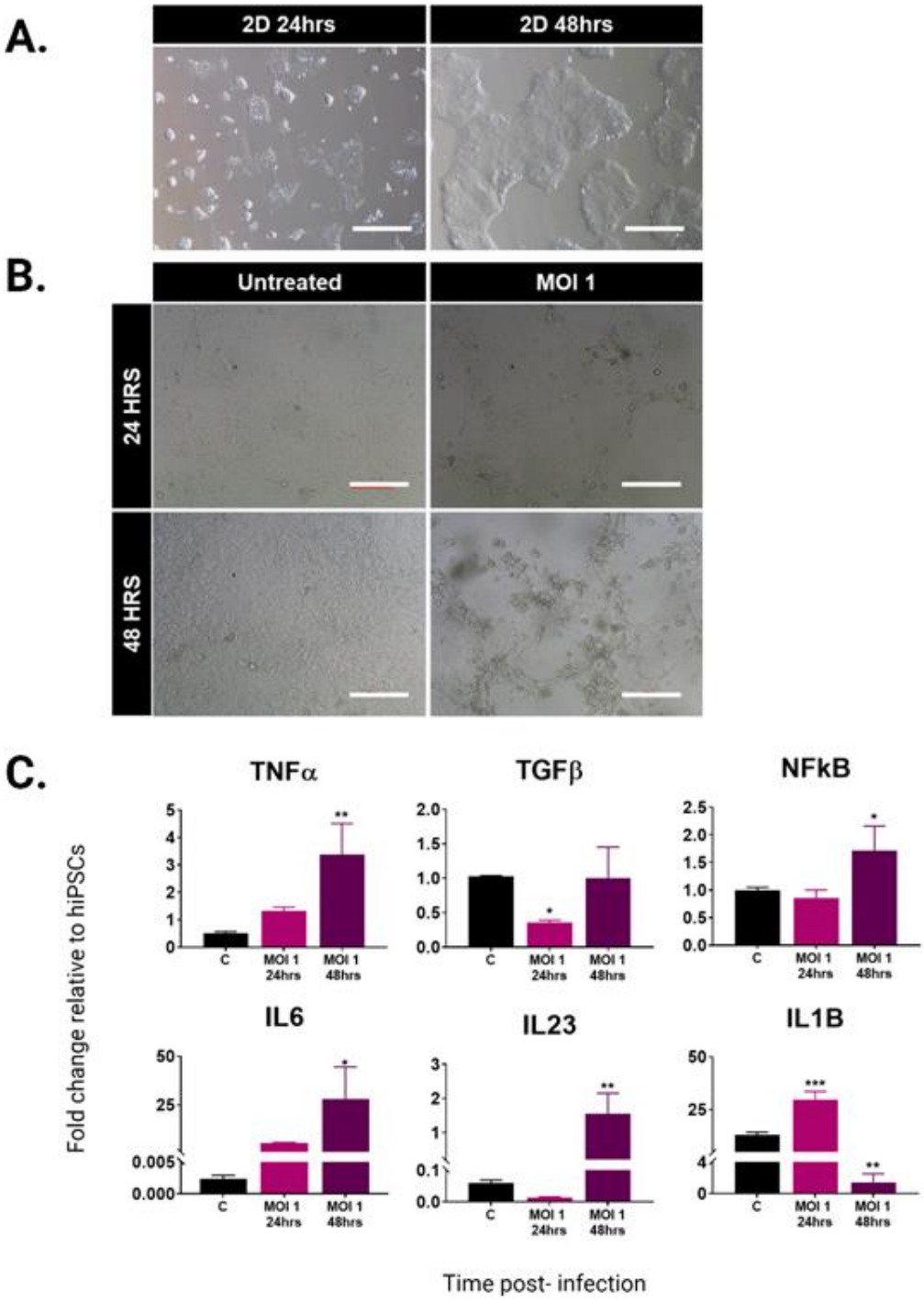


Figure 5.23 Implementation of selected infection parameters in mutant AT2 cells.

Brightfield images of SFTPC$^{WT/MUT}$ AT2 cells A) 24 and 48hrs after cultured in 2D conditions B) Untreated AT2 cells compared to the virus infected cells using the selected MOI 1, 24 and 48hrs post exposure. Scale bare= 200µm C) qPCR shows gene expression changes in inflammation markers after infecting the cells with H1N1 at MOI 1. Data shows the mean CT values ± standard error of the mean. One-way-ANOVA statistical analysis for significance was performed in comparison to the untreated control, corresponding to "C". *p≤0.05, ** p≤0.01, and *** p≤0.001.

5.2.4 In vitro infection of AT2 cells with influenza A (H1N1) virus

As was pointed out in the introduction to this thesis, influenza virus infections and individuals' genetic susceptibility play an essential role in IPF acute exacerbations. The next section describes the principal findings of infecting IPF-susceptible AT2 cells and their isogenic controls with influenza A virus H1N1.

Alveolar organoids derived from the three generated cell lines, $SFTPC^{WT/WT}$, $SFTPC^{MUT/MUT}$, and $SFTPC^{WT/MUT}$, were dissociated and processed to obtain single-cell AT2 cells. As previously described in this chapter, the cells were seeded in 2D culture using a seeding density of $1.5x10^5$ cells/cm^2 in maturation media supplemented with CloneR2. Following incubation for 48hrs to allow attachment and expansion, the AT2 cells were inoculated with 300µl of initial influenza H1N1 virus suspension for an MOI 1. **Figure 5.24A** shows how even if using the same seeding parameters, $SFTPC^{WT/WT}$ AT2 cells were 90% confluent by the time of the infection, whereas $SFTPC^{MUT/MUT}$ ones had only reached 60% confluency when the inoculum was added. Interestingly, $SFTPC^{WT/MUT}$ were fully confluent on a tight monolayer when infected.

Figure 5.24B demonstrates the effect of influenza virus infection in $SFTPC^{WT/WT}$ AT2 cells cell morphology. 24hrs post viral exposure, the cells were balling up and detaching from the monolayer. These observations were only amplified by the 48hrs time point. On the other hand, the untreated control showed the highest appearance of cell debris by the 24hrs time point with fewer cells floating 48hrs post-infection. The morphological changes observed on $SFTPC^{MUT/MUT}$ AT2 cells were somehow expected (**Figure 5.24C**). The untreated cells continued to proliferate in the 48hrs following the infection. However, the wells were highly uneven, as some areas were fully confluent, and others had only a few cells attached. Upon infection, the edges of the cell clusters started to lift from the well, and floating cells were identified on the culture media by the 24hrs timepoint. By the experiment's end, only the wells' edges had attached cells. Finally, $SFTPC^{WT/MUT}$ AT2 cells' morphological changes due to the viral infection were not as evident as in the other two cell lines. No debris was noticeable 24hrs post-infection, and for the 48hrs time point, areas of the monolayer were starting to lift, and dead cells were floating in the media. The untreated control cells continued to proliferate, forming tight clusters of cells forming crater-like structures (**Figure 5.24D**). Sample and control cells were harvested 24 and 48hrs after the infection, as described in Chapter 2.

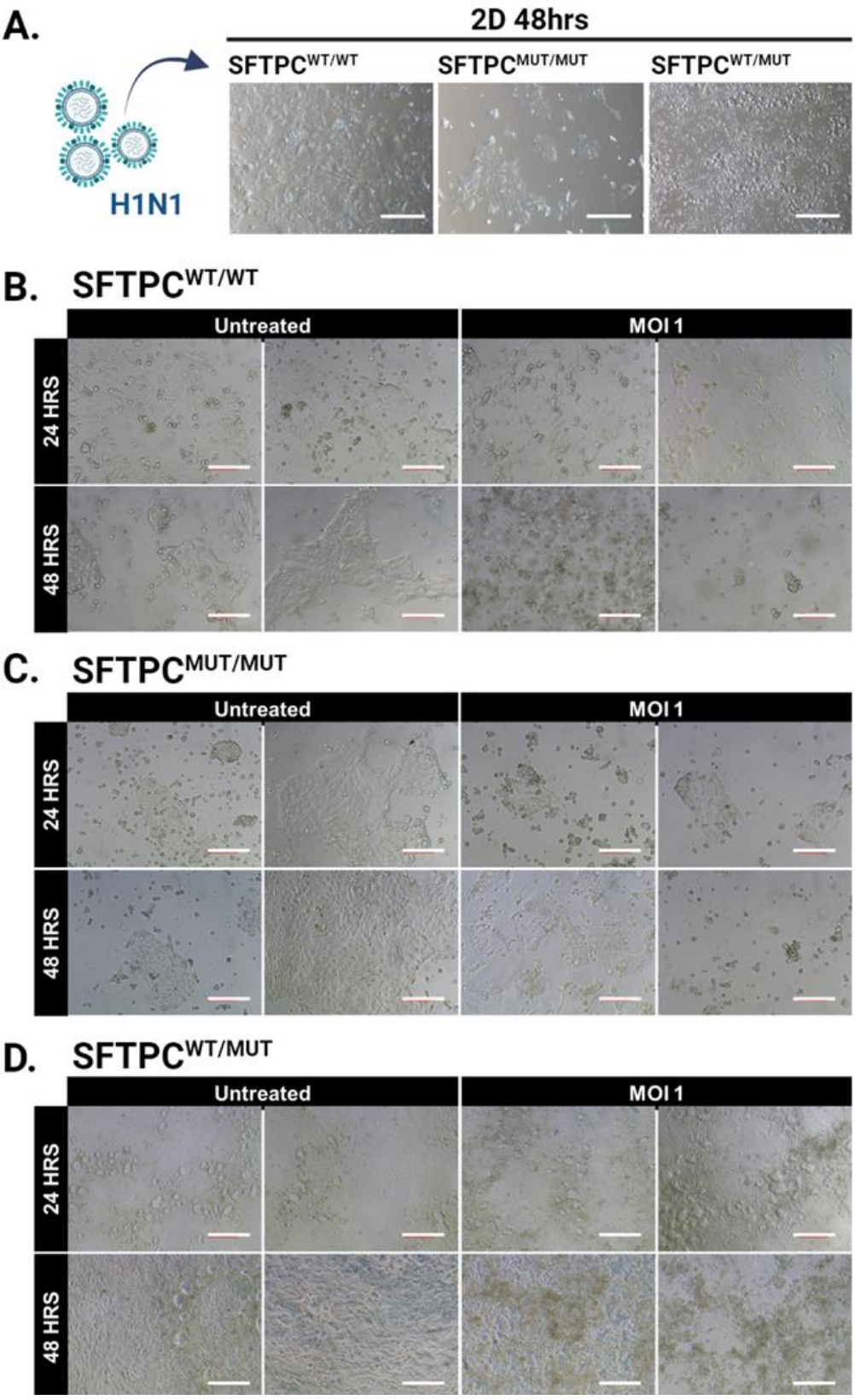

A.
2D 48hrs
SFTPC^WT/WT
SFTPC^MUT/MUT
SFTPC^WT/MUT
H1N1
B. SFTPC^WT/WT
Untreated
MOI 1
24 HRS
48 HRS
C. SFTPC^MUT/MUT
Untreated
MOI 1
24 HRS
48 HRS
D. SFTPC^WT/MUT
Untreated
MOI 1
24 HRS
48 HRS

Figure 5.24 Morphological changes of hiPSC-derived AT2 cells 24hrs and 48hrs post-infection with influenza virus H1N1.

AT2 cells generated from SFTPC$^{WT/WT}$, SFTPC$^{MUT/MUT}$, and SFTPC$^{WT/MUT}$ cell lines were transferred from 3D to 2D and seeded at 500.000 cells/cm^2. A) Brightfield images of cells in 2D culture 48 hours after seeding. A comparison between the untreated cells and the influenza virus infected cells with an MOI 1 is shown using two fields of view in B) SFTPC$^{WT/WT}$ C) SFTPC$^{MUT/MUT}$, and D) SFTPC$^{WT/MUT.}$ Scale bar= 200µm.

5.2.5 Transcriptomic profiling of healthy and disease cell lines in response to influenza A (H1N1) virus infection

To investigate the changes in gene expression triggered on each cell line after exposure to influenza A virus H1N1, SFTPC$^{WT/WT}$, SFTPC$^{MUT/MUT}$, and SFTPC$^{WT/MUT}$ AT2s were infected, and samples were harvested 24 and 48 hrs after the infection as described above. The following section of this thesis describes the transcriptome profiling of the wild-type and mutant cell lines in H1N1 infected and untreated samples at two different time points.

Bulk-mRNA sequencing and gene ontology analysis

As an initial assessment, to illustrate the global differences in the transcriptomic profile between the infected and uninfected samples, the RNAseq data was plotted using a PCA plot. PCA indicated that the global transcriptome of the infected samples was easily distinguished from the uninfected ones (**Figure 5.25A**). This primary separation of samples by infection status was accompanied by a closer similarity between the homozygous cell lines SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ samples, within both the infected and uninfected groups. Furthermore, the identification of the statistically significant DEGs upregulated by infection status is illustrated in the volcano plot in **Figure 5.25B**, shown as P-value versus Log2 Fold change. Out of the 24,476 variables identified, 102 genes were significantly upregulated by the H1N1 infection against 743 genes in the untreated controls. This analysis indicated that the in vitro-infected AT2s could mount an initial response when exposed to the virus.

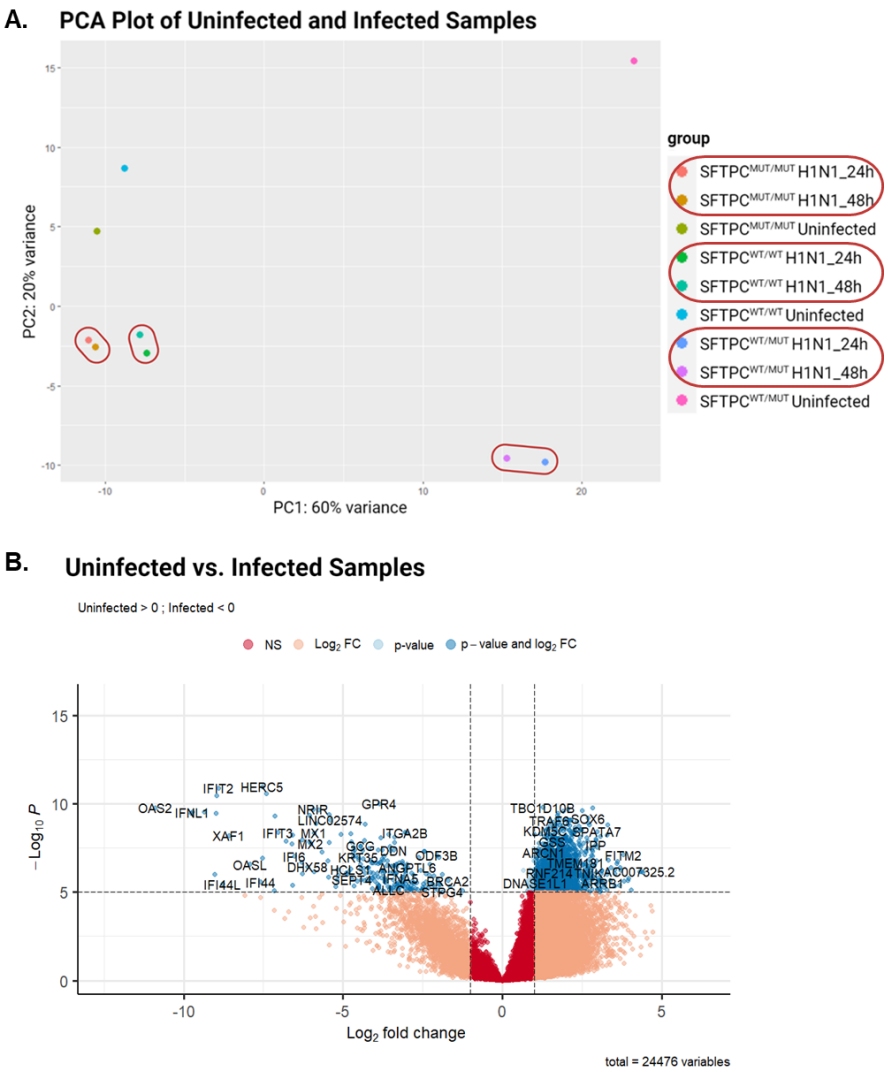


Figure 5.25 In-vitro H1N1 viral infection triggers initial response in hiPSC-derived AT2s.
A) Principal component analysis (PCA) RNA-seq data from infected (24 and 48 hours) and uninfected control AT2 cells shows the primary separation of samples by infection status. B) Volcano plot for AT2 cells demonstrating significant expression of transcripts ($P \leq 0.05$) based on treatment with H1N1 virus. Genes with statistically significant differential expression lie above the horizontal dashed line. Genes whose expression is increased in the infected samples are located to the left of zero on the X-axis (102 genes), while the genes increased in the uninfected samples are located to the right (743 genes). n=3.

To further investigate the impact of infection, a hierarchical analysis of the significant DEGs was performed, along with functional enrichment analysis to identify predicted biological processes. The analysis revealed the relevant Gene Ontology (GO) terms with significant gene ratios induced in the infected samples. The top GO terms included "response to bacterium", "positive regulation of cytokine production", and "cytokine-mediated signalling pathway", indicating an immune response to the infection (**Figure 5.26A**). Additionally, specific terms related to H1N1 infection, such as "response to

virus", "defence response to virus", and "lymphocyte-mediated immunity", were also among the top enriched GO terms. Furthermore, the top 500 DEGs in the infected samples were analysed to identify their association with Disease Ontology terms. Out of these 500 genes, 68% were found to be related to the term "Influenza", and 54% were associated with "Viral infectious disease" (**Figure 5.26B**).

To gain a more comprehensive insight into the primary cellular alterations triggered by infection, a deeper analysis shown in heatmap **Figure 5.26C** was made using the normalised top 15 DEGs in both infected and uninfected samples. It is worth noting that this particular number of DEGs was selected in an arbitrary manner for the purpose of this analysis. Among the significantly upregulated genes in the infected sample were genes that encode for interferons (IFNL1, IFNL2 and IFNL3) and the interferon-induced multiprotein complex (IFIT1, IFIT2, IFI44 and IFI44L), which are known to play a role in the antiviral immune response. Additionally, viral response modulators such as IDO, OASL, and BST2, as well as immunoregulatory chemokines like CXCL10 and CXCL11, were also upregulated in the infected samples.

Among the DEGs explored within the samples, particular attention was directed towards cell surface integrins, primarily due to their pivotal role in lung fibrogenesis [294]. Integrins have the capacity to directly initiate pro-fibrotic pathways by activating latent TGFβ [73]. This potent pro-fibrotic cytokine, in turn, exerts its influence on the underlying mesenchyme and upregulates the expression of αvβ6 integrins, establishing a positive feedback loop that significantly contributes to the progression of lung fibrogenesis [172]. In **Figure 5.26D** heatmap, the DEGs with the smallest P-values in both sample groups were the integrin beta subunit-encoding genes ITGB5, ITGB4, and ITGB1. This analysis of the top 20 differentially expressed integrin genes (number arbitrarily selected) revealed that only four were upregulated upon infection. These included ITGA10, ITGA2B, and ITGA7, which encode integrin alpha chains, as well as ITGB1BP2, which encodes an integrin-binding protein. Additionally, ITGAV and ITGB8, two integrin subunits known for their involvement in TGFβ activation, exhibited high expression levels in both the uninfected and infected samples.

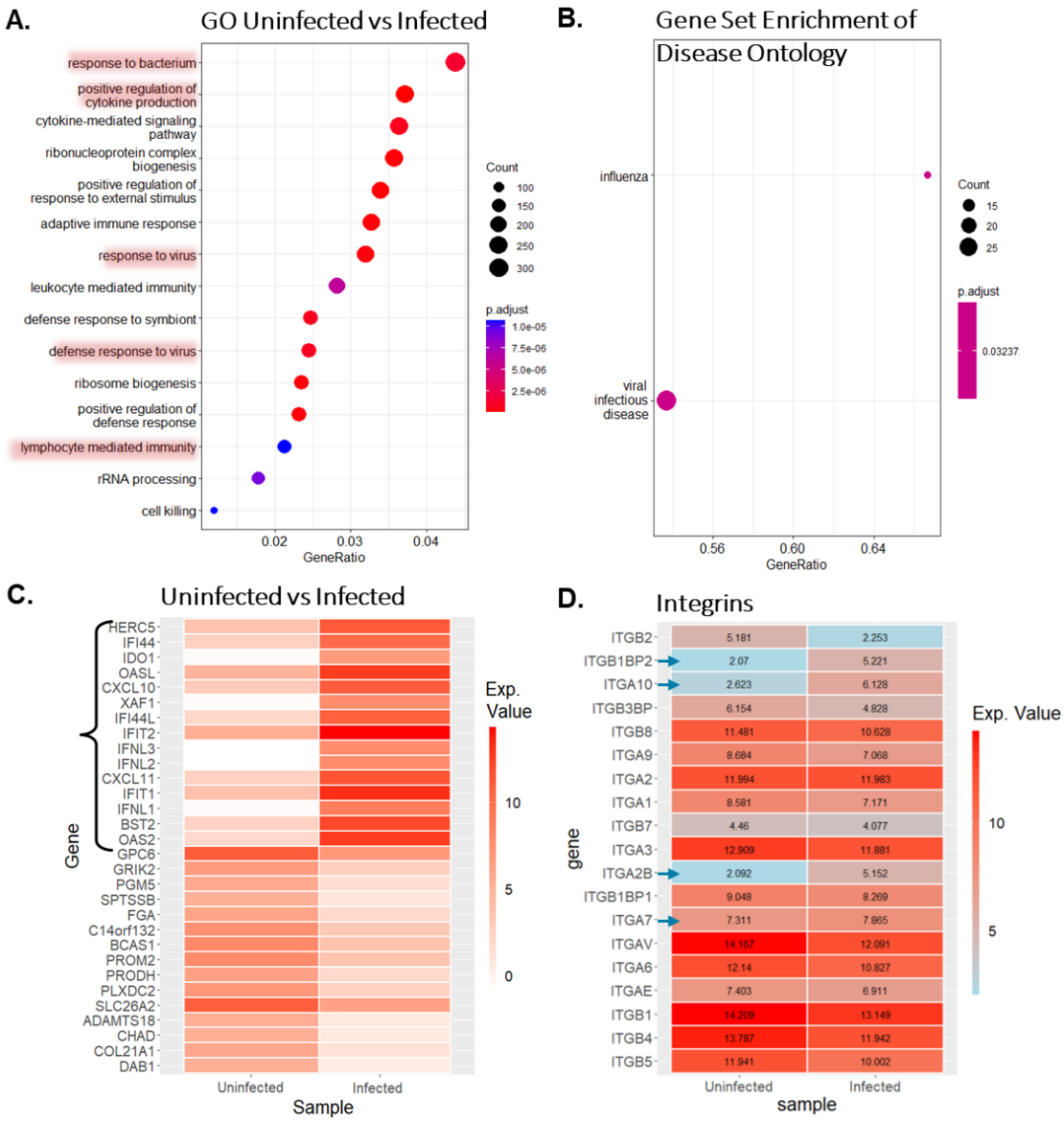


Figure 5.26 Generated AT2s mount a viral defence response.

A) Dot plot illustrating Gene Ontology (GO) enrichment analysis output of uninfected versus infected AT2 cells. B) Dot plot of gene set enrichment for Disease Ontology terms in the top 500 differentially expressed genes (DEGs) in infected AT2 cells. Node size is proportional to the number of genes overlapping the gene ontology term, and node colour indicate the enrichment P-value. C) Heatmap comparing relative expression of the top 15 DEGs in uninfected and infected AT2 cells. D) Heatmap of the top 20 integrin genes differentially expressed in uninfected and infected AT2 cells ordered by significance, with the smallest p-values at the bottom. Blue arrows indicate the genes that exhibit upregulation following exposure to the virus.

To examine the transcriptomic differences between wild-type and SFTPC-mutant AT2 cells in response to the influenza infection, DEGs were plotted in the heatmaps and volcano plots in **Figure 5.27**. Initially, to elucidate the magnitude of the AT2s response to the in-vitro infection, a comparison was done across the $SFTPC^{WT/WT}$ untreated and

24 hours post-infection samples. The analysis confirmed the response of the wild-type cells to the viral agent, as revealed by the 3,600 DEGs shown in **Figure 5.27A** heatmap. Next, the response of SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ cell lines to H1N1 infection was compared to assess potential differences induced by the SFTPC mutation. In the 24-hour infection samples, 1,406 DEGs were identified between these two cell lines, which decreased to 1,270 DEGs in the 48-hour infection samples. The volcano plots were then generated to visualise these DEGs, providing a clearer representation of the data (**Figure 5.27B**). It was evident that the wild-type SFTPC$^{WT/WT}$ cells exhibited a more pronounced response to the infection compared to the SFTPC$^{MUT/MUT}$ cells at both time points. These findings suggested that the SFTPC mutation could exert a significant influence on the cellular response to infection; therefore, an in-depth analysis was needed.

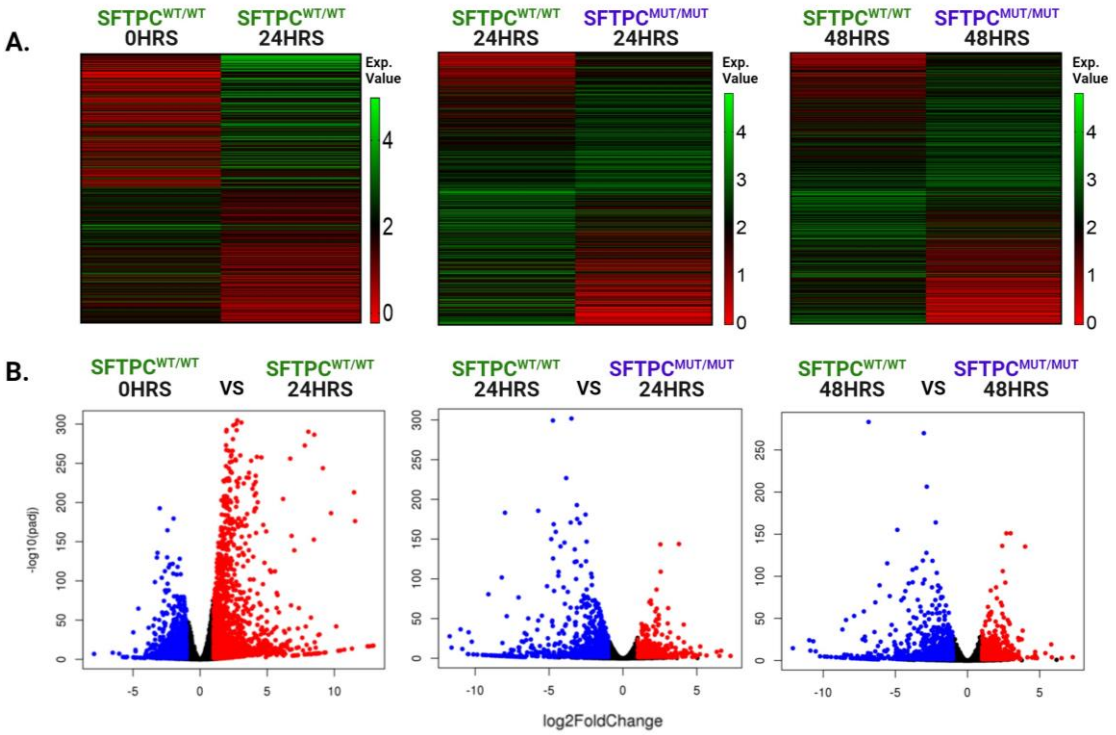


Figure 5.27 Differential gene expression in H1N1 infected vs uninfected AT2 cells.

A) Heatmaps represent 2D hierarchical clustering of genes across wild-type and IPF AT2 cells and show differentially expressed genes on infection samples. Left, top 3,600 most differentially expressed genes (DEGs) between uninfected and infected SFTPC$^{WT/WT}$ samples. Middle and right, top 1,406 DEGs between infected SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ 24 hours and 48 hours after infection. B) Volcano plots showing the same DEGs between WT and IPF AT2 cells. Volcano plot on the left shows SFTPC$^{WT/WT}$ significantly up-regulated genes in uninfected in blue and infected samples in red. Volcano plots on middle and right show significantly up-regulated genes in SFTPC$^{WT/WT}$ in blue, significantly up-regulated genes in SFTPC$^{MUT/MUT}$ in red and non-significant genes in black.

