Towards defining an antiviral role for BPIFB1 in the airway epithelium

Abdulaziz Alkhoshaiban

A thesis submitted towards the degree of
Doctor of Philosophy (PhD)

August 2022

Department of Infection, Immunity and Cardiovascular Disease
University of Sheffield

Declaration

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

Uniazah, August 2022

Dedication

Words cannot express my esteemed admiration and respect for the heroes in my life, my mother, Roqiyah and my brothers; Ali, Rehab, Mohammad, Adeeb, Hamad and Shada, who have always supported me throughout my whole educational journey. Mum and brothers, I keep pushing because of you, and you are the reason I can be brave through my struggles, hardships, and pain. I love all of you so much!

Special appreciation and thanks must be given to Khawla, my wife. Her love, dedication and constant confidence in me have alleviated all my burdens. Your presence and consistent encouragement have been a source of motivation during our life abroad and the demanding process of completing my research. While you have struggled too, during your illness and operation, it has been a wonderful comfort and relief knowing that you were willing to support me in any way you could.

This is also for my beloved sons; Saleh and Elyas, and my lovely son Yosef who was born during the difficult time of my PhD (and COVID-19 period). You are, and always will be my inspiration, as you always give me the strength during any difficult time.

I further would like to dedicate this work also to my relatives, family friends and PhD study friends.

Abstract

By analogy to other members of the wider BPI-fold contain protein family, it has been assumed that BPI fold containing family B member 1 (BPIFB1), a secretory protein produced by the respiratory tract, functions in host defence but functional data supporting this suggestion remains limited. *In vitro* studies have been hampered by the fact that the gene is only expressed in differentiated primary airway cells grown at the air liquid interface (ALI). Our previous studies have shown that BPIFB1 is found in a subpopulation of "goblet cells" in the airway, and IL-13 treatment induces a "mucous" phenotype in the cells by inducing BPIFB1. A systematic genomic and structural analysis of BPIFB1 was performed and temporal and spatial analysis was undertaken using published whole genome and scRNA-Seq data. Further, my results demonstrated that gene expression analysis of primary mouse tracheal epithelial cells (mTECs) showed that BPIFB1 increases during 14 days of ALI differentiation, and western blotting confirmed production of BPIFB1 as an N-glycosylated protein. Microarray analysis of mTEC ALI cultures showed gene signatures of mucociliary differentiation alongside specific membrane associated proteins such as cadherin-related family members. A series of FLAG-tagged expression clones including wild type, an N-glycosylation mutant and truncated proteins corresponding to the N- and C-terminal BPI domains, were shown to produce recombinant murine BPIFB1 proteins. Influenza A virus (IAV) is a significant respiratory pathogen and lab based preliminary data demonstrated that *Bpifb1* is reduced in lungs following IAV infection. Therefore, we hypothesised that BFIFB1 contributes to the defence against IAV infections, which can be studied *in vitro*. IAV infection assays were developed in mTECs and gene expression studied by array analysis, with the aim to identify regions of BPIFB1 associated with viral infection using generated recombinant BPIFB proteins. In parallel to the mTEC studies, the HBEC3-KT cell system, a human immortalized cell line that expresses *BPIFB1* as it undergoes mucociliary differentiation, was used to generate a CRISPR based *BPIFB1* knock out cell line. Infection studies using recombinant BPIFB1 protein during IAV infection suggested a protective effect against viral susceptibility. Taken together, my study has established mTECs and HBEC3-KT cells as suitable model systems to further investigate the protective role of BPIFB1 in viral IAV infection.

Acknowledgments

I am grateful to Allah, the Almighty, who has given me the strength, wisdom and patience to complete my research and for granting me this opportunity and the determination to see this research through.

Over the last four years, I have been privileged to study in the Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, under the guidance and supervision of Professor Colin Bingle. Firstly, I would like to offer my utmost respect and gratitude to Professor Bingle for his endless support of my PhD studies and research, for his motivation and immense scientific knowledge, and for affording me the opportunity to study under his guidance. His patience and advice have helped me greatly during my research and thesis writing.

Further thanks must go to my second supervisor, Dr. Lynne Bingle, for the instrumental advice, guidance and scientific contributions which were provided throughout this intense but thoroughly rewarding journey. There could not have been a superior combination of mentor and supervisor for me during my PhD studies. My gratitude to her is immense.

A huge thank you to all members of the Bingle group and PhD friends who made every day of my PhD studies enjoyable and a meaningful experience. Thank you to Dr. Priyanka Anuhan, Mr. Zeyad Alharbi and Mr. Fawaz Aljuhani, for all their help and being wonderful friends and PhD colleagues! A heartfelt thank you to Mr. Miraj Kobad Chowdhury for helping and supporting me when I was struggling with my work and for motivational talks.

I must also convey my thanks to the members of the Infection, Immunity and Cardiovascular Disease Department, for their technical, academic and deeply appreciated moral support.

I would like to express my gratitude to our collaborating researchers Dr Helen Marriott from the University of Sheffield for providing wild-type mice and Prof James Stewart from the University of Liverpool for providing the pVR1255 vector and IAV PR8 strain virus for this project.

Finally, my earnest appreciation goes to my sponsors in Saudi Arabia, Qassim University and the Saudi Arabian Cultural Bureau in the United Kingdom, for their financial support during this period.

Conference Presentations

I have presented a poster at Conference: "Lung injury and Repair" 10th -12th September 2019, **British Association for Lung Research (BALR)**, Cambridge University. Towards defining a role for BPIFB1 in infection and disease progression following infection with influenza A virus.

I presented a poster at **School Research Day**; 17/18 June 2019, The Octagon Centre, Sheffield Medical School, Sheffield, UK. Towards defining a role for BPIFB1 in infection and disease progression following infection with influenza A virus.

I have presented a poster at Conference: "Host-microbe interactions and their influence on lung disease onset and severity" 7th -9th September 2020, **European Respiratory Society (ERS)**, Virtual International Congress. Using mouse tracheal epithelial cells (mTECs) to uncover an antiviral role for BPIFB1.

I have presented a poster at Conference: "Drivers of lung disease: from the micro- to the macro-environment" 24 -25 June 2021, **British Association for Lung Research (BALR),** Virtual conference. In *vitro* approaches to defining an antiviral role for BPIFB1 in the airway epithelium.

Table of Contents

Table of Figures

List of tables

Abbreviations

aa	amino acid
ACTA2	smooth muscle actin alpha 2
ADU	arbitrary densitometry units
AEC	airway epithelial cells
AGER	advanced glycation endproduct–specific receptor
ALI	Air liquid interface
APS1	autoimmune disease polyglandular syndrome type 1
ASL	airway surface liquid
ATI	Alveolar type 1 cells
ATII	Alveolar type 2 cells
BAL	bronchoalveolar lavage
BCA	Bicinchoninic acid
BPI	Bactericidal/permeability-increasing protein
BPIF	BPI (bacterial/permeability-increasing protein) fold-containing'
BPIFA1	BPI fold containing family A member 1
BPIFB1	BPI fold containing family B member 1
BSA	bovine serum albumin
C-term	C-terminal
Cas9	CRISPR-associated (Cas) protein 9
CD	Cluster of differentiation
CDK4	cyclin-dependent kinase 4
CETP	Cholesteryl ester transfer protein
CF	cystic fibrosis
cgRNA	complementry guided RNA
COPD	chronic obstructive pulmonary disease
COVID-19	Coronavirus Disease 2019
CPM	Counts per milion
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
cRNA	Complementary ribonucleic acid
CXCL	C-X-C motif Chemokine Ligand
DAPT	N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester
DC	dendritic cell
DEG	Differntially expressed genes
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
dNTPs	Deoxynucleotide triphosphates
E.coli	*Escherichia coli*
EBI	European Bioinformatic Institute
EGF	epidermal growth factor
eGFP	enhanced green fluorescence protein
ELISA	enzyme-linked immunosorbent assay
ERK	extracellular signal-regulated kinase

EST	expressed sequence tag
FACS	fluorescence-activated cell sorting
FBS	Foetal bovine serum
FOXJ1	Forkhead Box J1
G-CSF	Granulocyte colony-stimulating factor
GFP	green fluorescent protein
GI	Gastrointestinal
GlcNAc	N-Acetylglucosamine
GM-CSF	Granulocyte-macrophage colony-stimulating factor
gRNA	genomic RNA
GTEx	Genotype-Tissue Expression
H	harvest
HA	haemagglutinin
hBD	human β-defensins
HBEC	human bronchial epithelial cells
HBS	HEPES buffered saline
HBSS	Hank's Balanced Salt Solution
HEK293	Human embryonic kidney 293
HPA	Human Protein Atlas
hpi	Hours post infection
hTERT	human Telomerase
IAV	Influenza A Virus
IFITM	IFN-inducible transmembrane
IFN	interferons
Ifnl2	Interferon Lambda 2
IFNγ	Interferon gamma
Iigp1	Interferon inducible GTPase 1
IL	Interleukin
ILC2	activated type 2 innate lymphoid cells
ILD	interstitial lung disease
IPF	idiopathic pulmonary fibrosis
ISGs	interferon-stimulated genes
kb	Kilobase
kDa	Kilodalton
KO	Knock Out
KSFM	Keratinocyte Serum-Free Media
LB	Luria Bertani media
LBP	Lipopolysaccharide Binding Protein
LC	Liquid Chromatography
LogFC	log2 fold change
LPLUNC	Long PLUNC
LPS	Lipopolysaccharide
LRT	Lower Respiratory Tract
LTA	Lipoteichoic Acid

M1	Matrix 1 protein
M2	Membrane 2 protein
MAT	Mature lung
MCC	mucociliary clearance
MD	Multiple dimension
MDS	Multidimensional scaling plots
MEK	MAPK/ERK kinase
MHC	Major Histocompatibility Complex
MIF	Macrophage Migration Inhibitory Factor
Mme	Metalloendopeptidase
MOI	Multiplicity of infection
Mpeg1	Macrophage-Expressed Gene 1
MS	Mass Spectrometry
MSA	Mulltiple Sequence Alignment
mTEC	Mouse Trachea Epithelial Cells
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2H-tetrazolium
MUC	mucin
mut	mutation
N-term	N-terminal
NA	Neuraminidase
NCBI	National Center for Biotechnology Information
NLRP3	NOD-like receptor family member NOD-, LRR- and pyrin domain-
NOS	Nitric Oxide Synthase
NP	Nucleoprotein
NPC	Nasopharyngeal Carcinomas
NS	Non-structural protein
NS2	Non-structural protein 2
nTPM	normalized Transcripts Per Million expression
Oas2	Oligoadenylate Synthetase 2
Olfm4	Olfactomedin 4
ORF	Open Reading Frame
PA	Polymerase acidic protein
PB	Polymerase basic protein
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
pDCs	plasmacytoid Dendritic Cells
PDZ	PSD95,Dlg1, Zo-1 domain
PER2	light-elicited circadian rhythm protein Period2
PI	Propidium iodide
PKR	Protein Kinase RNA
PLTP	Phospholipid Transfer Protein
PLUNC	Palate, Lung, and Nasal Epithelium Clone
PNEC	Pulmonary neuroendocrine cells
PNGase	Peptide-N-Glycosidase F enzyme

PR8	Influenza A virus strain A/Puerto Rico/8/1934 (H1N1)
pRRs	Pattern Recognition Receptors
RIG-I	Retinoic acid-inducible gene I
RNA-Seq	RNA sequencing
RT-PCR	Real-Time Polymerase Chain Reaction
SA	Sialic Acid
SCGB1A1	Secretoglobin Family 1A Member 1
scRNA-Seq	single-cell RNA Sequencing
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEM	Standard Error of the Mean
Slfn	Schlafen
SLPI	Secretory Leukocyte Protease Inhibitor
SMG	Submucosal Glands
SPDEF	SAM pointed domain containing ETS transcription factor
SPLUNC	short PLUNC
STAT3	Signal Transducer And Activator Of Transcription 3
TAC	Transcriptome Analysis Console
TAE	Tris-Acetate-EDTA
TBE	Tracheobronchial epithelial
TBS	Tris Buffered Saline
TBS-T	Tris Buffered Saline-Tween
TH2	T helper type 2
TIMP1	Tissue inhibitor 1 of metalloproteinases
TLR	Toll-like receptors
TMM	Trimmed Mean of M values
TNF-α	Tumour Necrosis Factor alpha
TPM	Transcripts Per Million expression
tSNE	T-distributed Stochastic Neighbourhood Embedding
UMAP	Uniform Manifold Approximation and Projection
Uniprot	Universal Protein resource database
URT	Upper Respiratory Tract
UV	Ultraviolet
VIM	Vimentin
vRNP	viral Ribonucleoprotein
wt	wild-type
X31	Influenza A virus strain A/HongKong/1/68 (H3N2)

Chapter 1

Introduction

1.1 The Respiratory System

The respiratory system is constituted of organs and tissues involved in breathing, which is the process of exchange of oxygen and carbon dioxide. It is divided into the upper respiratory tract (URT) (nasal cavity, pharynx, larynx), the conducting airways forming the lower respiratory tract (LRT) (trachea, bronchi, bronchioles) and a respiratory zone comprised of a large number of alveoli mediating gas exchange (Figure **1.1**) (Marib, 2001).

The entire respiratory tract, from the nasal cavity to the gas-exchanging alveoli is lined with respiratory epithelium in a continuous manner. In the respiratory systems of humans, cartilaginous airways extend from the parenchyma of the lungs as far as the bronchioles, and these are associated with fluid-secreting submucosal glands (SMGs), that secrete proteins involved in host-defence including the gel-forming mucins (Figure **1.2**A). The respiratory system is a complicated organ system which forms more than just a barrier to the external environment and a means to achieve gas exchange; it contains multiple specialised cells that are able to respond to changes induced by environmental or microbial stimuli and are proficient in mediating immune and neural responses. However, a number of features of the respiratory epithelium still remain incompletely understood, especially during chronic disease states, where it displays significant intrinsic functional differences from that seen in normal conditions.

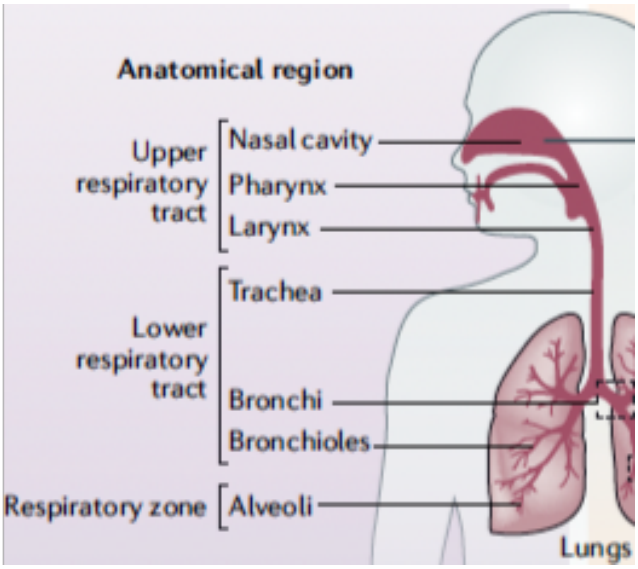


Figure 1.1: Anatomical overview of the respiratory tract. The schematic outlines components of the upper and lower respiratory tracts and the respiratory zone (Hewitt and Lloyd, 2021), available via the PubMed Central (PMC) Open access Subset for unrestricted research re-use.

1.1.1 Human Airway Epithelium

As outlined above the airway epithelium is made of a continuous cellular layer. However, it differs in phenotype throughout the respiratory tract. The larger human airways are lined by a pseudostratified columnar epithelium, a single layer of cells located above the basement membrane, that have the appearance of being multicellular (Figure **1.2**A). In the terminal bronchioles this epithelium becomes cuboidal, with the cells having a more cube-like shape. The respiratory epithelium is composed of multiple cell types (Breeze and Wheeldon, 1977; Widdicombe, 2019), and these cell types can be grouped into three primary categories: basal cells, ciliated cells, and secretory cells, including club (Clara) and goblet cells (Spina, 1998). However, more recent molecular studies have uncovered enormous cellular heterogeneity within the respiratory epithelium, describing for example new cell populations such as pulmonary ionocytes and new cellular subtypes or cellular differentiation states, such as deuterosomal or mucous ciliated cells (Deprez *et al.*, 2020; Travaglini *et al.*, 2020; Davis and Wypych, 2021).

In larger airways, the epithelium consists mainly of ciliated, secretory, basal and undifferentiated cells (Figure **1.2**A) and is paramount in facilitating easy airflow in and out of the alveoli (Knight and Holgate, 2003), this allows the essential gas exchanging function of the peripheral lung to take place. The epithelium also plays an important role in protecting the lung against inhaled particles, toxins, allergens and infectious pathogens. Ciliated and secretory cells work together to allow mucociliary clearance (MCC) to protect the host from pathogen attack (Knight and Holgate, 2003; Puchelle *et al.*, 2006). The epithelium secretes a large number of mediators (for example antimicrobial proteins, mucins, and cytokines) that defend against foreign particles.

The multiple cell types found within the airways are derived from a progenitor cell pool. The precise lingeage relationship between these different cells is complex and is still not fully resolved. This complexity has been accentuated by the advent opf single cell sequencing studies (Travaglini *et al.*, 2020; Deprez *et al.*, 2020).

Many of the cells have the capacity to change their phenotypes under different conditions. There may also be some species-specific differences in these relaitonships, as for example, not all airway cell types are common between humans and mice (Travaglini *et al.*, 2020).

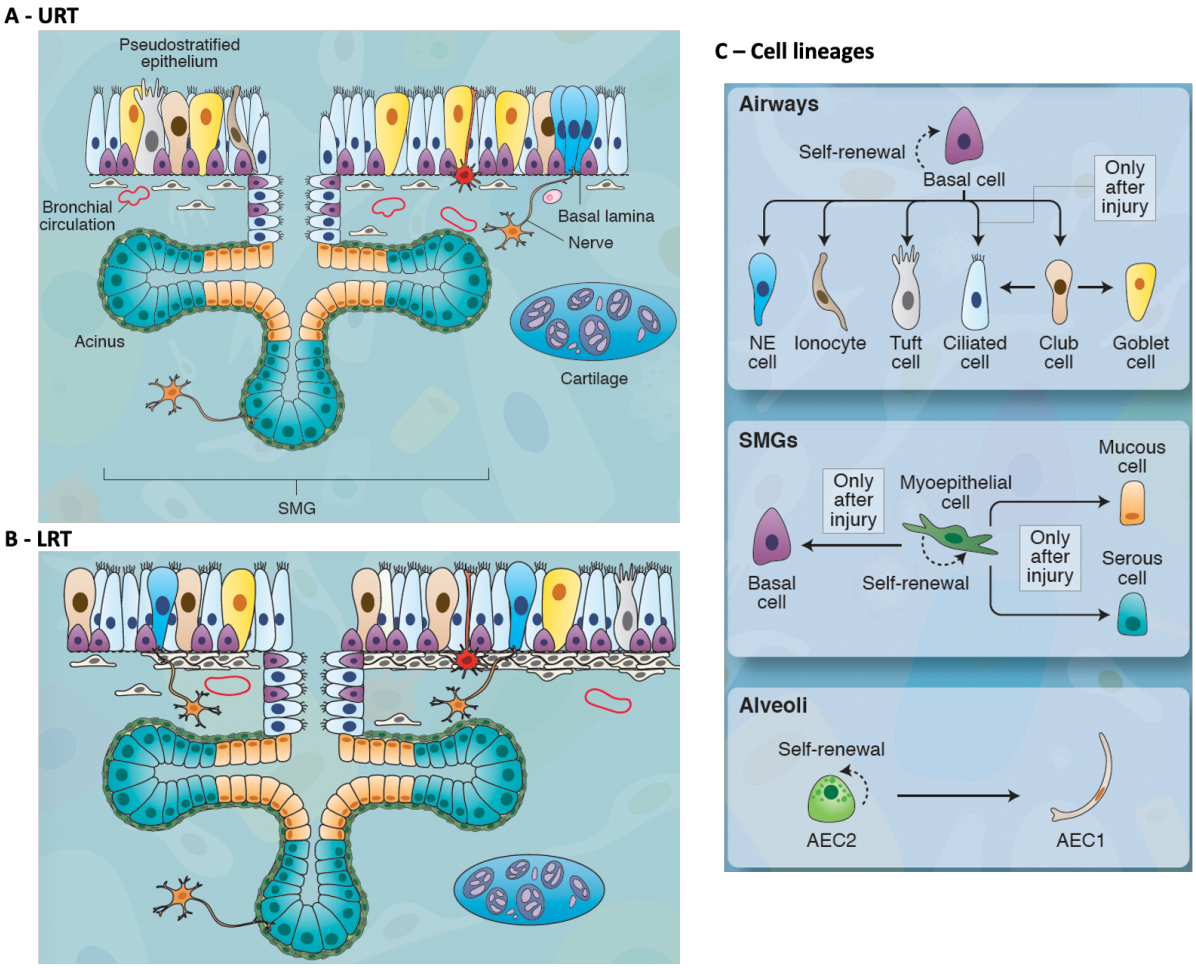


Figure 1.2: The cell types constituting the human airway epithelium. A) The large airway epithelium or upper respiratory tract (URT) has a higher presence of secretory (goblet) cells, whereas in the small airway epithelium or lower respiratory tract (LRT) **(B)** these shift to the club cell type. Although the cell types are similar across the respiratory tract, the small airway epithelium has a higher number of ciliated cells. **C)** Overview of cell types in the respiratory epithelium. See text for more information on the specific cell types. SMG - submucosal glands. Adapted from (Hogan and Tata, 2019), copyright permission licence 1335034.