To further elucidate how the SFTPC genotype could be influencing the cellular response to H1N1, gene set enrichment analysis was performed on the upregulated genes from the SFTPC$^{WT/WT}$ and SFTPC$^{MUT/MUT}$ infection samples to determine the relevant GO terms of predicted biological processes enriched on each cell line (**Figure 5.28**). Among the top 15 GO terms observed in the SFTPC$^{WT/WT}$ cells, the most significant ones with the highest gene ratios included "regulation of defence response," "response to bacterium," and "immune effector process." In contrast, the main GO terms identified in the SFTPC$^{MUT/MUT}$ cells encompassed "adaptive immune response," "signalling receptor regulator activity," and "signalling receptor regulator activity." Notably, the GO term "adaptive immune response" was also identified among the most significant GO terms in the wild-type cell line, and despite being the 8th highest GO term, it exhibited the same gene ratio as in the mutant cell line.

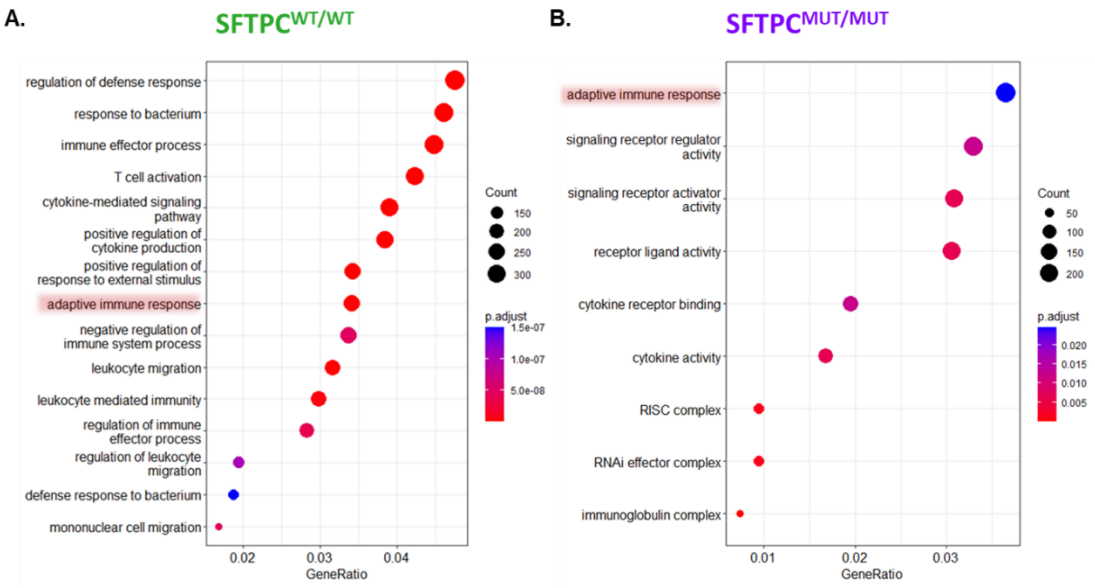


Figure 5.28 Output of Gene Ontology (GO) analysis of infected AT2 cells varies depending on their SFTPC genotype.

A) Dot plot illustrates the functional enrichment of predicted biological processes induced in infected SFTPC WT/WT AT2 cells. B) Dot plot of the functional enrichment of predicted biological processes induced in infected SFTPC MUT/MUT AT2 cells. Node size is proportional to the number of genes overlapping the gene ontology term, and the colour indicates the enrichment P-value.

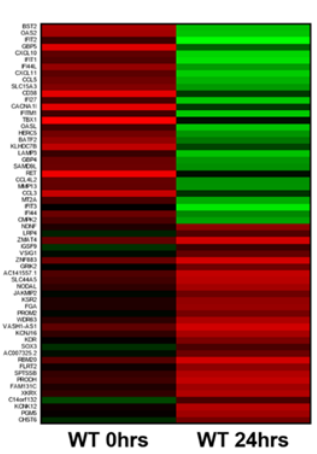
Having demonstrated the broader differences in the transcriptomic profile between SFTPC genotypes, an alternate set of heatmaps was generated to present the top 30 significantly upregulated and downregulated genes between WT and IPF infection samples (**Figure 5.29**). When comparing the wild-type uninfected control vs 24hrs infection sample, the top 30 upregulated genes included genes involved in the defence

response to virus. Many of these genes overlapped with the top 15 DEGs illustrated in **Figure 5.26C**. This included the Interferon-induced multiprotein complex genes IFIT1, IFIT2, IFI44 and IFI44L, plus additional ones like IFIT3 and IFI27. Likewise, the viral response modulators, BST2 and OAS2 were also among the top DEGs identified in the $SFTPC^{WT/WT}$ infection samples. Upregulation of CXCL10 and CXCL11 genes encoding for chemoattractants and activators of peripheral immune cells were also identified in these samples, along with other members of this category, CCL3, CCL4L2, and CCL5. Additionally, genes involved in the innate immune response, like the integral and associated lysosomal membrane proteins LAMP3 and SLC15A3, were also upregulated. Conversely, as part of the top 30 downregulated genes were genes implicated with signal transduction and cell adhesion. These included genes with protein kinase activity like GRK2, KSR2 and KDR, or membrane proteins such as PROM2, VSIG1 and IGSF9.

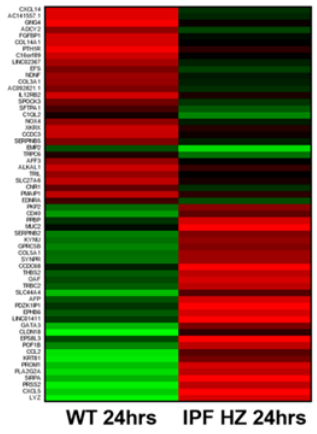
Further analysis of the DEGs between $SFTPC^{WT/WT}$ and $SFTPC^{MUT/MUT}$ infection samples 24 and 48hrs post-infection was also conducted. It was observed that a significant proportion of the top 30 upregulated genes in the 24-hour infection samples were also present in the 48-hour infection samples. A similar pattern was observed for the downregulated genes, with 13 out of 30 genes being consistently downregulated at both time points. Among the top upregulated were genes associated with immunomodulatory processes and defence against a variety of pathogens, including CXCL14, C1QL2 and SFTPA1. Additionally, ADCY2 and NOX4, encoding for enzymes that catalyse the formation of secondary messengers and reactive oxygen species, were part of this group too. Other genes identified encode for structural proteins, like membrane proteins that regulate cell attachment and surface expression, like EFS and EMP2, or collagen chains COL3A1 and COL14A1. Conversely, both the 24-hour and 48-hour infection samples exhibited downregulation of genes such as CLDN18 and PROM1, which are involved in the organisation of the apical plasma membrane in epithelial cells and the obliteration of the intercellular space. Interestingly, none of the top 15 DEGs identified in the overall comparison of infected samples (**Figure 5.26C**) showed high expression in the $SFTPC^{MUT/MUT}$ samples. This further supports the observation that the magnitude of the response to the infection mainly originates from the $SFTPC^{WT/WT}$ cells, indicating that the SFTPC mutation may somehow impair the cellular response to the influenza infection.

Hierarchical analysis of the identified GO terms related to the predicted biological processes induced in the infected mutant AT2 cells compared to the wild-type cells is

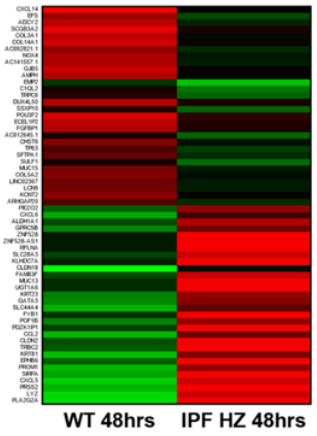
shown to the right of each aforementioned heatmap (**Figure 5.29**). When comparing the untreated wild-type cell line to the 24 hours post-infection sample, the three main overrepresented processes identified were "cell adhesion", "apoptotic process", and "positive regulation of GTPase activity". Additionally, immune and defence response processes, including IFN and NFkB signalling pathways, were prominently represented. These processes were also among the main differentially expressed GO terms when compared to the SFTPC-mutant cell lines. However, the mutant samples exhibited a distinct enrichment of processes associated with the extracellular matrix, such as "extracellular matrix organization", "extracellular matrix disassembly", and "collagen catabolic process". Furthermore, expression of signal transduction bioprocesses, including "cell-cell signalling" and "chemokine-mediated signalling pathway", was increased in the infected mutant samples.



Rank	Process Name	P-value
1	GO:0007155~cell adhesion	0
2	GO:0006915~apoptotic process	0
3	GO:0043547~positive regulation of GTPase activity	0
4	GO:0006954~inflammatory response	0
5	GO:0043066~negative regulation of apoptotic process	0
6	GO:0006955~immune response	0
7	GO:0045087~innate immune response	0
8	GO:0051607~defense response to virus	0
9	GO:0030198~extracellular matrix organization	0
10	GO:0042981~regulation of apoptotic process	0
11	GO:0030335~positive regulation of cell migration	0
12	GO:0051092~positive regulation of NF-kappaB transcription factor activity	0
13	GO:0060333~interferon-gamma-mediated signaling pathway	0
14	GO:0060337~type I interferon signaling pathway	0
15	GO:0045071~negative regulation of viral genome replication	0



1	GO:0070098~chemokine-mediated signaling pathway	0
2	GO:0030198~extracellular matrix organization	0
3	GO:0070374~positive regulation of ERK1 and ERK2 cascade	0
4	GO:0007267~cell-cell signaling	0
5	GO:0030335~positive regulation of cell migration	0
6	GO:0006955~immune response	0
7	GO:0022617~extracellular matrix disassembly	1.69E-13
8	GO:0006954~inflammatory response	4.54E-13
9	GO:0030574~collagen catabolic process	9.84E-13
10	GO:0006915~apoptotic process	2.96E-12
11	GO:0043066~negative regulation of apoptotic process	3.30E-12
12	GO:0043065~positive regulation of apoptotic process	5.76E-12
13	GO:0071222~cellular response to lipopolysaccharide	1.14E-11
14	GO:0007165~signal transduction	2.22E-11



1	GO:0008285~negative regulation of cell proliferation	0
2	GO:0007267~cell-cell signaling	0
3	GO:0030198~extracellular matrix organization	0
4	GO:0006954~inflammatory response	4.17E-14
5	GO:0006955~immune response	6.86E-13
6	GO:0055114~oxidation-reduction process	1.46E-12
7	GO:0043547~positive regulation of GTPase activity	1.61E-12
8	GO:0008284~positive regulation of cell proliferation	2.24E-12
9	GO:0006915~apoptotic process	2.42E-12
10	GO:0007155~cell adhesion	2.60E-12
11	GO:0045893~positive regulation of transcription, DNA-templated	4.41E-12
12	GO:0006508~proteolysis	8.74E-12
13	GO:0007165~signal transduction	2.76E-11
14	GO:0007186~G-protein coupled receptor signaling pathway	3.23E-07

Figure 5.29 AT2 cells top differentially expressed genes and biological processes induced upon influenza infection.

Heatmaps show the top 30 significantly upregulated or downregulated genes across wild-type and IPF AT2 cells when infected with H1N1. To the right is the corresponding functional enrichment analysis of biological processes for up-regulated genes on infected samples showing the top 14 Gene Ontology (GO) terms identified. Biological processes are ordered by descending P values (–log10P value). Top: Uninfected $SFTPC^{WT/WT}$ vs 24hrs infected $SFTPC^{WT/WT}$, middle: $SFTPC^{WT/WT}$ vs $SFTPC^{MUT/MUT}$ 24hrs infection, and bottom: $SFTPC^{WT/WT}$ vs $SFTPC^{MUT/MUT}$ 48hrs infection.

The transcriptomic profile associated with the SFTPC genotype was further scrutinised for select markers in the infected samples, as shown by their fold enrichment in **Figure 5.30** bar plots. The expression of inflammatory cytokine genes IL1 and TNF, which are associated with lung repair following influenza-induced injury, was examined [287]. As expected, the expression of IL1A, IL1B, and TNFAIP3 was significantly higher in the infected samples from all three cell lines. Among these, IL1A exhibited the most significant upregulation upon infection, and the $SFTPC^{WT/MUT}$ cell line showed the largest difference between infected and uninfected samples. Similarly, the expression of the influenza virus-inducible genes LAMP3 and LY6E was also upregulated in the infected samples for every cell line, except for the $SFTPC^{WT/MUT}$ cell line [295, 296]. This cell line showed upregulated LY6E expression in the uninfected sample and exhibited the lowest levels of LAMP3 expression. Furthermore, the tumour suppressor P53, which is directly associated with the transcriptional control of transitional AT2 cells, was upregulated in every uninfected sample, and its expression in the infected samples was highest in the $SFTPC^{MUT/MUT}$ cell line [238, 239]. Additionally, the plasminogen activator inhibitor (PAI-1) gene, which has been associated with the antifibrinolytic state in the alveoli, and the plasminogen activator (PLAU) gene were significantly upregulated in all the uninfected samples [297]. In contrast, the PLAU receptor (PLAUR) gene was significantly upregulated in the infected samples, especially in the $SFTPC^{MUT/MUT}$ cell line. The expression levels of IL11 and IL13, two interleukins that have been implicated in the development of fibrosis in IPF patients and alveolar organoid cultures, were also analysed (**Figure 5.31**) [221, 298]. Interestingly, it was observed that $SFTPC^{WT/WT}$ cells exhibited a significant upregulation of IL11 upon infection, while infected $SFTPC^{MUT/MUT}$ cells showed higher expression levels of IL13.

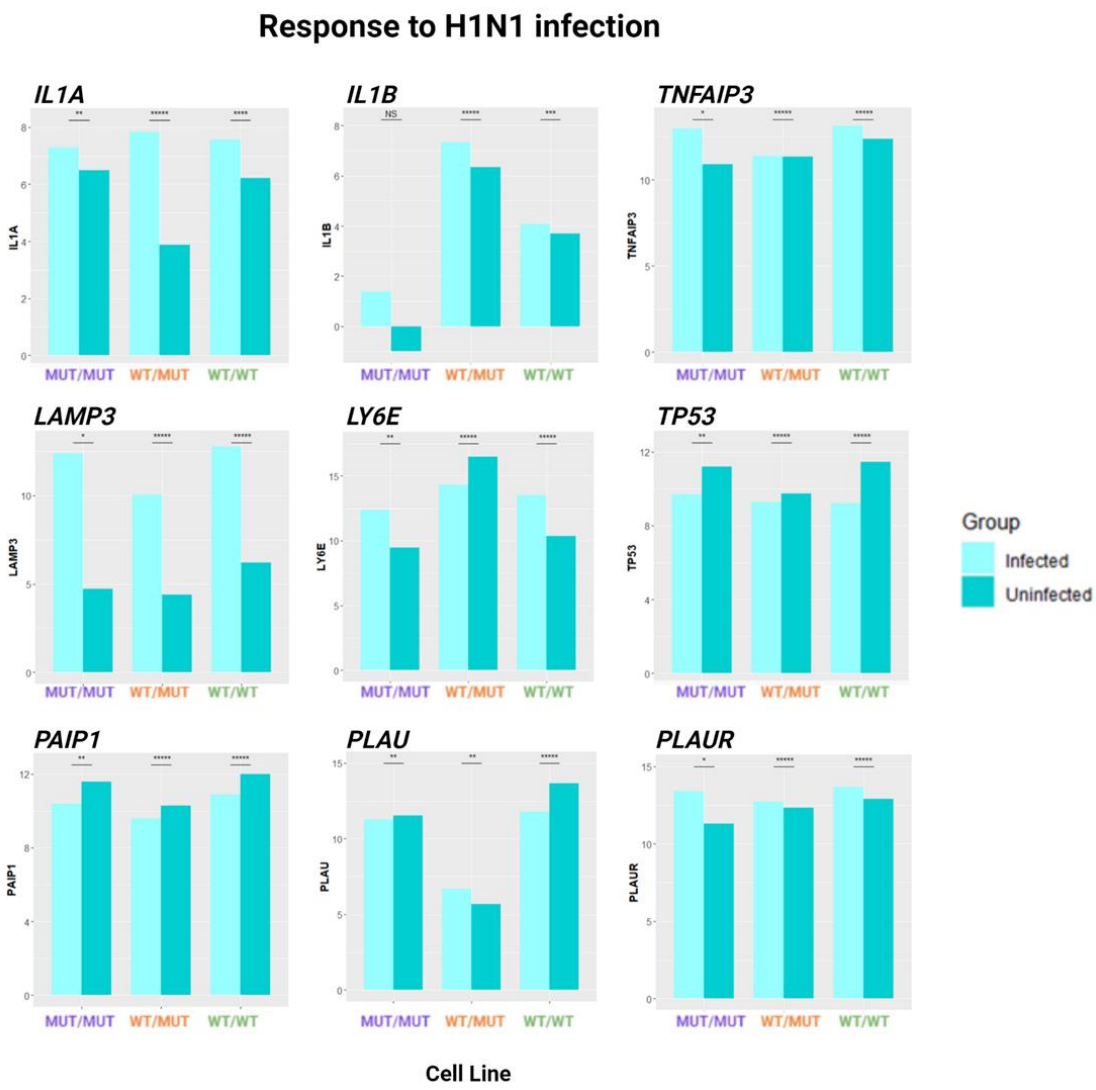


Figure 5.30 Infection mediated responses in AT2 cells.

Bar plots showing log2 normalised expression for select infection and lung-related genes differentially expressed between infected (light blue) and uninfected (dark blue) samples without significant differences due to SFTPC genotype. Top row shows inflammatory cytokines associated with lung repair IL1 and TNFα. Middle row shows influenza virus inducible genes LAMP3 and LY6E, as wells as the tumour protein 53. Bottom row shows genes associated with antifibrinolytic state in the alveoli PAI-1, PLAU and PLAUR. Statistical analysis for significance was performed between samples sets in comparison to the untreated control *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001.

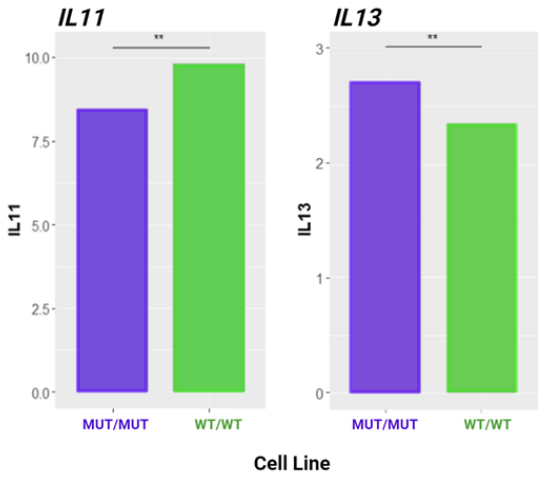


Figure 5.31 Infected SFTPC$^{MUT/MUT}$ AT2 cells upregulate IL13.

Bar plots of IL11 and IL13 normalised expression in infected SFTPC MUT/MUT (purple) and WT/WT (green) AT2 cells. Statistical analysis for significance was performed between samples groups ** p≤0.01.

Activation of the unfolded protein response (UPR) and endoplasmic reticulum (ER) stress in AT2 cells are significant features observed in IPF that have been associated with abnormal processing of surfactant proteins and infections. ER stress is also considered to contribute to AT2 cells' apoptosis and fibrosis progression in these patients [79, 299]. Hence, it was crucial to investigate whether SFTPC$^{MUT/MUT}$ cells exhibit increased expression of known ER stress markers in response to influenza compared to SFTPC$^{WT/WT}$ AT2 cells (**Figure 5.32**). The findings from this comparison yielded contradictory results. The ER stress sensors ATF6, PERK, and IRE1 were significantly upregulated in the wild-type cell line, while their downstream transcription factors, XBP1 and bZIP, showed higher expression in the mutant cell line. Regarding ER-associated degradation (ERAD) components, HS90 and AMER1 were significantly upregulated in the SFTPC$^{MUT/MUT}$ samples, while OS9, SEL1L, SYVN1, and UBXN genes were upregulated in the SFTPC$^{WT/WT}$ samples. Furthermore, the mutant cell line showed higher expression of transcription factors CEBPD and SREBF2, which are associated with phospholipid biosynthesis. On the other hand, the wild-type cell line exhibited upregulation of ER proteins INSIG1 and SCAP, which are also involved in lipid biosynthesis and play a crucial role in the regulation of pulmonary surfactant.

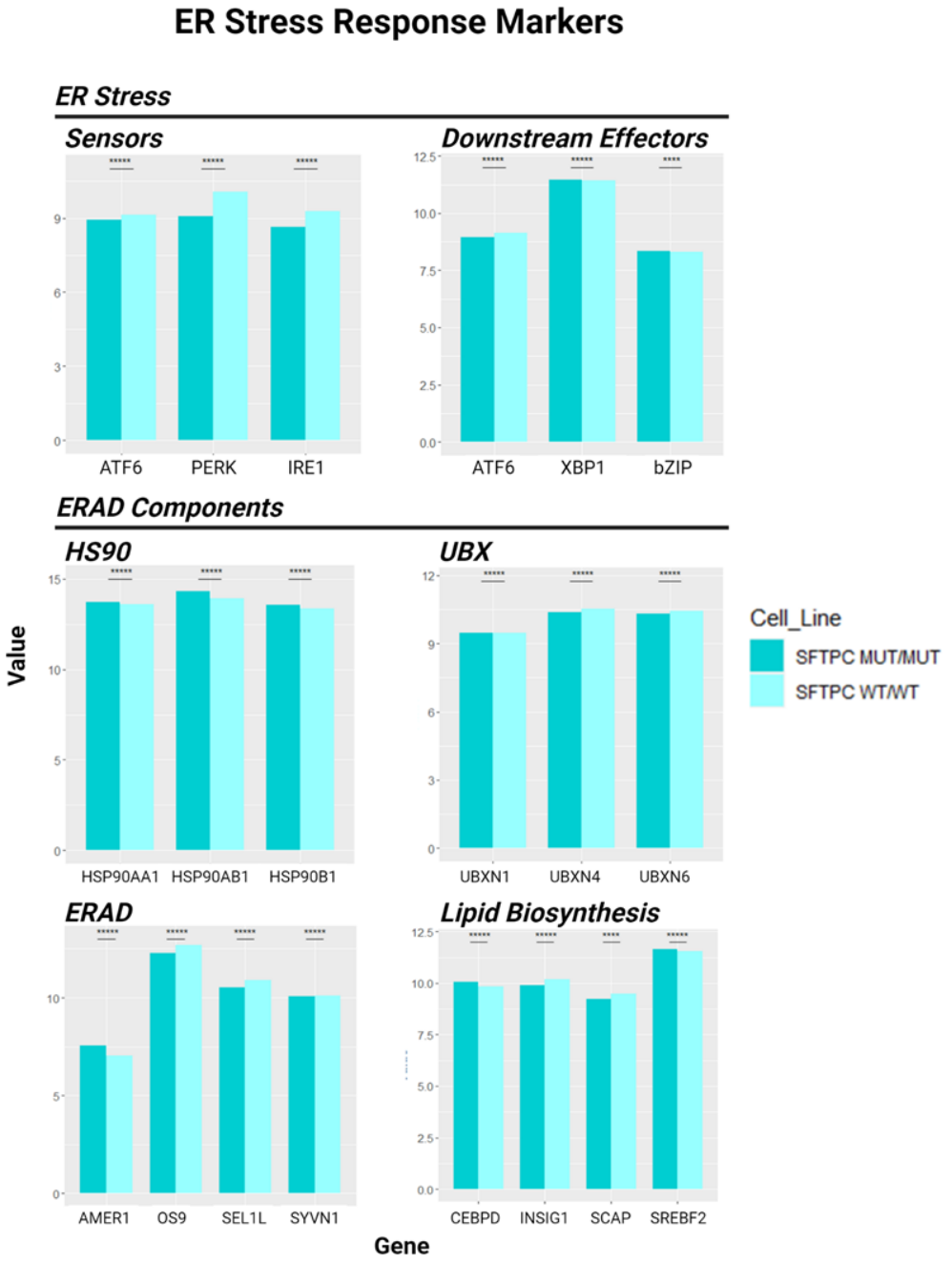


Figure 5.32 ER stress activity in infected SFTPC$^{MUT/MUT}$ and SFTPC$^{WT/WT}$ AT2 cells.

Multiple key components of the unfolded protein response (UPR) were differentially expressed between SFTPC$^{MUT/MUT}$ and SFTPC$^{WT/WT}$ AT2 cells when infected with the H1N1 virus. Bar plots showing normalised expression of Endoplasmic reticulum stress (ER) sensors and downstream effectors (top panel) and endoplasmic reticulum-associated degradation (ERAD) components and lipid biosynthesis (bottom panel). Statistical analysis for significance was performed between samples groups *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001.

Through single-cell experiments, transitional epithelial states have been identified in the injured distal lung tissues of fibrosis patients, indicating a loss of specialised alveolar epithelial cells and the proximalisation of the distal lung. These transitional

cells display a unique expression profile, with markers characteristic of both AT1 and AT2 cells, as well as markers specific to conducting airways. They have been identified by different names, such as transitional AT2 cells, aberrant basaloid cells, and alveolar basal intermediate state, among others [237, 300, 301]. To evaluate whether the infection induced proximalisation of the infected AT2 cells, specific markers of proximal airway epithelial cells were examined (**Figure 5.33**). Remarkably, the genes FOXJ1, PAX9, TP63, MMP10, and SOX2 showed significant upregulation in both uninfected homozygous SFTPC$^{MUT/MUT}$ and SFTPC$^{WT/WT}$ cell lines. However, in the SFTPC$^{WT/MUT}$ cell line, these markers were upregulated only upon infection. On the other hand, KRT5 exhibited upregulation in the uninfected SFTPC$^{MUT/MUT}$ cell line and in the infected SFTPC$^{WT/WT}$ cell line.

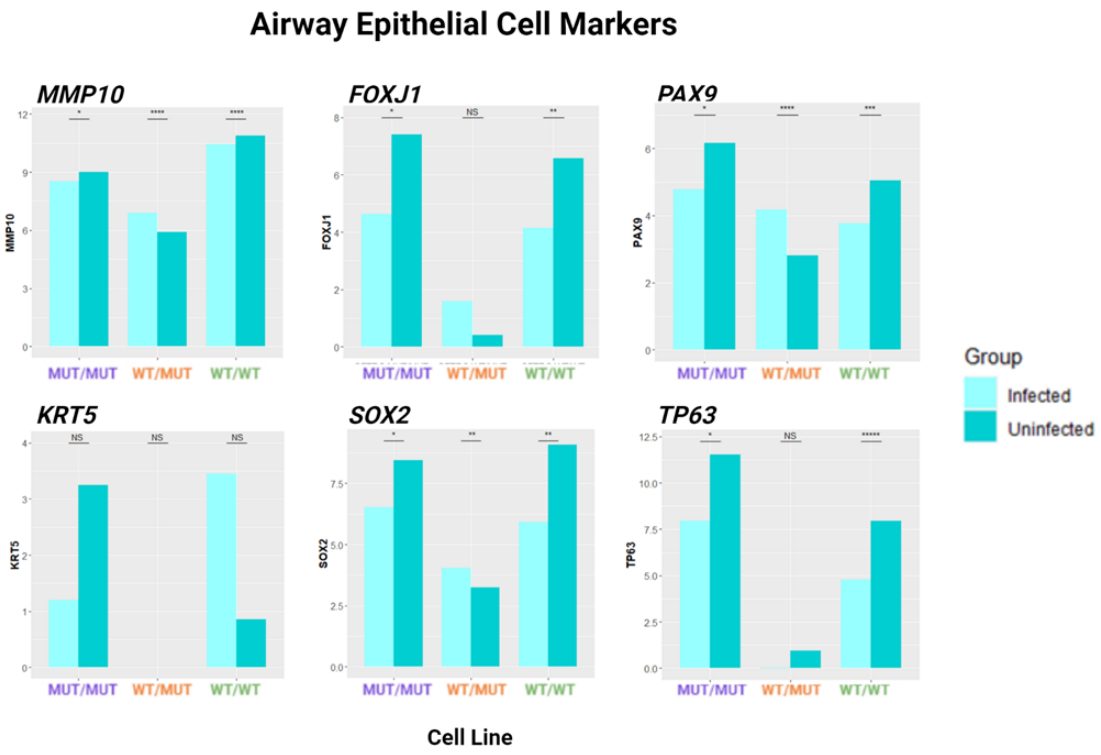


Figure 5.33 Infection does not induce proximalisation of hiPSC-derived AT2 cells.

Bar plots showing log2 normalised expression for select transcripts for airway epithelial cell markers differentially expressed between infected (light blue) and uninfected (dark blue) samples covering the whole spectrum of the SFTPC genotype. Statistical analysis for significance was performed between samples groups *p≤0.05, ** p≤0.01, *** p≤0.001 and ****p≤0.0001.