1.1.1.1 Basal Cells

Basal cells are pyramidal in shape and are particularly numerous in the conducting airways making up between 6 - 31 % of the epithelial cell population depending on the location (Ayers and Jeffery, 1988) (Boers et al., 1998). In the trachea, basal cells are situated amongst secretory and ciliated cells (Rock et al., 2009), and they are spread throughout all of the airway (Evans et al., 2001).

Basal cells attach to other superficial epithelial cells through desmosomes, and to the basement membrane of epithelial cells through hemidesmosomes, suggesting that basal cells are significant in maintaining the integrity of the epithelial surface (Crystal, 2014). They can be identified through expression of transformation-related protein 63 (Trp-63), cytokeratin 5 (KRT5) or cytokeratin 14 (KRT14) (Figure **1.3**) (Rock *et al.*, 2010). However, more recent studies have suggested that basal cells exist as at least two transcriptionally different subpopulations, identified in human and mouse airways (Ghosh *et al.*, 2011; Travaglini *et al.*, 2020).

The major function of basal cells is to aid epithelial regeneration following injury by conversion into most other differentiated epithelial and therefore they act as key modulators of respiratory homeostasis (Figure **1.2**C and Figure **1.3**) (Boers *et al.*, 1998; Hong *et al.*, 2004; Davis and Wypych, 2021). They can be considered to form a stem cell population that ultimately gives rise to all of the other cells in the airway epithelium. Therefore, it is not surprising that transcriptional dysregulation of basal cells through environmental factors like smoking has been have been found to contribute to chronic obstructive pulmonary disease (COPD) (Crystal, 2014; Ryan *et al.*, 2014; Goldfarbmuren *et al.*, 2020) and lung cancer (Ryan *et al.*, 2014).

1.1.1.2 Club Cells

Dome-shaped club cells are important secretory cells located in the airway amongst ciliated cells and they dominate the epithelium in the respiratory bronchioles (Boers *et al.*, 1999). Club cells are located in the proximal and distal airways, and require the transforming growth factor activin receptor-like kinase 5 (Alk5) for differentiation (Xing *et al.*, 2010). Club cells possess stem cell functions and can perform epithelial tissue repair through dedifferentiation into basal cells, followed by subsequent rediffierention into other cell types. (Reynolds *et al.*, 2000; Tata *et al.*, 2013; Zheng *et al.*, 2017). They can also directly differentiate into goblet and ciliated cells. Club cells perform a secretory function in the bronchioles, by expressing and secreting SCGB1A1 (their traditional marker protein) as well as other host defence proteins and thus facilitating an intrinsic immune response (Figure **1.3**) (Evans *et al.*, 1978; Rawlins *et al.*, 2009). Predictably, dysregulation of club cell homeostasis has been shown to contribute to a range of respiratory conditions like cystic fibrosis (CF) asthma, idiopathic pulmonary fibrosis (IPF) and COPD (Rokicki *et al.*, 2016; Zuo *et al.*, 2020).

1.1.1.3 Ciliated Cells

The most common cell type in the airway epithelium is the columnar ciliated cell. These cells account for more than 50% of the population (Spina, 1998), and contain numerous mitochondria located directly beneath the apical surface that providing the energy required to maintain the normal cilia function. In the airways, ciliated cells function primarily to propel broncho-tracheal secretions upwards towards the pharynx through the process known as the mucociliary clearance.

Ciliated cells can arise from either basal, club or goblet cells (Ruiz García *et al.*, 2019) and it has been historically considered that ciliated epithelial cells are terminally differentiated (Ayers and Jeffery, 1988). These cells are distributed throughout the entire airway epithelium, where at least two transcriptional distinct subsets along the proximal-distal axis have been identified (Travaglini *et al.*, 2020). Recently, an additional cell population has been described constituting a transitional cell state prior to full ciliated differentiation, and these cells are termed deuterosomal cells (Ruiz García *et al.*, 2019). They represent an intermediate progenitor cell type, which are characterised by expression of the key regulator for centriole amplification (Revinski *et al.*, 2018). Ciliated cell fate is mediated by the concerted action of MYB proto-oncogene (Ito *et al.*, 2020), Notch signalling (Morimoto *et al.*, 2012), geminin coiled-coil domain containing (GMNC) (Whitsett, 2018) and Forkhead Box J1 (FOXJ1), the last of these being required for cilia formation (You *et al.*, 2004) (You *et al.*, 2004). A common employed cilia marker is β-tubulin IV (Renthal *et al.*, 1993).

Cells with characteristic markers of both mucosal and ciliated cells have been identified both *in vitro* and *in vivo*, suggesting that the ciliated cells are able to transform into mucous cells, under certain conditions, for example following IL-13 exposure (Seibold, 2018). Additionally, ciliated cells have been described to be highly susceptible towards viral infections based on their expression of viral receptors. For example they have high expression of cadherin-related family member 3 (CDHR3), known to be targeted by viruses such as Rhinovirus-C (Griggs *et al.*, 2017).

1.1.1.4 Goblet Cells

Goblet cells are interspersed amongst the ciliated cells and contain mucin-rich granules that release mucus onto the luminal surface of the airways, and together they facilitate effective mucociliary clearance (Figure **1.2**) (Breeze and Wheeldon, 1977). A normal human trachea

contains mucosal cells, and secretions from the cells extruding through pores and pits that are fixed in situ, and they often project a fair distance into the lumen (Bonser and Erle, 2017; Ermund *et al.*, 2017). It has been suggested that goblet cells do not completely discharge their contents before beginning to synthesize mucin again, rather, they elaborate mucus either continuously or on several occasions (Hansson, 2019; Whitsett *et al.*, 2019).

Goblet cells differentiate from club cells, like their ciliated counterparts, however the differentiation is controlled by transcriptional pathways involving SAM pointed domain containing ETS transcription factor (SPDEF) and forkhead box A3 (FOXA3) (Rajavelu *et al.*, 2015; Whitsett, 2018). Recent studies have discovered two functionally different goblet cell subsets: the goblet-1 type, responsible for mucus production containing MUC5AC and MUC5B, and goblet-2 cells, which has a more antibacterial defence role by secreting zymogen granule protein 16 (ZG16B) orthologues (Figure **1.3**) (Montoro *et al.*, 2018). Both, qualitative and quantitative alterations in mucus secretion have been implicated in a number of lung diseases (Chen *et al.*, 2009; Roy *et al.*, 2014; Evans *et al.*, 2015; Vieira Braga *et al.*, 2019).

1.1.1.5 Pulmonary neuroendocrine cells (PNEC)

Pulmonary neuroendocrine cells (PNEC) are rare cells present in the respiratory epithelium, where they are able to sense airway activity and appropriately secrete bioactive amines and neuropeptide (Branchfield *et al.*, 2016). PNEC products modulate immune responses through mast cell recruitment or cytokine secretion, or a nervous response leading to vasodilation or bronchoconstriction (Atanasova and Reznikov, 2018). Specific PNEC markers are achaete-scute family BHLH transcription factor 1 (ASCL1) or calcitonin related polypeptide alpha (CALCA) (Hogan and Tata, 2019; Mou *et al.*, 2021).

Although PNEC constitute only a very small proportion of the airway epithelium (0.3-0.5%), these cells form the key mediators and communicators between the airway epithelia, the neuronal and immune system (Mou *et al.*, 2021). To briefly illustrate their immune modulatory function: the neuropeptide calcitonin gene related peptide (CGRP) produced by PNECs has been found to recruit activated type 2 innate lymphoid cells (ILC2), which are common lymphoid progenitor cells lacking the antigen specific B or T cell receptor. ILC2s produce type 2 cytokines such as IL-4 and IL-13 and regulate respiratory virus infections or the allergic immune response. Whereas PNEC derived γ-aminobutyric acid (GABA)

stimulates goblet cell hyperplasia, resulting in excessive mucus production in response to pathogens, toxins or cigarette smoke (Sui *et al.*, 2018).

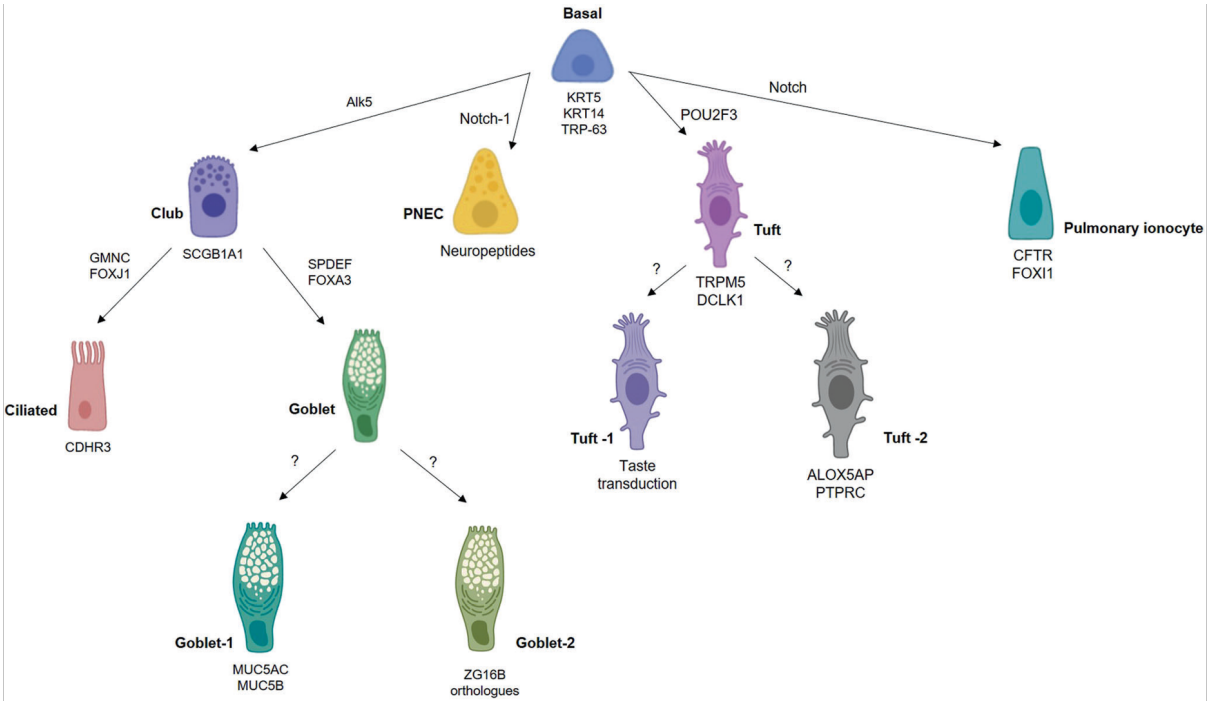


Figure 1.3: The cellular heterogeneity of the airway epithelium Basal cells are the principal stem cells of the respiratory epithelial and are able to differentiate into most other epithelial cell types. Outlined transcriptional networks (indicated on arrows) govern specific epithelial cell fates have been fully explored, however newly recognised cell types and subtypes require more investigation. See text for more information on the individual cell types and their associated functions (Davis and Wypych, 2021), copyright permission licence 1344732.

1.1.1.6 Rare cells

The airway epithelium also contains other rarer cells . Tuft cells have been recently described as critical components of the immune response and require the transcription factor POU Class 2 Homeobox 3 (POU2F3) for differentiation. Moreover, Taste 1 Receptor Member 3 (TAS1R3) and the mechanistic target of rapamycin (mTORC1) have been shown to regulate homeostatic tuft cell differentiation and abundance (Gerbe and Jay, 2016; Howitt *et al.*, 2016; Huang *et al.*, 2018; Howitt *et al.*, 2020; Strine and Wilen, 2022). They are not found exclusively in the respiratory tract (Montoro *et al.*, 2018), rather this cell type has been described across various organs such as the digestive tract (Luciano *et al.*, 1981; Höfer *et al.*, 1996; Luciano and Reale, 1997; Bezençon *et al.*, 2008), olfactory epithelium (Fletcher *et al.*, 2017), urethra (Deckmann and Kummer, 2016), thymus (Bornstein *et al.*, 2018) and auditory

tube (Krasteva *et al.*, 2012). The name tuft cell is derived from the distinct microvillous tuft on the apical side, however other names based on other locations and appearances have been coined, such as brush and microvillous cells (Schneider *et al.*, 2019). Recent scRNA studies determined the existence of three airway tuft cell subsets: immature tuft cell, tuft-1 and tuft-2 cells (Figure **1.3**) (Montoro *et al.*, 2018). Tuft cells express transient receptor potential cation channel subfamily M member 5 (TRPM5) and doublecortin like kinase 1 (DCLK1) and reportedly mediate various functions from chemosensory (tuft-1) to neuronal and immunological functions (tuft-2) (Bezençon *et al.*, 2008; Von Moltke *et al.*, 2016; Gerbe and Jay, 2016; Sharma *et al.*, 2017; Schneider *et al.*, 2019). This is consistent with the findings of arachidonate 5-lipoxygenase activating protein (ALOX5AP) protein and tyrosine phosphatase receptor type C (PTPRC) expression, which are associated with leukotriene synthesis and immunomodulation (Montoro *et al.*, 2018).

Pulmonary ionocytes in the respiratory airways have been a very recent discovery (Montoro *et al.*, 2018; Plasschaert *et al.*, 2018). They have been described to originate from basal or tuft-like cells and constitute less then 2 % of all airway epithelial cells (AECs) (Goldfarbmuren *et al.*, 2020). These cells require Notch signalling for differentiation and express Forkhead Box I1 (FOXI1) and high levels of CF transmembrane conductance regulator (CFTR) (Figure **1.3**) (Montoro *et al.*, 2018; Plasschaert *et al.*, 2018). The latter causes CF development when it is mutated (Miah *et al.*, 2019).

1.1.1.7 Submucosal glands

Submucosal glands (SMGs) contribute to the liquid film, the airway surface liquid (ASL) which coats and protects all airway surfaces and which contains a large number of secreted proteins, including the gel-forming proteins, MUC5B and MUC5AC, electrolytes and water. In order to allow for a large number of mucus-producing cells, without displacing the ciliated cells required for MCC, the glands are formed as tubuloacinar structures, containing a single collection duct, followed by branching secretory tubules lined with mucous cells, which end in serous acini (Figure **1.2**A) (Widdicombe and Wine, 2015). The glands are lined with myoepithelial multipotent stem cells, which are positive for markers like smooth muscle actin alpha 2 (ACTA2) or tumour protein 63 (TP63), which facilitate contraction and thus expel the secretions of the glands (Makarenkova and Dartt, 2015). Acinar cells are either of serous

nature and are positive for demilune cell and parotid protein (DCPP1) (Maruyama *et al.*, 2016), or the mucous kind that stain for MUC5B (Liu *et al.*, 2002).

1.1.1.8 Alveolar epithelial cells

The two types of specialised epithelial cells that constitute the distal alveolar lung region carry out distinct functions. Alveolar type 1 cells (ATI cells) have a large surface area and facilitate gas exchange through the capillaries that they overlay. Specific molecular cellular markers are Podoplanin (PDPN) (Rishi *et al.*, 1995) and advanced glycation endproduct–specific receptor (AGER) (Treutlein *et al.*, 2014). Alveolar type 2 cells (ATII cells) produce and secrete pulmonary surfactant that is critical to lower surface tension in the lung and to prevent alveolar collapse. ATII cells can detected using markers such as Surfactant Protein C (SFTPC) (Sun *et al.*, 2021) or dendritic cell lysosomal associated membrane glycoprotein (DC-LAMP) (Salaun *et al.*, 2004) (Figure **1.2**C). These cells also have stem cell capabilities for self-renewal within the alveolar space.

1.1.1.9 Mesenchymal cells

Besides epithelial cells, cell of mesenchymal origin are found associated with the airways (Fernandes *et al.*, 2006). These are smooth muscle cells, characterised by the expression of myosin heavy chain 11 (MYH11) and ACTA2 (Dobnikar *et al.*, 2018) and peribronchial fibroblasts, positive for the marker proteins glioma-associated oncogene homolog zinc finger 1 (GLI1) and leucine rich repeat containing G protein-coupled receptor 6 (LGR6) (Fang *et al.*, 2019). ATII alveolar associated fibroblasts, can be identified using the marker platelet-derived growth factor receptor-α (PDGFRA) (Shiraishi *et al.*, 2019) and pericytes by platelet-derived growth factor receptor-β (PDGFRB) (Muhl *et al.*, 2020).

1.1.1.10 Immune cells

Alongside the respiratory epithelium, recruited or residing immune cells such as dendritic cells, lung based-neutrophils, alveolar resident macrophages, interstitial macrophages, basophils, eosinophils or lymphoid T and B cells mediate the respiratory immune response.

1.1.1.11 Differences in murine versus human airways architecture

Mouse models and airway tissue samples have been used extensively in research studies to gain a better understanding of the respiratory epithelium, both in healthy and compromised conditions. Although these have proven to be invaluable tools for research, especially when human studies cannot be carried out, for example due the lack of samples, and practical or ethical reasons, it needs be acknowledged that murine airway architecture differs to the human in several areas. For example, unsurprisingly, considering the organism size, the mouse tidal lung volume is much smaller and respiratory bronchioles are absent from murine airways (Pack *et al.*, 1980; Mercer *et al.*, 1994a). With respect to cell types and distribution, goblet cells are found much less abundantly, and cartilage and basal cells are absent from mouse intralobar airways (Mercer *et al.*, 1994b; Boers *et al.*, 1998). Furthermore, SMGs are present only in the upper trachea in murine airways (Tata *et al.*, 2018). Human PNECs are mostly individual located and scattered across the airways and not in the clusters the are seen in mouse airways (Mercer *et al.*, 1994b; Rock *et al.*, 2010). Furthermore, due to the living environment of rodents and microbial exposure, the immune cell repertoire differs between murine and human models (Beura *et al.*, 2016; Tao and Reese, 2017; Abolins *et al.*, 2017).