In order to identify the overlap between genes associated with pulmonary fibrosis and those differentially expressed in infected SFTPC$^{MUT/MUT}$ AT2 cells, a compilation of genes related to pulmonary fibrosis and IPF was gathered from two available online

resources: the Human Genome Epidemiology (HuGE) Navigator and the Online Mendelian Inheritance in Man (OMIM) database (**Figure 5.34**). The HuGE Navigator offers a disease-centred view on genetic association studies and is maintained by the Centers for Disease Control and Prevention (CDC). As of June 2023, it had catalogued 112 genes associated with the IPF indexed disease term (https://phgkb.cdc.gov/PHGKB/hNHome) [302]. Similarly, OMIM provides a comprehensive repository of human genes and genetic phenotypes, developed by the National Center for Biotechnology Information (NCBI) under the supervision of Dr Ada Hamosh (https://omim.org) [303, 304]. Notably, among the AT2-infected samples, the expression of the top 14 IPF-related genes and top 12 pulmonary fibrosis-related genes from the OMIM database was observed. Within the pulmonary fibrosis category, MMP7, RPA1, HPS1, and PARN displayed the highest expression levels in the infected samples. As for the IPF genes, PTGFRN, TGFB1, ELMOD2, and CAV1 exhibited the highest expression levels. Interestingly, SFTPC and PTGS2 showed upregulation upon infection, indicating their potential involvement in the disease process. Additionally, MUC5B and TERC, which were found in both categories, showed upregulation upon infection.

Moreover, the top 20 IPF and pulmonary fibrosis-related genes from the HuGE database were also identified in the infected cells. Among these genes, EGFR, DSP, IFNGR2, IFNGR1, TOLLIP, and TGF-β1 demonstrated the highest expression levels. Notably, IL1B, IL6, MUC5B, and TNF showed upregulation upon infection in both the pulmonary fibrosis and IPF categories. Furthermore, ABCA3 was among the genes with the highest expression in the pulmonary fibrosis category, while TP53 was prominent in the IPF category.

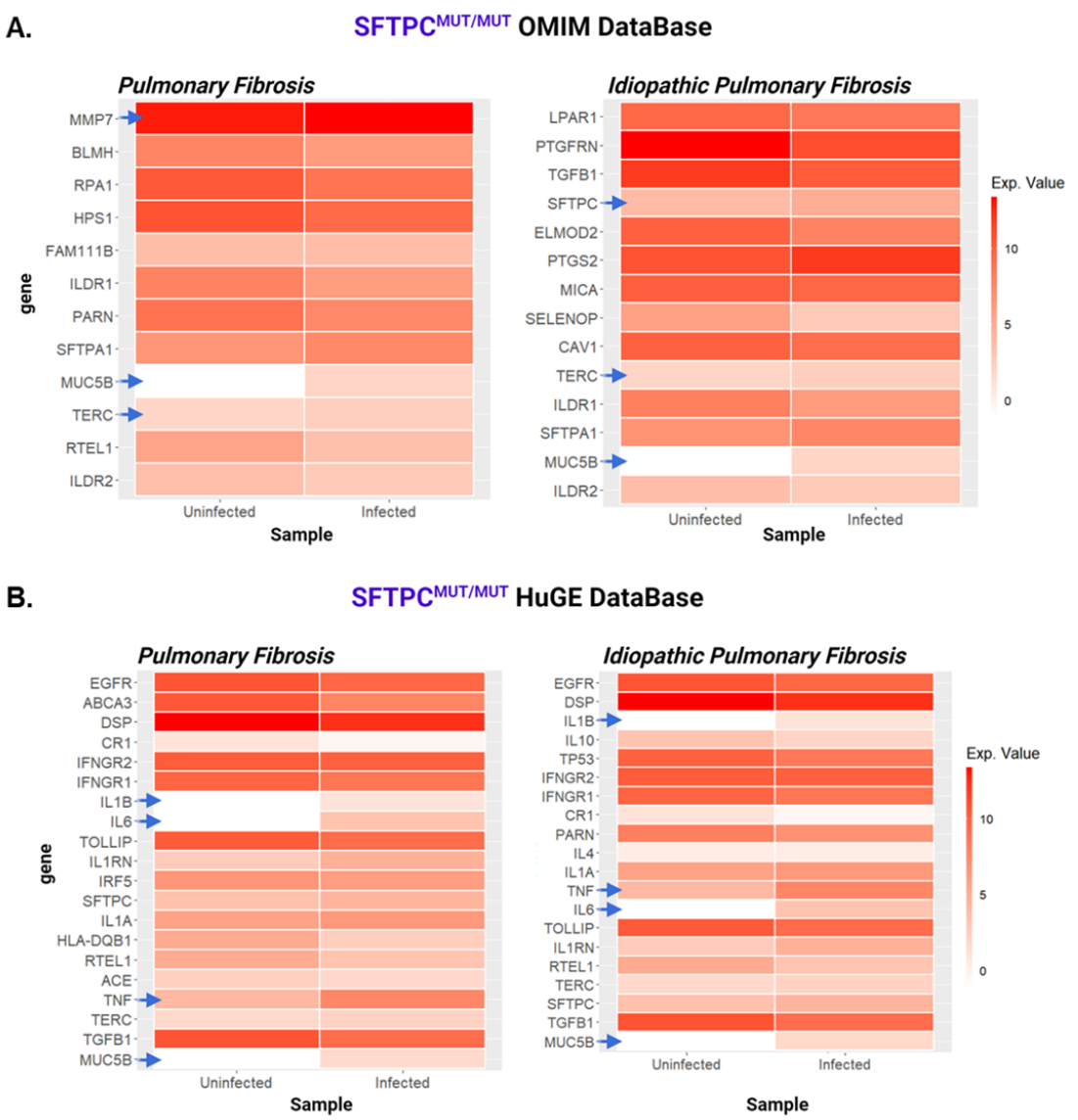


Figure 5.34 Expression of genes linked to pulmonary fibrosis and idiopathic pulmonary fibrosis changes when SFTPC$^{MUT/MUT}$ AT2s are infected.

Heatmaps showing the relative gene expression of transcripts among the top 20 differentially expressed genes in uninfected and infected SFTPC MUT/MUT AT2s that overlap with known fibrosis-associated genes compiled from the disease-centred databases A) OMIM and B) HuGE Navigator.

These findings suggest that the infected AT2 cells with SFTPC$^{MUT/MUT}$ genotype exhibit gene expression signatures that resemble features of both pulmonary fibrosis and IPF. This provides a valuable platform to investigate the impact of infections on the development and progression of IPF.

Chapter 6 Discussion

6.1 Generation of IPF-patient derived hiPSCs

From the establishment of human Embryonic Stem Cells (hESCs) by Thomson et al. in 1998 to the groundbreaking discovery of human induced pluripotent stem cells (hiPSCs) by Takahashi et al. in 2007, these pioneering advancements have proven to be instrumental novel platforms with far-reaching applications encompassing disease modelling, drug development, and regenerative medicine [141, 305]. The remarkable potential of these cells arises from their inherent capacity to provide a sustainable source due to their self-renewal properties, coupled with their ability to differentiate into any tissue throughout the human body. The utilisation of these pluripotent stem cells has paved the way to explore opportunities to replicate disease conditions in vitro, providing valuable insights into pathophysiological mechanisms. Additionally, they have accelerated drug screening processes, enabling the identification of potential therapeutic. Moreover, their regenerative potential holds great promise for treating a spectrum of medical conditions by generating healthy, functional tissues for transplantation [306].

In the context of this thesis, an effort was undertaken to establish a platform tailored for modelling Idiopathic Pulmonary Fibrosis (IPF) using patient-derived hiPSCs. The approach involved processing skin biopsies obtained from six IPF patients to isolate fibroblasts for subsequent reprogramming. The skin biopsies were carefully dissected and cultivated on tissue culture plates to facilitate the isolation and cultivation of dermal fibroblasts. Within the initial 7 to 10 days of culturing, the first signs of fibroblast emergence were observed from the skin biopsy explants. In the subsequent 20-day period, these outgrowing cells were prepared for further passaging and expansion. When confluent, the obtained fibroblasts exhibited a distinctive alignment, forming bundles characterised by elongated, spindle-like cell bodies with rounded oval cell nuclei [307].

Somatic stem cells have been successfully reprogrammed into pluripotent stem cells by introducing key transcription factors, such as Oct3/4, Sox2, Klf4, and c-Myc. This process has employed viral vectors, such as retroviruses and lentiviruses, which integrate into the host cell's genetic material, thereby elevating the risk of tumorigenesis and random genetic integrations [141, 308, 309]. Recognising these safety concerns, exploring non-integrating methodologies has emerged as a promising

alternative (Okita, 2008). These non-integrating approaches involve the transient expression of the reprogramming factors using diverse vehicles, including adenovirus vectors, plasmids, or proteins. Nevertheless, it is essential to note that these non-integrating techniques have shown significant inefficiencies [310]. Among these methods, the employment of Sendai Virus (SeV) has consistently demonstrated a reprogramming efficiency of approximately 1% [204]. This virus replicates its genome within the cytoplasm of infected cells and maintains stable expression of the reprogramming factors.

Utilising the SeV-based reprogramming approach, three distinct IPF patient-derived fibroblast cell lines were subjected to reprogramming. The process involved employing the same seeding densities and utilising a consistent multiplicity of infection (MOI) ratio of 5:5:3 for the SeV vectors containing KOS, hc-MYC, and hKlf4. After the successful derivation of hiPSC colonies, cells were transitioned from commercial TeSR-E8 medium to the in-house E8 media, widely employed within the Hannan Lab to maintain hiPSCs. This methodology allowed the successful establishment of three $SFTPC^{WT/MUT}$ hiPSC lines. These lines were designated IPF1, IPF2, and IPF5, carrying the heterozygous SFTPC mutations Y113C, M71V and Y113C, respectively.

The rapid evolution of gene editing technologies has significantly accelerated the capacity to explore known human diseases using in vitro culture-based models. Notably, the emergence of engineered nucleases like Zing-Finger nucleases (ZNF), transcription activator-like effectors (TALENs), CRISPR/Cas9, and the innovation of base editor technologies has allowed researchers to generate patient-derived cell lines, creating their genetically corrected counterparts, and introduce an array of mutations and reporter elements [311–315]. A fundamental pillar in disease modelling is the use of appropriate controls. In this regard, the inception of isogenic controls has been essential in establishing genotype-phenotype correlations, serving as a discerning tool, effectively separating intrinsic variations linked to patients' genetic backgrounds that could potentially contribute to observed phenotypic differences in conditions such as IPF from the genuine effects of the targeted mutations [316, 317].

Designed nucleases have introduced significant improvements in the efficiency of gene editing. This can be attributed to their capability to precisely induce double-strand breaks (DSBs) at predetermined genomic sites. This process promotes the homology-directed repair (HDR) pathway, thereby facilitating the insertion of tailored modifications into the cellular genome [318]. Among the array of gene editing tools

available, CRISPR/Cas9 has been widely used due to its straightforward design, versatile applications, heightened specificity, efficiency, and cost-effectiveness. Consequently, CRISPR/Cas9 has surpassed alternative nuclease systems, including ZNF and TALENs [319]. Nonetheless, it is crucial to consider potential off-target sites where nuclease activity might inadvertently occur when designing these tools [320]. Furthermore, research has demonstrated that the Cas9 enzyme exhibits a certain degree of tolerance towards mismatches between the guide RNA sequence and the target sequence, leading to the generation of notable off-target effects across the genome [321, 322]. Hence, the emergence of advanced genome editing technologies, including base editing and prime editing, offers an option to streamline the genetic modification process and generate isogenic controls [323]. DNA base editors operate with components from the CRISPR system, coupling a catalytically impaired Cas9 nuclease with a single-stranded DNA deaminase enzyme and, in certain instances, incorporating a DNA glycosylase inhibitor. This combination efficiently introduces point mutations without introducing double-strand breaks (DSBs), mitigating the generation of undesirable editing by-products. Notably, base editors can install transition mutations and can be mainly categorised into cytosine base editors (CBEs), responsible for converting a C•G base pair into a T•A base pair, and adenine base editors (ABEs), engineered to transform an A•T base pair into a G•C base pair [324].

Having optimised the transfection parameters using the DN-100 nucleofector program and the P3 buffer, this thesis proceeded with employing the cytosine base editor AncBE4max for the creation of the (G>A) corrected $SFTPC^{WT/WT}$ cell line and the adenine base editor ABEmax for the generation of the (A>G) mutant $SFTPC^{MUT/MUT}$ cell line. Employing the respective plasmids, cells underwent transfection, followed by transient puromycin treatment. The resultant cell populations were subsequently evaluated through restriction digestion using HpyCH4V to confirm the successful base editing. The puromycin-resistant clones were then screened using the same enzyme assay to determine the number of clones correctly edited by base editing. This selection was subsequently validated by PCR genotyping, confirming the presence of exclusively G nucleotides at the intended position (c.338A>G) in ABEmax-edited cells while confirming the successful correction of the mutation in AncBE4max-edited cells, resulting in the presence of only A nucleotides (c.338G>A). This multi-tiered approach yielded an efficiency of 72%, generating isogenic homozygous mutant and homozygous corrected cell lines. Furthermore, the cells were analysed to confirm their genomic stability and pluripotency characteristics.

6.2 Characterisation of differentiated IPF-patient-derived hiPSCs

The in vitro lung differentiation protocols available are mainly designed based on the insights gained through animal studies and in vitro experimentation to replicate the developmental signals and cues that guide the orchestration of the genetic program responsible for generating lung epithelium in humans [13]. Briefly, mirroring the emergence of one of the three germ layers, the initial step involved the directed differentiation of hiPSCs into definitive endoderm (DE) cells. These cells were then specified to form a region of the primitive gut tube, the anterior foregut endoderm (AFE). Subsequently, The AFE was then specified to allow its differentiation towards the lung fate and further maturation into alveolar epithelial cells. Therefore, these protocols allow for the controlled progression of cellular differentiation to model human alveolar epithelium development closely.

The initial protocol employed in this thesis for developing lung epithelial cells from hiPSCs derived from IPF patients centred around a 2D differentiation approach. This method encompassed the sequential differentiation of hiPSCs into DE, AFE, lung epithelium, and its maturation within a monolayer culture system. However, notable discrepancies emerged across different clones and cell lines when employing the 17-day differentiation protocol to generate lung progenitors. Morphological differences became increasingly evident as the differentiation process approached the culmination of the lung-specific stage in the protocol. Specifically, IPF5 clone 4 and IPF1 exhibited the formation of annular structures intertwined with surrounding monolayers, a characteristic not observed in the remaining differentiated cell lines. In addition to morphological differences, the gene expression profiles exhibited deviations from the ideal pattern during the differentiation. Notably, cells continued to express elevated levels of bipotent mesendodermal markers, MIXL and BRACHYURY, up to the foregut stage and even during the early stages of lung induction. Moreover, levels of CDX2, a hindgut marker, displayed significant upregulation across all differentiated cells throughout the lung induction process, with the most pronounced increase occurring on the final day of differentiation, lung day 12. This trend suggested the potential emergence of a shift in the differentiation trajectory toward an intestinal lineage as reported by several research groups [133, 135, 143].

Furthermore, during the initial differentiation experiments involving the newly derived IPF-hiPSCs, the process of characterising the resulting cells through immunocytochemistry yielded intriguing findings. Notably, IPF5 clones 4 and 5 exhibited an anomalous staining pattern for SFTPC, with localisation observed within

both the nucleus and the plasma membrane using an antibody targeting the residues 1-100 of the protein pro-peptide (abcam, ab90716). Similarly, within the monolayer, IPF1 and IPF2 manifested an unexpected cytoplasmic staining pattern for NKX2.1 (abcam, ab76013), although the spherical structures formed by IPF1 displayed the anticipated nuclear staining. It is noteworthy that this heterogeneous expression of NKX2.1, occurring in both the nucleus and cytoplasm, has also been observed by Jacob et al. (2017), where they suggest the change in localisation may reflect changes in NKX2.1 regulation and interaction with co-factors [145]. Furthermore, this staining pattern has also been documented by Tran et al. (2022), not only when applying the medium composition detailed by Jacob et al. but also when utilising the medium specified by Nikolic et al. (2017), where they observed the cytoplasmic expression of NKX2.1 [145, 325, 326].

The specificity of antibodies is of paramount importance in scientific research. Therefore, ensuring antibodies target the desired protein is essential to obtaining reliable and reproducible results [327]. The results obtained in this thesis should be interpreted with caution because the antibody specificity was not confirmed in our own experiments at the Hannan Lab. However, the NKX2.1 and SFTPC antibodies used have been widely reported in the literature for their use in in vitro and ex vivo models, indicating their suitability for the experiments conducted [328–340]. A comprehensive collection of examples showcasing their application in western blotting, immunocytochemistry, and immunohistochemistry can be found in the **Annex. 2-4**. For future references, to ensure antibody specificity and avoid interpretation issues due to cross-reactivity, antigen variability, or batch-to-batch variability, certain assays could be performed in the lab to confirm specificity through techniques such as western blotting, Enzyme-Linked Immunosorbent Assay (ELISA), and flow cytometry, while also employing appropriate controls and isotypes [341].

In addition to the differentiation of lung progenitors, its maturation was also attempted in monolayered culture conditions. A cocktail of trophic factors supported this maturation process over an additional 16-day interval. However, the outcomes revealed that the exposure to these factors failed to increase the maturation of the cells beyond the fourth day of maturation induction, as evidenced by the levels of SFTPC expression. Nevertheless, the induction of maturation resulted in heightened levels of endoplasmic reticulum (ER) stress markers, progressively increasing as the maturation period advanced, peaking towards the end. To be noted, this increase in ER stress levels was not linked to the genotype, as both the control cell line REBLPAT and the

patient-derived IPF5 cell line exhibited a comparable increasing trend and similar levels of ER stress signalling transmembrane proteins IRE-1 and ATF-6, as well as their downstream effectors XBP1 and EDEM. These observations suggest that the cells might have adapted to the ER stress imposed by the mutant SFTPC protein or that an external "second hit" might be required to trigger the disease phenotype [342].

Recognising the limitations of 2D cultures in adequately mimicking the intricate 3D environment of the human lung architecture and taking into account the growing body of evidence indicating that 3D in vitro culture conditions can not only stimulate the emergence of AT2 cells but also enhance their production of surfactant proteins and phospholipids, a decision was made to transition the lung progenitors into Matrigel droplets [145, 157, 158, 211, 217]. The primary objective was to assess AT2 cells' maturation from hiPSCs within a setting that more closely mirrors the native lung environment. Additionally, the transition sought to encourage the growth of the spherical structures previously identified in the monolayered cultures of IPF cell lines IPF1 and IPF5 clone 4. Intriguingly, these structures manifested spontaneously from the eighth day of lung induction and exhibited the desired phenotype with the highest expression levels of NKX2.1 and SFTPC.

To facilitate the transition of IPF-patient-derived putative lung progenitors to a 3D culture platform, the utilisation of Y-27632, a selective inhibitor of Rho-associated coiled-coil kinase (ROCKi), became essential to ensure cell survival upon dissociation from the monolayered cultures [343, 344]. However, applying the same approach to the control cell lines REBLPAT or LOPCK revealed that the use of Y-27632 alone did not promote the transferred cells' survival. Consequently, the potential of supplementing the maturation culture media with FGF7, a growth factor reported to enhance the growth of lung bud tip progenitors when cultured in conjunction with monothioglycerol, ascorbic acid, RA and CHIR-99021, was explored [217]. The addition of FGF7 yielded promising results, enabling the generation of organoids from the control cell lines and yielding notable impacts on marker expression profiles. Notably, FGF7 was found to contribute to a reduction in the levels of the pluripotency marker OCT4 and the hindgut marker CDX2, while concurrently elevating the expression levels of the endodermal foregut marker SOX17 and the lung-specific distal lung marker SOX9, in comparison with the original maturation media. Therefore, FGF7 was incorporated as part of the maturation media used hereinafter. Unfortunately, immunocytochemistry could not corroborate these findings due to the suboptimal quality of the generated images.

Despite the promising outcomes depicted in Chapter 3 regarding the successful generation of putative alveolar spheroids, achieved through the adaptation of the original differentiation protocol to a 3D-culture system and the introduction of FGF7, it was evident that these cells did not exhibit the AT2 cell signature characteristic of the stem cells of the adult distal lung [38]. Therefore, moving into Chapter 4, the focus was to investigate further and refine the alveolar differentiation platform to pursue the successful and reproducible generation of AT2 cells from IPF patient-derived hiPSCs. The ultimate objective was to establish an adequate for subsequent in vitro infection studies.

6.3 Suboptimal 3D culture differentiation platform

Following the interruption due to the COVID-19 pandemic, the ongoing characterisation of previously established 3D organoids continued while the genetically modified IPF patient-derived cell lines were ready for experimental purposes. Building on the findings from Chapter 3, it was well-established that the generated 3D organoids encompassed a heterogeneous cellular population.

The literature reported the feasibility of isolating NKX2.1 lung progenitors from cultures, using approaches based on reporter cell lines or reliance on cell surface markers such as carboxypeptidase M (CPM) and CD47/CD26 [157, 216]. Similarly, strategies for enriching SFTPC-positive cells had been showcased, encompassing the use of reporter cell lines, antibodies targeting surface proteins like the sodium-potassium co-transporter (NaPi2b) or HTII-280, and fluorescent probes for lamellar bodies such as Lysotracker [8, 162]. Based on this, the assessment of CPM and NaPi2b expression in the thawed IPF-hiPSC-derived organoids was achieved through immunocytochemistry. The confocal imaging of the organoids revealed the coexpression of SFTPC and NKX2.1 with their respective surrogate surface markers and further confirmed the heterogeneity of the cultures. These results pointed towards the need to introduce a sorting strategy. Nonetheless, the initial attempts to sort the cells within the organoids using antibodies against CPM and NaPi2b failed, primarily stemming from the remarkably low yield of cells following the staining process. This was attributed to the intricacies of the staining process and the lack of optimisation for cells previously embedded in Matrigel droplets. These initial struggles were resolved by incorporating 1% FBS/PBS as the washing reagent instead of the conventional PBS. This modification yielded an increasing number of cells processed for sorting.

However, this development did not correspondingly enhance the positivity rate amongst the sorted cells.

Once the genetically modified SFTPC$^{WT/WT}$, SFTPC$^{WT/MUT}$, and SFTPC$^{MUT/MUT}$ IPF-patient-derived hiPSC lines were ready, they were submitted to the differentiation protocol up to lung day 8 to elucidate the percentages of CPM and NaPi2b positive cells and to generate 3D cultures. Unfortunately, the outcomes revealed persistently low percentages, and very few CPM^{+} cells were recovered only from the SFTPC$^{MUT/MUT}$ cell line to be cultured in 3D for further maturation. The gene expression of the CPM^{+} cells growing within Matrigel droplets was compared to that of cells growing in the lung monolayer before the induction of maturation. Encouragingly, the distal airway cell markers SOX9, NKX2.1, SFTPC and SFTPB increased significantly when the cells were matured in the 3D platform. However, despite these promising findings, it was imperative to optimise the differentiation protocol to increase its yields in order to continue with the disease modelling.

6.4 Optimisation of the in vitro AT2 cells differentiation platform

The generation of definitive endoderm (DE) from hPSCs is orchestrated by the activation of various signalling pathways, encompassing Nodal, FGF, BMP, and Wnt [146, 147, 151]. Given this understanding, the initial focus for the optimisation centred on refining the DE stage of the differentiation. The initial approach included three days of differentiation with different combinations of Wnt, Activin, BMP4, and FGF2 (denoted as WABF, WAB, or WA), succeeded by two days of Wnt removal. The wild-type cell line SFTPC$^{WT/WT}$ BE31 was chosen for this initial optimisation. Gene expression analysis by qPCR revealed that WA and WAB did not induce an effective exit from pluripotency, as indicated by sustained NANOG expression, persisting until the fifth day of differentiation. Additionally, these conditions led to the formation of tightly packed colonies within the differentiating cultures. Furthermore, it was revealed that these conditions induced notably elevated levels of the hindgut marker CDX2. Conversely, WABF showcased the highest expression of endodermal foregut markers, including SOX17, FOXA2, and GATA4. To further corroborate the qPCR results for the expression of endodermal cell markers, flow cytometry analysis for cKIT and CXCR4 was also performed. The analysis confirmed WABF as the condition prompting the highest levels of these markers, with 60% positivity for CXCR4 and 52% for cKIT. However, the literature suggests that the expression levels of these markers should be above 90% to yield a robust population of lung progenitors in the subsequent stages

of differentiation [162, 225, 345]. Therefore, these conditions were not optimal for generating DE with the IPF-patient-derived hiPSCs.

In addition to the signalling pathways mentioned above, insulin signalling has also been associated with the induction of hiPSCs lines. The presence of tightly compacted hiPSCs that did not migrate out of the colonies upon extrinsic differentiation cues has been resolved by removing insulin from the cultures [226]. Furthermore, the PI3K pathway in insulin signal transduction has been found to antagonise Activin signalling. Consequently, its inhibition using LY294002 has also resulted in the successful induction of DE from hPSCs [223]. Based on this, three different conditions ((C1, C2, C3) were systematically evaluated to enhance the yield of DE cells. These conditions comprised distinct combinations of CHIR, FGF, Activin, and LY294002. The addition of the PI3K inhibitor was instrumental in facilitating the migration and differentiation of hiPSCs into DE. evidenced both morphologically and validated through FACS analysis of CXCR4 and cKIT expression. Specifically, condition C2, comprising CHIR, Activin, and LY294002, exhibited over 95% and 80% positivity for CXCR4 and cKIT, respectively. These results were further corroborated by qPCR analysis, which revealed a significant upregulation of critical endodermal markers—SOX17, FOX2, GATA4, cKIT, and CXCR4—by day 4 of differentiation. Concomitantly, the expression of the hindgut marker CDX2 was downregulated from the first day of the protocol. Therefore, condition 2 was elected to generate DE from the patient-derived hiPSCs.

Lung Progenitors

During in vivo lung development, NKX2.1 has been identified as the first gene locus known to be activated within the cells of the endodermal lung primordium [16]. The development of a reporter NKX2.1-GFP cell line from pluripotent stem cells, facilitated the targeted enrichment of lung progenitors expressing numerous transcription factors known to be integral to developing mouse lung [134, 346]. Nevertheless, the application of the lung differentiation methodology to hPSCSs had to be meticulously optimised for each cell line. This entailed modifications in the length of endoderm induction, the seeding density for endoderm replating, and the duration of TGF-β/BMP inhibition. These adaptations were essential for optimising the yield of $NKX2.1^{+}$ cells obtained at the end of differentiation [133, 216]

Applying some of these methodologies, an innovative approach was taken to refine the lung progenitors' differentiation protocol. The incorporation of a 3-day dual SMAD inhibition stage following the optimised 4-day DE differentiation, achieved via CHIR,

Activin, and LY294002. This step was conducted using either the DIFF1 differentiation medium (comprising RPMI, B27 supplement, NEEA, and P/S) or the DIFF2 medium (consisting of IMDM, Ham's F12, GlutaMAX, B27 supplement, 7.5% BSA, N2 supplement, AA2P, and 1-MTG) followed by the lung progenitor's differentiation for an additional 12 days [162]. SFTPC$^{WT/W}$T hiPSCs were differentiated using the 3 conditions were harvested at the end of the lung induction and stained with CPM for flow cytometry analysis as CPM has been used as a marker to isolate iPSC-derived NKX2.1 cells [157]. The expression of CPM was surprisingly high at 82% in cells cultivated within the original DIFF medium by day 8, all without the presence of Dorsomorphin and SB43154 (DS/SB). The condition most closely approximating these outcomes was DS/SB DIFF1 on day 10, demonstrating a CPM positivity of 22%. Based on these positive results, the DIFF1 condition was elected to proceed with sorting CPM-positive cells on lung day 8 for subsequent experiments.

However, upon further analysis, the evaluation of additional lung and other endodermal markers via qPCR revealed shocking results. The initially chosen condition, contrary to expectations, yielded the lowest levels of expression for key markers such as SOX9, SFTPC, NKX2.1, SLC34A2, and ABCA3. Furthermore, the DIFF1 conditions exhibited the highest induction of hepatic and hindgut markers—TTR, AFP, ALBUMIN, HNF4α, PDX1, and CDX2. Remarkably, these results were also observed when repeating the differentiation process with the other patient-derived SFTPC$^{WT/MUT}$ and SFTPC$^{MUT/MUT}$ cell lines. Interestingly, these observations are consistent with extensive reports in the literature. Notably, Kido et al. used single-cell RNA sequencing to reveal a close correlation between the CPM population and NKX2.1, demonstrating that it predominantly encompassed NKX2.1-GFP^{+} cells. Intriguingly, their study also revealed an association between this population and NKX2.1- hepatic cells [347]. Similarly, McCauley et al. encounter the same challenges when generating their distal airway spheres. Their work also unveiled the presence of non-lung endodermal lineages such as hepatic and gut epithelia, within their spheres., challenging the maintenance of their cultures over extended periods [348]. Furthermore, lineage tracing using DNA barcoding revealed heterogeneous lineages in the PSC-directed differentiation cultures from the Kotton Lab. They reported a phase of fate plasticity occurring after lung specification, leading to the retention of endodermal multipotency, hence the emergence of midgut and hindgut cell types [19].

In pursuit of alternative strategies to mitigate the emergence of non-lung endodermal lineages within our cultures and aiming to avoid the necessity of generating reporter

cell lines, we explored the potential of using CD47 as a sorting marker for NKX2.1^{+} lung progenitors. This glycoprotein exhibits a broad expression on the surface of lung epithelial cells both in vitro and in vivo [349]. Notably, CD47 has been reported as an attractive candidate for sorting NKX2-1^{+} lung progenitors with ~90% enrichment efficiency due to its high expression levels in lung progenitors compared with other cell types [216]. Despite these promising attributes, the introduction of the CD47-based approach to our culture platform yielded unexpected results. Nearly the totality of the cells within the organoids demonstrated positive CD47 staining, rendering it unsuitable for effectively enriching the NKX2.1^{+} lung progenitor population. Interestingly, when performing this experiment along with the CPM staining for control purposes, a new CPM-positive population was identified. Particularly, cells cultured under the DS/SB DIFF2 conditions exhibited a scarce yet highly CPM-positive (CPM^{high}) subpopulation. Upon further enrichment of this population, it was found that these CPM^{high} cells coincided with those expressing the highest levels of NKX2.1. Moreover, this enrichment was accompanied by a reduction in the expression of non-lung endodermal markers, including TTR, AFP, and ALB, when compared to the overall CPM-positive population. As a result, our focus shifted towards expanding this emerging CPM highly positive population within the monolayered cultures before proceeding to enrich the lung progenitor population using the surrogate surface marker CPM.

In addition to the incorporation of dual SMAD inhibition, researchers have explored the introduction of a reseeding step at various ratios following the endoderm induction but prior to transitioning to the TGFβ/BMP inhibition phase to generate anterior foregut endoderm cells. This strategy has been shown to enhance the emergence of NKX2.1 progenitors [142, 143, 216]. By implementing this replating step at a 1:3 ratio, a substantial increase was observed in the population of cells exhibiting CPM^{high} positivity within the SFTPC$^{WT/WT}$ hiPSC differentiation. Moreover, when comparing these CPM high-positive cells to the moderately positive sorted populations, significant upregulations were evident in the gene expression of NKX2.1, SFTPC, and CPM. These findings were further validated through immunocytochemistry, which confirmed the co-expression of NKX2.1 with its surrogate surface marker CPM, as well as SFTPC with NaPi2b. Additionally, the monolayer cultures displayed negative results for non-lung markers AFP and ALBUMIN, although TTR expression was still detected. These outcomes underscored the challenge of generating entirely homogeneous cultures of alveolar cells, with a propensity for some cells to deviate into other endodermal lineages, albeit at minimal levels. In line with this, a recent study illustrated the in vitro

differentiation of alveolar type I cells alongside a subset of contaminating non-lung endoderm cells expressing CDX2, ALB, and AFP, further highlighting the complexity of attaining purely homogenous alveolar cell cultures [350]. Moreover, the persistent TTR expression in alveolar type II cells has also been documented, reinforcing the necessity of a maturation phase to guide cells toward distal lung lineages.

Alveolar Epithelial Type 2 (AT2) Cells

In the initial differentiation of lung progenitors from IPF-patient-derived hiPSCs using the suboptimal DIFF1 conditions, it became evident that the monolayered cultures exhibited limited maturation of presumed AT2 cells and instead favoured the growth of non-lung endodermal contaminating lineages. However, transitioning the CPM^+ lung progenitors into a 3D culture environment resulted in heightened expression levels of SOX9 and SFTPC in comparison to their CPM^- counterparts. This observation resonates with the findings of Gotoh et al., who demonstrated that 3D cultures of CPM^+ cells facilitated more effective differentiation into AT2 cells as compared to the traditional 2D differentiation approaches [157]. During these experiments, CPM^+ cells maintained in Matrigel droplets were then sorted for NaPi2b to enrich the $SFTPC^+$ population, but the cell numbers recovered were extremely low. Importantly, the selected NaPi2b antibody from Cell Signaling proved unsuitable for FACS analysis, prompting the acquisition of a different antibody from NSJ Bioreagents. Consequently, the sorting procedures were repeated, and this time, both NaPi2b positive and negative populations were assessed via immunocytochemistry using an anti-SFTPC antibody. Surprisingly, both groups exhibited cells that displayed positive SFTPC staining; however, the NaPi2b+ group displayed particularly strong SFTPC staining. Given the promising results of the sorting strategy to enrich the AT2 cell populations, a third anti-NaPi2b (MX35) antibody widely used in the literature was sought to ensure a more consistent source of AT2 cells [8].

The new differentiation protocol, encompassing the optimised endoderm stage, replating of DE cells at a 1:3 ratio before the anteriorisation step utilising dual SMAD inhibition, and subsequent lung progenitor differentiation, was implemented to produce CPM^{high} positive cells. These cells were then matured within 3D Matrigel droplets, using maturation media supplemented with CHIR, FGF10, IBMX, c-AMP, Dexamethasone, and FGF7. The resultant alveolar organoids underwent maintenance for several passages before enrichment of the AT2 cell within them, employing NaPi2b and EpCAM co-staining [222]. The introduction of this novel sorting strategy yielded an incremental growth in the enriched cell positivity from 8% to 42%, over seven passages

within the established 3D culture conditions. However, in our cultured AT2 cells, the proximal-distal maturation enhancement, reported to be achieved through CHIR withdrawal/add-back promoting higher expression of maturation markers like SFTPC, was not manifested in our cultures [145, 351]. Remarkably, the withdrawal solely impeded organoid proliferation, consequently extending the expansion period before AT2 cell enrichment was feasible again. These outcomes align with existing reports of the essential role of CHIR in generating renewing distal tip progenitors, underlying the intricacies of maturation regulation [161]. However, this aspect could warrant further exploration, as it's possible that FGF7 might be effective for initiating organoid formation but might not be suitable for long-term maintenance, potentially due to its targeting of the stem cell population within the organoids.

AT2 cells derived from $SFTPC^{WT/WT}$ alveolar organoids were processed for immunostaining to better characterise their phenotypic characteristics. These cells exhibited robust nuclear staining for NKX2.1 alongside plasma membrane CPM positivity. Additionally, the presence of the epithelial cell marker ECAD and SFTPC was also confirmed. Intriguingly, the distribution of the NaPi2b surface marker positive cells was limited to less than 50% of the cell population. To better illustrate this finding, NaPi2b/EpCAM positive sorted cells were fixed and stained for the intracellular SFTPC and analysed using the Image Stream flow cytometer. Remarkably, this analysis revealed that a substantial 97% of the cells displayed positive SFTPC staining, whereas only 8% exhibited positivity for its surrogate surface marker NaPi2b. This phenomenon resonates with findings from Sun et al., wherein the SFTPCtdTomato reporter was employed to isolate AT2 cells within alveolospheres, preferentially enriching for the SFTPC-high expressing subgroup while not encompassing the rest of the AT2 cells displaying a more immature phenotype. As an alternative approach, the use of an ABCA3 reporter enabled the isolation of the entire putative AT2 cell population [220]. This could also explain how our maturation culture conditions can promote the emergence of highly expressing SFTPC-AT2 cells from a previously sorted NaPi2b negative population. In the context of this thesis, the enrichment of a more mature AT2 population represents a crucial attribute for enhanced modelling of chronic progressive conditions like IPF, as opposed to elevating the number of immature AT2 cells. Nevertheless, by employing genome-wide expression profiling of sorted AT2 cells sourced from both healthy and diseased organoids and subsequent comparison with the KeyGenes database, it has been indicated that the cells developed by the Snoeck group exhibit closer similarity to AT2 cells prevalent in the

second trimester of human lung development [219, 221]. Hence, the current landscape demands the formulation of maturation culture conditions capable of inducing the cells into a state of quiescence similar to that observed in human adult tissues. This strategic advancement could significantly enhance the in vitro generation of more mature AT2 cells, addressing a critical need in the field.

6.5 Transcriptomic profiling of the optimised differentiation platform

In order to gain deeper insights into the transcriptomic profile of the AT2 cells generated through the refined differentiation procedure utilising hiPSCs originating from an individual with IPF, mRNA sequencing was conducted. This approach aimed to detect variations in gene expression among samples collected at significant junctures throughout the differentiation journey. Employing the enhanced protocol, the IPF-patient-derived hiPSCs underwent differentiation, and samples were harvested for sequencing at distinct stages: definitive endoderm (DE), monolayered lung progenitors, CPM-positive high, moderate and negative lung progenitors, $EpCAM^{+}/NaPi2b^{+}$ AT2 cells, and the $EpCAM^{+}/NaPi2b^{-}$ cell population. It is important to highlight that the yields varied significantly among the three cell lines, primarily due to the optimisation process being completed using the $SFTPC^{WT/WT}$ hiPSC line. This was evident as the yields for CPM positive cells for the $SFTPC^{WT/WT}$ were around 30%, while for $SFTPC^{WT/MUT}$ and $SFTPC^{MUT/MUT}$, it was notably lower at around 2% and 7%, respectively. On the other hand, the NaPi2b+ population yielded 22% for the $SFTPC^{WT/WT,}$ and 5% and 8% for $SFTPC^{WT/MUT}$ and $SFTPC^{MUT/MUT}$, respectively.

Initial analysis conducted on the EpCAM+/NaPi2b+ AT2 cells derived from the patient-derived cell lines demonstrated a closer resemblance to the adult total lung sample used as a control in this thesis, rather than the foetal samples that are publicly available from the Kotton Lab. When comparing the gene expression profiles of the DE cells and the lung progenitors, it becomes evident that the key DEGs upregulated in the DE cells encompass SOX17, GATA6, GATA4, and FOXA2, which are established endodermal transcription factors well-documented in Xenopus and mouse model systems [23, 352, 353]. Accurate specification of this stage is of paramount importance, given its developmental identity's susceptibility to external cues. If exposed to signals similar to those from the posterior mesoderm, the endoderm could potentially adopt a hindgut fate, as evidenced by the presence of contaminating hepatic and intestinal markers observed in our differentiations and reported in the literature [216, 354].

On the other hand, in the Lung Progenitors sample, in addition to NKX2.1, the main DEGs included SOX9, SOX2, SHH, FGF10, and BMP4, which play crucial roles in initiating the expression of NKX2.1 within the lung primordium and are key in orchestrating the intricate epithelial-mesenchymal crosstalk that underlies the branching morphogenesis process [143, 144, 211, 230, 231]. Furthermore, the presence of all members of the PDGF family was noted in both populations, with PDGFA, PDGFC, and PDGFD showing upregulation in the lung progenitor's sample. PDGF signalling is widely recognised as essential during embryonic lung development for alveolar septation, demonstrated by in vivo studies using Pdgf-null mice, which exhibited arrested lung development resembling emphysema [355]. In line with this, a study by Barkauskas et al. showcased that Pdgfra+ fibroblasts provide crucial support for the growth and differentiation of AT2 cells in an alveolar organoid co-culture assay [38].

Moreover, the transcriptomic analysis provided clear evidence that the sorting strategy targeting CPM^{high} cells within the monolayered lung progenitor cultures indeed successfully enriched for cells that exhibited elevated expression levels of NKX2.1. Beyond this, the signature of the CPM^{high} cells unveiled a pattern that indicated the enriched population was predisposed to acquire an AT2 cell fate. This predisposition was underscored by the presence of markers associated with lamellar bodies (LAMP3, NAPSA), which are characteristic of mature AT2 cells. Additionally, an augmented lipid biosynthesis activity (LPCAT1) was also observed within this enriched population, further aligning with the functional attributes of AT2 cells [300]. Moreover, this enrichment approach exposed that the CPMhigh subpopulation exhibited higher expression of AT1 cell markers such as TSPAN13, COL4A1, AQP5, AGER, and HOPX in comparison to the lung progenitors. These findings suggest the potential to adapt the differentiation platform to not only favour the generation of AT2 cells but also to cultivate the emergence of the AT1 lineage. This adaptability could prove advantageous for co-culture applications, aiming to establish an in vitro environment that more faithfully mirrors the intricate alveolar niche in vivo for enhanced disease modelling [350].

Two other airway phenotypes were favoured by the culture conditions presented in this thesis. The multiciliated markers FOXJ1 and TUBB3 were observed in both lung progenitors and CPM^{high} populations, displaying modest upregulation upon enrichment. Similarly, the recently identified basal resting cell markers, including KRT19, IGFBP2, CYR61a, and BCAM, demonstrated pronounced expression within the lung progenitor

population, with their expression remaining unaffected by the CPM enrichment process [239]. This could be particularly interesting as scRNA-seq in IPF has revealed a substantial decline in AT1 and AT2 cells in the distal lung parenchyma accompanied by an increase in airway epithelial cells exhibiting aberrant basaloid characteristics on the surface of fibroblastic foci. These cells appear to possess a transcriptional profile that resembles that of the transitional intermediate state between AT2 and AT1 cells, also referred to in the literature as aberrant basaloid cells [356]. Overall, while markers from other airway lineages were also present, the broader expression pattern did not exhibit notable disparities favouring these particular lineages. Nevertheless, just as observed with the AT1 cells, the potential to generate these distinct airway cell types from the hiPSCs derived from IPF patients presents a promising avenue for disease modelling purposes.

When assessing the expression of key AT2 cell markers in the CPM^{high} cells against the NaPi2b positive cells grown under 3D culture maturation conditions, it was found that the transcriptomic profile of the later population had an increased expression of every AT2 cell marker assessed. Furthermore, six out of these upregulated genes, SFTPB, SFTPC, SLC34A2, NAPSA, CEBPD, and CLDN18 were found among the 8-gene signature for differentiating AT2 cells described in the literature [19, 235]. Moreover, the cells enriched for $NaPi2b^{+}$ cells within the organoid cultures also presented the AT2 maturation signature reported by Alysandratos et al. with an increase expression of SFTPA2, SFTPA1, PGC, LYZ, and CXCL5 [235].

Upon analysis of the expression of key markers specific to AT2 cells in the CPM^{high} population compared to the NaPi2b-positive cells cultivated under 3D culture maturation conditions, it was evident that the transcriptomic profile of the latter population exhibited higher expression of every AT2 cell marker under investigation. Remarkably, six of these upregulated genes – namely, SFTPB, SFTPC, SLC34A2, NAPSA, CEBPD, and CLDN18 – were included in the 8-gene signature that characteristic of differentiating AT2 cells described in the literature [233, 235]. Furthermore, the NaPi2b-positive enriched population within the organoid cultures also displayed the AT2 maturation signature that had been reported by Alysandratos et al., observed as the increased expression of SFTPA2, SFTPA1, PGC, LYZ, and CXCL5 upon enrichment [235]. These results provide were promising, as they demonstrate that the cells generated using the optimised protocol established in this thesis exhibit a mature phenotype that closely resembles the one achieved through the most widely employed protocol for generating this specific cell lineage in an in vitro setting [162].

Taking a more comprehensive look at the NaPi2b positive population, a comparison was made between these cells and the NaPi2b negative cells present in the organoid cultures. This comparative analysis yielded reassuring results. Notably, the population of AT2 cells generated in this thesis exhibited the potential for expansion due to their proliferative capacities. Interestingly, the NaPi2/EpCAM enrichment strategy employed selected for the more mature AT2 cells, characterised by a downregulation of markers associated with proliferation. This renders them particularly suitable for disease modelling in the context of IPF.

Nevertheless, it Is noteworthy that the enriched population also exhibited elevated levels of WNT signalling markers, such as NKD1, WNT7B, WNT6, WNT5B, and LEF1. These markers have been linked to the proliferation of AT2 cells in vivo, playing critical roles in late alveologenesis and alveolar repair [36, 357, 358]. This observation is in line with the concept that WNT signalling is instrumental in both the proliferation and maturation processes of AT2 cells, which has been used to promote proliferation or maturation in vitro cultures of AT2 cells [145]. Given that WNT signalling pathways have been identified as being highly induced in IPF epithelial cells, this intriguing observation opens a potential avenue for further exploration [300].

In context with datasets publicly available in the literature, a comparative analysis was conducted between the AT2 cells generated in this thesis and those generated in the Kotton lab, which utilised their SFTPCtdTomato reporter for enrichment [236]. By comparing the NaPi2b and tdTomato positive populations against their respective negative counterparts, it became evident that both sorting strategies effectively enriched for cells expressing higher levels of key AT2 markers, including SLC34A2, SFTPC, SFTPB, SFTPA2, SFTPA1, NAPSA, LAMP3, and ABCA3. However, the enrichment strategy employing NaPi2b yielded more pronounced differences between the negative and positive populations. Furthermore, the NaPi2b positive cells displayed a significant upregulation of LAMP3 and SFTPA1 when compared to the tdTomato cells, while the expression levels of other markers remained similar. Notably, the NaPi2b positive cells also exhibited a greater expression of proliferative markers in comparison to the tdTomato cells. These findings suggest that the AT2 cells generated in this thesis possess a higher potential for proliferation than those produced in the Kotton lab using the tdTomato reporter and within the population, some cells may present a more mature phenotype given the higher levels of the mature AT2 cell marker SFTPA1.

The comparison between the hiPSC-derived AT2 cells and their primary adult and foetal counterparts yielded highly encouraging results. Notably, the adult lung sample exhibited transcript levels of key AT2 markers, including SFTPC, SFTPB, SFTPA1, and ABCA3, that were comparable to those observed in the in vitro generated AT2 cells. In the case of the foetal samples, the latest gestational time point available was week 21, and this particular sample displayed the biggest resemblance to the transcript profile of the sorted AT2 cells. Interestingly, two DEGs, SLC34A2 and SFTPB, demonstrated relatively consistent expression patterns across all the analysed samples. This observation carries potential significance, especially considering that SLC34A2 is the gene responsible for encoding NaPi2b, the surrogate surface marker employed to isolate the AT2 cells generated in this thesis [8].

While analysing the expression of inflammatory markers within the AT2 cells generated from both $SFTPC^{WT/WT}$ and $SFTPC^{MUT/MUT}$ cell lines, it became evident that CTGF, TGFB2, TGFB1, PDGFC, IL6R, and IL17RD were among the DEGs with the highest expression levels in both AT2 cell populations. Interestingly, among these, TGFB1, IL6R, IL1B, and CTGF exhibited more pronounced upregulation in the $SFTPC^{MUT/MUT}$ AT2 cells. These findings are noteworthy given that these factors have been previously found to be released by AT2 cells in IPF, promoting fibroblast proliferation, migration and collagen synthesis [241, 359] . Furthermore, the stimulation of IPF fibroblasts with TGFβ and PDGF has been shown to increase the expression of IL6, acting as a mitogen for lung fibroblasts [360]. Likewise, the identification of the upregulated AT1-AT2 transitional marker, CLDN4, in the mutant cell line has also been detected in IPF patient lung [237–239]. Taken together, these results suggest that the phenotype of the $SFTPC^{MUT/MUT}$ AT2 cells, even in the absence of external stimuli, could closely resemble that of IPF epithelial cells.

In summary, different aspects of the differentiation protocol were characterised in order to develop a platform to generate AT2 cell in vitro with a relevant phenotype for downstream applications. This correspondence between the hiPSC-derived AT2 cell characteristics and IPF pathology identified hold noteworthy implications for advancing our understanding of the disease and its mechanisms at a more detailed cellular level. Furthermore, these insights provide a valuable foundation for the development of a robust cellular model that can mimic some IPF characteristics, aiding in the discovery of potential therapeutic interventions.

6.6 In vitro infection of AT2 cells for disease modelling

Infection experiments in this thesis were performed using healthy hPSCs-derived AT2s as a proof of concept to demonstrate that the differentiated cells could respond to infectious stimuli in vitro. The first Pseudomonas aeruginosa LPS-induced injury on healthy AT2s, grown in an air-liquid interface, modified the functionality of the AT2 cells, significantly decreasing the expression of SFTPC by >20-fold. This finding is consistent with the results from Katsura et al., who observed a dramatic loss or decrease in the expression of surfactant proteins SFTPC and SFTPB in their alveolosphere-derived AT2s when infected with SARS-CoV-2 [130]. Moreover, the first 24 hours post-infection showed an upregulation of TNFα and interferon-γ-inducing factors, and 24 hours later, a noticeable increase in COL1A1, COL3A1, FN, and LAMA1 was observed. These results indicate that the AT2 cells could mount a response to LPS. Similarly, a mere 30-minute exposure was sufficient to induce the upregulation of ERK1/ERK2, a signalling pathway whose modulation has been implicated in the pathogenesis of pulmonary fibrosis, increasing the expression of COL1A1 and IL6 in animal and cellular models [361].

The differentiation platform developed in this thesis for generating patient-derived AT2 cells required their maturation in 3D culture conditions. However, the complexity of the experimental planning increases when infecting cells embedded in Matrigel. Elucidating whether the organoids could be infected whilst immersed in the Matrigel droplets or if degradation of the droplet was needed to achieve a homogeneous infection of the AT2 cells in culture was crucial. The 3D environment turns the infection experiments into a manually laborious process, needing to control factors such as the organoids' size and density or the permeability of the ECM used to embed the organoids. Studies have reported that the size of the organoids is directly proportional to the levels of hypoxia in the centre of the sphere [362, 363]. This leads to the activation of apoptotic pathways and cell death with the inevitable accumulation of cellular debris in the centre of the organoids. With this in mind, distinguishing the exerted influence on the downstream signalling pathways due to the virus or hypoxia-apoptotic pathways could be challenging if heterogeneous-sized organoids were used for the infection experiments. Using same-size hypoxic organoids derived from IPF-hiPSCs could resemble a characteristic of a later stage of lung fibrosis, where the distal lung is hypoxic, making the infection modelling more relevant for fibrosis progression in an already genetically susceptible cell line [58]

In an attempt to adapt LPS infections to a 3D culture, alveolar organoids derived from IPF-patient-derived AT2 cells were generated using the original differentiation protocol. The results obtained from these experiments yielded conflicting outcomes between the IPF1 and IPF5 hiPSC cell lines. Generally, IPF1 demonstrated a more robust response to the 24-hour LPS exposure, exhibiting higher levels of TGFb, TNFa, and IL18 compared to IPF5. However, these responses were significantly downregulated after a 48-hour exposure. Similarly, the levels of ER stress markers exhibited contrasting results for both cell lines. These results indicated that the infections applied to a 3D culture platform would have required further optimisation. Therefore, for this thesis, a more simplified approach was adopted by infecting the IPF-patient-derived AT2 cells in immersed, monolayered culture conditions. When assessing the transfer of 3D-growing AT2 cells to the monolayered culture, it was observed that while some of the cells were strongly positive for NaPi2b staining, the culture did lose expression of this surface marker after being grown in 2D culture. However, all cells in culture retained the characteristic SFTPC expression of AT2 cells. Therefore, the approach was appropriate to demonstrate that the generated alveolar cells were a relevant source for in vitro infection modelling. In this context, to simulate conditions more closely resembling the in vivo alveolar environment, our platform could be adapted for infections in an air-liquid interface culture, as was initially done in this thesis using a healthy hPSC line. This approach could be applied to genetically modified IPF-patient-derived AT2 cells with slight modifications to the protocol and minimal infection optimisation.

The infection modelling in this thesis utilised the influenza virus strain A/PR/8/34 (H1N1) to infect the IPF-patient-derived AT2 cells. The viral quantification of the available stock was successfully performed using the Viral Tox Glo assay, and an MOI 1 was selected for experimental purposes. Salahudeen et al. (2020) demonstrated the successful infection of adult AT2 organoids by inoculating the dissociated cells with an estimated MOI 1 of H1N1 influenza, using the same A/PR/8/34 strain [131]. In this thesis, the assessment of the cellular response to the virus was conducted via qPCR and transcriptomic analysis. Initial expression analysis of infected $SFTPC^{WT/WT}$ and $SFTPC^{WT/MUT}$ cells showed the intriguing downregulation of TGFβ by both cell lines after 48 hours of infection. Further optimisation of the TMLC co-culture assay could prove beneficial for co-culturing with differentiated AT2 cells. This optimisation would help determine the viral dose response of AT2 cells through their TGFβ production,

allowing the selection of the ideal concentration for future infection experiments before transcriptomic analysis [273]

6.7 Transcriptomic profiling of the H1N1 infected AT2 cells

The initial transcriptomics analysis of the infected IPF-patient-derived AT2 cells demonstrated that the cells could mount the expected initial response when exposed to the H1N1 virus. The global transcriptome of the samples from the three infected cell lines—SFTPC$^{WT/WT}$, SFTPC$^{MUT/MUT}$, and SFTPC$^{WT/MUT}$—was easily distinguishable from the uninfected ones. The top GO terms identified in the analysis of the infected samples were "response to bacterium," "positive regulation of cytokine production," and "cytokine-mediated signalling pathway". Moreover, specific terms related to H1N1 infection, such as "response to virus," "defense response to virus," and "lymphocyte-mediated immunity," were also among the top enriched GO terms. In this regard, the top 30 DEGs included interferons (IFNL1, IFNL2, and IFNL3) and members of the interferon-induced multiprotein complex (IFIT1, IFIT2, IFI44, and IFI44L). The upregulation of these gene sets correlates with previous work implicating the importance of these pathways in H1N1 host response both in vitro and in vivo, such as that reported in AT2 cells derived from hiPSCs, where the infection with SARS-CoV-2 induced an epithelial-intrinsic interferon response [283]. Likewise, Youk et al. identified these genes to be increased compared to the uninfected control in AT2 organoids when infected with SARS-CoV [132]. Furthermore, among the top DEGs in our infected samples were viral response modulators such as IDO, OASL, and BST2, which were also identified by the later research group. Additionally, Lin et al., found IDO to be overexpressed in influenza virus (PR8)-infected nasal epithelial cells cultured in an air-liquid interface [364]. They found that viral infection resulted in the depletion of tryptophan via increased epithelial IDO activity, the major enzyme involved in the tryptophan metabolic pathway. Likewise, Yi et al., also observed the upregulation of the antiviral factor Tetherin (BST2) in response to influenza, which inhibits viral replication by triggering ER-stress-induced apoptotic signalling pathways [365].

Similarly, in response to the viral infection, the cells exhibited an augmented expression of chemokines aimed at promoting the activation and migration of immune cells towards the infection site. Notably, immunoregulatory chemokines CXCL10 and CXCL11 were among those that showed increased expression [366]. Moreover, the overexpression of the Lysosomal Associated Membrane Protein 3 (LAMP3) gene was observed, recognised as an inducible gene in the context of influenza A virus. This

gene plays a crucial positive role in the initial stages of viral infection by facilitating post-entry replication [367]. Intriguingly, specific integrin subunits displayed elevated levels following the viral infection. Integrins, acting as connectors between the cytoskeleton and the ECM, have the ability to activate fibroblasts, enhance macrophage phagocytosis, modulate endothelial barrier function, and directly initiate latent TGFβ-induced pro-fibrotic pathways [294]. Specifically, ITGA10, ITGA2B, and ITGA7, encoding integrin alpha chains, along with ITGB1BP2, an integrin-binding protein, showed increased expression. Interestingly, in the uninfected samples, an upregulation of ITGB8 was observed, a subunit recognised for its role in mediating the release of TGFβ from the regulatory Latency-associated peptide [273].