1.2 Protective functions of the respiratory epithelium

1.2.1 Barrier function of the airway epithelium

The peripheral and conducting segments of the airways in the human respiratory system have distinctly different roles in the innate immune defence and homeostasis of the host. The airway epithelium serves primarily as a physical barrier, preventing the internal milieu from being exposed to the external environment. The barrier function of the epithelium is facilitated by intercellular tight and adherens junctions (Pohl *et al.*, 2009), the mucociliary escalator, which is responsible for MCC, (Knowles and Boucher, 2002; Bustamante-Marin and Ostrowski, 2017) and secreted antimicrobial host defence products (Bals and Hiemstra, 2004).

Mucus and cilia are together responsible for the mucociliary function of the respiratory epithelium. Inhaled debris and pathogens are entrapped by the mucus, and the process of MCC then transports this towards the pharynx, away from the lungs, by the rhythmic beating of the cilia (Knowles and Boucher, 2002). Mucus hydration levels and the rate of ciliary

beating both determine the efficiency of mucus transport in the respiratory system (Puchelle *et al.*, 1995). There are over 200 proteins found in the respiratory mucus, which are secreted by both the submucosal glands and goblet cells (Joo *et al.*, 2015). Airway mucus is primarily composed of high molecular-weight glycoproteins such as mucins that interlink to make a barrier of mucus (Joo *et al.*, 2015; Radicioni *et al.*, 2016). MUC5B and MUC5AC are the most abundant mucins found in a normal respiratory tract (Rose and Voynow, 2006). The former is predominantly made by submucosal glands, while the latter are found in goblet cells (Okuda *et al.*, 2019). A number of inflammatory mediators such as neutrophil elastase, tumour necrosis factor-α (TNF-α), IL-17, IL-13, IL-1β, transforming growth factor β (TGF), and epidermal growth factor (EGF), and environmental factors, for example allergens, cigarette smoke and microbial pathogens, have been shown to cause mucus hypersecretion (Dohrman *et al.*, 1998; Casalino-Matsuda *et al.*, 2009).

The speed and efficiency of MCC is dependent on the thin ASL layer that lines the airway surface. ASL is composed of two layers, a lower periciliary layer with large glycoproteins that are bound to its membrane (Sheehan *et al.*, 2007), and an upper viscoelastic layer that is full of mucins and other proteins secreted from submucosal glands and goblet cells (Button *et al.*, 2012). CF patients secrete excess mucus due to goblet cell hyperplasia and metaplasia. They also experience airway obstruction due to hypertrophy of the submucosal glands (Simonin *et al.*, 2019). Persistent/recurrent infections of the respiratory tract can result from mucociliary dysfunction, as seen in COPD, ciliary dyskinesia, and CF patients (Sethi, 2000; Bhowmik *et al.*, 2009). Mucin glycoproteins interact with a number of pathogens of the respiratory tract, such as the influenza virus and *Staphylococcus aureus* (Sajjan *et al.*, 1992; Plotkowski *et al.*, 1993). Further, they protect the underlying epithelium from pathogens and environmental toxins by erecting a physical barrier that restricts their access to target receptors on host cells. In addition, they can act as non-productive decoys that capture infectious agents. For example, mucins contain sulfate and sialic acid residues, where the sulfation provides extra protection of the underlying epithelium against bacteria, and the sialic acids represent one example of a cell surface glycan "decoy" receptor frequently targeted by viruses such as the influenza A virus (Derrien *et al.*, 2010; Honigfort *et al.*, 2021). Pathogens bound to the mucus of the airways are usually cleared, however disruption of MCC can allow these pathogens to persist in the respiratory tract, potentially resulting in an epithelium-damaging inflammatory response.

Adherens junctions and tight junctions are significant contributors to the respiratory epithelium's barrier function. Adherens junctions mediate intercellular adhesion and promote tight junction formation, while tight junctions control ion and solute transport across the epithelium (Shin *et al.*, 2006). These junctions protect the respiratory epithelium from environmental insults and inhaled pathogens and they also serve signalling functions, regulating both cell differentiation and proliferation (Koch and Nusrat, 2009).

Bacterial or viral infections can result in temporary disruptions of the adherens and tight junctions (Kim *et al.*, 2005; Sajjan *et al.*, 2008) and repeated insults to junctional complexes affects the epithelial barrier function and also its' ability to heal and differentiate. Inflammatory mediators from the host, such as TNF-α and interferons, are expressed following infection and they can prolong the disruption of tight junctions, allowing the entry of pollutants and allergens following the eradication of an infection (Baker *et al.*, 2008; Tugizov, 2021).

1.2.2 Host defence proteins

The respiratory epithelium also functions as a biochemical barrier to pathogens as the epithelial cells secrete a large range of antimicrobial substances. Lysozyme, has antimicrobial effects against a broad spectrum of Gram-positive bacteria (Ibrahim *et al.*, 2002). In the presence of lactoferrin, lysozyme is able to protect against Gram-negative bacteria (Ellison and Giehl, 1991). In addition, lactoferrin is able to prevents microbes from growing (Ganz, 2002). It also exhibits antiviral properties against both DNA and RNA viruses by adhering to the pathogen or by causing the pathogen to adhere to a host cell (Ganz, 2003).

Protease inhibitors such as antichymotrypsin, α1-antiprotease, elastase inhibitor, and secretory leukocyte protease inhibitor (SLPI) are all produced by epithelial cells. These molecules limit the effect of innate immune cells and pathogenic proteases. During infection, the homeostasis of the airway epithelium is dependent on the balance between proteinases and antiproteinases. Lysozyme and SLPI levels decrease in the presence of bacterial infections in COPD patients, whereas lactoferrin levels remain constant (Parameswaran *et al.*, 2011). SLPI administration to CF patients lowers elastase and IL-8 activity in lung secretions (McElvaney *et al.*, 1992).

One of the most abundant antimicrobial peptides in the respiratory epithelium are the human β-defensins (hBD). These defend against a broad spectrum of viral and bacterial pathogens and belong to a class of small cationic antimicrobial peptides. They act through disruption of the microbial membrane, which allow influx of water (Ganz, 2003; Kalenik *et al.*, 2018). hBD1 is expressed constitutively, but hBD2, hBD3, and hBD4 are expressed in response to lipopolysaccharide (LPS) of bacteria via IL-1 and the activation of NFκB (Singh *et al.*, 1998; Becker *et al.*, 2000). hBD2 activity is attenuated in patients with CF due to raised concentrations of salt (Goldman *et al.*, 1997). hBD2 levels are significantly lower in the sputum and pharyngeal wash of former/current smokers than in that of non-smokers; respiratory epithelium exposure to cigarette smoke *in vitro* was shown to prevent hBD2 induction by bacteria (Herr *et al.*, 2009).

Another group of antimicrobials are the cathelicidins. The only human cathelicidin is LL37 (other species contain more) and it binds to LPS and leads to its inactivation (Nagaoka *et al.*, 2020). LL37 overexpression in CF mouse models led to an increase in the destruction of *Pseudomonas aeruginosa* and a reduction in the ability of the bacterium to colonize the respiratory epithelium (Bals *et al.*, 1998).

The respiratory epithelium also produces oxidants, such as hydrogen peroxide and nitric oxide (NO). Three distinct NO synthases (NOS1, NOS2, and NOS3) contribute to NO production in the respiratory epithelium. NOS1 and NOS3 are constitutively expressed, while NOS2 is induced by pro-inflammatory cytokines and viruses. The lack of expression of NOS2 is responsible for the raised rates of viral replication in patients with CF, therefore it has been proposed that NOS2 overexpression is protective against infections from viruses (Zheng *et al.*, 2004). Lactoperoxidase produces hypothiocyanite, a microbial oxidant that is effective at killing both Gram-negative and Gram-positive bacteria (Moskwa *et al.*, 2007).

1.2.3 Innate immunity of the respiratory system

The airways allow the continuous flow of air to the peripheral lung and the immune system in the lung is required to maintain this uninterrupted gas exchange, as inhaled microorganisms or debris can cause pathology if they successfully invade the respiratory tract. The immune system needs to rapidly eliminate pathogens, to prevent inflammatory response from occurring. A highly ordered, progressive immune response involving a multi-tiered host defence system counters pathogens that enter the respiratory tract (Johnston *et al.*, 2021).

Immunity is divided into innate and adaptive. Innate immunity is the first line of defence against pathogens and allergens, and functions non-specifically without the need for an antibody (adaptive immunity) mediated (Koyama *et al.*, 2008; Thaiss *et al.*, 2016) response.

The innate immune system exhibits specificity for different pathogen classes, mediated through leukocytes interactions and intrinsic epithelial mechanisms (Johnston *et al.*, 2021). Type 1 immune responses are directed primarily at intracellular pathogens, particularly viruses, and are critical in controlling respiratory virus infections. This response is triggered through microbial nucleic acids and interferons, which are released from neighbouring epithelial cells. Epithelial responses lead to induction of type I and type III interferons (IFNs) as well as interferon-stimulated genes (ISGs), which impair viral replication (Schoggins, 2019). Type 2 immune responses are directed primarily at multicellular parasites and fungal infections, which are stimulated by specific proteases and chitin, and responses include mucin hypersecretion through the IL-13 signalling pathway and chitinase release (Elieh Ali Komi *et al.*, 2018; Weatherhead *et al.*, 2020). Type 3 immune responses serve as a defence mechanism towards a wide range of extracellular microbial pathogens, such as bacteria and fungi, as well as viruses during their extracellular phase of infection and it triggered by LPS, lipoteichoic acid (LTA) or nucleic acids. The main stimuli include nucleic acids or bacterial wall components, such as lipopeptides and endotoxin. The defence responses to these stimuli are the release of antimicrobial peptides (AMPs) and a strong neutrophil recruitment (Tzou *et al.*, 2000). Together, this multi-tiered system of innate immunity provides a strong defence barrier against infections caused by fungi, protozoa, bacteria, and viruses such as Influenza A (IAV).

1.3 Influenza A virus

IAV, a member of the orthomyxovididae family of viruses, causes approximately 650,000 annual deaths through respiratory diseases (World Health Organization, 2018). IAV is an enveloped virus with a genome that is composed of single-stranded, negative-sense, segmented ribonucleic acid (RNA). Each virus contains eight segments, coding for 11 viral genes (Figure **1.4**) and facilitate various functions to generate and release newly synthesised viral particles. Polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2) and polymerase acidic protein (PA) are the three subunits that make up the viral RNA-dependent

RNA polymerase required for RNA synthesis and replication of IVA infected cells. The glycoproteins haemagglutinin (HA) facilitates binding to sialic acid containing receptors and subsequent viral entry; and neuraminidase (NA) is required for viral particle release and viral spread. Matrix 1 (M1) protein provides a scaffold function to maintain the viral shape and in combination with nuclear export protein (NEP), these proteins are responsible for movement of the viral RNA segments into the host cell. Membrane 2 (M2) protein forms a membrane located proton ion channel, essential for viral entry and exit. The RNA genome is bound by the viral nucleoprotein (NP), and the non-structural protein 1 (NS) is a virulence factor that inhibits host antiviral responses in infected cells (Figure **1.4**) (Sriwilaijaroen and Suzuki, 2012; Krammer *et al.*, 2018b).

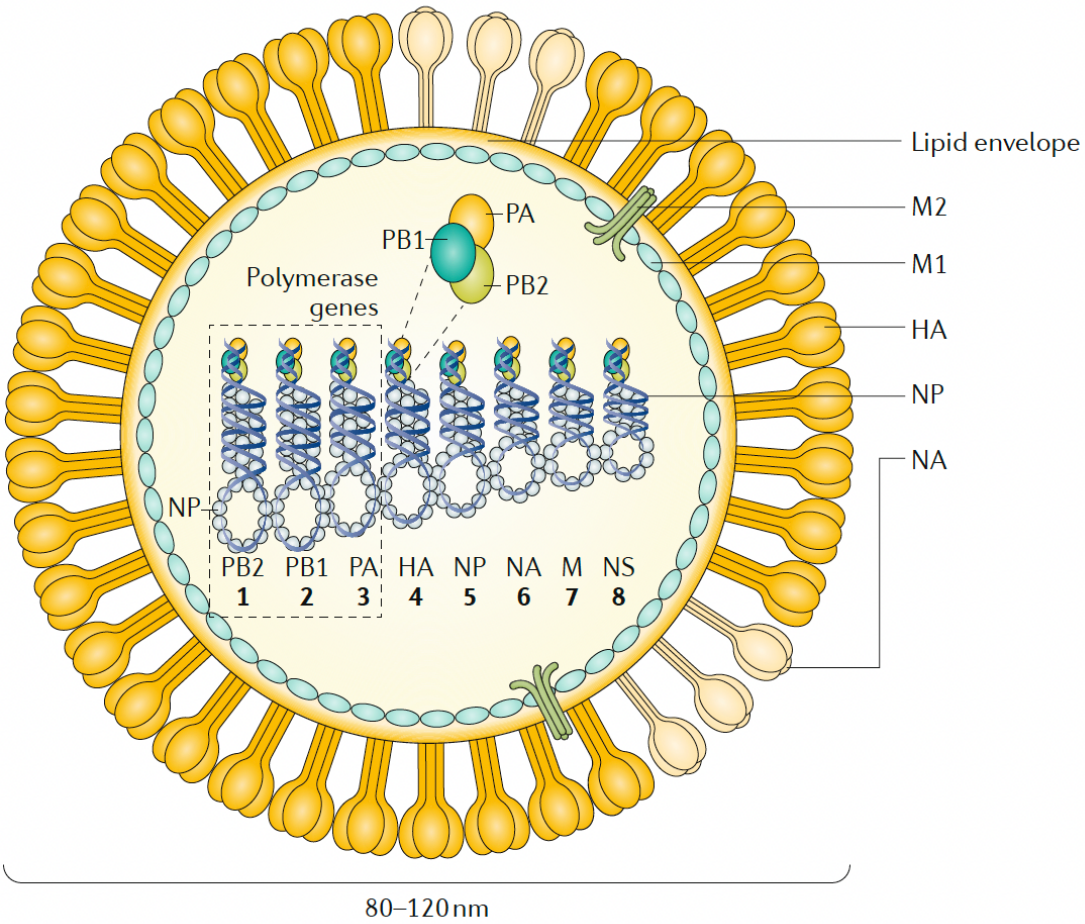


Figure 1.4: Structure of the Influenza A virus. IAV is a negative sense RNA virus, containing eight single-stranded RNA segments (1-8). The function of these segments and their proteins is outlined in the text. Figure from Krammer *et al.,* 2018, which has been made available via the PMC Open access Subset for unrestricted research re-use.

1.3.1 IAV Life Cycle

The life cycle of IAV can be separated into a number of stages: host cell entry; viral ribonucleoprotein (vRNP) entry into the host cell nucleus; replication and transcription of the viral genome; vRNP export from the host cell nucleus; and lastly virus assembly and budding at the plasma membrane of the host cell, providing the viral particle with a lipid membrane (Figure **1.5**) (Skehel and Wiley, 2000; Krammer *et al.*, 2018b).

A homotrimer of HA produces spikes on the lipid membrane of the virus, allowing it to adhere to sialic acid-containing molecules present on the host respiratory cell membrane, and subsequently triggering endosomal internalisation (Skehel and Wiley, 2000). The low endosomal pH induces a conformational change in the HA protein, causing the virus to fuse with the endosome membrane and vRNP release into the cytoplasm (Figure **1.5**)

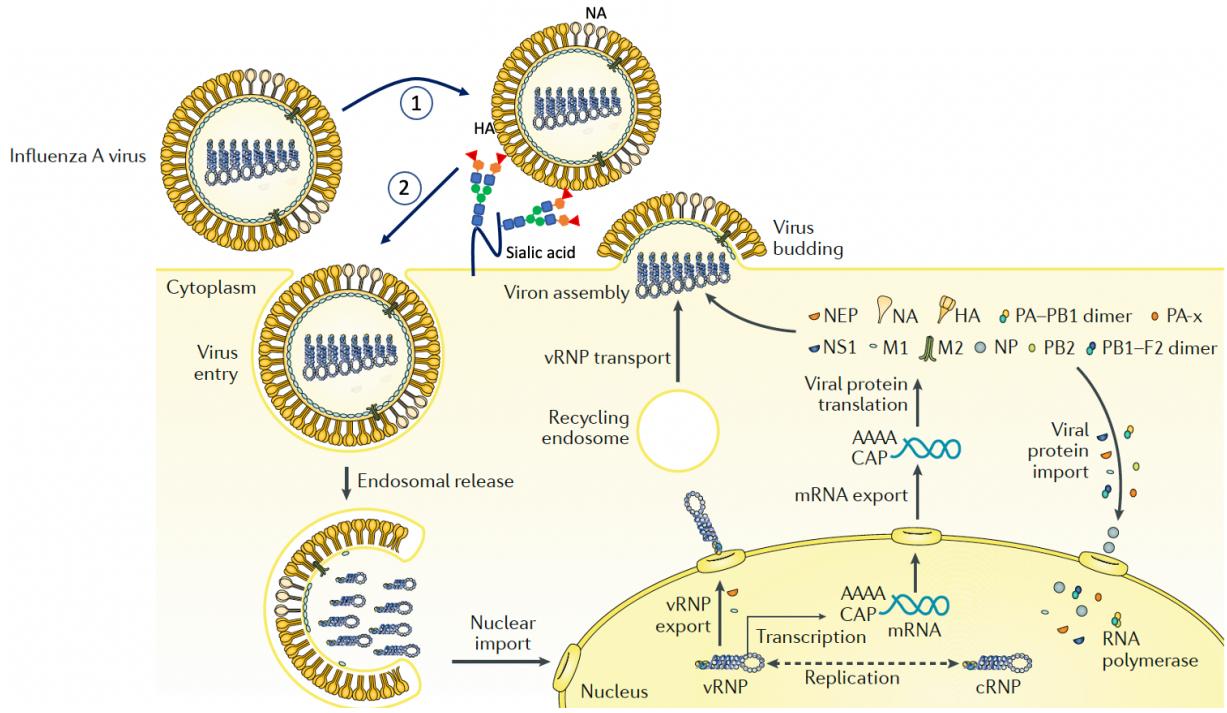


Figure 1.5: Life cycle of Influenza A particles. Viral haemagglutinin (HA) spikes (dark yellow 'head' and 'stalk' shaped proteins) facilitate host cell attachment (1) through sialic acid-containing molecules present on the host cell membrane and subsequent internalisation (2). cRNP - complementary ribonucleoprotein, vRNP - viral ribonucleoprotein, M1 - matrix protein, M2 - membrane protein; NA – neuraminidase, NEP - nuclear export protein, NP - nucleoprotein; NS1 - nonstructural protein, PB1, PB2 and PA - viral RNA polymerase subunits. Adapted figure from Krammer *et al.,* 2018, which has been made available via the PMC Open access Subset for unrestricted research re-use.

vRNP consists of PB1, PB2, PA RNA polymerase subunits in addition to NP. vRNP molecules are imported to the host cell nucleus, through the use of nuclear localization signals, utilising the cell's nuclear import machinery (Azzeh *et al.*, 2001; Crow *et al.*, 2004). Only negative-sense vRNPs leave the nucleus through the chromosomal maintenance 1 (CRM1) dependent pathway, and form viral particles prior to their exit from the host cell (Shapiro *et al.*, 1987). IAV uses the host plasma membrane to form these particles that exit cells through budding from the apical side of polarized cells. NA, M2, and HA move towards the plasma membrane on the apical side, and glycolipid and glycoprotein sialic acid residues need to be cleaved prior the budding process to facilitate the exit, which occurs through NA (Palese *et al.*, 1974).Viral particles with few/no vRNPs may be formed, but viral proteins like NA, M2, and HA must be present in the bilayer for the viral protein to form (Nayak *et al.*, 2009).