Upon closer examination of the differences between wild-type and SFTPC mutant cells it was identified that the response of $SFTPC^{WT/WT}$ cells to the infection was more pronounced compared to $SFTPC^{MUT/MUT}$ cells across both analysed infection time points, indicating that the SFTPC mutation may somehow impair the cellular response to the influenza infection. Notably, among the top 15 GO terms observed in the $SFTPC^{WT/WT}$ the most significant ones included "regulation of defence response," "response to bacterium," and "immune effector process." In contrast, the main Go terms in the $SFTPC^{MUT/MUT}$ cells were "adaptive immune response," "signalling receptor regulator activity," and "signalling receptor activator activity." Despite the absence of the virus defense response as the primary upregulated biological process in the mutant cells, this suggests alternative mechanisms by which the cells addressed the influenza infection. The analysis of the differentially expressed genes (DEGs) between uninfected and infected $SFTPC^{WT/WT}$ cells revealed a predominant emphasis on virus defense response. Notably, the interferon-induced multiprotein complex genes IFIT1, IFIT2, IFI44, and IFI44L were within the top 15 DEGs in infected samples, alongside additional genes like IFIT3 and IFI27. Correspondingly, cytokines such as CCL3, CCL4L2, and CCL5 exhibited upregulation in infected wild-type AT2 cells. Importantly, these monocyte chemoattractants have been identified as being released following infection of primary mouse alveolar epithelial cells with influenza A virus strain PR/8, coinciding with robust monocyte transepithelial migration [349, 368]

On the other hand, when analysing the DEGs between the infected $SFTPC^{WT/WT}$ and $SFTPC^{MUT/MUT}$. Mutant AT2 cells also upregulated genes involved in immunoregulatory and inflammatory processes. Among these genes, SFTPA1, a collectin crucial for defending against respiratory pathogens, has been linked to the inhibition of influenza virus infection in A549 lung epithelial cells [368]. Furthermore, alterations in SFTPA

levels have been indicative of pulmonary function decline in patients with IPF during antifibrotic treatment [369]. The IPF genetic background might also impact the cell response to the virus, as evidenced by the upregulation of CXCL14 in mutant cells compared to wild-type cells. Notably, the plasma levels of CXCL14 protein have also been identified to be significantly higher in IPF patients in comparison to those of healthy patients [370]. Moreover, CXCL14 was identified, among other genes (SLC40A1, RNASE3, CCR3, and RORA), through examination of IPF lung biopsies and in vitro stimulated fibroblasts, to be significantly associated with overall survival in IPF patients [371]. Likewise, the impact of infection on mutant cells extended to the integrity of the epithelial barrier and the organisation of the extracellular matrix, as demonstrated by altered expression of the claudin and collagen genes CLDN18, COL3A1, and COL14A1. These collagen genes form part of a 5-gene signature linked to the occurrence and prognosis of IPF. Notably, a study identified reduced expression of the tight junction protein involved in the organisation of the apical plasma membrane in epithelial cells and the obliteration of the intercellular space, CLDN18 in AT2 cells adjacent to fibroblastic foci in IPF lung tissue [372, 373]. Likewise, Youk et al,. also demonstrated a significantly reduced cytokeratin genes, cytoskeleton, cell-cell adhesion genes in AT2 organoids infectes with SARS-CoV-2 further underscoring the importance of these findings [132]. Furthermore, the upregulation of the NOX4 gene in infected mutant samples aligns with the strong expression of this protein in hyperplastic AT2 cells within IPF lungs [[374].

Analysing pathways that have previously been related to the pathogenesis of IPF, it was found that In the infected samples from all three cell lines, the levels of IL1A, IL1B, and TNFAIP3 exhibited significant elevation. These findings align with Katsura et al.'s research, wherein influenza-induced injury led to the induction of L-1b and TNFa expression within the AT2 cell niche, vivo leading to the regeneration of the damaged alveolar area by the neighboring AT2 cells [287]. Likewise, the lysosomal Associated Membrane Protein 3 (LAMP3) levels were upregulated across the infected samples from all three cell lines. This influenza A virus-inducible gene has demonstrated a constructive role in the post-entry stages of influenza A virus replication, facilitating the nuclear accumulation of influenza nucleoprotein during early viral infection [367, 375]. The gene TP53, intricately linked to the transcriptional regulation of transitional AT2 cells, showed upregulation in each uninfected sample, potentially associated with downstream NFkB activation [238].Interestingly, SFTPCWT/WT cells displayed a significant upregulation of IL11 upon infection. This finding resonates with Strikoudis et

al.'s work, which shows the essential role of IL11 in instigating fibrosis in HPS mutant fibrotic organoids [221]. Conversely, SFTPC$^{MUT/MUT}$ cells exhibited higher expression levels of IL13. This observation aligns with research involving IPF patients, where an increase in IL13 was detected in bronchoalveolar lavage fluids, and treatment of AT2 organoids with IL13 led to a reduction of SFTPC+ cells, revealing its impact on type 2 pneumocyte stem cell activity [298].

Analysing the expression of the Unfolded protein response and ER stress markers in both infected and uninfected samples from wild-type and mutant cell lines, intriguing patterns emerged. Notably, the ER stress sensors ATF6, PERK, and IRE1 exhibited significant upregulation in the wild-type samples. Alongside these sensors, ER-associated degradation (ERAD) components including OS9, SEL1L, and SYVN1, along with ER proteins INSIG1 and SCAP, involved not only in lipid biosynthesis but also playing a pivotal role in regulating pulmonary surfactant, were also upregulated. Interestingly, Guo et al. documented the activation of the unfolded protein response in mouse alveolar epithelial cells induced during birth, accompanied by a heightened expression of transcription factors regulating surfactant protein and lipid biosynthesis [301]. However, the expression patterns of the analysed ER markers did not exhibit a straightforward correlation with the mutation status, as one might anticipate from existing literature. For instance, Korfei et al. identified significantly elevated protein levels of ER stress mediators within lung homogenates and AT2 cells from IPF lungs compared to those from COPD and healthy donor lungs [376]. Similarly, Lawson et al., found prominent expression of UPR/ ER stress markers in AT2 cells in the lungs of IPF patients with SFTPC mutation-associated fibrosis colocalised with Herpes virus protein expression [299].

AT2 cells possess inherent regenerative capabilities that allow them to proliferate and differentiate into ATI cells, contributing to alveolar restoration during normal repair processes. However, this regenerative potential becomes compromised in the face of injury, causing these cells to adopt markers reminiscent of proximal airway cells, known as bronchiolarization [38, 377, 378]. In order to determine whether our AT2 cells exhibited signs of this aberrant repair mechanism in response to influenza infection based on their SFTPC genotype, we evaluated characteristic markers associated with bronchiolarization. Remarkably, FOXJ1, PAX9, TP63, MMP10, and SOX2 exhibited significant upregulation in both uninfected homozygous SFTPC$^{MUT/MUT}$ and SFTPC$^{WT/WT}$ cells. Intriguingly, in the SFTPCWT/MUT cell line, these markers displayed increased expression only upon infection. Conversely, KRT5 displayed

elevated expression in the uninfected SFTPC$^{MUT/MUT}$ cell line as well as in the infected SFTPC$^{WT/WT}$ cell line. This finding aligns with the observations of Xi et al., who noted a hypoxic subpopulation with activated Notch signalling, suppressed SFTPC, and transdifferentiation into a Krt5-positive basal-like state in human AT2 cells from fibrotic lungs [379].

In order to comprehensively characterise the changes induced by the influenza virus within the SFTPC$^{MUT/MUT}$ cell line, a comparison was conducted between the top DEGs observed in the in vitro infected AT2 cells and their corresponding counterparts in the human disease-centred databases HuGE Navigator and OMIM. Among the highly expressed genes in the infected samples were MMP7, RPA1, HPS1, and PARN. Notably, both MMP7 along with MMP1 have been identified as overexpressed in the lung microenvironment in IPF and other chronic lung diseases. Hence, elevated MMP7 levels have been suggested as an indicator of asymptomatic interstitial lung disease, potentially reflecting disease progression and it has been included as part of a biomarker profile associated with progressive IPF [380, 381]. Upon infection, several genes including IL1B, IL6, MUC5B, and TNF exhibited upregulation within both categories analysed, pulmonary fibrosis and IPF. In the context of organoids infected with the flu virus, IL1 and TNFα triggered a localised response through direct interaction with surface receptors on SFTPC+ AE2 cells, facilitated by MYD88, subsequently activating NFkB. This inflammatory response after alveolar epithelial injury was pivotal in promoting lung repair [287]. Notably, TGFβ1 displayed the highest expression levels, consistent with the process wherein alveolar epithelial cell injury culminates in the infiltration of fibroblasts and inflammatory cells, leading to the release and activation of pro-fibrotic mediators like TGFβ and PDGF, ultimately resulting in matrix synthesis and accumulation [247]. Surprisingly, genes such as SFTPC, MUC5B, and TERC exhibited upregulation upon infection, contrary to expectations from existing literature. Huang et al., demonstrated the transcriptome shift in infected AT2 cells moving toward an inflammatory phenotype characterised by NF-κB signalling upregulation and loss of the mature alveolar program [382]. Likewise, other groups also revealed a substantial reduction or decrease in the expression of surfactant proteins SFTPC and SFTPB after influenza-induced injury [130, 383].

The findings presented in this study establish the hiPSC-derived AT2 cells as a robust and effective in vitro model for investigating H1N1 infections. The results demonstrate the susceptibility of both wild-type and SFTPC mutant hiPSC-derived AT2 cells to successful H1N1 infection. These generated AT2 cells not only exhibit this

susceptibility but also showcase a desirable initial response upon viral exposure, characterised by the upregulation of interferons and interferon-induced proteins, pivotal defence mechanism aimed at hindering viral replication. The capacity of hiPSC-derived alveolar organoids to faithfully recapitulate crucial aspects of viral infections further underscores their potential as a valuable tool for deciphering intricate cellular responses to infections, paving the way for exploring therapeutic interventions. Importantly, it is worth noting that the applicability of this model extends beyond the confines of IPF, as the protocol for generating AT2 cells can be adapted to study various environmental exposures in vitro, broadening its potential applications.

6.8 Study limitations

The efficacy of the developed platform is intricately linked to the quality of the differentiation protocol in place. Similarly, the proficiency of the protocol relies upon the depth of our understanding of lung developmental biology. The protocol optimised in this thesis has demonstrated its capability to effectively generate hiPSC-derived AT2 cells with relative genetic backgrounds for IPF disease modelling. Nonetheless, certain concerns necessitate addressing. For instance, there is a challenge concerning the stability of hiPSC-derived cells when cultivated in vitro, as they exhibit greater plasticity than primary cells. Consequently, they are markedly susceptible to deviations induced by variations in culture conditions [384]. In the context of this thesis, it was observed that cells undergoing foregut endodermal differentiation exhibited the potential to follow an alternate differentiation route leading to non-lung lineages corresponding to hindgut fates. To counter this issue, it is necessary to deepen our knowledge of the temporal developmental signalling pathways essential for specifying lung lineages to tailor better the cues provided during in vitro cultures, thereby minimising this diversion.

Although hiPSC-derived AT2 cells are more favourable to genetic modifications, one of their drawbacks is their immature phenotype, which complicates the study of age-related pathogenesis in IPF. Chapter 4 of this thesis detailed the optimisation of the differentiation platform to generate AT2 cells that could be maintained in 3D cultures to induce a relatively more mature phenotype. Although the alveolar organoids contain differentiated AT2 cells specific to the alveolar compartment and share characteristics of their adult counterparts, the transcriptomic signature still resembles that of foetal-like cells. Similar observations have been obtained from PSC-derived lung organoids, where the cellular composition primarily reflects foetal-stage cells and falls short of fully recapitulating the characteristics of adult counterparts. This highlights the persistent

challenges of achieving in vitro cellular maturation [160, 161, 217, 219]. These limitations are recognised as significant obstacles when employing human PSC-derived cells for disease modelling and regenerative medicine applications [385]. Moreover, the functional capabilities of AT2 cells within the organoids tend to diminish over time. This is evident from the necessity to continuously enrich the cultures for NaPi2b-positive cells to maintain a more homogenous population of SFTPC-producing cells. This effort is essential even if the entire population still retains its AT2 cell identity to a certain extent [220]. Similarly, the absence of the native alveolar microenvironment, encompassing crucial elements like the ECM, blood vessels, and immune cells, presents a barrier to conducting thorough investigations regarding the intricate interactions involving AT2 cells within their niches. Thus, a promising avenue for further exploration lies in enhancing the organoid culture platform by including additional cell types such as fibroblasts and immune cells. This could offer a partial recapitulation of the alveolar niche, enabling more faithful modelling of the complex in vivo inflammatory responses [386].

Moreover, the culture of AT2 cells within a 3D environment introduces additional challenges. The diffusion of nutrients through the Matrigel droplets can lead to heterogeneity within the larger organoids, potentially resulting in necrotic areas at their core. Furthermore, the undefined composition of Matrigel, coupled with its inherent batch-to-batch variability, can impact the cultures' homogeneity and affect reproducibility between experiments. This extracellular matrix (ECM), derived from the mouse Engelbreth-Holm-Swarm sarcoma, exhibits a complex composition enriched with collagens, heparan sulfate proteoglycans, laminins, and growth factors [387]. Its incorporation marked a significant advancement, as it allowed for the replacement of mouse embryonic fibroblasts (MEFs) as feeder layers in the culture and maintenance of pluripotent stem cells. However, newer and more suitable alternatives for the culture of hPSCs have been utilised, natural proteins and peptides like vitronectin and laminin, along with synthetic polymers such as poly-ethylene glycol (PEG) [388]. Adopting these emerging techniques can potentially customise the composition and stiffness of the ECM embedding the cultured cells to satisfy specific experimental needs.

In conclusion, the developed platform utilising hiPSC-derived AT2 cells from IPF patients is a valuable tool for in vitro disease modelling of the alveolar cellular compartment. Even if imperfect, the model offers substantial opportunities to study patient-specific mutations within a relevant 3D context. The ability to generate AT2 cells carrying mutations associated with IPF has facilitated examining how infections

impact the phenotype of both genetically normal and mutant cells. This platform can potentially advance the analysis of signalling pathways involved in the host-pathogen response, specifically in acute exacerbations within the context of IPF. This reductionist approach can contribute significantly to our understanding of the complex dynamics during such occurrences, the genetic foundations of respiratory illnesses, and their intricate connections with infecting pathogens.

6.9 Future work

Recent advancements in technology have yielded fresh insights through newly published papers and data. Utilizing embryonic stem cells or induced pluripotent stem cells and their differentiation into distinct lung cell types, these studies have enhanced our comprehension of the identities of lung cell progenitors and the intricate regulatory networks dictating cell fate determinations [216, 234, 389–393]. Unfortunately, the timing of these discoveries prevented their integration into the framework of my PhD design. Nevertheless, they proved valuable in setting a standard for our model, further validating its capabilities. Additionally, these findings guided us toward prospective avenues of exploration and enhancement opportunities, contributing to their significance in shaping the direction of our work.

For instance, modifying the DIFF3 culture medium employed in this thesis to differentiate lung progenitors yielded a noteworthy increase in the obtained CPM+ cells after enrichment. Using the same modified medium could enhance the yield of NaPi2b+ cells. Furthermore, there's the potential to cultivate the purified AT2 cells within a modified ECM, varying the matrix substrates and altering collagen concentrations and synthetic components. This strategy could more accurately replicate the intricate microenvironment of the native distal lung, thus promoting enhanced cell maturation. Moreover, the identification of a transitional AT1-AT2 state in the single-cell RNA profiles by multiple research groups, characterised by the presence of both AT2 and AT1 transcripts along with their coexpression alongside conducting airway and other alveolar markers, support the hypothesis that epithelial cells within the remodelled distal lung of IPF adopt mixed differentiation states [237–239]. Therefore, there is an opportunity to investigate culture conditions capable of inducing such a phenotype in the AT2 cells generated within this thesis, holding potential for advancing IPF disease modelling.

Moreover, a comprehensive co-culture system derived entirely from hiPSCs offers the potential for extended investigations encompassing various epithelial and non-

epithelial cell types. This inclusive approach could encompass the incorporation of hiPSC-derived macrophages, endothelial cells, and other immune cells. This advancement stands to unveil the intricate dynamics of cell interactions within a setting that mimics the in vivo environment. Consequently, it has the capacity to elucidate how these interactions are influenced by factors such as injury, inflammation, infection, and the natural aging process. In the context of this study, such a co-culture framework could facilitate a thorough exploration of host responses to infectious agents like H1N1 and other pathogens that impact the respiratory system. Additionally, the differentiated AT2 cells and their co-cultures could be seamlessly integrated into 'organ-on-a-chip' setups. This integration could be extended to link these systems, culminating in the creation of 'body-on-a-chip' models. These models have the potential to emulate physiological conditions with greater fidelity, enabling more precise and sophisticated analyses [394, 395].

Furthermore, the cells can also be cultured in a cytostretcher system to replicate the mechanical forces of lung and simulate the mechanical stress occurring in the alveolar compartment during respiration under both normal physiological conditions and in the presence of disease. Research conducted on animal models utilising Cdc42 or Sin3a null AT2 cells demonstrated the heightened mechanical tension resulting from dysfunctional alveolar cells. This elevated tension has been associated with the development of a fibrotic pattern that extends from the periphery to the centre, closely resembling the fibrosis progression observed in IPF. This insight underscores the potential of subjecting hiPSC-derived AT2 cells to varying degrees of mechanical tension to create a gradient of fibrosis. This gradient could potentially offer a more accurate IPF model compared to the widely used bleomycin-induced model, which primarily elicits acute fibrosis, divergent from the gradual age-related or injury-triggered pulmonary fibrosis characteristic of IPF [112, 113].

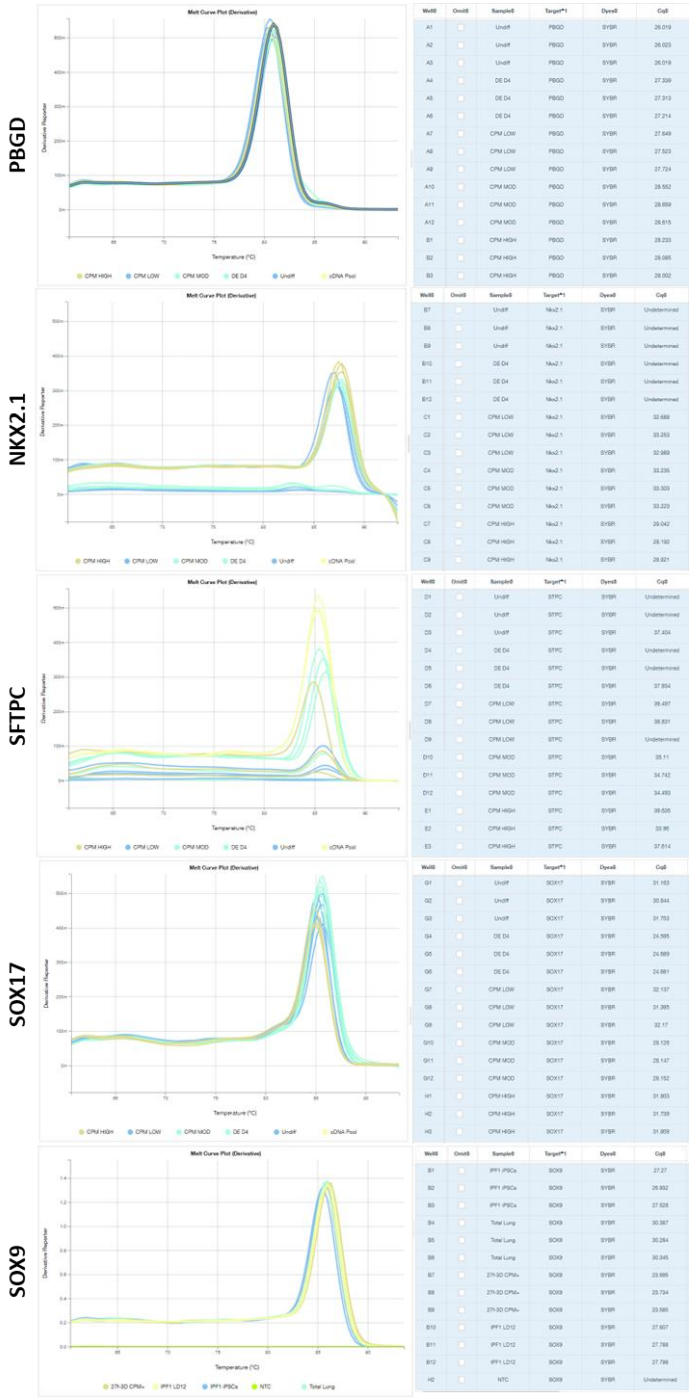
Incorporating these improvements into the platform developed within this thesis holds the promise of unveiling critical aspects of IPF pathogenesis. Ultimately, this effort could significantly contribute to the development of targeted therapeutic approaches for IPF patients, aiming to limit fibrosis progression in response to infections.

COVID-19 Impact Statement

The pandemic significantly impacted my PhD project and my mental health. The lockdown due to COVID-19 hampered my progress in the lab, especially during the second half of my PhD, with a downtime away from the lab and generation of data for nearly 8 months. Just before the pandemic, we were settling into a new building, so we dedicated some time to cleaning and organising the labs, calibrating incubators and finally bringing stem cells into culture with several drawbacks until growth rates and cell cultures had stabilised. At this point, the government established a national lockdown, and the lab had to shut down again. Once we were allowed back to the lab, we had to deal with several contaminations and issues that hindered the well-constructed experimental planning for re-opening the lab, having to adjust on the go. Overall, this period made it impossible to achieve the milestones set for the second year of my PhD, having to work extra hard to produce meaningful data without making up for the lost time.

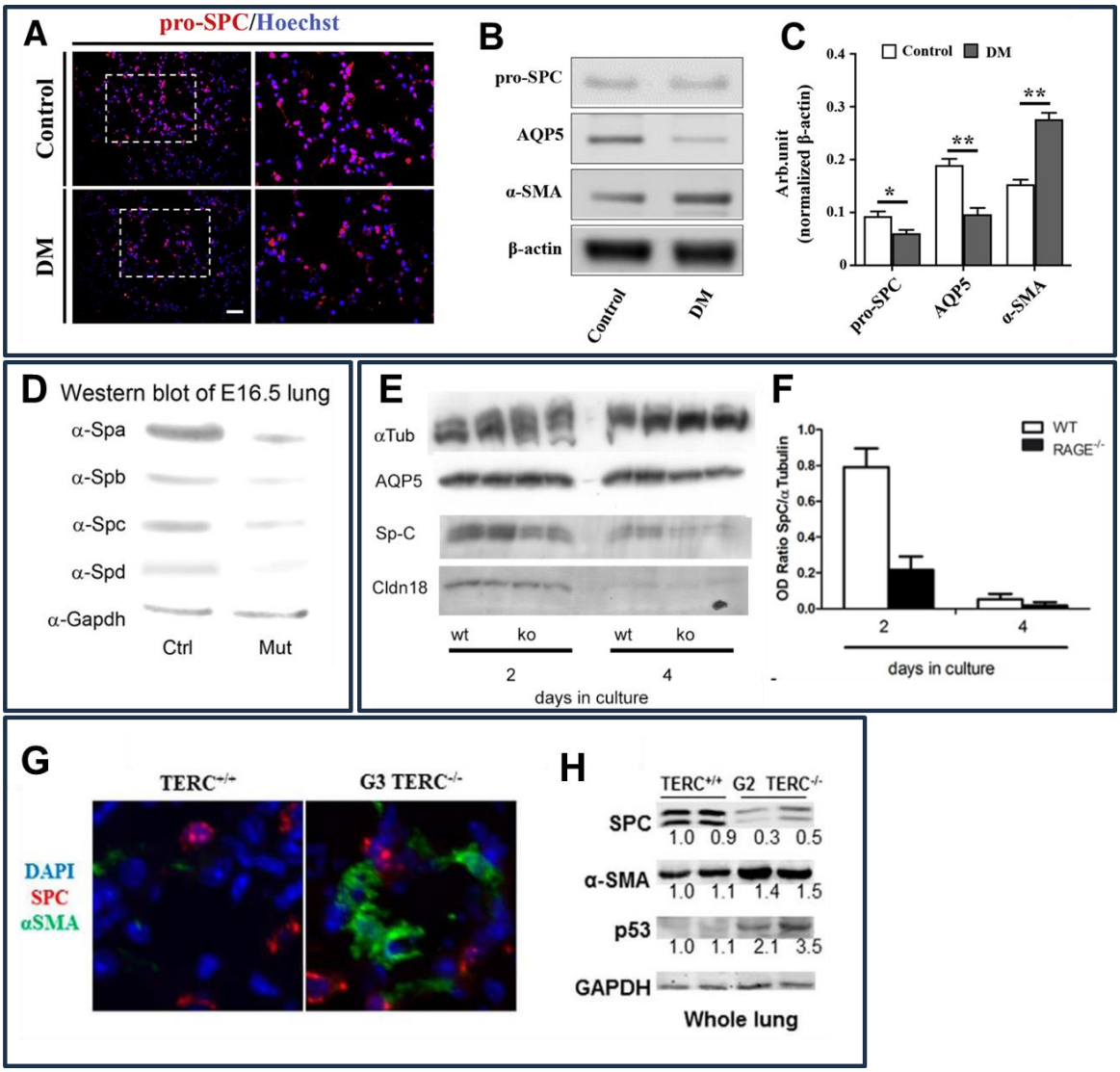
My research project and the expected outcomes changed during the pandemic. Initially, the plan was to identify molecules/pathways that could be targeted with an overexpression or knockdown strategy to further assess their impact on the platform we have been optimising. With the pandemic, we decided to change the approach and used the same platform to perform infection experiments without altering the expression of any target. This way, as soon as I got back to the lab, we focused on finishing the platform optimisation alongside the training to work with the virus for the infection experiments. We still had to modify the experimental planning on the go; for example, instead of doing single-cell RNA sequencing, we opted for bulk RNA sequencing due to the time the optimisation of the former would have needed. The designed contingency plan worked well, and the new objectives were accomplished.

Annex



Annex. 1 Primers' dissociation curves.

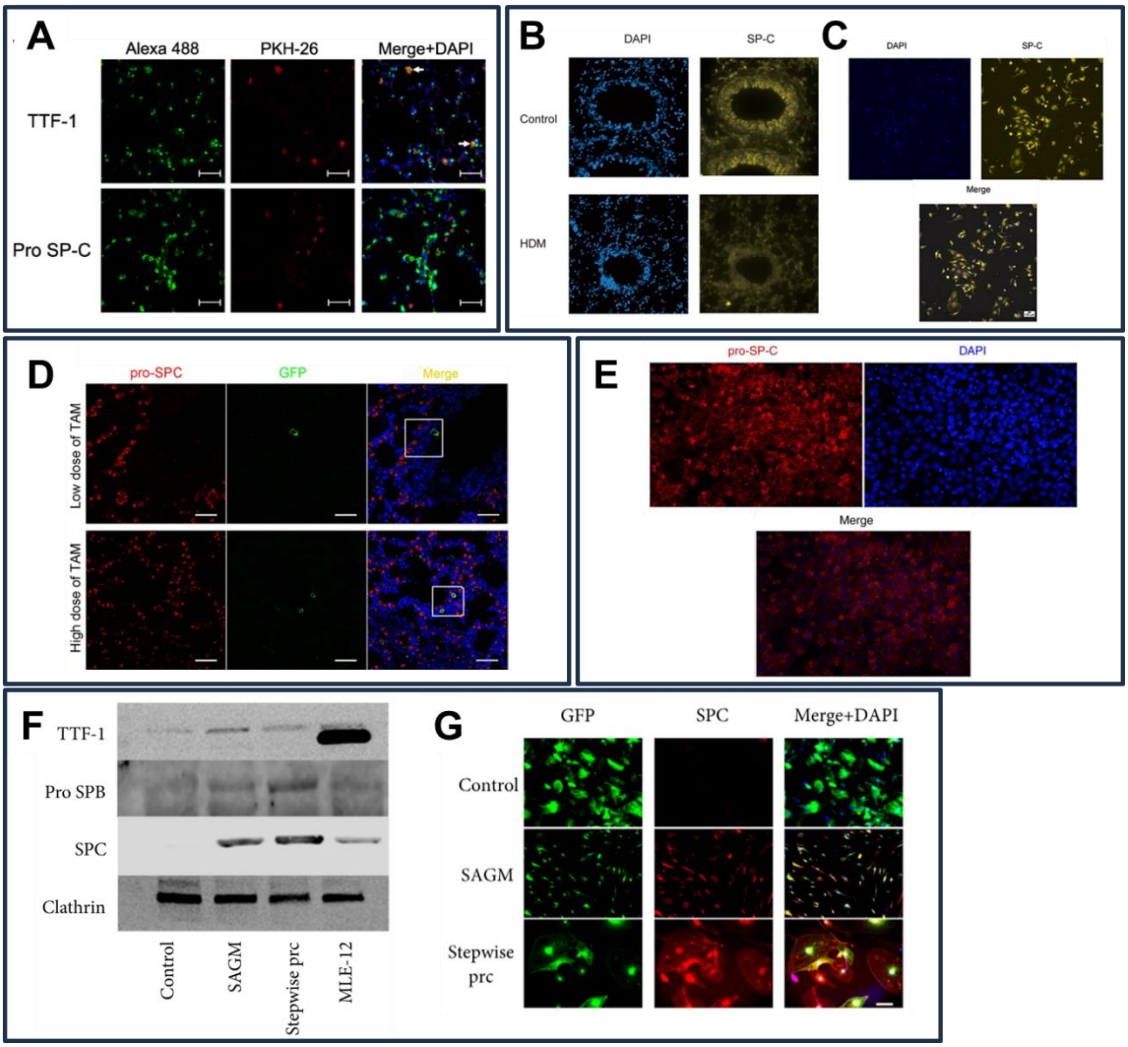
Dissociation curves depicting representative genes (PBGD, NKX2.1, SFTPC, SOX17, SOX9) demonstrate the amplification of a single product. These curves were generated from data collected using plates run on the Applied Biosystems® 7500 Fast Real-Time PCR Systems.



Annex. 2 Surfactant protein C Ab90716 antibody.

Literature data sets generated using the Ab90716 pro-surfactant protein C antibody. A-C) Wang et al. investigated the cellular-biological mechanisms of the pulmonary fibrotic change in gut-lung microbiota dysbiosis induced by diabetes mellitus (DM). A- immunofluorescence staining from control and DM mouse lung tissue. B- Western blot showing pro-surfactant protein C (pro-SPC), C- Bar chart shows the ratios of the band arbitrary units of the corresponding gene expression normalised to β-actin [337]. D) Western blot analysis showing expression of genes encoding surfactant proteins Spa, Spb, Spc, and Spd in the lungs mutant Pbx1 mice embryos at E16.5. Samples were prepared from control and mutant mice, using Gapdh as the internal control [396]. E-F) E- Western blot analysis for advanced glycation end products (RAGE) mutant and control mice. Alveolar epithelial cell lysates were prepared at day 2 and 4 post-isolation. F- Densitometric analysis of the bands in the blot [339]. G-H) G- pro-SPC

immunofluorescence staining of lung tissue of TERC-null mice. H-Western blot analysis of G2 TERC-null and control mice for SPC, α-SMA, and p53. Gapdh was used as internal control [338].

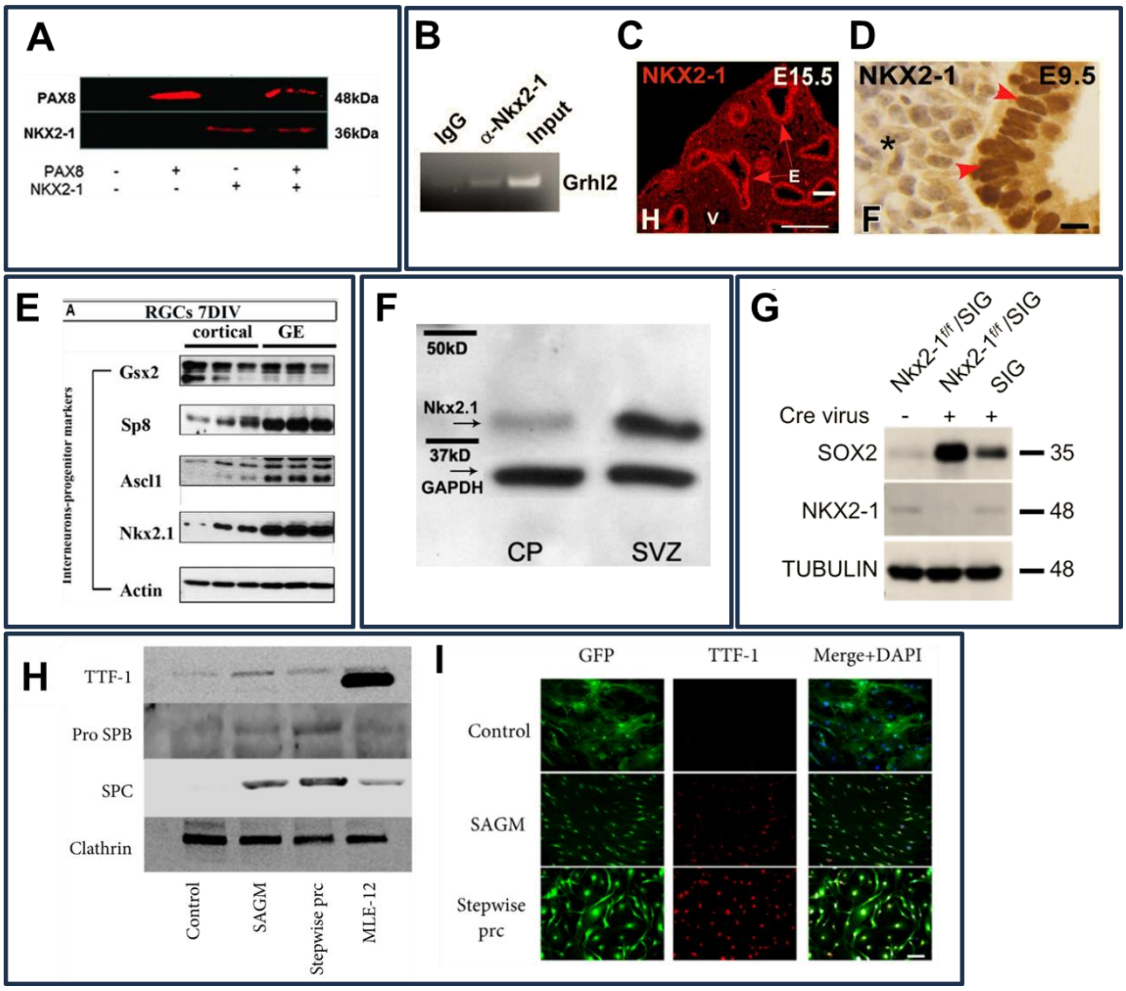


Annex. 3 Surfactant protein C Ab170699 antibody.

Examples of literature data sets generated using the Ab170699 pro-surfactant protein C antibody. A) Fujioka et al., used lung sections from SCID-beige mice administered intravenously with human adipose tissue-derived stem cells (ADSCs) to observe the expression of lung epithelial cells markers TTF-1 and pro-SPC [334]. B-C) B-Immunofluorescent staining for pro-SPC in lung sections from a murine model of asthma, established by the administration of House dust mite (HDM) extract C-Immunofluorescence staining of pro-SPC in primary AT2 cells [340]. D) Immunofluorescence for GFP and pro-SPC in lung slices from a tamoxifen inducible Follistatin like-1 (Fstl1)-reporter mouse strain [336]. E) Immunofluorescence staining for pro-SP-C in mouse lung epithelial (MLE)-12 cells stimulated with various

concentrations of bleomycin to induces endoplasmic reticulum stress [335]. F-G) F- Western blot analysis for expression of TTF-1, Pro SPB, and SPC of ADSCs cultured in control medium (Control), SAGM or with stepwise protocol, to demonstrate their differentiation into type 2 alveolar cells. G- Immunofluorescence staining of ADSCs with anti-SPC [329].

a



Annex. 4 NKX2.1 ab76013 antibody.

Examples of literature data sets generated using the Ab76013 NKX2.1 antibody. A) Ma et al., established hESCs expressing human thyroid transcription factors PAX8-eGFP and NKX2-1-mCherry using reporter gene tagged pEZ-lentiviral vectors. Western blot shows cells expressing either PAX8 or NKX2-1, or both [330]. B-D) B- ChIP-PCR analysis where NKX2.1 antibody was used to immunoprecipitate chromatin fragments bound by NKX2.1. C- Immunofluorescence staining of NKX2.1 in lungs from mouse embryos in day E15.5. D- Immunohistochemical analyses of NKX2.1 protein expression of mouse embryos in day E9.5 in the nuclei of lung epithelial cells [332]. E)

Expression of Nkx2.1 as markers of interneurons expressed in cortical and ventral telencephalon (GE) enriched cortical radial glia cells (RGCs) after 7 days in vitro (DIV) [331]. F) Western blot analysis of Nkx2.1 expression in the subventricular zone (SVZ) and cortical plate (CP) of a 21-week-old foetal brain [333]. G) Western blot analysis for SOX2 (35 kDa), Nkx2-1 (48 kDa), and tubulin (48 kDa, loading control) in lung cells from SIG and Nkx2-1f/f-SIGf/f mice (NKX2.1 loss-of-function alleles) with or without Cre virus [328]. H-I) H- Western blot analysis for expression of TTF-1, Pro SPB, and SPC of ADSCs cultured in control medium (Control), SAGM or with stepwise protocol, to demonstrate their differentiation into type 2 alveolar cells. I- Immunofluorescence staining of ADSCs with anti-TTF1 (also known as NKX2.1) [329].

References

[1] Maher TM, Bendstrup E, Dron L, et al. Global incidence and prevalence of idiopathic pulmonary fibrosis. *Respir Res* 2021; 22: 197.

[2] Strongman H, Kausar I, Maher TM. Incidence, Prevalence, and Survival of Patients with Idiopathic Pulmonary Fibrosis in the UK. *Adv Ther* 2018; 35: 724–736.

[3] Lawson WE, Loyd JE, Degryse AL. Genetics in Pulmonary Fibrosis—Familial Cases Provide Clues to the Pathogenesis of Idiopathic Pulmonary Fibrosis. *Am J Med Sci* 2011; 341: 439–443.

[4] Wang Y, Zhou W. Candidate genes of idiopathic pulmonary fibrosis: current evidence and research. *Appl Clin Genet* 2016; 9: 5.

[5] Selman M. Role of Epithelial Cells in Idiopathic Pulmonary Fibrosis: From Innocent Targets to Serial Killers. *Proc Am Thorac Soc* 2006; 3: 364–372.

[6] Baron RM, Choi AJS, Owen CA, et al. Genetically manipulated mouse models of lung disease: potential and pitfalls. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2012; 302: L485–L497.

[7] Cuevas-Ocaña S, Jin Ye Y, Aushev M, et al. A Cell-Based Optimised Approach for Rapid and Efficient Gene Editing of Human Pluripotent Stem Cells. *Int J Mol Sci* 2023; 24: 10266.

[8] Korogi Y, Gotoh S, Ikeo S, et al. In Vitro Disease Modeling of Hermansky-Pudlak Syndrome Type 2 Using Human Induced Pluripotent Stem Cell-Derived Alveolar Organoids. *Stem Cell Reports* 2019; 12: 431–440.

[9] Bauer Y, White ES, de Bernard S, et al. MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. *ERJ Open Res*; 3. .2017. DOI: 10.1183/23120541.00074-2016.

[10] Moodley YP, Scaffidi AK, Misso NL, et al. Fibroblasts Isolated from Normal Lungs and Those with Idiopathic Pulmonary Fibrosis Differ in Interleukin-6/gp130-Mediated Cell Signaling and Proliferation. *Am J Pathol* 2003; 163: 345–354.

[11] Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; 45: 613–620.

[12] Hannan NRF, Sampaziotis F, Segeritz C-P, et al. Generation of Distal Airway Epithelium from Multipotent Human Foregut Stem Cells. *Stem Cells Dev* 2015; 24: 1680–1690.

[13] Whitsett JA, Kalin T V., Xu Y, et al. Building and Regenerating the Lung Cell by Cell. *Physiol Rev* 2019; 99: 513–554.

[14] Burri PH. Fetal and Postnatal Development of the Lung. *Annu Rev Physiol*; 46. -1984. DOI: 10.1146/annurev.ph.46.030184.003153.

[15] Zaret KS. From Endoderm to Liver Bud: Paradigms of Cell Type Specification and Tissue Morphogenesis. In: *Current Topics in Developmental Biology*. 2016. .2016. DOI: 10.1016/bs.ctdb.2015.12.015.

[16] Lazzaro D, Price M, Felice M De, et al. The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development* 1991; 113: 1093–1104.

[17] Rawlins EL. Lung Epithelial Progenitor Cells: Lessons from Development. *Proc Am Thorac Soc* 2008; 5: 675–681.

[18] Rankin SA, Zorn AM. Gene Regulatory Networks Governing Lung Specification. *J Cell Biochem* 2014; 115: 1343–1350.

[19] Hurley K, Ding J, Villacorta-Martin C, et al. Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. *Cell Stem Cell* 2020; 26: 593-608.e8.

[20] Frank DB, Penkala IJ, Zepp JA, et al. Early lineage specification defines alveolar epithelial ontogeny in the murine lung. *Proceedings of the National Academy of Sciences* 2019; 116: 4362–4371.

[21] Herriges M, Morrisey EE. Lung development: orchestrating the generation and regeneration of a complex organ. *Development* 2014; 141: 502–513.

[22] Rackley CR, Stripp BR. Building and maintaining the epithelium of the lung. *Journal of Clinical Investigation*; 122. .2012. DOI: 10.1172/JCI60519.

[23] Zorn AM, Wells JM. Vertebrate Endoderm Development and Organ Formation. *Annu Rev Cell Dev Biol* 2009; 25: 221–251.

[24] Sinner D, Rankin S, Lee M, et al. Sox17 and β-catenin cooperate to regulate the transcription of endodermal genes. *Development*; 131. .2004. DOI: 10.1242/dev.01176.

[25] Goss AM, Tian Y, Tsukiyama T, et al. Wnt2/2b and β-Catenin Signaling Are Necessary and Sufficient to Specify Lung Progenitors in the Foregut. *Dev Cell*; 17. .2009. DOI: 10.1016/j.devcel.2009.06.005.

[26] Domyan ET, Ferretti E, Throckmorton K, et al. Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. *Development*; 138. .2011. DOI: 10.1242/dev.053694.

[27] Que J, Choi M, Ziel JW, et al. Morphogenesis of the trachea and esophagus: Current players and new roles for noggin and Bmps. *Differentiation*; 74. .2006. DOI: 10.1111/j.1432-0436.2006.00096.x.

[28] Sekine K, Ohuchi H, Fujiwara M, et al. Fgf10 is essential for limb and lung formation. *Nat Genet*; 21. .1999. DOI: 10.1038/5096.

[29] Isago H, Mitani A, Mikami Y, et al. Epithelial expression of YAP and TAZ is sequentially required in lung development. *Am J Respir Cell Mol Biol*; 62. .2020. DOI: 10.1165/rcmb.2019-0218OC.

[30] Chen F, Cao Y, Qian J, et al. A retinoic acid-dependent network in the foregut controls formation of the mouse lung primordium. *Journal of Clinical Investigation*; 120. .2010. DOI: 10.1172/JCI40253.

[31] Tompkins DH, Besnard V, Lange AW, et al. Sox2 activates cell proliferation and differentiation in the respiratory epithelium. *Am J Respir Cell Mol Biol*; 45. .2011. DOI: 10.1165/rcmb.2010-0149OC.

[32] Rockich BE, Hrycaj SM, Shih HP, et al. Sox9 plays multiple roles in the lung epithelium during branching morphogenesis. *Proc Natl Acad Sci U S A*; 110. .2013. DOI: 10.1073/pnas.1311847110.

[33] Gouveia L, Betsholtz C, Andrae J. PDGF-A signaling is required for secondary alveolar septation and controls epithelial proliferation in the developing lung. *Development (Cambridge)*; 145. .2018. DOI: 10.1242/dev.161976.

[34] Zepp JA, Zacharias WJ, Frank DB, et al. Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. *Cell* 2017; 170: 1134-1148.e10.

[35] Wang Y, Frank DB, Morley MP, et al. HDAC3-Dependent Epigenetic Pathway Controls Lung Alveolar Epithelial Cell Remodeling and Spreading via miR-17-92 and TGF-β Signaling Regulation. *Dev Cell*; 36. .2016. DOI: 10.1016/j.devcel.2015.12.031.

[36] Zacharias WJ, Frank DB, Zepp JA, et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* 2018; 555: 251–255.

[37] Liggins G, Howie R. A controlled trial of antepartum glucocorticoid treatment for prevention of the respiratory distress syndrome in premature infants. *Pediatrics* 1972; 50: 512–525.

[38] Barkauskas CE, Cronce MJ, Rackley CR, et al. Type 2 alveolar cells are stem cells in adult lung. *Journal of Clinical Investigation* 2013; 123: 3025–3036.

[39] Zepp JA, Morrisey EE. Cellular crosstalk in the development and regeneration of the respiratory system. *Nat Rev Mol Cell Biol* 2019; 20: 551–566.

[40] Borthwick DW, Shahbazian M, Krantz QT, et al. Evidence for stem-cell niches in the tracheal epithelium. *Am J Respir Cell Mol Biol*; 24. .2001. DOI: 10.1165/ajrcmb.24.6.4217.

[41] AVERY ME. Surface Properties in Relation to Atelectasis and Hyaline Membrane Disease. *Arch Pediatr Adolesc Med* 1959; 97: 517.

[42] Crapo JD, Barry BE, Gehr P, et al. Cell number and cell characteristics of the normal human lung. *Am Rev Respir Dis* 1982; 126: 332–7.

[43] Yang J, Hernandez BJ, Alanis DM, et al. The development and plasticity of alveolar type 1 cells. *Development (Cambridge)*; 143. .2016. DOI: 10.1242/dev.130005.

[44] Macklin CC. THE PULMONARY ALVEOLAR MUCOID FILM AND THE PNEUMONOCYTES. *The Lancet* 1954; 263: 1099–1104.

[45] Clements JA. Surface Tension of Lung Extracts. *Exp Biol Med* 1957; 95: 170–172.

[46] Thane K, Ingenito EP, Hoffman AM. Lung regeneration and translational implications of the postpneumonectomy model. *Translational Research* 2014; 163: 363–376.

[47] King RJ. Pulmonary surfactant. *J Appl Physiol* 1982; 53: 1–8.

[48] Hsia CCW, Hyde DM, Weibel ER. Lung Structure and the Intrinsic Challenges of Gas Exchange. In: *Comprehensive Physiology*. Hoboken, NJ, USA: Wiley, pp. 827–895.

[49] Weaver TE, Conkright JJ. FUNCTIONS OF SURFACTANT PROTEINS B AND C, www.annualreviews.org (2001, accessed 18 October 2022).

[50] Williams MC, Hawgood S, Hamilton RL. Changes in lipid structure produced by surfactant proteins SP-A, SP-B, and SP-C. *Am J Respir Cell Mol Biol*; 5. .1991. DOI: 10.1165/ajrcmb/5.1.41.

[51] Voorhout WF, Weaver TE, Haagsman HP, et al. Biosynthetic routing of pulmonary surfactant proteins in alveolar type II cells. *Microsc Res Tech* 1993; 26: 366–373.

[52] Wright JR, Clements JA. Metabolism and turnover of lung surfactant. *American Review of Respiratory Disease*; 136. .1987. DOI: 10.1164/ajrccm/136.2.426.

[53] Whitsett JA, Wert SE, Weaver TE. Alveolar Surfactant Homeostasis and the Pathogenesis of Pulmonary Disease. *Annu Rev Med* 2010; 61: 105–119.

[54] Whitsett JA, Alenghat T. Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nature Immunology* 2015; 16: 27–35.

[55] Sever N, Miličić G, Bodnar NO, et al. Mechanism of Lamellar Body Formation by Lung Surfactant Protein B. *Mol Cell* 2021; 81: 49-66.e8.

[56] Ryu JH, Moua T, Daniels CE, et al. Idiopathic Pulmonary Fibrosis: Evolving Concepts. *Mayo Clin Proc* 2014; 89: 1130–1142.

[57] Cavazza A, Rossi G, Carbonelli C, et al. The role of histology in idiopathic pulmonary fibrosis: An update. *Respir Med* 2010; 104: S11–S22.

[58] Plantier L, Crestani B, Wert SE, et al. Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. *Thorax*; 66. .2011. DOI: 10.1136/thx.2010.151555.

[59] Spagnolo P, Sverzellati N, Rossi G, et al. Idiopathic pulmonary fibrosis: An update. *Ann Med* 2015; 47: 15–27.

[60] Froidure A, Marchal-Duval E, Homps-Legrand M, et al. Chaotic activation of developmental signalling pathways drives idiopathic pulmonary fibrosis. *European Respiratory Review*; 29. .2020. DOI: 10.1183/16000617.0140-2019.

[61] Nakamura Y, Suda T. Idiopathic pulmonary fibrosis: Diagnosis and clinical manifestations. *Clin Med Insights Circ Respir Pulm Med*; 9. .2016. DOI: 10.4137/CCRPM.S39897.

[62] Collard HR, King TE, Bartelson BB, et al. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*; 168. .2003. DOI: 10.1164/rccm.200211-1311OC.

[63] Flaherty KR, Mumford JA, Murray S, et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med*; 168. .2003. DOI: 10.1164/rccm.200209-1112OC.

[64] Haschek WM, Witschi H. Pulmonary Fibrosis—A possible mechanism. *Toxicol Appl Pharmacol* 1979; 51: 475–487.

[65] Raghu G, Collard HR, Egan JJ, et al. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *Am J Respir Crit Care Med* 2011; 183: 788–824.

[66] Krauss E, Gehrken G, Drakopanagiotakis F, et al. Clinical characteristics of patients with familial idiopathic pulmonary fibrosis (f-IPF). *BMC Pulm Med*; 19. .2019. DOI: 10.1186/s12890-019-0895-6.

[67] Raghu G, Rochwerg B, Zhang Y, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *Am J Respir Crit Care Med* 2015; 192: e3–e19.

[68] García-Sancho C, Buendía-Roldán I, Fernández-Plata MR, et al. Familial pulmonary fibrosis is the strongest risk factor for idiopathic pulmonary fibrosis. *Respir Med*; 105. .2011. DOI: 10.1016/j.rmed.2011.08.022.

[69] Marshall RP. Adult familial cryptogenic fibrosing alveolitis in the United Kingdom. *Thorax* 2000; 55: 143–146.

[70] Lee HL, Ryu JH, Wittmer MH, et al. Familial idiopathic pulmonary fibrosis: Clinical features and outcome. *Chest*; 127. .2005. DOI: 10.1378/chest.127.6.2034.

[71] Sgalla G, Iovene B, Calvello M, et al. Idiopathic pulmonary fibrosis: pathogenesis and management. *Respir Res* 2018; 19: 32.

[72] Maher TM, Wells AU, Laurent GJ. Idiopathic pulmonary fibrosis: multiple causes and multiple mechanisms? *European Respiratory Journal* 2007; 30: 835–839.

[73] Tatler AL, Jenkins G. TGF-β activation and lung fibrosis. In: *Proceedings of the American Thoracic Society*. 2012. .2012. DOI: 10.1513/pats.201201-003AW.

[74] Selman M, King TE, Pardo A. Idiopathic Pulmonary Fibrosis: Prevailing and Evolving Hypotheses about Its Pathogenesis and Implications for Therapy. *Ann Intern Med* 2001; 134: 136.

[75] Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. *New England Journal of Medicine* 2018; 378: 1811–1823.

[76] Sisson TH, Mendez M, Choi K, et al. Targeted injury of type II alveolar epithelial cells induces pulmonary fibrosis. *Am J Respir Crit Care Med*; 181. .2010. DOI: 10.1164/rccm.200810-1615OC.

[77] Beers MF, Nureki SI, Mulugeta S. The Role of Epithelial Cell Quality Control in Health and Disease of the Distal Lung. In: *Lung Epithelial Biology in the Pathogenesis of Pulmonary Disease*. 2017. .2017. DOI: 10.1016/B978-0-12-803809-3.00008-7.

[78] van Moorsel CHM, van der Vis JJ, Grutters JC. Genetic disorders of the surfactant system: Focus on adult disease. *European Respiratory Review*; 30. .2021. DOI: 10.1183/16000617.0085-2020.

[79] Korfei M, Ruppert C, Mahavadi P, et al. Epithelial Endoplasmic Reticulum Stress and Apoptosis in Sporadic Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2008; 178: 838–846.

[80] Kaur A, Mathai SK, Schwartz DA. Genetics in Idiopathic Pulmonary Fibrosis Pathogenesis, Prognosis, and Treatment. *Front Med (Lausanne)* 2017; 4: 154.

[81] Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2020; 201: 564–574.

[82] Pardo A, Selman M. The interplay of the genetic architecture, aging, and environmental factors in the pathogenesis of idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol*; 64. .2021. DOI: 10.1165/RCMB.2020-0373PS.

[83] Nishioka Y. Pathogenesis of IPF. In: *Idiopathic Pulmonary Fibrosis*. Tokyo: Springer Japan, pp. 43–58.

[84] Molyneaux PL, Maher TM. The role of infection in the pathogenesis of idiopathic pulmonary fibrosis. *European Respiratory Review* 2013; 22: 376–381.

[85] Williams KJ, Robinson NE, Lim A, et al. Experimental Induction of Pulmonary Fibrosis in Horses with the Gammaherpesvirus Equine Herpesvirus 5. *PLoS One* 2013; 8: e77754.

[86] Stoolman JS, Vannella KM, Coomes SM, et al. Latent infection by γherpesvirus stimulates profibrotic mediator release from multiple cell types. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2011; 300: L274–L285.

[87] Lok SS, Haider Y, Howell D, et al. Murine gammaherpes virus as a cofactor in the development of pulmonary fibrosis in bleomycin resistant mice. *European Respiratory Journal* 2002; 20: 1228–1232.

[88] Molyneaux PL, Cox MJ, Wells AU, et al. Changes in the respiratory microbiome during acute exacerbations of idiopathic pulmonary fibrosis. *Respir Res* 2017; 18: 29.

[89] O'Dwyer DN, Ashley SL, Gurczynski SJ, et al. Lung Microbiota Contribute to Pulmonary Inflammation and Disease Progression in Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019; 199: 1127–1138.

[90] Burman A, Tanjore H, Blackwell TS. Endoplasmic reticulum stress in pulmonary fibrosis. *Matrix Biology* 2018; 68–69: 355–365.

[91] Huang WJ, Tang XX. Virus infection induced pulmonary fibrosis. *J Transl Med* 2021; 19: 496.

[92] Jolly L, Stavrou A, Vanderstoken G, et al. Influenza Promotes Collagen Deposition via αvβ6 Integrin-mediated Transforming Growth Factor β Activation. *Journal of Biological Chemistry* 2014; 289: 35246–35263.

[93] Qiao J, Zhang M, Bi J, et al. Pulmonary fibrosis induced by H5N1 viral infection in mice. *Respir Res* 2009; 10: 107.

[94] Torres-González E, Bueno M, Tanaka A, et al. Role of endoplasmic reticulum stress in age-related susceptibility to lung fibrosis. *Am J Respir Cell Mol Biol*; 46. .2012. DOI: 10.1165/rcmb.2011-0224OC.

[95] Naik PN, Horowitz JC, Moore TA, et al. Pulmonary Fibrosis Induced by - Herpesvirus in Aged Mice Is Associated With Increased Fibroblast Responsiveness to Transforming Growth Factor- . *J Gerontol A Biol Sci Med Sci* 2012; 67: 714–725.

[96] Shojaee A, Siner JM, Zinchuk A, et al. Viral Pneumonia is Associated with Increased Risk and Earlier Development of Post-Inflammatory Pulmonary Fibrosis. *medRxiv* 2021; 2021.03.08.21252412.

[97] Beers MF, Moodley Y. When Is an Alveolar Type 2 Cell an Alveolar Type 2 Cell? A Conundrum for Lung Stem Cell Biology and Regenerative Medicine. *Am J Respir Cell Mol Biol* 2017; 57: 18–27.

[98] Matute-Bello G, Frevert CW, Martin TR. Animal models of acute lung injury. *American Journal of Physiology - Lung Cellular and Molecular Physiology*; 295. .2008. DOI: 10.1152/ajplung.00010.2008.

[99] Morrisey EE, Hogan BLM. Preparing for the First Breath: Genetic and Cellular Mechanisms in Lung Development. *Developmental Cell*; 18. .2010. DOI: 10.1016/j.devcel.2009.12.010.

[100] Coulombe P, Paliouras GN, Clayton A, et al. Endothelial Sash1 Is Required for Lung Maturation through Nitric Oxide Signaling. *Cell Rep* 2019; 27: 1769-1780.e4.

[101] Zhuo Y, Hoyle GW, Shan B, et al. Over-expression of PDGF-C using a lung specific promoter results in abnormal lung development. *Transgenic Res* 2006; 15: 543–555.

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

[103] Bridges JP, Ikegami M, Brilli LL, et al. LPCAT1 regulates surfactant phospholipid synthesis and is required for transitioning to air breathing in mice. *Journal of Clinical Investigation*; 120. .2010. DOI: 10.1172/JCI38061.

[104] Lawson WE, Polosukhin V V., Stathopoulos GT, et al. Increased and Prolonged Pulmonary Fibrosis in Surfactant Protein C-Deficient Mice Following Intratracheal Bleomycin. *Am J Pathol* 2005; 167: 1267–1277.

[105] Claussen CA, Long EC. Nucleic acid recognition by metal complexes of bleomycin. *Chem Rev*; 99. .1999. DOI: 10.1021/cr980449z.

[106] Blum RH, Carter SK, Agre K. A clinical review of bleomycin—a new antineoplastic agent. *Cancer*; 31. .1973. DOI: 10.1002/1097-0142(197304)31:4<903::AID-CNCR2820310422>3.0.CO;2-N.

[107] Moeller A, Ask K, Warburton D, et al. The bleomycin animal model: A useful tool to investigate treatment options for idiopathic pulmonary fibrosis? *Int J Biochem Cell Biol* 2008; 40: 362–382.

[108] Walters DM, Kleeberger SR. Mouse models of bleomycin-induced pulmonary fibrosis. *Current Protocols in Pharmacology*. .2008. DOI: 10.1002/0471141755.ph0546s40.

[109] Chua F, Gauldie J, Laurent GJ. Pulmonary Fibrosis. *Am J Respir Cell Mol Biol* 2005; 33: 9–13.

[110] Borzone G, Moreno R, Urrea R, et al. Bleomycin-induced chronic lung damage does not resemble human idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*; 163. .2001. DOI: 10.1164/ajrccm.163.7.2006132.

[111] Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* 2014; 507: 190–194.