The interaction between viral HA and the host cell receptors is important in the pathophysiology of respiratory diseases caused by the virus, as IAV particles target host cells through the α2,6- and α2,3-type receptors. Glycosylated oligosaccharides that end with a sialic acid residue are target entry points for IAV (Connor *et al.*, 1994; Matrosovich *et al.*, 1997). α2,8, α2,6, and α2,3 linkages are formed through species-/cell-specific sialyl-transferases, which join these residues to glycans (Li *et al.*, 2011; Liu *et al.*, 2022).

1.3.2 The essential role of innate immunity in the defence against Influenza A virus mediated infections

Airborne IAV commences its invasion of the host through the respiratory tract, and in the majority of cases, infections are limited to the respiratory system. IAV entering the respiratory tract of a host needs to firstly evade the mucus that coats the respiratory epithelium. If successful, it then invades the respiratory epithelial cells, after which it is able to spread to the immune cells of the respiratory system (Perrone *et al.*, 2008).

Viral RNA found in infected cells is detected as foreign by a range of pattern recognition receptors (PRRs). Viral infections are detected by the innate immune system by PRRs through pathogen-associated molecular patterns, which are found in the pathogen or are formed during infection (Janeway, 1989; Medzhitov, 2001). This leads to the secretion of multiple host defence proteins including chemokines, pro-inflammatory cytokines and type I IFNs such as IFNα and IFNβ (Pestka *et al.*, 2004; Ivashkiv and Donlin, 2014).

At least three distinctly different types of PRRs, along with their respective ligands, are able to recognise the influenza virus. These are the: NOD-like receptors (i.e. LRR-, NOD-, and pyrin domain-containing 3 (NLRP3)), retinoic acid-inducible gene-I (RIG-I), encoded by the gene *DDX54,* and Toll-like receptors (i.e. TLR8 (single-stranded RNA), TLR7 (single-stranded RNA), and TLR3 (double-stranded RNA)) (Pang and Iwasaki, 2012).

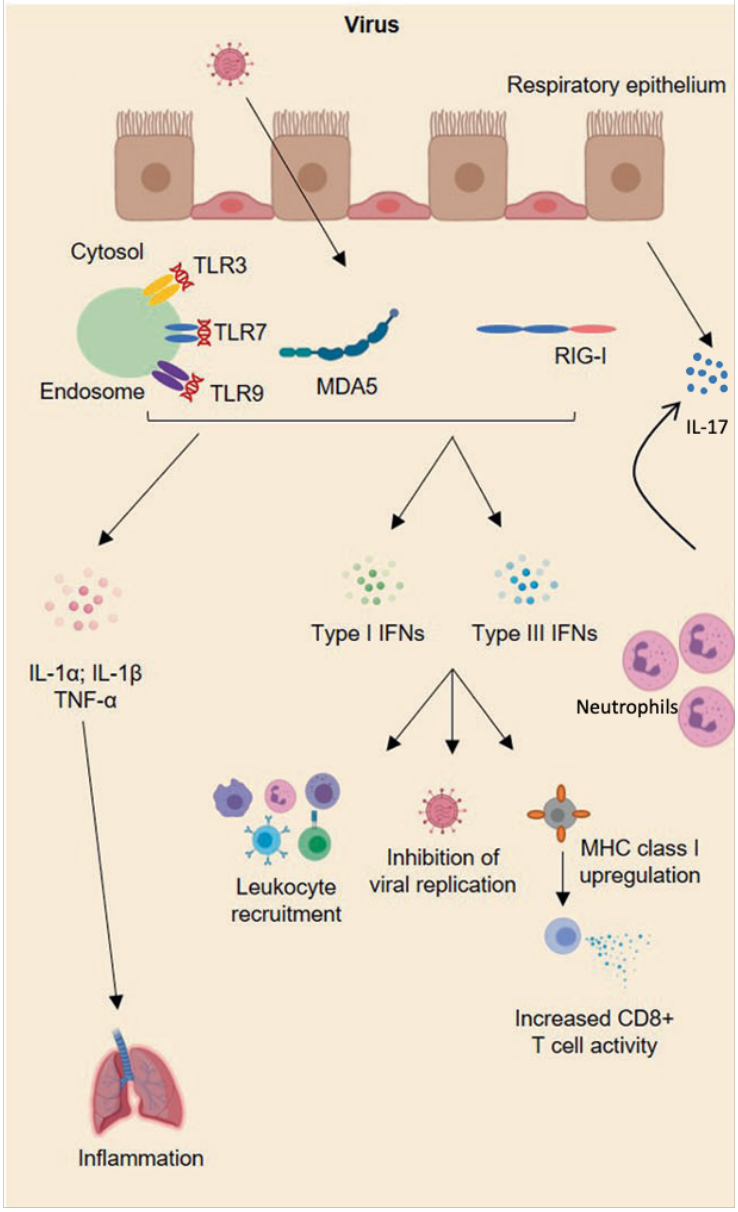


Figure 1.6: The immune response upon viral infection. Virus attacks are countered by intracellular virus PRRs such as toll-like receptors (TLR) and rig-like helicases (MDA5 and RIG-I), which leads to release of type 1 and 3 IFNs. This in turn stimulates leukocyte recruitment, inhibits viral replication, and stimulates MHC class I levels, leading to increased $CD8^+$ T cell activity. In addition, airway epithelial cells secrete proinflammatory cytokines, such as IL-1α, IL-1β, and TNF-α (Davis and Wypych, 2021), copyright permission licence 1344732.

NLRP3 and RIG-I recognize viruses in the cytosol of cells that have been infected, whereas TLR8, TLR7 and TLR3 recognize viral RNA in sentinel cell endosomes after it has been taken up (Alexopoulou et al., 2001). Additionally, TLR3 potentially recognizes RNA structures, which have not yet been identified, from dying influenza-infected cells that have undergone phagocytosis (Kawai and Akira, 2008; Jiang *et al.*, 2011). Further, the non-lethal doses of IAV are recognized by $CD8^+$ and $CD4^+$ T-cells, and $CD8^+$ T-cells in conjunction with leukocyte infiltration and chemokine expression, are reduced in the absence of TLR3 following a viral challenge, confirming a role for TLR3 in generating T-cell immunity (Heer *et al.*, 2007).

TLR3 promotes the recruitment of the adaptive and innate immune systems, both of which may damage the host, and induces signals that limit viral replication. TLR7 recognizes the single-stranded RNA genomes within the virions after they are taken up by endosomes in plasmacytoid dendritic cells (pDCs) (Lund *et al.*, 2004; Diebold *et al.*, 2004).

RIG-I is essential for the detection of viruses and the production of type I IFN in alveolar macrophages, conventional dendritic cells (DCs), and infected epithelial cells (Kato *et al.*, 2005). RIG-I binds to short double strands with 5' triphosphate moieties, found in the cytosol following viral replication (Pichlmair *et al.*, 2006; Baum *et al.*, 2010; Kell and Gale Jr, 2015).

Bronchial epithelial cells in humans also express NLRP3 (Pothlichet *et al.*, 2013), and NLRP3 inflammasomes have significant functions in defending against IAV (Allen *et al.*, 2009). NLRP3 was shown to be essential in recruiting leukocytes to the lungs following a challenge with a high viral dose of IAV (A/PR8 strain) (Thomas *et al.*, 2009). Activation of NLRP3 inflammasomes raises the degree of tolerance in tissues challenged with high-dose IAVs, which occurs through tissue repair and cellular recruitment in the respiratory tract (Allen *et al.*, 2009; Thomas *et al.*, 2009).

ISGs form a critical response to viral infections. Upon viral infection, a cascade of sensors, adaptors, signal transducers and transcription factors lead to IFN type I and III expression and secretion in the infected cell. These act on neighbouring cells by activating the JAK/STAT pathway, resulting in transcriptional induction of ISGs encoding antiviral effectors (Schoggins, 2014).

A large number of ISGs such as IFN-inducible transmembrane (IFITM) proteins, MX proteins, and protein kinase RNA (PKR) perform different functions in preventing the spread of influenza within the host (Schoggins, 2014). IFITM proteins limit IAV replication by

preventing virus-host membranes fusing following the attachment and endocytosis of the virus (Brass *et al.*, 2009). MX proteins were some of the first ISGs to be identified with this function (Staeheli *et al.*, 1986; Haller and Kochs, 2020). PKR, a threonine/serine kinase, adheres to viral double-stranded RNA and prevents translation, thus resulting in lower viral protein and total cellular synthesis rates, and therefore effectively lowering viral replication rates (Kumar *et al.*, 1994).

Cytokines such as IL-1, IL-6 and IL-18 are a part of the innate immune system, and they protect the host from infections by promoting the response of the adaptive immune system. IL-18 stimulates the production of IFNγ by $CD8^+$ T-cells (Denton *et al.*, 2007), whereas IL-1 promotes DC migration to the mediastinal lymph nodes from the lungs (Pang *et al.*, 2013). In contrast, IL-10 is a cytokine that impairs the defence mechanisms of viruses and is produced in lungs infected by viruses. Therefore, pro-inflammatory cytokines that promote the adaptive immune system's reduction of the viral burden can also induce host resistance to viruses (Liu *et al.*, 2016).

In addition, host defence proteins play a significant role as pathogen-recognizers in the innate immune response. For example, the antimicrobial and antiviral peptide LL-37 is produced by neutrophils, macrophages and certain epithelial cells, and acts through inhibition of viral replication and reduction of virus-induced pro-inflammatory cytokine generation (Tripathi *et al.*, 2013). Another example are the surfactant proteins A (SP-A) or D (SP-D), which confer a dual role as anti-infectious agents by mediation of clearance of IAV particles from the mucosa or as immune modulators (Watson *et al.*, 2021).

1.3.3 Influenza A research models

As highlighted earlier, respiratory tract infections caused by IAV are a massive burden on human health and are one of the leading cases of deaths worldwide. Therefore, it is clear that research to fully understand viral pathogenesis and human and animal host factors involved in infection may lead to the development of more effective prophylactic or curative treatments.

Disease models are either established *in vivo* involving a whole organism for example models of human influenza infection in humans (Lambkin-Williams *et al.*, 2018; Sherman *et al.*, 2019), mice or ferrets (Radigan *et al.*, 2015; Mansell and Tate, 2018; Greaney *et al.*, 2020;

Veldhuis Kroeze *et al.*, 2021), or the more commonly used *in vitro* models as outlined further below in more detail. However, research studies using *in vivo* or *in vitro* models are limited by factors such as ethical concerns, practical challenges, or incomplete representation of the respiratory epithelium and associated immune environment in cell-based studies and therefore need to be chosen and designed to specific experimental requirements.

Additionally, when designing experimental studies of *in vivo* or *in vitro* nature, factors need to be considered such as donor availability, gender, age and life-style circumstances such as smoking, anatomic source of the material, culturing methods including supplemented factors and differentiation states of *in-vivo* cultures.

Multiple respiratory cell culture models have been developed, which allow to study airway cell type – virus interactions, and a large number are based on immortalised cell lines (Rijsbergen *et al.*, 2021; Barron *et al.*, 2021). An example is the HBEC3-KT cell line, a normal human bronchial epithelial cells immortalized with CDK4 and hTERT, which has been employed to study IAV infection susceptibility (Ramirez *et al.*, 2004). This approach provides a valuable tool if processes like replication mechanism or direct viral infection are investigated, but have their limitations in studies of infection dynamics, cellular response, and innate immune response due to the nature of a homogenous cell culture model lacking the architecture and cellularity of the respiratory tract (Rijsbergen *et al.*, 2021). Alternatively, a more advanced experimental system aiming to bridge the gap between cumbersome *in vivo* and limited *in vitro* models are lung slices. This translational *ex vivo* tool allows for studies of infection stimuli and progression or response to drug compounds (Liu *et al.*, 2019; Viana *et al.*, 2022).

More advanced *in vitro* models involve primary respiratory human airway epithelial cells (HAEC) in differentiated or undifferentiated states (Swain *et al.*, 2010; Müller *et al.*, 2013; Gowers *et al.*, 2018). These cells are obtained from various anatomical respiratory tract locations such as nasal, tracheal or bronchial origin (Forrest *et al.*, 2005) and can be obtained from resections, biopsies and brushings. Human cell samples are commercially available from multiple suppliers or can be obtained from patients (Müller *et al.*, 2013). HAECs can be cultured immediately as undifferentiated cultures, however they can only be passaged a few cycles (Lopez-Souza *et al.*, 2003; Gowers *et al.*, 2018). For functional proximal airway or alveolar models, primary cells are cultured at an air-liquid interface (ALI). These cultures are increasingly thought to represent a suitable analogue for the respiratory barrier tissues as they mimic mucus formation, mucociliary flow, and a range of physiologically relevant responses

(Huh *et al.*, 2010; Villenave *et al.*, 2013; Gard *et al.*, 2021). ALI cultures are usually established by seeding cells on top of a permeable transwell insert, with a pore size of 0.4 μm, which has a basolateral lower chamber and an apical chamber where the cells are located (Müller *et al.*, 2013). The cells grown submerged in expansion media remain undifferentiated and are not polarised, lack ciliated or goblet cells and therefore lack important characteristics of an airway model. In order to stimulate differentiation of HAECs, media is removed from the apical side, exposing cells to the air, while replenishing the basolateral side with differentiation medium containing specific growth factors. Over a period of 2-4 weeks, a polarised, a pseudostratified epithelium is formed containing a population of basal, goblet and ciliated cells (Pezzulo *et al.*, 2011; Dvorak *et al.*, 2011; Villenave *et al.*, 2013).

Besides human AECs, a well-established murine model are primary mouse tracheal epithelial cells (mTEC) (Eenjes *et al.*, 2018). These can be extracted from specific engineered transgenic mouse models and thus provide a valuable tool for functional studies. In addition to human or mouse derived cell lines, ALI cell cultures can also be generated from other species (Wu *et al.*, 2016b; Li *et al.*, 2016; Rijsbergen *et al.*, 2022) and primary ALI cultures could be co-cultured with inflammatory cells to mimic a closer host-pathogen scenario (Yonker *et al.*, 2017). However, this model still has limitations as a continuous airflow, blood flow or the complete immune cell environment is still absent (Dvorak *et al.*, 2011; Aydin *et al.*, 2021).

If the restricted life span of primary cells is an experimental limitation, stem cell derived cultures might be the better choice. They can be obtained from adult lungs by means of endobronchial biopsies or airway brushings, which are expanded in culture, embryonic tissue, or as induced pluripotent stem cells (Vazin and Freed, 2010; Butler *et al.*, 2016; Chen *et al.*, 2017). Direct co-culture of biopsy samples with 3T3-J2 feeder cells in presence of the Rho-associated protein kinase inhibitor, Y-27632, results in expansion of a greater number of basal epithelial stem cells (Butler *et al.*, 2016). Methods have also been developed to expand the number of basal cells without the use of feeder cells (Mou *et al.*, 2016; Butler *et al.*, 2016).

Whereas the adult lung source carries the least ethical restraints. Stem cell models can be used to generate 3D cultures such as organoid structures which have the potential to self-organise and renew (Chen *et al.*, 2017; Gowers *et al.*, 2018; Lehmann *et al.*, 2019; Porotto *et al.*, 2019). This model system is still less well explored but shows great potential to mimic

the human respiratory tract and can be maintained for a long period of time (Luo *et al.*, 2021; Wang *et al.*, 2022).

1.4 Bactericidal/ Permeability-Increasing Protein (BPI) protein family

As outlined previously, the innate immune system serves as the first line of defence against invading pathogens and induces immediate activation of effectors due to detection of conserved molecular patterns such as LPS present in gram-negative bacteria, or gram-positive bacteria cell wall component LTA. These act as potent inducers of the inflammatory response and are well conserved across a wide range of species. Detection of bacterial pathogens trigger pro-inflammatory responses such as TNF-α (Schultz and Weiss, 2007).

Bactericidal/permeability-increasing protein (BPI) and lipopolysaccharide binding protein (LBP), function as innate immune system LPS sensors and have evolved from a common ancestor through gene duplication (Gray *et al.*, 1993; Beamer *et al.*, 1997; Elsbach and Weiss, 1998). Cholesteryl ester transfer protein (CETP), and phospholipid transfer protein (PLTP) are lipid transfer proteins are that are related to these proteins and share conserved functional domains (Schumann *et al.*, 1994; Bruce *et al.*, 1998). CETP and PLTP are secreted by the liver and are involved in fatty acid transport and have limited function in bacterial pathogenesis (Tall, 1993; Yamashita *et al.*, 2001; Beamer, 2003). However, they all share structural features of four conserved proline and two cysteine residues, as well a C-terminal octapeptide, suggesting that these residues are important in secondary structure formation and activity (Hubacek *et al.*, 1997). In addition, the genes show a similarity of between 45 and 65% (Schumann *et al.*, 1994), and almost identical exons/intron junctions, resulting in almost identical exon sizes (Hubacek *et al.*, 1997). Except for CETP on chromosome 16, all genes are located on chromosome 20, more specifically 20q11.23 (Gray *et al.*, 1993; Hubacek *et al.*, 1997).

1.4.1 LBP and BPI protein structure and function

Human LBP shares 45% similarity with BPI, and is constitutively secreted from hepatocytes, lung and intestinal epithelial cells (Vreugdenhil *et al.*, 1999; Dentener *et al.*, 2000). *LBP* expression is stimulated during acute phase of infections through interleukins (IL-1, IL-6)

and steroids (Dexamethasone) (Schumann *et al.*, 1990; Grube *et al.*, 1994; Kirschning *et al.*, 1997).

LBP mediates acute phase immunologic responses upon gram-negative bacterial infections by binding and presenting LPS to cell surface pattern recognition receptors such as CD14 and TLR4. LBP is required for a rapid acute phase response, however it does not facilitate LPS clearance from the circulation (Muta and Takeshige, 2001; Ciesielska *et al.*, 2021). However LBP has been reported to mediate an immune response towards gram-positive pneumococci through binding of pneumococcal cell wall multimers, exhibiting a novel lipid-independent binding capability (Weber *et al.*, 2003).

Immediately upstream of *LBP* is the *BPI* gene. Both, BPI and LBP, are key mediators of the immune response to LPS, but one main difference is that BPI renders LPS non-inflammatory, whilst the LBP detects and relays the presence of low amounts of LPS to the host. Thus, these two proteins can be seen as being anti-inflammatory and pro-inflammatory respectively (Krasity *et al.*, 2011).

BPI encodes for a 55 kDa protein, which contains a 31aa N-terminal signal peptide and two conserved cysteine residues, required for disulphide bridge formation located in the LPS/lipid N-terminal domain, and overall displays a pseudo-symmetrical structure (Figure **1.7**). The N-terminal domain is essential for its biological activity and anti-angiogenic effects, which has been demonstrated to inhibit angiogenesis through specific induction of apoptosis in endothelial cells. Whereas the C-terminus is responsible for opsonophagocytosis and disposal of LPS/BPI complexes (WJ van der Schaft *et al.*, 2002; Yamagata *et al.*, 2006). Notably, the cytotoxic activity of BPI towards a multitude of Gram-negative bacteria species may be explained by a strong affinity of the very basic N-terminal half for the negatively charged lipopolysaccharides of the outer bacterial envelope (Gazzano-Santoro *et al.*, 1992).

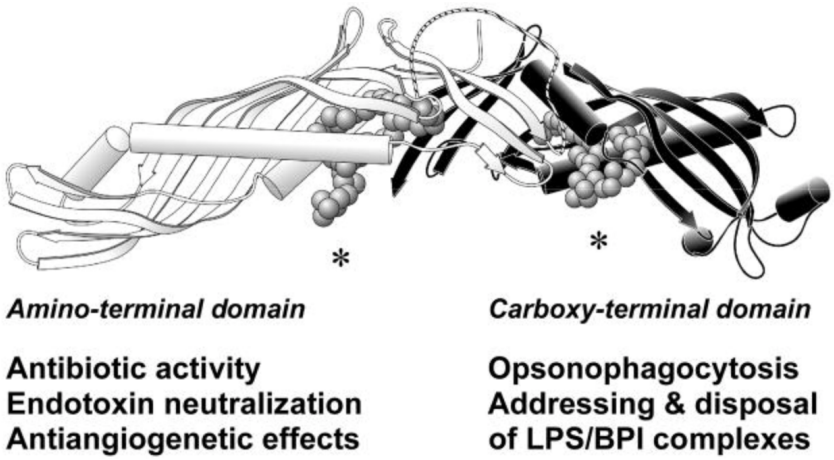


Figure 1.7: Structure and antibacterial functions of the BPI protein. BPI ribbon diagram of the structure which is bound to host phospholipids (indicated by asterisks). Alpha helices are indicated as cylinders, and beta-sheets as flat arrows. Flexible regions are depicted as lines. The N-terminus (light colour, aa 1-220) functions in endotoxin neutralisation, whereas the C-terminus (dark colour, aa 250-456) main task is to carry out antibacterial functions. Both domains are joined by a 21 aa centrally located proline-rich linker region (Schultz and Weiss, 2007). Copyright permission licence number 5510841311359.

1.4.2 The PLUNC / BPIF protein family

The PLUNC (Palate, Lung, and Nasal Epithelium Clone) protein family has been described in the past twenty five years, with members of the family predominantly being found in the respiratory, nasopharyngeal, and oral epitheliums (Bingle *et al.*, 2004). The first family member was identified in a study investigating murine palate closure in the developing mouse (Weston *et al.*, 1999). The human *PLUNC* orthologue was cloned and characterised (Bingle and Bingle, 2000; Bingle *et al.*, 2011a) and this led to the identification of seven other *PLUNC* genes, located in the same region on chromosome 20 (Bingle and Craven, 2002). These genes encode presumptive secreted proteins all containing an N-terminal signal sequences. These proteins were predicted to be structurally similar to BPI and LBP, suggesting analogous functions in host defence (Bingle and Craven, 2002; Krasity *et al.*, 2011). Based on the genomic structure and peptide length, the *PLUNC* genes were divided in two categories: “short” PLUNC (SPLUNC) and “long” PLUNC (LPLUNC) (Bingle and Craven, 2002).

Despite showing a low homology with each other, they maintain a highly conserved exon / intron structure and protein length. SPLUNC genes contain between 8-9 exons and yield about 250 aa proteins, whereas LPLUNC genes are made up of 15-16 exons and lead to 480

aa proteins (Bingle *et al.*, 2011a; Bingle *et al.*, 2011c). The shorter SPLUNC proteins are homologous to and only contain the N-terminal BPI domain, whilst the long LPLUNC proteins are homologous to both N- and C-terminal BPI domains and contain both.

PLUNC proteins have been reported to be rapidly evolving and are highly divergent, with members of this family predominantly found placental animals (Bingle *et al.*, 2004). Some PLUNC gene were identified independently generating multiple names for the same genes such as parotid secretory protein (PSP) (Bingle *et al.*, 2011b), bactericidal/permeability-increasing protein-like 1, 2 and 3 (BPIL1-3) (Mulero *et al.*, 2002), lung-specific X protein (LUNX) (Wheeler *et al.*, 2002).

As the PLUNC gene family enlarged, a systematic nomenclature was developed, which reflected the origin and ancestry of the proteins (Bingle *et al.*, 2011b). This scheme uses the root symbol BPIF for 'BPI (bacterial/permeability-increasing protein) fold-containing' instead of PLUNC. Family members with a single BPI-domain, became BPIFAs and those that have two BPI-domains became BPIFBs (Bingle *et al.*, 2011c) In this scheme SPLUNC1 became BPIFA1 and LPLUNC1 became BPIFB1. These two proteins are secreted by the respiratory epithelium at high levels, whereas the other BPIF proteins are not found to be as significantly expressed within the respiratory system (Barnes *et al.*, 2008; Bingle *et al.*, 2010).

The BPIF protein family has undergone rapid molecular change, reflecting their putative roles in host defence (Bingle *et al.*, 2004). The fact that the genes are clustered close together on a chromosome provides evidence of gene duplication; however, the relatively low homology between the paralogues suggests that the genes have mutated individually (Bingle *et al.*, 2004).

The BPIF protein family is present in all mammals (Bingle *et al.*, 2011b),. The human *BPIF* gene family is located on chromosome 20q11, and encodes eight functional proteins, whereas it is found on chromosome 2 in mice, with 13 functional members (Figure **1.8**). In rat, the gene locus is found on chromosome 3, while in cattle 14 members have been identified, located on chromosome 13 (Bingle and Bingle, 2000; LeClair *et al.*, 2001; Bingle and Craven, 2002; Bingle *et al.*, 2004; Musa *et al.*, 2012)

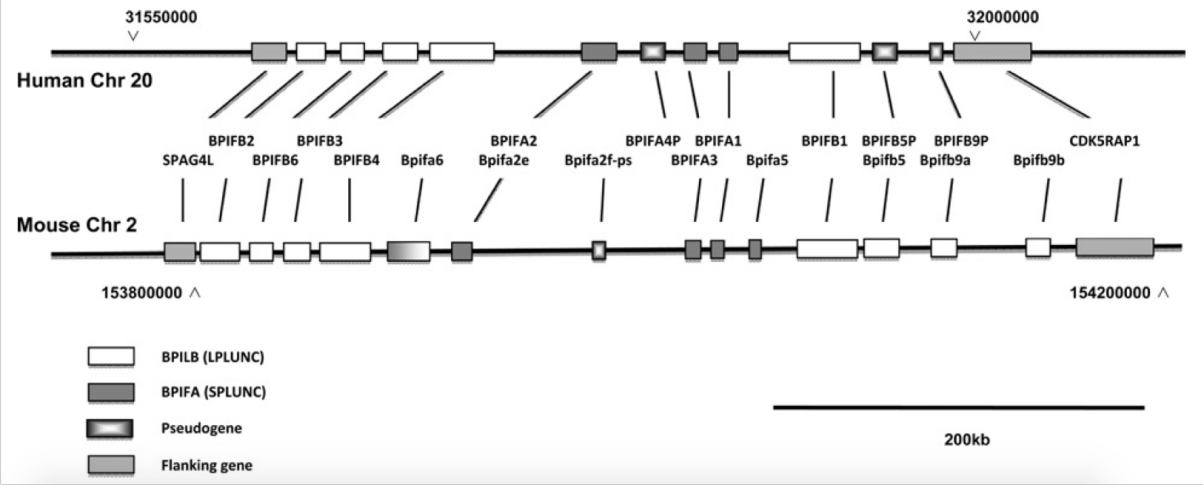


Figure 1.8: Organisation of the *BPIF* gene locus. Schematic representation of the human and mouse PLUNC loci. The upper panel represents the human locus on chromosome 20, and the lower panel depicts the mouse locus on chromosome 2, and respective positions of the locus are indicated by black arrowheads displaying nucleotide numbers. Lines between the two loci indicate orthologous relationships. *BPIFA* (*SPLUNC*) genes are illustrated in grey, while *BPIFB* (*LPLUNC*) genes are highlighted in with boxes. Pseudogenes (P or ps) are indicated by shaded boxes. Both loci are flanked by the unrelated genes *SPA4L* and *CDK5RAP1* (Bingle *et al.*, 2011a) copyright permission licence 6587841311359.

1.4.3 BPI fold containing family B member 1 (BPIFB1)

BPIFB1 was originally identified by expressed sequence tag (EST) and bioinfomatic analysis, and the complementary DNA (cDNA) was cloned from tracheal tissue (Bingle and Craven, 2002; Barnes *et al.*, 2008). Additionally, *BPIFB1* was also found as a highly expressed transcript in EST studies of nasopharyngeal epithelial genes (Zhang *et al.*, 2003), in the respiratory epithelium of CF patients (Scheetz *et al.*, 2004) and in the embryonic mouse node (Hou *et al.*, 2004).

BPIFB1 is found most abundantly in the trachea, in the submucosal glands of the larger airways, including some smaller airways, as well as the epithelium of the large airways and trachea, but it is essentially absent in epithelial cells of the peripheral lung and alveoli (Bingle *et al.*, 2010). Notably, BPIFB1 has been further found to be weakly expressed in salivary glands, the duodenum, and the stomach (Musa *et al.*, 2012; Alves *et al.*, 2017).

Interestingly, in the lung, BPIFB1 and the other main pulmonary family member BPIFA1, are found in different cell types. BPIFB1 is found in goblet cells of the respiratory system, whereas BPIFA1 is found in serous cells (Bingle *et al.*, 2010). Furthermore, BPIFB1 is also found in minor glands of the oral and nasal cavities and biological fluids like nasal secretions,

bronchoalveolar lavage (BAL), and sputum (Bingle *et al.*, 2010) (Casado *et al.*, 2005; Wu *et al.*, 2005; Nicholas *et al.*, 2006).

BPIFB1 is the most studied two-BPI domain containing member of the BPIF-family. Its distribution in the respiratory tract, in combination with the observation that *BPIFB* genes are highly expressed in the nasal regions, as well as its structural relation to other innate immune proteins, supports the hypothesis that the protein functions in the immune system, where it may regulate the removal of debris and potential pathogens from the respiratory tract.

1.4.3.1 Expression and biological functions of BPIFB1

Functional analysis of BPIFB1 to date is still very limited. The majority of studies have focused on its tissue distribution, expression levels and potential implications in lung disease, immune defence (Bingle *et al.*, 2012; Shum *et al.*, 2013; Gao *et al.*, 2015; Saferali *et al.*, 2015; De Smet *et al.*, 2018), or malignant conditions (González-Arriagada *et al.*, 2012; Wei *et al.*, 2018a; Wang *et al.*, 2019b).

It has been suggested that BPIFB1 levels are likely to reflect respiratory epithelial homeostasis. Although BPIFB1 levels have been observed to be altered in certain illnesses, the spatial expression of the protein is preserved in severe lung disease (Bingle *et al.*, 2012; Gao *et al.*, 2015; De Smet *et al.*, 2018). BPIFB1 levels are found elevated in patients with severe COPD. In these patients, epithelial cells develop into goblet cells as a result of increased irritation of the airways. Notably, BPIFB1 positively correlates with goblet cell hyperplasia, which has been found to be inversely proportional to airflow limitation, and BPIFB1 was shown to be localised to MUC5B-rich regions, suggesting that it is linked to goblet cells (Gao *et al.*, 2015). Moreover, COPD patients demonstrated higher amounts of glycosylated and secreted BPIFB1 compared to non-smokers as well non-COPD smokers. BPIFB1 levels in COPD patients who smoked were shown to correlate positively with lung function changes during a four-year follow-up (Gao *et al.*, 2015). These findings are consistent with studies establishing increased BPIFB1 levels in COPD patients and positive correlation with disease severity (Bingle *et al.*, 2012; Bingle *et al.*, 2013; De Smet *et al.*, 2018). Thus, BPIFB1 levels are likely to reflect respiratory epithelial homeostasis; they may be upregulated in COPD due to chronic infections, smoking, and increased immune responses (Ghafouri *et al.*, 2002; Titz *et al.*, 2015; Ghosh *et al.*, 2016). BPIFB1 staining was still limited to the epithelium in respiratory tract, suggesting that the exact spatial expression of

the protein is preserved in severe lung disease (De Smet *et al.*, 2018). This is in agreement with the initial studies carried out by Bingle *et al.* (2012) and Bingle *et al.* (2013).

Another lung disorder exhibiting markedly increased levels of BPIFB1 is CF, and it has been suggested that the observed increased production of BPIFB1 may be a response to the epithelial remodelling associated with this condition (Bingle *et al.*, 2012).

BPIFB1 may influence respiratory homeostasis though an association with mucin regulation within the lung. It has been reported to co-localise to MUC5B-rich regions and goblet cells (Gao *et al.*, 2015). However, BPIFB1 only demonstrates limited overlapping localisation with the other major gel-forming mucin MUC5AC (Bingle *et al.*, 2013). In line with this finding is the observation that *Bpifb1* knockout mice display an increased MUC5B expression by approximately three-fold, suggesting that the protein is associated with MUC5AC and MUC5B function (Donoghue *et al.*, 2017; Rogers *et al.*, 2022).

Another angle of BPIFB1 function might be mediated by a role as an auto-antigen. It has been reported that there is a significant link between BPIFB1 and the autoimmune disease polyglandular syndrome type 1 (APS1), an immune-cell dysfunction displaying severe endocrine gland and gastrointestinal dysfunctions (Guo *et al.*, 2018). Screening of a large cohort of APS1 samples showed the presence of autoantibodies to BPIFB1 in 9.6% of all tested cases, which represented 100% of ASP1 samples with clinical interstitial lung disease (ILD). Investigation of ILD specific samples showed a 12 or 14.6% (idiopathic vs connective tissue disease-associated, respectively) BPIFB1 autoantibody presence (Shum *et al.*, 2013). However, it could be also argued that inflammation and recruitment of antigen presenting cells may cause the appearance of antibodies to lung proteins. It is also unclear if such antiantibodies themselves are causativ.

Another potential protective function of BPIFB1 has been suggested from a murine model studying bacterial or inflammation-induced acute lung injury. In conjunction with light-elicited circadian rhythm protein Period2 (PER2), it was shown that alveolar barrier function was improved during Pseudomonas aeruginosa-induced acute lung injury. Moreover, during a genome wide mRNA screen under these conditions, it was found that BPIFB1 forms a novel downstream target of intense light-elicited alveolar type 2 (ATII)-PER2 mediated lung protection (Oyama *et al.*, 2022).

Of note, the BPIFB1 knock out model has been used in a few studies, but overall KO mice have no significant phenotypes related to immune defence or respiratory function aside the

mentioned dysregulated mucin MUC5B protein levels by 3-fold detected in lung lavage fluid (https://www.mousephenotype.org/data/genes/MGI:2137431) (Hou *et al.*, 2004; Donoghue *et al.*, 2017; Rogers *et al.*, 2022). In addition, a derived knockout cell line was used in the host lab and did not display a gross phenotype.

Additionally, BPIFB1 has also been found to be differentially expressed during infection with *Vibrio cholerae* in the gut, where it decreased proinflammatory innate immune responses to *V. cholerae* and *Escherichia coli* LPS (Shin *et al.*, 2011). This effect of BPIFB1 was dose-dependent and arose in a TLR4-dependent manner. Immunostaining demonstrated expression of BPIFB in Paneth cells in cholera patients and as well as controls (Shin *et al.*, 2011).

1.4.3.2 BPIFB1 in cancer

Inflammation has been associated with certain cancers in patients (Moore *et al.*, 2010; Bremnes *et al.*, 2011; Ullman and Itzkowitz, 2011) due to chronic infections, and recent studies have found abnormal BPIFB1 levels in a range of cancers, suggesting a potential role of BPIFB1 in tumour development or progression (González-Arriagada *et al.*, 2012; Luo *et al.*, 2018; Wei *et al.*, 2018b; Zhou *et al.*, 2019; Li *et al.*, 2020).

BPIFB1 differential expression has been described in nasopharyngeal carcinomas (NPC), where it is markedly downregulated (Zhang *et al.*, 2003). From a functional point of view, BPIFB1 inhibits STAT3 activation and suppresses the NPC cell proliferation that is induced by IL-6 (Liao *et al.*, 2014). Additionally, this protein is able to prevent NPC cells from metastasising and invading other tissues (Wei *et al.*, 2018b). Furthermore, an additional study in this context reported that co-expression of BPIFB1 inhibits vitronectin (VTN)-mediated radioresistance (Wei *et al.*, 2018a). How BPIFB1 can affect NPC tumour growth has been explored more recently. For example, a recent finding showed that BPIFB1 negatively regulates expression of the glucose transporter 1 (GLUT1), which mediates tumour neovascularization (Jiang *et al.*, 2022). This is in agreement with the previous observation that BPIFB1 expression is found strongly downregulated NPC tumour tissues (Zhou *et al.*, 2008). Furthermore, it was found that overexpression of BPIFB1 induced apoptosis and DNA damage in a NPC derived cell line, whereas silencing of BPIFB1 showed the reverse effect. It was shown that BPIFB1 overexpression generates a G0/G1 cell cycle arrest through regulation of the MEK/ERK signalling pathway, in addition to an augmented anti-

proliferative effect of chemotherapy drugs (Xu *et al.*, 2019). Therefore, BPIFB1 has been suggested to be a potential treatment target for NPC (Xu *et al.*, 2019).

BPIF protein expression patterns have been suggested to have the potential to be used as an auxiliary tool for the diagnosis of high-grade mucoepidermoid carcinoma of the salivary gland (González-Arriagada *et al.*, 2012). This particular cancerous tissue showed a positive staining for BPIFA1 and BPIFB1 in respective 90% or 93% of cases, mainly in mucous cells, mucin plugs, and intermediate cells across all investigated grades (González-Arriagada *et al.*, 2012). To date there have been no functional studies of BPIF proteins in salivary gland cancers.

1.4.3.3 BPIFB1 protein structure

As previously mentioned, BPIFB1 contains two BPI domains and shares structural features of the BPI protein family (Bingle *et al.*, 2011b). The structure of BPIFB1 has not been solved. However, using human BPIFB1 sequence with the Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) online tool (Kelley *et al.*, 2015). (http://www.sbg.bio.ic.ac.uk), it is possible to generate a predicted structure, which highlights the two distinct BPI domain containing folds, illustrated as N- and C-terminus (Figure **1.9**). The Phyre engine uses a library of known protein structures taken from the Structural Classification of Proteins (SCOP) and the Protein Data Bank (PDB) database. It scans a user submitted query against a non-redundant sequence database and constructs a profile which is deposited in the 'fold library' and allows for prediction of its secondary structure. This profile and constructed secondary structure are then scanned against the 'fold library', which returns an alignment score. The top 10 scoring alignments are then used to construct full three-dimensional models of the query (Murzin *et al.*, 1995; Berman *et al.*, 2000).

The secondary structure elements are highlighted in red and purples for alpha helices and beta-sheets respectively. The conserved disulphide bond is illustrated in green and is located in the N-terminal part of the protein (Figure **1.9**). Both domains form hydrophobic pockets contained by two alpha helices, which would allow for lipid binding similar to what has been reported for BPI (Beamer *et al.*, 1997).

The universal protein resource (UniProt) database allows for visualisation of the amino acid (aa) position of the main protein features and permits comparison between human and other

species orthologues such as mouse (Figure **1.10**). The human BPIFB1 protein is a 484 aa long protein with an N-terminal signal peptide, which directs the protein for secretion, a disulphide bridge between amino acid 158 and 201, and three predicted N-glycosylation sites at aa48, aa264 and aa401 (Figure **1.10**A).

Figure 1.9: Predicted BPIFB1 protein structure. Human BPIFB1 was modelled using the Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) online tool (Kelley *et al.*, 2015) (http://www.sbg.bio.ic.ac.uk). The model illustrates the N- and C-terminal BPI domains containing regions, a linker region, as well alpha helices in red and a beta-sheets in purple. The conserved disulphide bound is shown in green.