[112] Wu H, Yu Y, Huang H, et al. Progressive Pulmonary Fibrosis Is Caused by Elevated Mechanical Tension on Alveolar Stem Cells. *Cell* 2020; 180: 107-121.e17.

[113] Yao C, Guan X, Carraro G, et al. Senescence of Alveolar Stem Cells Drives Progressive Pulmonary Fibrosis. *SSRN Electronic Journal*. .2019. DOI: 10.2139/ssrn.3438364.

[114] Giangreco A, Reynolds SD, Stripp BR. Terminal bronchioles harbor a unique airway stem cell population that localizes to the bronchoalveolar duct junction. *Am J Pathol* 2002; 161: 173–182.

[115] Lewis DI. Animal experimentation: Implementation and application of the 3Rs. *Emerging Topics in Life Sciences*; 3. .2019. DOI: 10.1042/ETLS20190061.

[116] Sanderson MJ. Exploring lung physiology in health and disease with lung slices. *Pulm Pharmacol Ther* 2011; 24: 452–465.

[117] Bai Y, Krishnamoorthy N, Patel KR, et al. Cryopreserved human precision-cut lung slices as a bioassay for live tissue banking a viability study of bronchodilation with bitter-taste receptor agonists. *Am J Respir Cell Mol Biol*; 54. .2016. DOI: 10.1165/rcmb.2015-0290MA.

[118] Akram KM, Yates LL, Mongey R, et al. Live imaging of alveologenesis in precision-cut lung slices reveals dynamic epithelial cell behaviour. *Nat Commun* 2019; 10: 1178.

[119] Placke ME, Fisher GL. Adult peripheral lung organ culture--a model for respiratory tract toxicology. *Toxicol Appl Pharmacol* 1987; 90: 284–98.

[120] Alsafadi HN, Staab-Weijnitz CA, Lehmann M, et al. An ex vivo model to induce early fibrosis-like changes in human precision-cut lung slices. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2017; 312: L896–L902.

[121] Roach KM, Sutcliffe A, Matthews L, et al. A model of human lung fibrogenesis for the assessment of anti-fibrotic strategies in idiopathic pulmonary fibrosis. *Sci Rep*; 8. .1 December 2018. DOI: 10.1038/S41598-017-18555-9.

[122] Roach KM, Castells E, Dixon K, et al. Evaluation of Pirfenidone and Nintedanib in a Human Lung Model of Fibrogenesis. *Front Pharmacol*; 12. .12 October 2021. DOI: 10.3389/FPHAR.2021.679388/FULL.

[123] Temann A, Golovina T, Neuhaus V, et al. Evaluation of inflammatory and immune responses in long-term cultured human precision-cut lung slices. *Hum Vaccin Immunother*; 13. .2017. DOI: 10.1080/21645515.2017.1264794.

[124] Dobbs LG, Gonzalez R, Williams MC. An improved method for isolating type II cells in high yield and purity. *Am Rev Respir Dis* 1986; 134: 141–5.

[125] Mason RJ, Williams MC, Greenleaf RD, et al. Isolation and properties of type II alveolar cells from rat lung. *Am Rev Respir Dis* 1977; 115: 1015–26.

[126] Barkauskas CE, Chung M-I, Fioret B, et al. Lung organoids: current uses and future promise. *Development* 2017; 144: 986–997.

[127] Yu W, Fang X, Ewald A, et al. Formation of Cysts by Alveolar Type II Cells in Three-dimensional Culture Reveals a Novel Mechanism for Epithelial Morphogenesis. *Mol Biol Cell* 2007; 18: 1693–1700.

[128] Barkauskas CE, Cronce MJ, Rackley CR, et al. Type 2 alveolar cells are stem cells in adult lung. *Journal of Clinical Investigation* 2013; 123: 3025–3036.

[129] McQualter JL, Yuen K, Williams B, et al. Evidence of an epithelial stem/progenitor cell hierarchy in the adult mouse lung. *Proc Natl Acad Sci U S A*; 107. .2010. DOI: 10.1073/pnas.0909207107.

[130] Katsura H, Sontake V, Tata A, et al. Human Lung Stem Cell-Based Alveolospheres Provide Insights into SARS-CoV-2-Mediated Interferon Responses and Pneumocyte Dysfunction. *Cell Stem Cell* 2020; 27: 890-904.e8.

[131] Salahudeen AA, Choi SS, Rustagi A, et al. Progenitor identification and SARS-CoV-2 infection in human distal lung organoids. *Nature* 2020; 588: 670–675.

[132] Youk J, Kim T, Evans K V., et al. Three-Dimensional Human Alveolar Stem Cell Culture Models Reveal Infection Response to SARS-CoV-2. *Cell Stem Cell* 2020; 27: 905-919.e10.

[133] Green MD, Chen A, Nostro MC, et al. Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nat Biotechnol* 2011; 29: 267–273.

[134] Longmire TA, Ikonomou L, Hawkins F, et al. Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. *Cell Stem Cell* 2012; 10: 398–411.

[135] Mou H, Zhao R, Sherwood R, et al. Generation of Multipotent Lung and Airway Progenitors from Mouse ESCs and Patient-Specific Cystic Fibrosis iPSCs. *Cell Stem Cell* 2012; 10: 385–397.

[136] Ali NN, Edgar AJ, Samadikuchaksaraei A, et al. Derivation of type II alveolar epithelial cells from murine embryonic stem cells. *Tissue Eng*; 8. .2002. DOI: 10.1089/107632702760240463.

[137] Rippon HJ, Ali NN, Polak JM, et al. Initial observations on the effect of medium composition on the differentiation of murine embryonic stem cells to alveolar type II cells. *Cloning Stem Cells*; 6. .2004. DOI: 10.1089/1536230041372328.

[138] Wang D, Haviland DL, Burns AR, et al. A pure population of lung alveolar epithelial type II cells derived from human embryonic stem cells. *Proceedings of the National Academy of Sciences* 2007; 104: 4449–4454.

[139] Samadikuchaksaraei A, Cohen S, Isaac K, et al. Derivation of distal airway epithelium from human embryonic stem cells. *Tissue Eng*; 12. .2006. DOI: 10.1089/ten.2006.12.867.

[140] Van Haute L, De Block G, Liebaers I, et al. Generation of lung epithelial-like tissue from human embryonic stem cells. *Respir Res*; 10. .2009. DOI: 10.1186/1465-9921-10-105.

[141] Takahashi K, Tanabe K, Ohnuki M, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* 2007; 131: 861–872.

[142] Huang SXL, Green MD, de Carvalho AT, et al. The in vitro generation of lung and airway progenitor cells from human pluripotent stem cells. *Nat Protoc* 2015; 10: 413–425.

[143] Ghaedi M, Calle EA, Mendez JJ, et al. Human iPS cellâ€“derived alveolar epithelium repopulates lung extracellular matrix. *Journal of Clinical Investigation* 2013; 123: 4950–4962.

[144] Hannan NRF, Sampaziotis F, Segeritz C-P, et al. Generation of Distal Airway Epithelium from Multipotent Human Foregut Stem Cells. *Stem Cells Dev* 2015; 24: 1680–1690.

[145] Jacob A, Morley M, Hawkins F, et al. Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell* 2017; 21: 472-488.e10.

[146] Kubo A, Shinozaki K, Shannon JM, et al. Development of definitive endoderm from embryonic stem cells in culture. *Development* 2004; 131: 1651–1662.

[147] D'Amour KA, Agulnick AD, Eliazer S, et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol*; 23. .2005. DOI: 10.1038/nbt1163.

[148] Gadue P, Huber TL, Paddison PJ, et al. Wnt and TGF-β signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc Natl Acad Sci U S A*; 103. .2006. DOI: 10.1073/pnas.0603916103.

[149] Loh KM, Ang LT, Zhang J, et al. Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. *Cell Stem Cell* 2014; 14: 237–252.

[150] Xu X, Browning VL, Odorico JS. Activin, BMP and FGF pathways cooperate to promote endoderm and pancreatic lineage cell differentiation from human embryonic stem cells. *Mech Dev* 2011; 128: 412–427.

[151] Vallier L, Touboul T, Chng Z, et al. Early Cell Fate Decisions of Human Embryonic Stem Cells and Mouse Epiblast Stem Cells Are Controlled by the Same Signalling Pathways. *PLoS One* 2009; 4: e6082.

[152] Mou H, Zhao R, Sherwood R, et al. Generation of Multipotent Lung and Airway Progenitors from Mouse ESCs and Patient-Specific Cystic Fibrosis iPSCs. *Cell Stem Cell* 2012; 10: 385–397.

[153] Mucenski ML, Nation JM, Thitoff AR, et al. β-Catenin regulates differentiation of respiratory epithelial cells in vivo. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2005; 289: L971–L979.

[154] Liu Y, Hogan BLM. Differential gene expression in the distal tip endoderm of the embryonic mouse lung. *Gene Expression Patterns* 2002; 2: 229–233.

[155] Rankin SA, Han L, McCracken KW, et al. A Retinoic Acid-Hedgehog Cascade Coordinates Mesoderm-Inducing Signals and Endoderm Competence during Lung Specification. *Cell Rep* 2016; 16: 66–78.

[156] Gonzales LW, Guttentag SH, Wade KC, et al. Differentiation of human pulmonary type II cells in vitro by glucocorticoid plus cAMP. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2002; 283: L940–L951.

[157] Gotoh S, Ito I, Nagasaki T, et al. Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem Cell Reports* 2014; 3: 394–403.

[158] Yamamoto Y, Gotoh S, Korogi Y, et al. Long-term expansion of alveolar stem cells derived from human iPS cells in organoids. *Nat Methods* 2017; 14: 1097–1106.

[159] Konishi S, Gotoh S, Tateishi K, et al. Directed Induction of Functional Multi-ciliated Cells in Proximal Airway Epithelial Spheroids from Human Pluripotent Stem Cells. *Stem Cell Reports* 2016; 6: 18–25.

[160] Miller AJ, Dye BR, Ferrer-Torres D, et al. Generation of lung organoids from human pluripotent stem cells in vitro. *Nat Protoc* 2019; 14: 518–540.

[161] Miller AJ, Hill DR, Nagy MS, et al. In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor Cells Derived from Human Pluripotent Stem Cells. *Stem Cell Reports* 2018; 10: 101–119.

[162] Jacob A, Vedaie M, Roberts DA, et al. Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. *Nat Protoc* 2019; 14: 3303–3332.

[163] Chen G, Gulbranson DR, Hou Z, et al. Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* 2011; 8: 424–429.

[164] Hay DC, Fletcher J, Payne C, et al. Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *Proc Natl Acad Sci U S A*; 105. .2008. DOI: 10.1073/pnas.0806522105.

[165] Hannan NRF, Fordham RP, Syed YA, et al. Generation of Multipotent Foregut Stem Cells from Human Pluripotent Stem Cells. *Stem Cell Reports* 2013; 1: 293–306.

[166] Uenishi G, Theisen D, Lee J-H, et al. Tenascin C Promotes Hematoendothelial Development and T Lymphoid Commitment from Human Pluripotent Stem Cells in Chemically Defined Conditions. *Stem Cell Reports* 2014; 3: 1073–1084.

[167] Travis WD, King TE, Bateman ED, et al. American thoracic society/European respiratory society international multidisciplinary consensus classification of the

idiopathic interstitial pneumonias. In: *American Journal of Respiratory and Critical Care Medicine*. 2002. .2002. DOI: 10.1164/ajrccm.165.2.ats01.

[168] van Moorsel CHM, van Oosterhout MFM, Barlo NP, et al. Surfactant Protein C Mutations Are the Basis of a Significant Portion of Adult Familial Pulmonary Fibrosis in a Dutch Cohort. *Am J Respir Crit Care Med* 2010; 182: 1419–1425.

[169] van Moorsel CHM, ten Klooster L, van Oosterhout MFM, et al. *SFTPA2* Mutations in Familial and Sporadic Idiopathic Interstitial Pneumonia. *Am J Respir Crit Care Med* 2015; 192: 1249–1252.

[170] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*; 4. .2009. DOI: 10.1038/nprot.2009.86.

[171] Lai TY, Cao J, Ou-Yang P, et al. Different methods of detaching adherent cells and their effects on the cell surface expression of Fas receptor and Fas ligand. *Sci Rep*; 12. .2022. DOI: 10.1038/s41598-022-09605-y.

[172] Tatler AL, Goodwin AT, Gbolahan O, et al. Amplification of TGFβ Induced ITGB6 Gene Transcription May Promote Pulmonary Fibrosis. .2016. DOI: 10.1371/journal.pone.0158047.

[173] Abe M, Harpel JG, Metz CN, et al. An Assay for Transforming Growth Factor-β Using Cells Transfected with a Plasminogen Activator Inhibitor-1 Promoter-Luciferase Construct. *Anal Biochem* 1994; 216: 276–284.

[174] Green IJ. Serial propagation of influenza B (Lee) virus in a transmissible line of canine kidney cells. *Science (1979)*; 138. .1962. DOI: 10.1126/science.138.3536.42.

[175] Cuevas Ocaña S, Serna Valverde AL, Reed L, et al. Next generation of gene-editing technologies for respiratory research and medicine. 2022. .2022. DOI: 10.1183/13993003.congress-2022.2815.

[176] Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* 2001; 25: 402–408.

[177] Tan YY, Ji ZL, Zhao G, et al. Decreased SCF/C-kit signaling pathway contributes to loss of interstitial cells of cajal in gallstone disease. *Int J Clin Exp Med*; 7.

[178] Van Riet S, Ninaber DK, Mikkers HMM, et al. In vitro modelling of alveolar repair at the air-liquid interface using alveolar epithelial cells derived from human induced pluripotent stem cells. DOI: 10.1038/s41598-020-62226-1.

[179] Calcagno AM, Chewning KJ, Wu CP, et al. Plasma membrane calcium ATPase (PMCA4): A housekeeper for RT-PCR relative quantification of polytopic membrane proteins. *BMC Mol Biol*; 7. .2006. DOI: 10.1186/1471-2199-7-29.

[180] Munk RB, Sugiyama K, Ghosh P, et al. Antigen-independent IFN-γ production by human naïve CD4+ T cells activated by IL-12 plus IL-18. *PLoS One*; 6. .2011. DOI: 10.1371/journal.pone.0018553.

[181] Wu WJH, Kim M, Chang LC, et al. Interleukin-1β secretion induced by mucosa-associated gut commensal bacteria promotes intestinal barrier repair. *Gut Microbes*; 14. .2022. DOI: 10.1080/19490976.2021.2014772.

[182] Dekkers JF, Alieva M, Wellens LM, et al. High-resolution 3D imaging of fixed and cleared organoids. *Nat Protoc* 2019; 14: 1756–1771.

[183] Priddle H, Allegrucci C, Burridge P, et al. Derivation and characterisation of the human embryonic stem cell lines, NOTT1 and NOTT2. *In Vitro Cell Dev Biol Anim*; 46. .2010. DOI: 10.1007/s11626-010-9290-6.

[184] International Standing Committee on Human Cytogenetic Nomenclature., Shaffer LG, McGowan-Jordan Jean, et al. ISCN 2013 : an international system for human cytogenetic nomenclature (2013). 2013; 140.

[185] Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*; 43. .2015. DOI: 10.1093/nar/gkv007.

[186] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: Tool for the unification of biology. *Nature Genetics*; 25. .2000. DOI: 10.1038/75556.

[187] Carbon S, Douglass E, Good BM, et al. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res*; 49. .2021. DOI: 10.1093/nar/gkaa1113.

[188] Yu G, Wang LG, Han Y, et al. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*; 16. .2012. DOI: 10.1089/omi.2011.0118.

[189] Schriml LM, Munro JB, Schor M, et al. The Human Disease Ontology 2022 update. *Nucleic Acids Res*; 50. .2022. DOI: 10.1093/nar/gkab1063.

[190] Beers MF, Mulugeta S. SURFACTANT PROTEIN C BIOSYNTHESIS AND ITS EMERGING ROLE IN CONFORMATIONAL LUNG DISEASE. *Annu Rev Physiol* 2005; 67: 663–696.

[191] Glasser SW, Korfhagen TR, Perme CM, et al. Two SP-C genes encoding human pulmonary surfactant proteolipid. *Journal of Biological Chemistry*; 263. .1988. DOI: 10.1016/s0021-9258(19)81519-3.

[192] Tredano M, Griese M, Brasch F, et al. Mutation of SFTPC in Infantile Pulmonary Alveolar Proteinosis with or without Fibrosing Lung Disease. *Am J Med Genet*; 126 A. .2004. DOI: 10.1002/ajmg.a.20670.

[193] Mulugeta S, Nureki SI, Beers MF. Lost after translation: Insights from pulmonary surfactant for understanding the role of alveolar epithelial dysfunction and cellular quality control in fibrotic lung disease. *Am J Physiol Lung Cell Mol Physiol*; 309. .2015. DOI: 10.1152/ajplung.00139.2015.

[194] Johansson J. Structure and properties of surfactant protein C. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1998; 1408: 161–172.

[195] Willander H, Askarieh G, Landreh M, et al. High-resolution structure of a BRICHOS domain and its implications for anti-amyloid chaperone activity on lung surfactant protein C. *Proceedings of the National Academy of Sciences* 2012; 109: 2325–2329.

[196] Johansson J. Structure and properties of surfactant protein C. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1998; 1408: 161–172.

[197] Maguire JA, Mulugeta S, Beers MF. Endoplasmic reticulum stress induced by surfactant protein C BRICHOS mutants promotes proinflammatory signaling by epithelial cells. *Am J Respir Cell Mol Biol*. .2011. DOI: 10.1165/rcmb.2009-0382OC.

[198] Brasch FE, Griese M, Tredano M, et al. Interstitial lung disease in a baby with a de novo mutation on the SFTPC gene. *European Respiratory Journal*; 24. .2004. DOI: 10.1183/09031936.04.00000104.

[199] Hawkins A, Guttentag SH, Deterding R, et al. A non-BRICHOS SFTPC mutant (SP-CI73T) linked to interstitial lung disease promotes a late block in macroautophagy disrupting cellular proteostasis and mitophagy. *Am J Physiol Lung Cell Mol Physiol*; 308. .2015. DOI: 10.1152/ajplung.00217.2014.

[200] Hetz C. The unfolded protein response: Controlling cell fate decisions under ER stress and beyond. *Nature Reviews Molecular Cell Biology*; 13. .2012. DOI: 10.1038/nrm3270.

[201] Oslowski CM, Urano F. Measuring ER stress and the unfolded protein response using mammalian tissue culture system. In: *Methods in Enzymology*. NIH Public Access, pp. 71–92.

[202] Schröder M, Kaufman RJ. The mammalian unfolded protein response. *Annu Rev Biochem* 2005; 74: 739–89.

[203] Beers MF, Hawkins A, Maguire JA, et al. A Nonaggregating Surfactant Protein C Mutant Is Misdirected to Early Endosomes and Disrupts Phospholipid Recycling. *Traffic* 2011; 12: 1196–1210.

[204] FUSAKI N, BAN H, NISHIYAMA A, et al. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy, Series B* 2009; 85: 348–362.

[205] Cameron HS, Somaschini M, Carrera P, et al. A common mutation in the surfactant protein C gene associated with lung disease. *Journal of Pediatrics*; 146. .2005. DOI: 10.1016/j.jpeds.2004.10.028.

[206] Miller AJ, Hill DR, Nagy MS, et al. In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor Cells Derived from Human Pluripotent Stem Cells. *Stem Cell Reports* 2018; 10: 101–119.

[207] Ye J, Coulouris G, Zaretskaya I, et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012; 13: 134.

[208] Taylor SC, Nadeau K, Abbasi M, et al. The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends Biotechnol* 2019; 37: 761–774.

[209] Murry CE, Keller G. Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. *Cell* 2008; 132: 661–680.

[210] Coraux C, Nawrocki-Raby B, Hinnrasky J, et al. Embryonic stem cells generate airway epithelial tissue. *Am J Respir Cell Mol Biol*; 32. .2005. DOI: 10.1165/rcmb.2004-0079RC.

[211] Huang SXL, Islam MN, O'Neill J, et al. Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat Biotechnol* 2014; 32: 84–91.

[212] Wong AP, Bear CE, Chin S, et al. Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nat Biotechnol* 2012; 30: 876–882.

[213] Crane AM, Kramer P, Bui JH, et al. Targeted correction and restored function of the CFTR gene in cystic fibrosis induced pluripotent stem cells. *Stem Cell Reports*; 4. .2015. DOI: 10.1016/j.stemcr.2015.02.005.

[214] Firth AL, Dargitz CT, Qualls SJ, et al. Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells. *Proc Natl Acad Sci U S A*; 111. .2014. DOI: 10.1073/pnas.1403470111.

[215] Hawkins F, Kotton DN. Embryonic and Induced Pluripotent Stem Cells for Lung Regeneration. *Ann Am Thorac Soc* 2015; 12: S50–S53.

[216] Hawkins F, Kramer P, Jacob A, et al. Prospective isolation of NKX2-1–expressing human lung progenitors derived from pluripotent stem cells. *Journal of Clinical Investigation* 2017; 127: 2277–2294.

[217] Dye BR, Hill DR, Ferguson MA, et al. In vitro generation of human pluripotent stem cell derived lung organoids. *Elife* 2015; 4: 1–25.

[218] McCauley KB, Hawkins F, Serra M, et al. Efficient Derivation of Functional Human Airway Epithelium from Pluripotent Stem Cells via Temporal Regulation of Wnt Signaling. *Cell Stem Cell* 2017; 20: 844-857.e6.

[219] Chen Y-W, Huang SX, de Carvalho ALRT, et al. A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nat Cell Biol* 2017; 19: 542–549.

[220] Sun YL, Hurley K, Villacorta-Martin C, et al. Heterogeneity in Human Induced Pluripotent Stem Cell-derived Alveolar Epithelial Type II Cells Revealed with ABCA3/SFTPC Reporters. *Am J Respir Cell Mol Biol* 2021; 65: 442–460.

[221] Strikoudis A, Cieślak A, Loffredo L, et al. Modeling of Fibrotic Lung Disease Using 3D Organoids Derived from Human Pluripotent Stem Cells. *Cell Rep* 2019; 27: 3709-3723.e5.

[222] Gonzalez RF, Allen L, Gonzales L, et al. HTII-280, a Biomarker Specific to the Apical Plasma Membrane of Human Lung Alveolar Type II Cells. *Journal of Histochemistry & Cytochemistry* 2010; 58: 891–901.

[223] McLean AB, D'Amour KA, Jones KL, et al. Activin A Efficiently Specifies Definitive Endoderm from Human Embryonic Stem Cells Only When Phosphatidylinositol 3-Kinase Signaling Is Suppressed. *Stem Cells* 2007; 25: 29–38.

[224] Vallier L, Mendjan S, Brown S, et al. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* 2009; 136: 1339–1349.

[225] Gouon-Evans V, Boussemart L, Gadue P, et al. BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nat Biotechnol*; 24. .2006. DOI: 10.1038/nbt1258.

[226] Siller R, Naumovska E, Mathapati S, et al. Development of a rapid screen for the endodermal differentiation potential of human pluripotent stem cell lines. *Sci Rep* 2016; 6: 37178.

[227] Yin BWT, Kiyamova R, Chua R, et al. Monoclonal antibody MX35 detects the membrane transporter NaPi2b (SLC34A2) in human carcinomas. *Cancer Immun* 2008; 8: 3.

[228] Gonzales LW, Guttentag SH, Wade KC, et al. Differentiation of human pulmonary type II cells in vitro by glucocorticoid plus cAMP. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2002; 283: L940–L951.

[229] Alysandratos K-D, Garcia-de-Alba C, Yao C, et al. Culture impact on the transcriptomic programs of primary and iPSC-derived human alveolar type 2 cells. *JCI Insight*; 8. .10 January 2023. DOI: 10.1172/jci.insight.158937.

[230] Pepicelli C V, Lewis PM, McMahon AP. Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Curr Biol* 1998; 8: 1083–6.

[231] Lebeche D, Malpel S, Cardoso W V. Fibroblast growth factor interactions in the developing lung. *Mech Dev* 1999; 86: 125–136.

[232] Laresgoiti U, Niko MZ, Rao C, et al. Lung epithelial tip progenitors integrate glucocorticoid- and STAT3-mediated signals to control progeny fate. *Development (Cambridge)*; 143. .2016. DOI: 10.1242/dev.134023.

[233] Hurley K, Ding J, Villacorta-Martin C, et al. Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. *Cell Stem Cell* 2020; 26: 593-608.e8.

[234] Sikkema L, Ramírez-Suástegui C, Strobl DC, et al. An integrated cell atlas of the lung in health and disease. *Nat Med* 2023; 29: 1563–1577.

[235] Alysandratos K-D, Garcia-de-Alba C, Yao C, et al. Culture impact on the transcriptomic programs of primary and iPSC-derived human alveolar type 2 cells. *JCI Insight* 2023; 8: 2022.02.08.479591.

[236] Alysandratos KD, Russo SJ, Petcherski A, et al. Patient-specific iPSCs carrying an SFTPC mutation reveal the intrinsic alveolar epithelial dysfunction at the inception of interstitial lung disease. *Cell Rep* 2021; 36: 109636.

[237] Choi J, Park JE, Tsagkogeorga G, et al. Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. *Cell Stem Cell* 2020; 27: 366-382.e7.

[238] Strunz M, Simon LM, Ansari M, et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun* 2020; 11: 3559.

[239] Kobayashi Y, Tata A, Konkimalla A, et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat Cell Biol* 2020; 22: 934–946.

[240] Jiang P, Gil de Rubio R, Hrycaj SM, et al. Ineffectual Type 2–to–Type 1 Alveolar Epithelial Cell Differentiation in Idiopathic Pulmonary Fibrosis: Persistence of the KRT8 ^{hi} Transitional State. *Am J Respir Crit Care Med* 2020; 201: 1443–1447.

[241] Crestani B, Cornillet P, Dehoux M, et al. Alveolar type II epithelial cells produce interleukin-6 in vitro and in vivo: Regulation by alveolar macrophage secretory products. *Journal of Clinical Investigation*; 94. .1994. DOI: 10.1172/JCI117392.

[242] Ng B, Dong J, D'Agostino G, et al. Interleukin-11 is a therapeutic target in idiopathic pulmonary fibrosis. *Sci Transl Med*; 11. .2019. DOI: 10.1126/scitranslmed.aaw1237.

[243] Zhang J, Wang D, Wang L, et al. Profibrotic effect of IL-17A and elevated IL-17RA in idiopathic pulmonary fibrosis and rheumatoid arthritis-associated lung disease support a direct role for IL-17A/IL-17RA in human fibrotic interstitial lung disease. *Am J Physiol Lung Cell Mol Physiol*; 316. .2019. DOI: 10.1152/ajplung.00301.2018.

[244] Adams TS, Schupp JC, Poli S, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv*; 6. .1 July 2020. DOI: 10.1126/SCIADV.ABA1983.

[245] Habermann AC, Gutierrez AJ, Bui LT, et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv*; 6. .2020. DOI: 10.1126/sciadv.aba1972.

[246] Parimon T, Yao C, Stripp BR, et al. Alveolar epithelial type II cells as drivers of lung fibrosis in idiopathic pulmonary fibrosis. *Int J Mol Sci*; 21. .2020. DOI: 10.3390/ijms21072269.

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

[248] O'Dwyer DN, Ashley SL, Gurczynski SJ, et al. Lung Microbiota Contribute to Pulmonary Inflammation and Disease Progression in Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019; 199: 1127–1138.

[249] Han MLK, Zhou Y, Murray S, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med* 2014; 2: 548–556.

[250] Moghoofei M, Mostafaei S, Kondori N, et al. Bacterial and viral coinfection in idiopathic pulmonary fibrosis patients: the prevalence and possible role in disease progression. *BMC Pulm Med* 2022; 22: 60.

[251] Juarez MM, Chan AL, Norris AG, et al. Acute exacerbation of idiopathic pulmonary fibrosis-a review of current and novel pharmacotherapies. *J Thorac Dis* 2015; 7: 499–519.

[252] Ryerson CJ, Cottin V, Brown KK, et al. Acute exacerbation of idiopathic pulmonary fibrosis: shifting the paradigm. *European Respiratory Journal* 2015; 46: 512–520.

[253] Song JW, Hong SB, Lim CM, et al. Acute exacerbation of idiopathic pulmonary fibrosis: Incidence, risk factors and outcome. *European Respiratory Journal*; 37. .2011. DOI: 10.1183/09031936.00159709.

[254] Dickson RP, Erb-Downward JR, Martinez FJ, et al. The Microbiome and the Respiratory Tract. *Annu Rev Physiol* 2016; 78: 481–504.

[255] Weng D, Chen XQ, Qiu H, et al. The role of infection in acute exacerbation of idiopathic pulmonary fibrosis. *Mediators Inflamm*; 2019. .2019. DOI: 10.1155/2019/5160694.