These features and positions are largely conserved in the murine orthologue, with the exception that there are four presumptive N-glycosylation sites, compared to three in hBPIFB1 with only two of these being conserved (aa264 and aa401 in hBPIFB1, and aa263 and aa400 for Bpifb1) (Figure **1.10**B).

A – BPIFB1 (human)

Feature key	Position(s)	Description	Graphical view	Length
Signal peptide	1 – 21	Sequence analysis		21
Chain (PRO_0000017180)	22 – 484	BPI fold-containing family B member 1		463
Glycosylation	48	N-linked (GlcNAc...) asparagine Sequence analysis		1
Disulfide bond	158 ↔ 201	By similarity		
Glycosylation	264	N-linked (GlcNAc...) asparagine Sequence analysis		1
Glycosylation	401	N-linked (GlcNAc...) asparagine Sequence analysis		1

B – Bpifb1 (mouse)

Feature key	Position(s)	Description	Graphical view	Length
Signal peptide	1 – 21	Sequence analysis		21
Chain (PRO_0000017181)	22 – 474	BPI fold-containing family B member 1		453
Glycosylation	153	N-linked (GlcNAc...) asparagine Sequence analysis		1
Disulfide bond	157 ↔ 200	By similarity		
Glycosylation	160	N-linked (GlcNAc...) asparagine Sequence analysis		1
Glycosylation	263	N-linked (GlcNAc...) asparagine Sequence analysis		1
Glycosylation	400	N-linked (GlcNAc...) asparagine Sequence analysis		1

Figure 1.10: Comparison of BPIFB1 structural features between human and mouse. Human BPIFB1 (Q8TDL5) **(A)** and murine Bpifb1 (Q61114) **(B)** features have been visualised using the universal protein resource (UniProt) database, highlighting the amino acid positions for protein features such as the N-terminal signal peptide, disulphide bond formation and N-glycosylation positions (https://www.uniprot.org/).

There is biological evidence for the existence of a range of glycosylated forms of BPIFB1 from BAL, sputum, saliva, nasal secretions and differentiated tracheobronchial epithelial (TBE) cells cultured at the air-liquid interface (ALI) (Casado *et al.*, 2005; Wu *et al.*, 2005; Nicholas *et al.*, 2006; Bingle *et al.*, 2010; Bingle *et al.*, 2012; Musa *et al.*, 2012). Enzymatic assays assessing BPIFB1 protein migration patterns before and after cleavage of N-glycans demonstrated altered molecular weights, thus suggesting that N-glycosylation modifications were present (Figure **1.11**). However, no conclusion could be drawn concerning which sites have been post-translationally modified. Also, no studies have investigated if BPIFB1 N-glycosylation is required for biological function (Bingle *et al.*, 2010).

ALI HBECs secrete more BPIFB1 as they undergo differentiation; in healthy subjects, proteolytically cleaved products have not been found in the airway-lining fluid or TBE secretions (Bingle *et al.*, 2010). BPIFB1 was identified as a major secretory product of TBE cells in acute lung injury models; its secretion was shown to decrease with IL-4 treatment (Candiano *et al.*, 2007). These findings support the view that BPIFB1 functions within the milieu that protects the host's local mucosal surfaces (Bingle *et al.*, 2010).

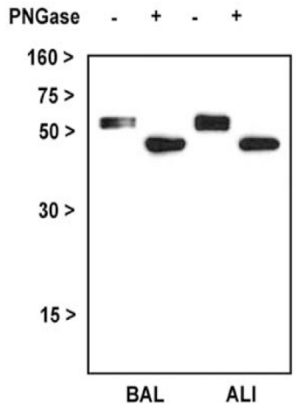


Figure 1.11: Human BPIFB1 protein is N-glycosylated. Protein samples from human bronchoalveolar lavage (BAL), and apical washes from primary airway cells at the air-liquid interface (ALI) culture, were digested with PNGase (+) to enzymatically remove N-linked glycosylation sites. Samples were run on a western blot and probed with an antibody against hBPIFB1. Following treatment with PNGase a smaller sized product can be seen due to the loss of N-glycosylation, resulting in an altered molecular weight and associated migration pattern (Bingle et al., 2010). This article is distributed under a Creative Commons license.

The best characterised BPIF protein to date is BPIFA1, but recently more effort is made to extend research studies to BPIFB1, which is also constitutively expressed across most of the respiratory epithelium. However, it is still not entirely clear if BPIFA1 and BPIFB1 carry out redundant, compensatory roles and are expressed in a more mutually exclusive manner, especially in an abnormal lung health conditions (Musa *et al.*, 2012; Bingle *et al.*, 2012; De Smet *et al.*, 2018; Vieira Braga *et al.*, 2019; Saferali *et al.*, 2020).

1.5 Aims and objectives

A number of recent studies have found that BPIFB1 is a significant contributor to host immunity within the respiratory system. However, no specific role for BPIFB1 has been identified. Our central hypothesis is that BPIFB1 may be involved in the homeostasis of the immune system within the airways.

There is limited functional data supporting specific biological roles for BPIFB1. As highlighted above, there is more data on the function of BPIFA1. Our lab has shown that BPIFA1 may function as regulator of IAV infection (Britto and Cohn, 2015; Akram *et al.*, 2018). Furthermore, our lab has also reported that *BPIFB1* and *BPIFA1* levels are modulated in the lungs of mice challenged with MHV-68 a δ-herpes virus (Leeming *et al.*, 2015). To

begin to address the potential antiviral function of BPIFB1 in preliminary studies the lab challenged *Bpifb1*$^{-/-}$ mice with both viruses. Data from these experiments suggested that *Bpifb1*$^{-/-}$ mice have higher viral loads when challenged with either IAV or MHV68 δ-herpes virus (Leeming *et al.*, 2015). Due to the structural similarity between these two proteins and their sites of expression in the respiratory tract, this project was designed to determine whether BPIFB1 influences the rate of infection and disease progression, following infection with IAV. Specifically, this thesis was designed to test the hypothesis that BPIFB1 functions as a modulator of respiratory tract homeostasis and provides a protective function against infection with IAV.

To achieve this aim, the objectives of the study are:

(i) Extend the knowledge of BPIFB1 and its potential function using available online bioinformatic tools with regards to gene homology and evolution, protein structure and published single-cell RNA sequence (scRNA-Seq) data in health and disease conditions

(ii) To use cell and microarray data-based validation of the ALI culturing model of mTECs as potential experimental system for IAV infection studies

(iii) Generation of murine BPIFB1 constructs in the VR1255 mammalian expression vector and validation of expression of recombinant BPIFB1

(iv) Generation and functional validation of a BPIFB1 knock out cell line using a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) approach to establish a biological assay for IAV infection

(v) Assessment of a protective role of BPIFB1 during IAV infection

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Chemical compounds and reagents

If not specified otherwise, compounds or reagents were purchased from Sigma-Aldrich.

2.1.2 Bacterial media and agar selection plates

Luria Bertani (LB) was prepared using a ready to dissolve mix containing 10 g/L Bacto-Tryptone (Becton, Dickinson and Company), 5 g/L Bacto-yeast extract (Becton, Dickinson and Company), 10 g/L NaCl and 1 g/L glucose. LB selection media was prepared with either 100 µg/mL Ampicillin or 50 µg/mL Kanamycin (Gibco). LB agar selection plates were prepared by using LB media and supplementing it with 15 g/L of agar. After autoclaving, LB agar was prepared supplemented with appropriate antibiotics prior pouring, as stated above.

2.1.3 Antibodies

Primary antibody dilutions were prepared as outlined in Table *2.1* in 5% milk/TBST supplemented 0.02% (w/v) with sodium azide. Secondary antibody dilutions were prepared as primary antibodies as outlined in Table *2.1*, without NaN_3 supplementation.

Table 2.1: Primary and secondary antibody list including dilution factors and species.

antibody	clonal	host	dilution	supplier
Primary				
Bpifb1	polyclonal	rabbit	1:200	in house
Bpifa1	polyclonal	rabbit	1:200	in house
Anti-Flag M2	monoclonal	mouse	1:1000	Sigma-Aldrich
BPIFB1	polyclonal	rabbit	1:200	in house
OLMF4 (D1E4M)	polyclonal	rabbit	1:1000	CellSignalling
Secondary				
anti-mouse HRP	polyclonal	goat	1:2000	Dako
anti-rabbit HRP	polyclonal	goat	1:2000	Dako

2.1.4 Plasmids

The vectors used in this study were the cloning vector pCRII-TOPO (Invitrogen, Appendix I), the expression vector pVR1255 (kindly provided by Professor James Stewart from the Institute of Infection and Global Health, University of Liverpool, Appendix II), pBSK(+) Simple mBPIFB1 (Biomatik, Appendix III) containing the four N-glycosylation mutations, abrogating N-glycosylation (Appendix IV, pBluescript SK (Biomatik) for generation of BPIFB1 N- or C-terminal truncations Appendix V and VI respectively), the CRISPR targeting plasmid pSpCas9(BB)-2A-GFP (PX458) (Addgene, Plasmid #48138) (Appendix VII) was a gift from Feng Zhang (Addgene plasmid # 48138 ; http://n2t.net/addgene:48138 ; RRID:Addgene_48138) (Ran *et al.*, 2013c) and the eGFP expression plasmid pEGFP-N1 (Clontech) (Appendix VIII).

2.1.5 Primers

Gene expression analysis of selected genes was carried out by amplifying equal amounts of cDNA by endpoint polymerase chain reaction (PCR) and *Oaz1* served as an internal positive control. Primers were generated by Sigma, received lyophilized and were resuspended using sterile water, at a final stock primer concentration of 1 µg/µl, before further dilution to 0.1 µg/µl, which served as PCR working stock concentrations (*Table 2.4*). All primer pairs used for RT-PCR crossed intron/exon boundaries.

2.1.6 Buffers

2X SDS sample buffer: 100 mM Tris-Cl (pH 6.8), 2% (w/v) sodium dodecyl sulfate (SDS), 0.2% (w/v) bromophenol blue, 20% glycerol, 100 mM dithiothreitol (DTT), protease inhibitor tablet (Roche).

DNA loading buffer (10x): 0.00125 g/mL Bromophenol Blue, 40% sucrose, 60 % glycerol, 0.1M EDTA.

Phosphate buffered saline (PBS) buffer: pH 7.4; 137 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, 1.76 mM KH_2PO_4.

PI/Triton X-100 staining solution: to 10 ml of 0. 1 % (v/v) Triton X-100 in PBS add 2 mg DNAse-free RNAse A and 0.4 ml of 500 µg/ml PI (Roche, 11348639001).

Running buffer: 25 mM Tris, 192 mM Glycine, 1% SDS

Tris-buffered saline (TBS): 10 mM Tris, 150 mM NaCl

Tween supplemented TBS (TBST): 0.1% Tween in TBS.

Transfer buffer: 10 % (w/v) Methanol, 25 mM Tris, 192 mM Glycine, 1% SDS

2.1.7 Influenza A viruses

The mouse-adapted A/PR/8/34 (PR8, H1N1) Influenza virus strain, which contains NS segments of PR8 fused with an eGFP fluorescent reporter gene was generated in Madin-Darby Canine Kidney Cells (MDCK) cells. The virus was produced by Yoshi Kawaoka (Fukuyama *et al.*, 2015). Purified viral aliquots were provided by Professor James Stewart from the Institute of Infection and Global Health, University of Liverpool.

2.2 DNA Techniques

2.2.1 Cloning strategy for murine BPIFB1

Amplification of full-length murine BPIFB1 was performed using cDNA prepared from RNA extracted from mTEC ALI cell cultures, isolated from the CD-68 BL6 mouse strain. PCR amplification reactions were performed using the primer pair mB1F1NotI/mB1RSTOP (Table 2.2), which introduced the following sequences into the PCR product: Forward primer: *Not1* site; Kozak sequence; ATG start codon; Reverse primer: BamH1 site; STOP codon and a FLAG tag sequence. PCR amplification was performed in a final volume of 20 µl as outlined below (Appendix IX).

Table 2.2: Primer sequences used to clone the murine BPIFB1 gene. Red nucleotides represent the NotI restriction enzyme sequence; grey indicates the Kozak sequence; green marks the ATG start codon; Blue represents the BamHI restriction sequence; orange represents the STOP codon and purple marks the FLAG-tag sequence.

Primer name	Gene	Primer Sequence 5'→3'
mB1F1NotI	mBPIFB1	ATGCGGCCGCCGCCGCCACCATGGCCGGCCCGTGGATTAT
mB1RSTOP		GCGGATCCTTACTTGTCATCGTCGTCCTTGTAGTCGGAGGAGGCTGGAGTGAGCTTGA

2.2.2 Polymerase chain reaction (PCR)

PCR amplification reactions were performed in a final volume of 20 µl, by combining the following reagents: 1 µl cDNA, 1 µl forward primer, 1 µl reverse primer, 7 µl nuclease-free water and 10 µl of Maxima Hot Start Green PCR Master mix (2X) (Thermo Scientific). The negative control contained the regular reaction mix but only 1 µl of Nuclease-free water instead of cDNA template. PCR reactions were carried out in a thermocycler (DNA Engine), using the following cycle conditions. An initial denaturation step of 95°C for 4 min, followed by 35 cycles of 95°C denaturation for 1 min; 60°C annealing for 1 min and 72°C extension for 1 min. After completion of the 35 cycles, the PCR products were extended using a final extension step of 72°C for 7 min, and the PCR products were run on agarose gels alongside a 1 kb DNA hyperladder (Bioline) to determine the size of the PCR products.

2.2.3 DNA gel electrophoresis

PCR amplified products were separated and visualised by gel electrophoresis. 10 µl of each reaction was run on a 1% (w/v) agarose gel containing ethidium bromide (0.5 µg/ml), in Tris-acetate-EDTA (TAE buffer) (Fisher-Bio-Reagents). Agarose gels were prepared by dissolving 1g of agarose in 100 ml of TAE buffer using a microwave (medium power). The agarose was then cooled in a fume hood to around 55ºC and ethidium bromide solution added to a final concentration of 0.5 µg/ml, subsequently poured into a gel casting tray and allowed to set at room temperature (RT) for around 45 min. For electrophoresis, gels were run at 100 volts (V) for 30 min, using the Bio-Rad gel electrophoresis tank containing TAE running buffer. After completion, gels were examined under UV light in a Bio-Rad Chemi-Doc^{TM} XRS^{+} system. Digital images were taken, and the size of the PCR products were determined by comparison to the 1 kb DNA hyperladder (Bioline).

2.2.4 Transformation of bacteria

Cloning reactions were transformed into competent bacterial cells for amplification and subsequent plasmid extraction. 3 µl of the reaction mixture was combined with 17 µl of competent *E. coli* cells (TOP10 One Shot cells) (Invitrogen), gently mixed and incubated for 30 min on ice. The plasmid and competent *E. coli* cell mixture was subject to heat-shock for 45 sec at 42°C, and immediately placed on ice for 5 minutes. 100 µl of Super Optimal broth with catabolite repression (SOC) medium (Invitrogen) was gently added to the cells and incubated for 1 hr at 37°C while shaking. A 100 µl aliquot of the sample was plated onto a LB agar plate, containing 50µg/ml Ampicillin. The plate was incubated overnight at 37°C and transformed colonies were picked and re-inoculated into 2 of ml Lysogeny broth (LB) medium, containing 50 µg/ml Ampicillin, and cultured overnight in a shaking incubator at 37°C, and 225 rpm.

2.2.5 Preparation and purification of plasmid DNA

Plasmids were amplified in *E. coli* bacteria, and recombinant constructs extracted using the Isolate II Plasmid Mini kit (Bioline) for Minipreparations, and Maxipreps were performed using Qiagen Plasmid Maxiprep kit and protocol according to the manufacturer's instructions. The eluted plasmid DNA yield was quantified using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific) and stored at -20°C for further use.

2.2.6 Restriction endonuclease reactions

Plasmid DNA was digested with restriction endonucleases (Promega) in a 10 µl final reaction volume. Generally, 2ug of plasmid DNA was combined with 1µL of restriction endonuclease and 1µL of 10x reaction buffer (Promega). Reactions were incubated overnight at 37ºC prior agarose gel resolution alongside the 1 kb DNA hyperladder (Bioline) on 1% or 2% agarose gels containing ethidium bromide.

2.2.7 DNA extraction from agarose gels and ligation reactions

DNA samples which were resolved by agarose gel electrophoresis post endonuclease digestion for subcloning, were visualised on a UV transilluminator. DNA fragments of interest were isolated from the agarose gel with a scalpel and recovered using the GeneJET Gel Extraction and DNA Cleanup Micro Kit (Thermo Scientific). Concentrations of extracted DNA products were determined using the NanoDrop ND-1000 (ThermoScientific). For subcloning purposes, ligation reactions of digested plasmid DNA were performed in 10 µl reaction volumes. Respective DNA was incubated for 18 hours at 15ºC in presence of the destination vector / or cloning fragment, 1 µl T4 DNA ligase 10x buffer (Promega) and 1 µl T4 DNA ligase (Promega).

2.2.8 DNA sequencing

Plasmid DNA and recombinant constructs were sequenced to confirm that inserts were generated correctly and did not show any sequence alteration. For sequencing of pCRII-TOPO vectors, 10 µl of the recombinant plasmid DNA construct (100 ng/µl), together with 1pmol/ µl of both the Sp6 reverse, and T7 forward primer (*Table 2.3*), were submitted to the University of Sheffield, Core Genomics facility and sequenced using the BIgDye v3.1 chain termination kit in an Applied Biosystems systems 3730 DNA sequence analyser (www.genetics.group.shef.ac.uk). DNA sequences were analysed using the FinchTV sequencing programme downloaded from: (https://digitalworldbiology.com/FinchTV). Plasmids with the correct cloned sequence were used in the subsequent sub-cloning procedures as well as served as templates for PCR.

Plasmid DNA of VR1255 constructs were extracted using a Midi-prep kit and was sequenced at the University of Sheffield Core Genomic facility by using a small aliquot of plasmid DNA (10 µl at a concentration of 100 ng/ µl), submitted together with the relevant forward and reverse primer sequences as outlined in *Table 2.3*. The DNA sequence chromatogram was analysed using the FinchTV DNA analysis software.

Table 2.3: Sequencing primers

Plasmid DNA	Forward primer (1pmol/ µl)	Reverse primer (1pmol/ µl)
pCRII-TOPO	5'-TAATACGACTCACTATAGGG-3' (T7)	5'-ATTTAGGTGACACTATAG-3' (Sp6)
pVR1255	5´-AATAGCTGACAGACTAACAGACTG-3´	5´-GAGTGAGCTGATACCGCTC-3´

2.3 RNA Techniques

2.3.1 RNA isolation

RNA was isolated from cultured cells using the Direct-zol RNA MicroPrep (Zymo Research, R2062) according to the manufacturer's instructions. Briefly, cells were treated with TRI-reagent and the lysate was transferred to a spin column, followed by an in-column DNase I digestion step, to prevent any contamination from genomic DNA (gDNA). After buffer washes, the RNA was eluted from the column using nuclease-free water. RNA concentrations were determined using the NanoDrop ND-1000 (ThermoScientific), and samples were stored at -20ºC.

2.3.2 RNA quantification

The assessment of the extracted RNA quality was performed using 1µL in a Nanodrop ND-1000 instrument (ThermoScientific), which determined quality and concentration of the RNA. The accepted value for the 260nm:280nm ratio was greater than 1.7 for all samples used.