[256] McMillan TR, Moore BB, Weinberg JB, et al. Exacerbation of established pulmonary fibrosis in a murine model by gammaherpesvirus. *Am J Respir Crit Care Med* 2008; 177: 771–80.

[257] Moore BB, Moore TA. Viruses in Idiopathic Pulmonary Fibrosis. Etiology and Exacerbation. *Ann Am Thorac Soc* 2015; 12 Suppl 2: S186-92.

[258] Thyagarajan B, Bloom JD. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife*; 3. .2014. DOI: 10.7554/elife.03300.

[259] Shieh W-J, Blau DM, Denison AM, et al. 2009 Pandemic Influenza A (H1N1). *Am J Pathol* 2010; 177: 166–175.

[260] Nakajima N, Sato Y, Katano H, et al. Histopathological and immunohistochemical findings of 20 autopsy cases with 2009 H1N1 virus infection. *Modern Pathology* 2012; 25: 1–13.

[261] Nayak DP, Hui EKW, Barman S. Assembly and budding of influenza virus. *Virus Res*; 106. .2004. DOI: 10.1016/j.virusres.2004.08.012.

[262] Xue J, Chambers BS, Hensley SE, et al. Propagation and characterization of influenza virus stocks that lack high levels of defective viral genomes and hemagglutinin mutations. *Front Microbiol* 2016; 7: 326.

[263] Hirst GK. The agglutination of red cells by allantoic fluid of chick embryos infected with influenza virus. *Science (1979)*; 94. .1941. DOI: 10.1126/science.94.2427.22.

[264] Killian ML. Hemagglutination assay for the avian influenza virus. *Methods Mol Biol* 2008; 436: 47–52.

[265] Killian ML. Hemagglutination assay for influenza virus. *Methods Mol Biol* 2014; 1161: 3–9.

[266] Hierholzer JC, Killington RA. Virus isolation and quantitation. *Virology Methods Manual* 1996; 25–46.

[267] Payne S. Methods to Study Viruses. In: *Viruses*. 2017. .2017. DOI: 10.1016/b978-0-12-803109-4.00004-0.

[268] Cooper PD. The Plaque Assay of Animal Viruses. *Adv Virus Res*; 8. .1962. DOI: 10.1016/S0065-3527(08)60689-2.

[269] Aslantürk ÖS. In Vitro Cytotoxicity and Cell Viability Assays: Principles, Advantages, and Disadvantages. In: *Genotoxicity - A Predictable Risk to Our Actual World*. 2018. .2018. DOI: 10.5772/intechopen.71923.

[270] Thorne N, Inglese J, Auld DS. Illuminating Insights into Firefly Luciferase and Other Bioluminescent Reporters Used in Chemical Biology. *Chemistry and Biology*; 17. .2010. DOI: 10.1016/j.chembiol.2010.05.012.

[271] Smee DF, Hurst BL, Evans WJ, et al. Evaluation of cell viability dyes in antiviral assays with RNA viruses that exhibit different cytopathogenic properties. *J Virol Methods*; 246. .2017. DOI: 10.1016/j.jviromet.2017.03.012.

[272] Munger JS, Huang X, Kawakatsu H, et al. The integrin αvβ6 binds and activates latent TGFβ1: A mechanism for regulating pulmonary inflammation and fibrosis. *Cell*; 96. .1999. DOI: 10.1016/S0092-8674(00)80545-0.

[273] Annes JP, Chen Y, Munger JS, et al. Integrin αvβ6-mediated activation of latent TGF-β requires the latent TGF-β binding protein-1. *Journal of Cell Biology*; 165. .2004. DOI: 10.1083/jcb.200312172.

[274] Meylan E, Tschopp J. Toll-Like Receptors and RNA Helicases: Two Parallel Ways to Trigger Antiviral Responses. *Molecular Cell*; 22. .2006. DOI: 10.1016/j.molcel.2006.05.012.

[275] Alexopoulou L, Holt AC, Medzhitov R, et al. Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3. *Nature*; 413. .2001. DOI: 10.1038/35099560.

[276] Katzen J, Wagner BD, Venosa A, et al. A SFTPC BRICHOS mutant links epithelial ER stress and spontaneous lung fibrosis. *JCI Insight*; 4. .5 February 2019. DOI: 10.1172/jci.insight.126125.

[277] Nureki SI, Tomer Y, Venosa A, et al. Expression of mutant Sftpc in murine alveolar epithelia drives spontaneous lung fibrosis. *Journal of Clinical Investigation* 2018; 128: 4008–4024.

[278] Naikawadi RP, Disayabutr S, Mallavia B, et al. Telomere dysfunction in alveolar epithelial cells causes lung remodeling and fibrosis. *JCI Insight*; 1. .2016. DOI: 10.1172/jci.insight.86704.

[279] Vannella KM, Luckhardt TR, Wilke CA, et al. Latent Herpesvirus Infection Augments Experimental Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2010; 181: 465–477.

[280] Shatskaya E V, Kovner A V, Potapova O V, et al. Study of SMAD-Dependent Signal Pathway in the Development of Early Pulmonary Fibrosis in Mice Infected with Influenza A/H1N1 Virus. *Bull Exp Biol Med* 2017; 162: 647–649.

[281] Vaughan AE, Brumwell AN, Xi Y, et al. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature*; 517. .2015. DOI: 10.1038/nature14112.

[282] Kumar PA, Hu Y, Yamamoto Y, et al. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. *Cell*; 147. .2011. DOI: 10.1016/j.cell.2011.10.001.

[283] Abo KM, de Aja JS, Lindstrom-Vautrin J, et al. Air-liquid interface culture promotes maturation and allows environmental exposure of pluripotent stem cell-derived alveolar epithelium. *JCI Insight*; 7. .2022. DOI: 10.1172/jci.insight.155589.

[284] Hui KPY, Ching RHH, Chan SKH, et al. Tropism, replication competence, and innate immune responses of influenza virus: an analysis of human airway organoids and ex-vivo bronchus cultures. *Lancet Respir Med* 2018; 6: 846–854.

[285] Zhou J, Li C, Sachs N, et al. Differentiated human airway organoids to assess infectivity of emerging influenza virus. *Proceedings of the National Academy of Sciences* 2018; 115: 6822–6827.

[286] Sachs N, Papaspyropoulos A, Zomer-van Ommen DD, et al. Long-term expanding human airway organoids for disease modeling. *EMBO J*; 38. .15 February 2019. DOI: 10.15252/embj.2018100300.

[287] Katsura H, Kobayashi Y, Tata PR, et al. IL-1 and TNFα Contribute to the Inflammatory Niche to Enhance Alveolar Regeneration. *Stem Cell Reports* 2019; 12: 657–666.

[288] Porotto M, Ferren M, Chen YW, et al. Authentic modeling of human respiratory virus infection in human pluripotent stem cell-derived lung organoids. *mBio*; 10. .1 May 2019. DOI: 10.1128/mBio.00723-19.

[289] Ciancanelli MJ, Huang SXL, Luthra P, et al. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. *Science (1979)*; 348. .2015. DOI: 10.1126/science.aaa1578.

[290] Lim HK, Huang SXL, Chen J, et al. Severe influenza pneumonitis in children with inherited TLR3 deficiency. *Journal of Experimental Medicine*; 216. .2019. DOI: 10.1084/jem.20181621.

[291] Lee CM, Cho SJ, Cho WK, et al. Laminin α1 is a genetic modifier of TGF-β1-stimulated pulmonary fibrosis. *JCI Insight*; 3. .2018. DOI: 10.1172/jci.insight.99574.

[292] Bertram S, Glowacka I, Steffen I, et al. Novel insights into proteolytic cleavage of influenza virus hemagglutinin. *Reviews in Medical Virology*; 20. .2010. DOI: 10.1002/rmv.657.

[293] Severson WE, McDowell M, Ananthan S, et al. High-throughput screening of a 100,000-compound library for inhibitors of influenza a virus (H3N2). *J Biomol Screen*; 13. .2008. DOI: 10.1177/1087057108323123.

[294] Burgstaller G, Oehrle B, Gerckens M, et al. The instructive extracellular matrix of the lung: Basic composition and alterations in chronic lung disease. *European Respiratory Journal*; 50. .2017. DOI: 10.1183/13993003.01805-2016.

[295] Zhou Z, Xue Q, Wan Y, et al. Lysosome-associated membrane glycoprotein 3 is involved in influenza A virus replication in human lung epithelial (A549) cells. *Virol J* 2011; 8: 384.

[296] Mar KB, Rinkenberger NR, Boys IN, et al. LY6E mediates an evolutionarily conserved enhancement of virus infection by targeting a late entry step. *Nat Commun*; 9. .2018. DOI: 10.1038/s41467-018-06000-y.

[297] Gralinski LE, Bankhead A, Jeng S, et al. Mechanisms of severe acute respiratory syndrome coronavirus-induced acute lung injury. *mBio*; 4. .2013. DOI: 10.1128/mBio.00271-13.

[298] Glisinski KM, Schlobohm AJ, Paramore S V., et al. Interleukin-13 disrupts type 2 pneumocyte stem cell activity. *JCI Insight*; 5. .24 February 2020. DOI: 10.1172/jci.insight.131232.

[299] Lawson WE, Crossno PF, Polosukhin V V., et al. Endoplasmic reticulum stress in alveolar epithelial cells is prominent in IPF: association with altered surfactant protein processing and herpesvirus infection. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2008; 294: L1119–L1126.

[300] Xu Y, Mizuno T, Sridharan A, et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016; 1: 90558.

[301] Guo M, Du Y, Gokey JJ, et al. Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat Commun* 2019; 10: 37.

[302] Yu W, Clyne M, Khoury MJ, et al. Phenopedia and genopedia: Disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics*; 26. .2009. DOI: 10.1093/bioinformatics/btp618.

[303] Amberger JS, Bocchini CA, Schiettecatte F, et al. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res*; 43. .2015. DOI: 10.1093/nar/gku1205.

[304] Amberger JS, Bocchini CA, Scott AF, et al. OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res*; 47. .2019. DOI: 10.1093/nar/gky1151.

[305] Thomson JA, Itskovitz-Eldor J, Shapiro SS, et al. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science (1979)* 1998; 282: 1145–1147.

[306] Singh VK, Kalsan M, Kumar N, et al. Induced pluripotent stem cells: Applications in regenerative medicine, disease modeling, and drug discovery. *Frontiers in Cell and Developmental Biology*; 3. .2015. DOI: 10.3389/fcell.2015.00002.

[307] Vangipuram M, Ting D, Kim S, et al. Skin punch biopsy explant culture for derivation of primary human fibroblasts. *J Vis Exp*. .2013. DOI: 10.3791/3779.

[308] Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007; 448: 313–317.

[309] Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science (1979)*; 318. .2007. DOI: 10.1126/science.1151526.

[310] Okita K, Yamanaka S. Induced pluripotent stem cells: opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2011; 366: 2198–2207.

[311] Soldner F, Laganière J, Cheng AW, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset parkinson point mutations. *Cell*; 146. .2011. DOI: 10.1016/j.cell.2011.06.019.

[312] Silva G, Poirot L, Galetto R, et al. Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *Curr Gene Ther*; 11. .2011. DOI: 10.2174/156652311794520111.

[313] Hockemeyer D, Wang H, Kiani S, et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol*; 29. .2011. DOI: 10.1038/nbt.1927.

[314] Carlson DF, Fahrenkrug SC, Hackett PB. Targeting DNA with fingers and TALENs. *Molecular Therapy - Nucleic Acids*; 1. .2012. DOI: 10.1038/mtna.2011.5.

[315] Urbach A, Bar-Nur O, Daley GQ, et al. Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. *Cell Stem Cell*; 6. .2010. DOI: 10.1016/j.stem.2010.04.005.

[316] Inoue H, Nagata N, Kurokawa H, et al. IPS cells: A game changer for future medicine. *EMBO Journal*; 33. .2014. DOI: 10.1002/embj.201387098.

[317] Rouhani F, Kumasaka N, de Brito MC, et al. Genetic Background Drives Transcriptional Variation in Human Induced Pluripotent Stem Cells. *PLoS Genet*; 10. .2014. DOI: 10.1371/journal.pgen.1004432.

[318] Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*; 31. .2013. DOI: 10.1016/j.tibtech.2013.04.004.

[319] Khan SH. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Molecular Therapy - Nucleic Acids*; 16. .2019. DOI: 10.1016/j.omtn.2019.02.027.

[320] Doyle EL, Booher NJ, Standage DS, et al. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: Tools for TAL effector design and target prediction. *Nucleic Acids Res*; 40. .2012. DOI: 10.1093/nar/gks608.

[321] Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*; 31. .2013. DOI: 10.1038/nbt.2623.

[322] Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*; 31. .2013. DOI: 10.1038/nbt.2647.

[323] Anzalone A V., Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*; 38. .2020. DOI: 10.1038/s41587-020-0561-9.

[324] Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics*; 19. .2018. DOI: 10.1038/s41576-018-0059-1.

[325] Tran E, Shi T, Li X, et al. Development of human alveolar epithelial cell models to study distal lung biology and disease. *iScience*; 25. .2022. DOI: 10.1016/j.isci.2022.103780.

[326] Nikolić MZ, Caritg O, Jeng Q, et al. Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. *Elife*; 6. .30 June 2017. DOI: 10.7554/eLife.26575.

[327] Bradbury A, Plückthun A. Reproducibility: Standardize antibodies used in research. *Nature*; 518. .2015. DOI: 10.1038/518027a.

[328] Tata PR, Chow RD, Saladi SV, et al. Developmental History Provides a Roadmap for the Emergence of Tumor Plasticity. *Dev Cell* 2018; 44: 679-693.e5.

[329] Fukui E, Funaki S, Kimura K, et al. Adipose Tissue-Derived Stem Cells Have the Ability to Differentiate into Alveolar Epithelial Cells and Ameliorate Lung Injury Caused by Elastase-Induced Emphysema in Mice. *Stem Cells Int* 2019; 2019: 1–14.

[330] Ma R, Latif R, Davies TF. Human Embryonic Stem Cells Form Functional Thyroid Follicles. *Thyroid* 2015; 25: 455–461.

[331] Radonjic N V., Ortega JA, Memi F, et al. The complexity of the calretinin-expressing progenitors in the human cerebral cortex. *Front Neuroanat*; 8. .13 August 2014. DOI: 10.3389/fnana.2014.00082.

[332] Varma S, Cao Y, Tagne J-B, et al. The Transcription Factors Grainyhead-like 2 and NK2-Homeobox 1 Form a Regulatory Loop That Coordinates Lung Epithelial Cell Morphogenesis and Differentiation. *Journal of Biological Chemistry* 2012; 287: 37282–37295.

[333] Jakovcevski I, Mayer N, Zecevic N. Multiple origins of human neocortical interneurons are supported by distinct expression of transcription factors. *Cerebral Cortex*; 21. .2011. DOI: 10.1093/cercor/bhq245.

[334] Fujioka N, Kitabatake M, Ouji-Sageshima N, et al. Human Adipose-Derived Mesenchymal Stem Cells Ameliorate Elastase-Induced Emphysema in Mice by Mesenchymal–Epithelial Transition. *Int J Chron Obstruct Pulmon Dis* 2021; Volume 16: 2783–2793.

[335] Peng X, Li X, Li C, et al. NMDA receptor activation inhibits the protective effect of BM-MSCs on bleomycin-induced lung epithelial cell damage by inhibiting ERK signaling and the paracrine factor HGF. *Int J Mol Med*; 44. .13 May 2019. DOI: 10.3892/ijmm.2019.4195.

[336] Liu X, Liu Y, Yang Z, et al. Cell type specific expression of Follistatin-like 1 (Fstl1) in mouse embryonic lung development. *J Mol Histol* 2018; 49: 399–409.

[337] Wang G, Hu Y-X, He M-Y, et al. Gut-Lung Dysbiosis Accompanied by Diabetes Mellitus Leads to Pulmonary Fibrotic Change through the NF-κB Signaling Pathway. *Am J Pathol* 2021; 191: 838–856.

[338] Chen R, Zhang K, Chen H, et al. Telomerase Deficiency Causes Alveolar Stem Cell Senescence-associated Low-grade Inflammation in Lungs. *Journal of Biological Chemistry* 2015; 290: 30813–30829.

[339] Wolf L, Herr C, Niederstraßer J, et al. Receptor for advanced glycation endproducts (RAGE) maintains pulmonary structure and regulates the response to cigarette smoke. *PLoS One* 2017; 12: e0180092.

[340] Wang Z, Wu J, Jiang J, et al. KIF2A decreases IL-33 production and attenuates allergic asthmatic inflammation. *Allergy, Asthma & Clinical Immunology* 2022; 18: 55.

[341] Bordeaux J, Welsh AW, Agarwal S, et al. Antibody validation. *Biotechniques* 2010; 48: 197–209.

[342] Bridges JP, Xu Y, Na CL, et al. Adaptation and increased susceptibility to infection associated with constitutive expression of misfolded SP-C. *Journal of Cell Biology* 2006; 172: 395–407.

[343] Watanabe K, Ueno M, Kamiya D, et al. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat Biotechnol* 2007; 25: 681–686.

[344] Gauthaman K, Fong CY, Subramanian A, et al. ROCK Inhibitor Y-27632 Increases Thaw-Survival Rates and Preserves Stemness and Differentiation Potential of Human Wharton's Jelly Stem Cells After Cryopreservation. *Stem Cell Rev Rep*; 6. .2010. DOI: 10.1007/s12015-010-9184-8.

[345] Yasunaga M, Tada S, Torikai-Nishikawa S, et al. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nat Biotechnol*; 23. .2005. DOI: 10.1038/nbt1167.

[346] Herriges JC, Yi L, Hines EA, et al. Genome-scale study of transcription factor expression in the branching mouse lung. *Developmental Dynamics* 2012; 241: 1432–1453.

[347] Kido T, Koui Y, Suzuki K, et al. CPM is a useful cell surface marker to isolate expandable bi-potential liver progenitor cells derived from human iPS cells. *Stem Cell Reports*; 5. .2015. DOI: 10.1016/j.stemcr.2015.08.008.

[348] McCauley KB, Alysandratos K-D, Jacob A, et al. Single-Cell Transcriptomic Profiling of Pluripotent Stem Cell-Derived SCGB3A2+ Airway Epithelium. *Stem Cell Reports* 2018; 10: 1579–1595.

[349] Herold S, von Wulffen W, Steinmueller M, et al. Alveolar Epithelial Cells Direct Monocyte Transepithelial Migration upon Influenza Virus Infection: Impact of Chemokines and Adhesion Molecules. *The Journal of Immunology*; 177. .2006. DOI: 10.4049/jimmunol.177.3.1817.

[350] Burgess CL, Huang J, Bawa P, et al. Generation of human alveolar epithelial type I cells from pluripotent stem cells. *bioRxiv* 2023; 2023.01.19.524655.

[351] Rodrigues Toste de Carvalho AL, Strikoudis A, Liu H-Y, et al. Glycogen synthase kinase 3 induces multilineage maturation of human pluripotent stem cell-derived lung progenitors in 3D culture. *Development* 2018; 146: dev171652.

[352] Rankin SA, Kormish J, Kofron M, et al. A gene regulatory network controlling hhex transcription in the anterior endoderm of the organizer. *Dev Biol* 2011; 351: 297–310.

[353] Millien G, Beane J, Lenburg M, et al. Characterization of the mid-foregut transcriptome identifies genes regulated during lung bud induction. *Gene Expression Patterns*; 8. .2008. DOI: 10.1016/j.modgep.2007.09.003.

[354] Kumar M, Jordan N, Melton D, et al. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Dev Biol*; 259. .2003. DOI: 10.1016/S0012-1606(03)00183-0.

[355] Boström H, Willetts K, Pekny M, et al. PDGF-A signaling is a critical event in lung alveolar myofibroblast development and alveogenesis. *Cell*; 85. .1996. DOI: 10.1016/S0092-8674(00)81270-2.

[356] Adams TS, Marlier A, Kaminski N. Lung Cell Atlases in Health and Disease. *Annu Rev Physiol* 2023; 85: 47–69.

[357] Frank DB, Peng T, Zepp JA, et al. Emergence of a Wave of Wnt Signaling that Regulates Lung Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation. *Cell Rep* 2016; 17: 2312–2325.

[358] Nabhan AN, Brownfield DG, Harbury PB, et al. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science (1979)* 2018; 359: 1118–1123.

[359] Bonniaud P, Margetts PJ, Kolb M, et al. Adenoviral gene transfer of connective tissue growth factor in the lung induces transient fibrosis. *Am J Respir Crit Care Med*; 168. .2003. DOI: 10.1164/rccm.200210-1254OC.

[360] Glass WG, Argentieri RL, Bracht M, et al. Generation of bleomycin-induced lung fibrosis is independent of IL-16. *Cytokine* 2009; 46: 17–23.

[361] Liu Y, Wang Y, Lu F, et al. BTB and CNC homology 1 inhibition ameliorates fibrosis and inflammation via blocking ERK pathway in pulmonary fibrosis. *Exp Lung Res* 2021; 47: 67–77.

[362] Hubert CG, Rivera M, Spangler LC, et al. A three-dimensional organoid culture system derived from human glioblastomas recapitulates the hypoxic gradients and cancer stem cell heterogeneity of tumors found in vivo. *Cancer Res*; 76. .2016. DOI: 10.1158/0008-5472.CAN-15-2402.

[363] Papas KK, Avgoustiniatos ES, Tempelman LA, et al. High-Density Culture of Human Islets on Top of Silicone Rubber Membranes. *Transplant Proc* 2005; 37: 3412–3414.

[364] Lin Y-T, Lin C-F, Yeh T-H. Influenza A virus infection induces indoleamine 2,3-dioxygenase (IDO) expression and modulates subsequent inflammatory mediators in nasal epithelial cells. *Acta Otolaryngol* 2020; 140: 149–156.

[365] Yi E, Oh J, Kang HR, et al. BST2 inhibits infection of influenza A virus by promoting apoptosis of infected cells. *Biochem Biophys Res Commun*; 509. .2019. DOI: 10.1016/j.bbrc.2018.12.110.

[366] Sidahmed AME, León AJ, Bosinger SE, et al. CXCL10 contributes to p38-mediated apoptosis in primary T lymphocytes in vitro. *Cytokine*; 59. .2012. DOI: 10.1016/j.cyto.2012.05.002.

[367] Zhou Z, Xue Q, Wan Y, et al. Lysosome-associated membrane glycoprotein 3 is involved in influenza A virus replication in human lung epithelial (A549) cells. *Virol J*; 8. .2011. DOI: 10.1186/1743-422X-8-384.

[368] Al-Qahtani AA, Murugaiah V, Bashir HA, et al. Full-length human surfactant protein A inhibits influenza A virus infection of A549 lung epithelial cells: A recombinant form containing neck and lectin domains promotes infectivity. *Immunobiology* 2019; 224: 408–418.

[369] Yoshikawa T, Otsuka M, Chiba H, et al. Surfactant protein A as a biomarker of outcomes of anti-fibrotic drug therapy in patients with idiopathic pulmonary fibrosis. *BMC Pulm Med*; 20. .2020. DOI: 10.1186/s12890-020-1060-y.

[370] Rodriguez LR, Emblom-Callahan M, Chhina M, et al. Global Gene Expression Analysis in an in vitro Fibroblast Model of Idiopathic Pulmonary Fibrosis Reveals Potential Role for CXCL14/CXCR4. *Sci Rep*; 8. .2018. DOI: 10.1038/s41598-018-21889-7.

[371] Jia G, Chandriani S, Abbas AR, et al. CXCL14 is a candidate biomarker for Hedgehog signalling in idiopathic pulmonary fibrosis. *Thorax* 2017; 72: 780–787.

[372] Wan H, Huang X, Cong P, et al. Identification of Hub Genes and Pathways Associated With Idiopathic Pulmonary Fibrosis via Bioinformatics Analysis. *Front Mol Biosci*; 8. .2021. DOI: 10.3389/fmolb.2021.711239.

[373] Zou J, Li Y, Yu J, et al. Idiopathic pulmonary fibrosis is associated with tight junction protein alterations. *Biochim Biophys Acta Biomembr*; 1862. .2020. DOI: 10.1016/j.bbamem.2020.183205.

[374] Carnesecchi S, Deffert C, Donati Y, et al. A key role for NOX4 in epithelial cell death during development of lung fibrosis. *Antioxid Redox Signal* 2011; 15: 607–19.

[375] Dominguez-Bautista JA, Klinkenberg M, Brehm N, et al. Loss of lysosome-associated membrane protein 3 (LAMP3) enhances cellular vulnerability against proteasomal inhibition. *Eur J Cell Biol* 2015; 94: 148–61.

[376] Korfei M, Ruppert C, Mahavadi P, et al. Epithelial Endoplasmic Reticulum Stress and Apoptosis in Sporadic Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2008; 178: 838–846.

[377] Schruf E, Schroeder V, Le HQ, et al. Recapitulating idiopathic pulmonary fibrosis related alveolar epithelial dysfunction in a human iPSC-derived air-liquid interface model. *The FASEB Journal* 2020; 34: 7825–7846.

[378] Chapman HA, Li X, Alexander JP, et al. Integrin α6β4 identifies an adult distal lung epithelial population with regenerative potential in mice. *Journal of Clinical Investigation*; 121. .2011. DOI: 10.1172/JCI57673.

[379] Xi Y, Kim T, Brumwell AN, et al. Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. *Nat Cell Biol* 2017; 19: 904–914.

[380] Rosas IO, Richards TJ, Konishi K, et al. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med*; 5. .2008. DOI: 10.1371/journal.pmed.0050093.

[381] Verleden SE, Tanabe N, McDonough JE, et al. Small airways pathology in idiopathic pulmonary fibrosis: a retrospective cohort study. *Lancet Respir Med*; 8. .2020. DOI: 10.1016/S2213-2600(19)30356-X.

[382] Huang J, Hume AJ, Abo KM, et al. SARS-CoV-2 Infection of Pluripotent Stem Cell-Derived Human Lung Alveolar Type 2 Cells Elicits a Rapid Epithelial-Intrinsic Inflammatory Response. *Cell Stem Cell* 2020; 27: 962-973.e7.

[383] Kebaabetswe LP, Haick AK, Gritsenko MA, et al. Proteomic analysis reveals down-regulation of surfactant protein B in murine type II pneumocytes infected with influenza A virus. *Virology*; 483. .2015. DOI: 10.1016/j.virol.2015.03.045.

[384] Martins-Taylor K, Xu RH. Concise review: Genomic stability of human induced pluripotent stem cells. *Stem Cells*; 30. .2012. DOI: 10.1002/stem.705.

[385] Lancaster MA, Knoblich JA. Organogenesisin a dish: Modeling development and disease using organoid technologies. *Science*; 345. .2014. DOI: 10.1126/science.1247125.

[386] Fatehullah A, Tan SH, Barker N. Organoids as an in vitro model of human development and disease. *Nature Cell Biology*; 18. .2016. DOI: 10.1038/ncb3312.

[387] Vukicevic S, Kleinman HK, Luyten FP, et al. Identification of multiple active growth factors in basement membrane matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components. *Exp Cell Res*; 202. .1992. DOI: 10.1016/0014-4827(92)90397-Q.

[388] Aisenbrey EA, Murphy WL. Synthetic alternatives to Matrigel. *Nature Reviews Materials*; 5. .2020. DOI: 10.1038/s41578-020-0199-8.

[389] Chen HJ, Poran A, Unni AM, et al. Generation of pulmonary neuroendocrine cells and SCLC-like tumors from human embryonic stem cells. *J Exp Med* 2019; 216: 674–687.

[390] McCauley KB, Hawkins F, Kotton DN. Derivation of Epithelial-Only Airway Organoids from Human Pluripotent Stem Cells. *Curr Protoc Stem Cell Biol*; 45. .1 May 2018. DOI: 10.1002/cpsc.51.

[391] Nikolić MZ, Sun D, Rawlins EL. Human lung development: recent progress and new challenges. *Development*; 145. .15 August 2018. DOI: 10.1242/dev.163485.

[392] He P, Lim K, Sun D, et al. A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell* 2022; 185: 4841-4860.e25.

[393] Lim K, Donovan APA, Tang W, et al. Organoid modeling of human fetal lung alveolar development reveals mechanisms of cell fate patterning and neonatal respiratory disease. *Cell Stem Cell* 2023; 30: 20-37.e9.

[394] Miranda CC, Fernandes TG, Diogo MM, et al. Towards multi-organoid systems for drug screening applications. *Bioengineering*; 5. .2018. DOI: 10.3390/bioengineering5030049.

[395] Ronaldson-Bouchard K, Vunjak-Novakovic G. Organs-on-a-Chip: A Fast Track for Engineered Human Tissues in Drug Development. *Cell Stem Cell*; 22. .2018. DOI: 10.1016/j.stem.2018.02.011.

[396] Li W, Lin C, Shang C, et al. Pbx1 activates Fgf10 in the mesenchyme of developing lungs. *genesis* 2014; 52: 399–407.