2.3.3 cDNA synthesis

Genomic DNA free RNA was reverse transcribed into cDNA using the AMV Reverse Transcriptase kit (Promega). A reaction mix was generated by adding 0.5 µl OligoDT and 0.5 µl of random primers (Promega) to the DNAse treated RNAse sample in a final reaction volume of 19 µl using nuclease free water to top up. Samples were subsequently heated at 70℃ for 5 minutes and 6 µl mastermix containing 5 µl 5x AMV buffer, 0.5 µl 100uM dNTPs, 0.25 µl AMV RT enzyme and 0.25µl RNasin ribonuclease inhibitor was added to RNA samples before placing in a thermal cycler for cDNA synthesis. The thermal cycler settings were set to 42℃ for 1 hour, followed by 95℃ for 5 minutes. Samples were used either immediately for RT-PCR or stored long-term at -20℃. Alternatively, the FastGene Scriptase II ready mix (NIPPON Genetics LS64) was employed for later experiments according to the instructions of the manufacturer.

2.3.4 RT-PCR reactions

To quantify and detect gene expression end-point real-time polymerase chain reaction (RT-PCR) was conducted. The gene *Oaz1* was used in all RT-PCR reactions as an unrelated control and reference. Maxima Hot Start Green PCR Master Mix (ThermoFisher Scientific) was used to amplify genes using 1 µl template cDNA. Each reaction contained in addition 12.5 µl of Maxima Hot Start Green PCR Master Mix, 4.5 µl of nuclease free water, and 1 µl each of forward and reverse primers. *Table 2.4* details the RT-PCR primer pair information. PCR reactions were performed in a thermocycler using an initial enzyme activation step for 95ºC for 4 minutes, followed by the cycle block of a denaturation step at 95ºC for 1 minute, annealing at 60ºC for 1 minute and elongation at 72ºC for 1 minute with a final elongation stage at 72ºC for 7 minutes. PCR reactions for all genes were run for 35 cycles with the exception of *Bpifa1*, which was amplified for 30 cycles. PCR products were separated by gel electrophoresis and visualised using a Biorad ChemiDoc TM XRS+.

Table 2.4: Primer sequences for endpoint PCR reactions

Construct	**Size [bp]**	**Forward primer** (5′ -> 3′)	**Reverse primer** (3′ -> 5′)
Bpifb1	148	CCCTGACCAAGATCCTTGAA	GAGGCTGGAGTGAGCTTGAG
Bpifa1	127	GGTGCACAACATTGCTGAAT	CAAGAGGCAGGAGACTGAG
Tekt1	373	CAGTGCGAAGTGGTAGACG	TTCACCTGGATTTCCTCCTG
Oaz1	274	ACAGAGGAGCCGACGTCTAA	CCAAGAAAGCTGAAGGTTC
Ifnl2	200	GACAAGAACCCAAGCTGACC	ACCTCAGGTCCTTCTCAAGC
BPIFB1	207	CCCTGCCCAATCTAGTGAAA	TCACCTTCCCCTGTGAGTCG
BPIFA1	466	ATGCCCTCAGCAATGGCCTGC	GAGAGGCTGTCCAGAAGACC
TEKT1	198	TGGAGCTGTGTCGTCATGTC	GCATACACAGCACTTCGTCG
OAZ1	164	CCAACGACAAGACGAGGATT	AGCGAACTCCAGGAGAACTG
IP-10	190	CCAATTTTGTCCACGTGTTA	CCTCTGTGTGGTCCATCCTT

2.4 Cell Culture

2.4.1 Cell lines, culture conditions and media

Human embryonic kidney 293 (HEK293) cells were used for initial transfection studies of mBPIFB1 expression constructs and initial protein expression trials. This cell line was obtained from ATCC and maintained in DMEM media (Invitrogen) supplemented with 10% FBS (Invitrogen). Cells were passaged every 2-3 days as a 1:10 split.

Human bronchial epithelial cells (HBEC3-KT) immortalized with cyclin-dependent kinase 4 (CDK4) and human telomerase (hTERT) (Reddle, 1988; Lundberg *et al.*, 2002; Nakauchi *et al.*, 2019) were employed as a model for IAV infection studies. While HBECs are immortalized, they do not display cell phenotypes such as disruption of the p53 pathway, extensive copy number changes, lack of contact-inhibition or anchorage-independent growth (Ramirez *et al.*, 2004; Sato *et al.*, 2006). Most importantly, immortalized HBECs retain the ability to differentiate into structures found in the normal bronchial epithelium (Vaughan *et al.*, 2006; Delgado *et al.*, 2011). HBEC3-KT were maintained in keratinocyte serum-free media (KSFM) supplemented with 50 ng/ml human recombinant epidermal growth factor 1-53 (EGF 1-53), 5 µg/L bovine pituitary extract (BPE) and 100 U/ml penicillin (*Table 2.5*). Cells were maintained in a humidified environment at 37 ^{0} C with 5% CO_2. Confluent cells in T25 flasks were trypsinised for 3 minutes at 37 ^{0} C followed by neutralization of trypsin using soybean trypsin inhibitor. Detached cells were spun at 250 x g for 5 minutes and the supernatant was discarded. Cells were seeded at either 10^6 cells for a T25 flask, 6×10^4 cell in a 24 well format, or 3×10^4 cells in each transwell (Falcon 353095) for ALI culture in complete media. For ALI culture, cells were media supplied only from the basolateral side of the transwell with the PneumaCult-ALI complete media (StemCell) as outlined in *Table 2.5*. The medium was changed and replaced every second day, and the apical side of the ALI culture surface was washed with Hank's Balanced Salt Solution (HBSS) (Gibco). ALI cultures were differentiated for up to 21 days. To note, PneumaCult-ExPlus medium is a defined, serum- and BPE-free cell culture medium that supports more expansion of primary human airway epithelial cells at each passage, compared to other commercially available expansion media. This medium has been reported by the manufacturer to support at least two additional passages of cell expansion with better differentiation potential, defined as the ability to form a pseudostratified mucociliary epithelium at the ALI.

Table 2.5: HBEC3-KT media components and supplier information.

KSFM expansion medium		
Component	**Supplier**	**Final concentration**
KSFM	Gibco (17005034)	
Penicillin	Lonza (DE17-602e)	100 U/ml
EGF (1-53)	Sigma (E9644)	50 ng/ml
Bovine Pituitary Extract	Gibco (13028014)	5 ng/ml
PneumaCult™-ExPlus		
Component	**Supplier**	**Final concentration**
ExPlus basal medium	STEMCELL (05041)	
50x supplement	STEMCELL (05042)	1x
Hydrocortisone	STEMCELL (07925)	96 ng/mL
PneumaCult™-ALI complete media		
Component	**Supplier**	**Final concentration**
ALI basal medium	STEMCELL (05002)	
Supplement	STEMCELL (05003)	10x
Maintenance supplement	STEMCELL (05006)	100x

2.4.2 mTEC extraction and culture conditions

Primary mouse tracheal epithelial cells (mTEC) were extracted from tracheal epithelia derived from C57BL/6 mice in sterile conditions. Cells were isolated and cultured based on well-established protocols (You *et al.*, 2002; You and Brody, 2012), which have been further modified in our lab (Mulay, Akram et al. 2019). Briefly, resected tracheas were cleaned of adhered tissue, opened longitudinally and placed in sterile mTEC basic media containing antibiotics (*Table 2.6*). Tracheas were incubated overnight in 0.15% pronase solution (Roche) in mTEC basic media at 4ºC. Subsequently, pronase activity was neutralised with 10% FBS final concentration, and epithelial cells were gently dislodged from tracheas by gentle tube inversions. The pooled cell pellet collected by centrifugation was resuspended in cold 0.5 mg/ml pancreatic DNase I and 1 mg/ml bovine serum albumin (BSA) solution and incubated in a cell culture dish for 3-4 hours at 37°C, to allow the fibroblast population to attach to the surface. The non-adherent cells, enriched for epithelium, were collected, and resuspended in mTEC plus media (*Table 2.6*) and a small population was used for RNA extraction, serving as an original or wt cell reference. The remainder of the isolated epithelial cell population was seeded at the density of 3-3.5x10^4 cells per transwell (Falcon) insert.

Table 2.6: mTEC media and supplier information

Medium	Component	Supplier	Final concentration
mTEC extraction medium			
mTEC basic			
	DMEM/F12	Gibco (31330038)	conaining
		containing HEPES, L-Glutamine	
	Penicillin	ThermoFisher (15140122)	100 U/ml
	Streptomycin	ThermoFisher (15140122)	100 µg/ml
mTEC proliferation medium			
mTEC plus media			
	MTEC basic		
	Fetal Calf Serum	ThermoFisher (A5256701)	5%
	Insulin	Sigma (I3536)	10 µg/ml
	Transferrin	Sigma (T8158)	5 µg/ml
	Cholera Toxin	Sigma (C8052-0.5mg)	0.1 µg/ml
	Murine EGF	BD Biosciences (734-1304)	0.025 µg/ml
	Bovine Pituitary Extract	Gibco (13028014)	0.03 mg/ml
	Y-27632 (add fresh)	Tocris Bioscience (1254)	10 µM
	Retinoic Acid	Sigma (R2625-50mg)	0.01 µM
mTEC differentiation medium			
mTEC ALI media			
	MTEC basic		
	BSA	Sigma (A9418)	1mg/ml
	Insulin	Sigma (I3536)	5 µg/ml
	Transferrin	Sigma (T8158)	5 µg/ml
	Cholera Toxin	Sigma (C8052-0.5mg)	0.025 µg/ml
	Murine EGF	BD Biosciences (734-1304)	0.005 µg/ml
	Bovine Pituitary Extract	Gibco (13028014)	0.03 mg/ml
	Retinoic Acid	Sigma (R2625-50mg)	0.01 µM

2.4.3 Air Liquid Interface (ALI) culturing conditions

Extracted and expanded mTEC cells were seeded into transwells at a density of 3-3.5x10^4 and were cultured in submerged conditions using mTEC plus media in both chambers until they reached confluency, which usually took seven to eight days (*Table 2.6*). When confluent, this day was termed Day 0 (D0) and mTECs were differentiated into upper airway-like epithelium in ALI conditions for at least 14 days (D14) using mTEC ALI media in the basal chamber only (You *et al.*, 2002; You and Brody, 2012) (Figure 2.1 and Figure **4.1**). Media in the basal chamber was exchanged every two days and the apical cell surface washed with HBSS.

HBEC3-KT cells were similarly differentiated as outlined for mTEC cells (Figure 2.1). 3 × 10^4 cells were seeded in each transwell of a 24-well plate (Falcon 353504) in a submerged fashion using PneumaCult-ExPlus media (*Table 2.5*). After reaching confluency, cultures were subjected to ALI culture (D0) by applying PneumaCult-ALI complete media in the

lower basal chamber and exposing the apical surface to air for 14-21 days. Basal chamber media was changed every two days and apical cell surfaces washed with HBSS.

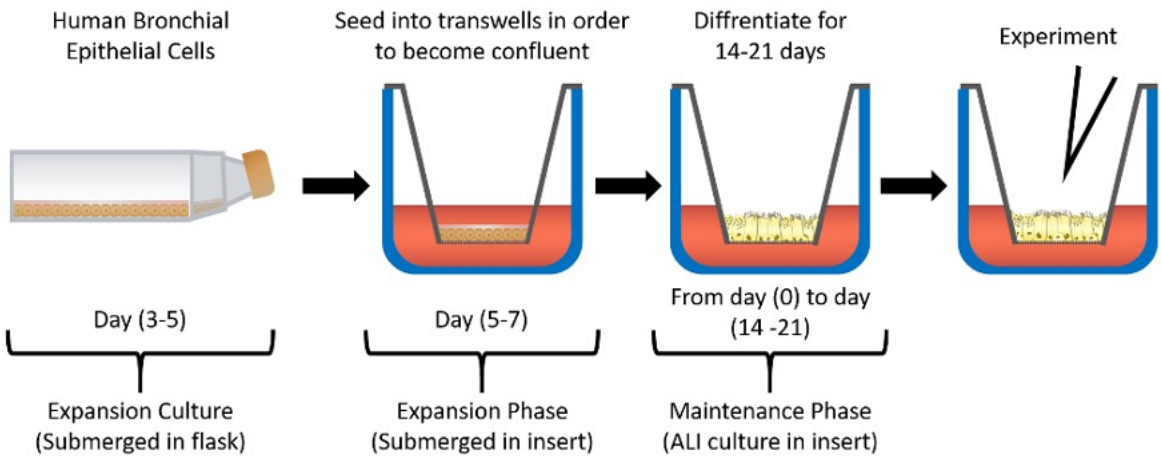


Figure 2.1: Schematic of ALI culturing and differentiation timeline. Human bronchial epithelial cells such as HBEC-KT were expanded in flasks and cultured in a submerged phase until reaching confluency. Cells were seeded into transwell inserts on the apical site and cultured until confluent. Differentiation into pseudo-stratified epithelium was initiated by elimination of the apical medium and replacement of the basal medium with differentiation medium, for consecutive 14-21 days. The image was provided by Zeyad Alharbi (unpublished).

2.4.4 Thawing and freezing of mammalian cells

Mammalian cell cultures were prepared for long term storage by trypsinisation, pelleted at 250x g for 5 minutes, and the collected cell pellet resuspend in 1 ml of FBS supplemented with 10% (v/v) DMSO, and immediately transferred to cryovials. Cells were frozen at -80°C overnight and then transferred to the vapour phase of liquid nitrogen.

2.4.5 Manual cell counting and assessment of viability

For cell seeding and cell number determination, cells were prepared as single cell solutions and 10 µl was applied to a Bright-Line hemacytometer (Sigma). Cells were counted manually from four independent areas of the chamber, average values calculated, and cell density determined as cell number per ml. Cell viability was determined by using the exclusion dye trypan blue, by mixing equal amount of cell suspension with a 0.4% trypan blue solution prior application to the counting chamber. Positive stained cells lost their integrity of the cell membrane and were deemed non-viable.

2.4.6 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) targeting strategy

Guide RNAs (gRNAs) for gene-editing of *BPIFB1* using the CRISPR-Cas9 technology were designed using the online tool E-CRISPR (http://www.e-crisp.org/E-CRISP/designcrispr.html) (Ran *et al.*, 2013a; Hsu *et al.*, 2013; Heigwer *et al.*, 2014). Three gRNAs were selected targeting exon 2 of *BPIFB1* containing the start codon (*Table 2.7*). Each double stranded guide RNAs was cloned in vector PX-458 to facilitate enhanced green fluorescence protein (eGFP) selection following successful plasmid delivery into the target cell. Vector PX-458 was linearized with the restriction enzyme *BbsI* and purified from the gel after agarose gel electrophoresis (Figure **2.2**). Individual gRNAs were annealed with the respective complementary sequences (*Table 2.7*) by thermal renaturation and ligated into the linearized plasmid using a T4-DNA ligase reaction. For detailed cloning steps and oligo design refer to the following protocol provided by Addgene: https://media.addgene.org/cms/filer_public/6d/d8/6dd83407-3b07-47db-8adb-4fada30bde8a/zhang-lab-general-cloning-protocol-target-sequencing_1.pdf.

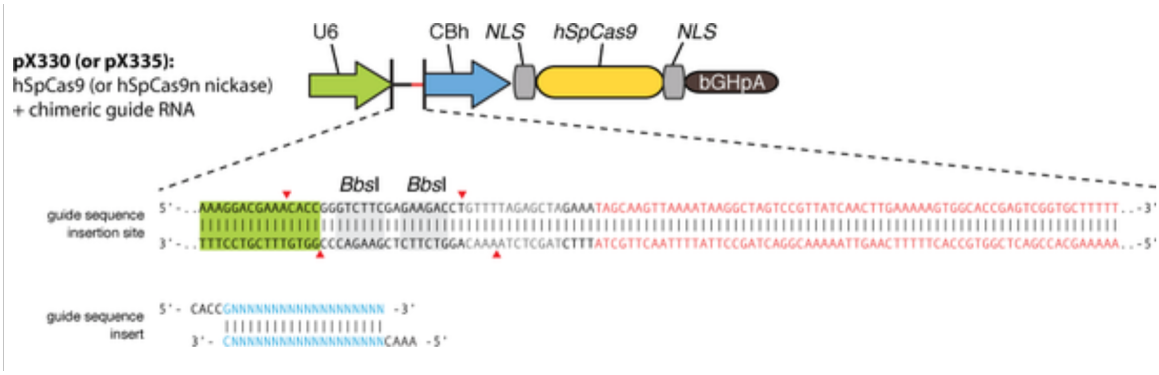


Figure 2.2: Schematic of the expression cassette of pX330 type vectors such as the pX-458. Two expression cassettes allow for co-expression of the *Streptococcus pyogenes* Cas9 (SpCas9) nuclease along with the incorporated guide RNA through the *BbsI* restriction sites, respectively. The annealed guide RNA / complementary oligo can be cloned scarlessly. The gRNA oligos are designed based on the target site sequence (20bp) and need to be followed on the 3' end by a 3bp NGG PAM sequence. Notably, it has been observed that addition of an additional guanine nucleotide to the 5' of the 20-bp guide can increase targeting efficiency (Hsu *et al.*, 2013) (Hsu *et al.*, 2013). Image from addgene: https://www.addgene.org/crispr/zhang/.

After ligation, plasmids were transformed by heat shock into competent *E. coli strain* Topo10 and selected on LB-agar plates containing 100 µg/ml ampicillin. Selected colonies were cultured in LB broth followed by a plasmid extraction using the Isolate II Plasmid Mini kit

(Bioline). Lastly, the generated plasmids were sequenced using Sanger's di-deoxy method employing the U6 primer for confirmation of the guide RNA insert. Sequences were analysed using the FinchTV software.

*Table 2.7: Single gRNA sequences for BPIFB1 CRISPR targeting approach. Reverse complementary sequences (cgRNA) for each respective sgRNA outlined below. The oligo design incorporated the sgRNA scaffold as highlighted in bold: 5' – **CACCG**NNNNNNNNNNNNNNNNNN – 3', 3' – **C**NNNNNNNNNNNNNNNNNN**CAAA** – 5'.*

sgRNA	sequence (5'-3')	target location
gRNA1	**CACCG**GTGAGCACAGAGGGACCTG	intron region exon 1-2
cgRNA1	**AAAC**CAGGTCCCTCTGTGCTCACC	
gRNA2	**CACCG**GGCATCATGGGCCTGCCCT	intron region exon 1-2
cgRNA2	**AAAC**AGGGCAGGCCCATGATGCCC	
gRNA3	**CAGGG**AGCTGGCTGCCATAAGACA	intron region exon 2-3
cgRNA3	**AAAC**TGTCTTATGGCAGCCAGCTC	

2.4.7 DNA transfection

2.4.7.1 Transient transfection using calcium phosphate protocol for mBPIFB1 protein production

To optimize conditions for HEK293 transfection to maximize mBPIFB1 protein production and secretion, HEK293 cells were seeded at a density of $3x10^5$ cells per well in a 24-well plate and incubated overnight. Two hours prior to transfection, the media was exchanged. In order to monitor transfection efficiency, pEGFP-N1 (0.1 µg) was used as positive control alongside FLAG-tag mouse BPIFB1-pVR1255 constructs (0.5 µg; wild-type and N-glycosylation mutant) in a co-expression approach. A DNA-$CaCl_2$ solution was prepared by combining 3.75 µl of 2.5 M $CaCl_2$ and sterile water in a final volume of 30 µl with 30 µl of 2X HEPES buffered saline (HBS). The mixture was incubated at room temperature for 15 mins before being added to the cells dropwise. The plate was swirled slowly by hand, incubated overnight and the media was changed the next day, and following further 24h culture the supernatant was collected.

After successful small-scale transfection, the optimized transfection conditions were repeated and expanded using T75 flasks. HEK293 cells were seeded at a density of 4.5 x 10^6 cells per flask and incubated for 24 hours. Two hours prior transfection, cells were co-transfected with

FLAG-tag mouse BPIFB1-pVR1255 constructs (15 µg; wild-type mBPIFB1, N-glycosylation mutant, N-terminal domain and C-terminal domain), and pEGFP-N1 (3.75 µg). A DNA-$CaCl_2$ mix containing 140.6 µl of 2.5 M $CaCl_2$ and sterile water in a final volume of 1.125 ml) was combined with 1.125 ml of 2X HBS, mixed by pipetting, and incubated for 15 mins at room temperature before being added dropwise to the cells. The flask was swirled slowly and then incubated overnight. After 24 hours, the media was exchanged with DMEM containing 10% (v/v) FBS. The following day, the media was discarded, the cells were washed 3 times with PBS, and incubated with serum-free and phenol red-free DMEM. The following day, the media was collected and fresh serum-free DMEM was added, and this process was repeated for an additional 3 days. All conditioned media samples that were collected were stored at -20°C.

2.4.7.2 Transient transfection using FuGENE HD reagent

Plasmid delivery to mammalian cell cultures was also performed using the non-liposomal FuGENE HD Transfection Reagent (Promega), which allows for transfection of plasmid DNA variety of cell lines with high efficiency and low toxicity in the presence of serum. Initial experiments were performed to optimise the FuGENE HD transfection reagent to DNA ratio, incorporating the manufacturers' recommended range of 1.5:1 to 4:1, as well as ratios of 3:1 and 2.5:1. For the cell line HBEC3-KT the ratio of 1:4 was used for plasmid delivery. FugeneHD transfection reagent was combined with either the empty vector pX458 (serving as background control) or with a combination of vectors expressing the individual sgRNAs (pX458-sgRNA) (Table *2.7*).

2.4.8 Fluorescence-activated cell sorting (FACS) and clonal expansion

Cells were microscopically observed at 48 hours post transfection for green fluorescence protein expression, indicating successful delivery of the vector and assumed expression of the Cas9 protein alongside sgRNAs. Transfected cells were selected based on fluorescence levels from vector driven eGFP co-expression. Cells were fluorescence-activated cell sorted (FACS) on the BDFACSAria Ilu cell sorter (Research Core Facility, the Medical School, University of Sheffield, UK) by Sue Clark (Flow Cytometry Core Facility technician). This allowed GFP positive and GFP negative cells from the transfected population to be identified,

separated and collected. Individual fluorescent cells were collected in 96 wells, for clonal expansion of a founder cells for 1-2 weeks. Upon reaching confluency, clonal populations were validated for the desired genotype by endpoint PCR.

2.4.9 Genomic DNA isolation and PCR validation for desired genotype

Clonal CRISPR target cell lines were validated by endpoint PCR for identify genomic mutations. Genomic DNA isolation was performed using DNAreleasy Advance (FastGene) lysis reagent, specially formulated to eliminate nucleic acid purification for PCR analysis. DNAreleasy Advance replaces time consuming and tedious extraction and purification methods and allows for the released DNA to be used directly in a PCR reaction or it can be stored at -20 °C for several months. HBEC-KT cells were grown to confluency in 96 wells and lysed with 30 µl DNAreleasy Advance lysis reagent. The lysate was transferred into a PCR tube and incubated in a thermocycler for at 65 °C for 5 min, following 95 °C for 5 min and finally 5 min at 20 °C. The reaction was spun and 1 µl used in a PCR reaction with 17 ul Maxima Hot Start Green PCR Master Mix (ThermoFisher Scientific) with the primer pair outlined in Table *2.8*. PCR reactions were carried out in a thermocycler (DNA Engine), using the following cycle conditions. An initial denaturation step of 95°C for 4 min, followed by 35 cycles of 95°C denaturation for 1 min; 60°C annealing for 1 min and 72°C extension for 1 min. After completion of the 35 cycles, the PCR products were extended using a final extension step of 72°C for 7 min, and the PCR products were run on agarose gels alongside a 1 kb DNA hyperladder (Bioline) to determine the size of the PCR products.

Table 2.8: PCR primer sequences for CRISPR target validation in HBEC-KT cells

Amplicon	size (bp)	Primer sequence	
		Forward (5' - 3)	Reverse (3' - 5')
wt	551	CAACAGCCTCTCCTGGACAC	CAGGTGAAGGCAGTGGGAG
HET/HOM	319		

2.5 Protein techniques

2.5.1 Immunoblot analysis

Conditioned media protein samples were collected over three days as described above and were analysed via western blotting. Samples were resolved on polyacrylamide gels of 12% or 15%, depending on the molecular weight of the protein of interest. Apical wash samples were either prepared by equal amount of total protein or by equal volume and combined with 2x lysis buffer. Samples were denatured at 90˚C for 5 minutes and run alongside a reference marker (BioLabs Color Protein, Broad Range 11-245 kDa standard) until the dye reached the front on the gel. Resolved gels were placed in transfer buffer and transferred to a methanol activated polyvinylidene difluoride membrane (PVDF) (Milipore), using a Western blotting transfer system (BioRad, Trans-Blot Turbo) at transfer conditions 1.0 A and 25 V for 30 minutes. Subsequently, membranes were blocked in a 5% skim milk / Tris buffered saline (TBS) solution for 1 hour at room temperature (RT) on while agitated. Next, membranes were incubated with primary antibodies (Table *2.1*) prepared in 5% skim milk / TBST on an orbital shaker at 4°C overnight. The following day, membranes were washed three times at 10-minute intervals with TBST prior secondary antibody incubation (Table *2.1*), prepared in 5% skim milk / TBST. The secondary antibody solution was incubated for 1 hour at RT, and membranes were washed three times at 10-minute intervals with TBST. The enhanced chemiluminescence (ECL) western blotting detection kit (Geneflow) was used to detect the HRP-conjugated secondary antibody, and luminescence signals captured on X-ray film (ThermoFisher Scientific) and processed using a Compact X4 automatic X-ray film processor (Xograph). For higher molecular weight protein bands, the Biorad ChemiDocTM XRS+ with Image Studio Software was used.

2.5.2 Determination of protein concentrations

Total protein concentrations of cell lysates or washes were determined colorimetrically using the Bradford reagent according to manufacturer's instructions. To calculate protein concentrations in lysates, a bovine serum albumin (BSA) standard curve was prepared, covering a range of concentrations from 0.0625 mg/ml to 1.4 mg/ml. Typically, 5 µl of BSA standards or lysate samples were added in triplicate to 250µl of Bradford reagent in a 96 well format. Samples were incubated for 1 minute at room temperature, and absorbance values of

samples were determined at a wavelength of 595 nm on a plate reader (VarioScan). Protein concentrations were then calculated using the linear regression curve equation derived from the BSA standard curve points.

2.5.3 Densitometric analysis

Quantitative analysis of Western blots bands was carried out using the LI-COR software (Image Studio Version 5.2.5). Scanned X-Ray images were analysed in the grey channel, and the spot pixel density of the individual bands were detected, associated average background subtraction of each position performed, and plotted as arbitrary densitometry units (ADU) using Excel software (Microsoft Office).

2.5.4 PNGase assay

The PNGase assay was carried out using the Peptide-N-Glycosidase F enzyme (PNGase F) enzyme (NEB) to cleave N-linked glycans from ALI supernatant proteins. Enzymatic reactions were prepared according to manufacturer's instructions. In brief, 1 µl of PNGase F enzyme was combined with 20 µg of supernatant sample in a total reaction volume of 20 µl. Reactions were incubated at 37°C for 1 h and subsequently lysed in 2x SDS lysis buffer and analysed by Western blotting.

2.6 Cell biological techniques

2.6.1 Live cell density assay

Cell proliferation was assessed using the MTS chemical (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium), which in the presence of phenazine methosulfate (PMS) or the more stable phenazine ethosulfate (PES) is bio-reduced by cells into a soluble coloured formazan product. This product is stable and non-toxic in cell media and can be measured by a colorimetric assay using the absorbance at 490 nm. The MTS assay was carried out in a 96-well plate format, by seeding 5000 cells in a 96-well plate (Greiner) format at 100 µl/well. Wells containing only culture medium without cells served as background control and were used to generate a average value for background subtraction. The proliferation assay was carried out over a period of 72 hours and a plate seeded for each

timepoint interval of 24 hours. At a given time point, 20 µl MTS solution containing PES at a final concentration of 0.33 mg/ml were added to each well, incubated for 4 hours in a cell culture incubator and absorbance recorded on a plate reader (Varioscan) at 490 nm. Data were expressed as percent cell growth or viability. In order to calculate cell viability and cell growth over time, absorbance values were processed from raw data by subtracting background absorbance (cell-free wells containing only media) and forming a mean over all replicates of the same condition. Cell proliferation was calculated using a reference plate reading at timepoint 0 and expressed as percentage increase.

2.6.2 Cell cycle analysis

Cell cycle analysis was carried out by means of DNA content quantification using the DNA intercalation dye propidium iodide (PI). Samples were produced as single cell solution and washed with a cold protein-free PBS buffer without Ca^{2+} or Mg^{2+} ions. The wash step was repeated twice, and cells pelleted at 200 x *g* for 5 minutes at 4°C. 2x 10^6 cells were prepared in 1 ml ice cold buffer and fixed by slowly adding the cell suspension dropwise to 9 ml of 70% ethanol in a 15 ml polypropylene centrifuge tube, while gently vortexing. The fixed cell solution was stored at 4°C between 12 - 24 hours. Cells were prepared for PI staining by centrifugation at 200 x *g* for 10 min at 4°C, the cell pellet was resuspended in 3 ml cold PBS and cell suspension transferred to a Falcon 12 x 75 mm polystyrene tube (Falcon 352235) using the nylon filter cap to remove cell clumps. Cells were pelleted at 200 x *g* for 10 min at 4°C and resuspend in 500 µl PI/Triton X-100 staining solution, and incubated at 37°C for 15 minutes, before tubes were transferred to ice or stored at 4°C protected from light. Data was acquired on the BDFacs Melody (BD) flow cytometer, cell cycle fitted using the BD FACSChorus Software, and graphs and statistical analysis plotted with GraphPad Prism 9.

2.6.3 Cytokine array

Conditioned media post PR8 infection of undifferentiated HBEC3-KT cells were collected at 2h, 8h, and 24h timepoints post viral challenge, and the entire sample of 1ml was applied to the cytokine membrane (Mouse Cytokine Array Panel A, R&D Systems). Incubation and wash steps were carried out according to manufacturers' instructions. The signal detection was generated by means of an HRP- Streptavidin conjugated antibody and detected using the

Biorad ChemiDocTM XRS+ equipment. Data processing and signal quantification was performed using the LI-COR software (Image Studio Version 5.2.5). The spot pixel density of the individual positions on the array in duplicate were detected, average background subtraction of each spot was performed, and signal reads extracted. Data were imported and processed in Microsoft Excel, normalising the array data per condition to the corresponding control using the average signal of the reference spots per array. The average signal pixel density of each positive detected cytokine duplicate pair and associated standard deviation was calculated in Excel. Spot intensities were presented as arbitrary units (mean spot intensity) (not fold increase over control) alongside the associated control and statistical significance was determined using Prism GraphPad 9 applying an unpaired t-test.

2.6.4 IAV virus infection of undifferentiated or differentiated ALI cultures

2.6.4.1 PR8 infection time course in undifferentiated HBEC3-KT cells

Undifferentiated parental HBEC3-KT were cultured in KSFM or ExPlus media and seeded at 6 x 10^4 cells per 24 well. Upon reaching confluency, the PR8 virus strain was added in a volume of 1000 µL respective culture medium at an MOI of 1, and incubated for 1 hour at 37°C. Subsequently, the inoculum was removed, the wells washed twice with pre-warmed HBSS, and 1000 µL of culture media added. Cell morphology and virus induced GFP co-expression was monitored by brightfield microscopy or the green-fluorescent channel, respectively, 2, 8 and 24 hours post infection (Zoe fluorescent microscope). Fluorescent negative and positive cells were either counted manually per field and displayed as cells per field along the standard deviation or fluorescent images of cells were quantified by mean fluorescence per field using ImageJ (Version 1.53k), background subtracted using a non-fluorescent control and the average mean fluorescence from three replicate experiments alongside the respective standard deviation displayed using GraphPad Prism 9.

2.6.4.2 PR8 infection of HBEC3-KT cells in presence of CRISPR clone derived washes

Undifferentiated HBEC3-KT cells were cultured in KSFM and CRISPR, and generated wt clone 7 apical washes or BPIFB1 knockout clone 13 washes applied at a concentration of 250 µg/ml total protein. Subsequently, cells were challenged with the PR8 virus strain at an MOI of 1. Cells were assessed microscopically (Zoe fluorescent microscope) by brightfield and

green-fluorescent imaging 24h post infection. Fluorescent negative and positive cells were either counted manually per field and displayed as cells per field along the standard deviation or recorded fluorescent images of cells were quantified by mean fluorescence per field using ImageJ (Version 1.53k), background subtracted using a non-fluorescent control and the average mean fluorescence from three indented experiments alongside the respective standard deviation displayed using GraphPad Prism 9.

2.6.4.3 PR8 infection of HBEC3-KT cells in presence of recombinant BPIFB1

Undifferentiated parental HBEC3-KT were cultured in KSFM and seeded at 6 x 10^4 cells per 24 well. Upon reaching confluency, the PR8 virus strain was added in a volume of 1000 µL respective culture medium at an MOI of 1, and co-incubated for 1 hour at 37°C with recombinant human BPIFB1 protein (13275-H08H, Sino Biological) between 0-2.86 µg/ml final protein concentration. Subsequently, the inoculum was removed, the wells washed twice with pre-warmed HBSS, and 1000 µL of culture media added. Cell morphology and GFP virus expression was monitored by brightfield microscopy or the green-fluorescent channel, respectively, 24 hours post infection (Zoe fluorescent microscope). Fluorescent negative and positive cells were either counted manually per field and displayed as cells per field along the standard deviation or recorded fluorescent images of cells were quantified by mean fluorescence per field using ImageJ (Version 1.53k), background subtracted using a non-fluorescent control and the average mean fluorescence from three indented experiments alongside the respective standard deviation displayed using GraphPad Prism 9.

2.6.5 Statistical analysis

Data are represented as means ± SD. Statistical differences between samples were assessed by one-way analysis of variance (ANOVA) with Tukey's test to compare more than two groups, or two-tailed student's t-test was used to compare two groups of interest. P-values less than 0.05 were considered as statistically significant (*$p < 0.05$). Significance levels represented on graphs are as follows: ns - not significant; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$. GraphPad Prism (version 9) was used to perform all statistical analysis, data analysis and preparation of graphs.

2.7 Bioinformatic analysis

2.7.1 Basic bioinformatic analysis

Sequences for BPIFB1 protein, RNA, and genomic DNA were retrieved from National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov) website. These sequences were aligned using Clustal Omega at European Bioinformatic Institute (EBI) or visualised using Uniprot (https://www.uniprot.org). Protein structures were predicted using the Phyre2 Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) online tool (Kelley *et al.*, 2015) or SWISS MODEL (https://swissmodel.expasy.org). PCR primers were designed using Primer-BLAST at NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast) (Ye *et al.*, 2012). Gene annotations were performed using BioGPS (http://biogps.org/)(Wu *et al.*, 2016a).

2.7.2 RNA-Seq data analysis

2.7.2.1 Raredon et al 2019 data set analysis

Using Lung Connectome (http://lungconnectome.net/), an interactive platform created by Raredon *et al.,* 2019 to allow transcriptional exploration of single cells gene expression in healthy adult lung across species, pre-clustered tSNE plots were generated to visualize the entire mouse and human single cell RNA sequencing datasets for both species. Each dataset was then searched for *BPIFB1* expression and fold change visualized in the form of a tSNE plot and violin plots on the interactive platform. Gene expression for each cell type in each species was further visualized with a heatmap (pkg. matplotlib) in order to demonstrate cell type expression of *BPIFB1* in both mouse and human lung tissue.

2.7.2.2 Deprez et al 2020 and Vieira Braga et al 2019 data set analysis

In order to identify which regions of the respiratory tract express *BPIFB1*, single cell RNA-Seq datasets were identified containing nasal, bronchial, and parenchyma samples. Two datasets, created by Deprez *et al.,* 2020 and by Vieira Braga *et al.,* 2019, contained analysis of cell populations from all three sections of the respiratory tract. The expression of *BPIFB1* was visualized using the cellxgene custom platform (https://www.covid19cellatlas.org/deprez19/) for the single cell RNA-Seq generated by Deprez *et al.,* 2019. The *BPIFB1* gene expression, breakdown of populations belonging to

each area of the respiratory tract, and distribution of sample collection were demonstrated using generated UMAP plots on the interactive platform. The percentage of cells in each part of the respiratory tract expressing *BPIFB1* fold change > 1 was calculated and visualized using pie charts. The percentage of cells expressing *BPIFB1* fold change > 1 as varied by sampling method via brushing or biopsy sampling methods were also visualized using pie charts. In order to further investigate how *BPIFB1* expression might vary based on location in the respiratory tract, the Lung Cell Atlas (https://asthma.cellgeni.sanger.ac.uk/), an interactive platform created by Vieira Braga et al. 2019 in order to examine cell populations throughout the respiratory tract, was used. Datasets containing nasal, bronchial, and parenchyma samples were visualized using UMAP plots to view clusters and *BPIFB1* expression. A heatmap was generated (via pkg matplotlib) in order to view how the expression of *BPIFB1* varies by cell type in different parts of the respiratory tract.

2.7.2.3 Miller et al 2020 and Beauhemin et al 2016 data set analysis

To identify whether *BPFIB1* expression is present in the developing lung, a single cell RNA-Seq dataset was identified containing cells collected from foetal lung tissue at week 11.5 -21 of development collected by Miller *et al.,* 2020. The expression of *BPIFB1* was visualized using the cellxgene interactive platform (https://www.covid19cellatlas.org/miller20/) for the single cell RNA-Seq generated by Miller *et al.,* 2020. The percentage of cells expressing *BPIFB1* fold change >1 in each cell type, timepoint, and location was visualized using a heatmap (via pkg matplotlib). In order to identify the gene expression of *BPIFB1* in mouse lung development as compared to human, data from Beauhemin *et al.,* 2016 was downloaded from their web interface (http://lungdevelopment.jax.org/) and visualized as a heatmap (via pkg matplotlib). This data provided valuable interrogation of *BPIFB1* expression during lung development from embryonic (E) day 9.5 -postnatal (P) day 56 in three mouse lines including A/J, C3H/HeJ, and C57BL5/J, as well as *BPIFB1* gene expression at different stages of lung development.

2.7.3 Clariom S mouse microarray study and analysis

2.7.3.1 Microarray data sample generation

The exposed apical surface of ciliated, polarized pseudo-stratified mTEC cultures at D0 or D14 was washed four times with pre-warmed HBSS and removed prior X31 infection at MOI of 1. The basolateral media in the lower chamber was replaced with 1mL of pre-warmed mTEC differentiation medium (Table *2.6*). The X31 virus strain was added in a volume of 200 µL mTEC differentiation medium to the apical side of the transwell culture, and incubated for 1 hour at 37°C. Subsequently, the inoculum was removed, the apical side of the ALI cultures washed once with pre-warmed HBSS, and the apical surface was maintained dry at 37°C for 23h. IAV infection levels were assessed through viral induced by mean of green fluorescence expression. RNA was extracted and collected by Fawaz Aljuhani and Priyanka Anujuan.

2.7.3.2 Chip preparation

Extracted RNA samples were assessed for quantity and quality using the NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific) and the Agilent Bioanalyser 600 (Agilent). A total of 200ng RNA was prepared for analysis per condition on the Clariom S mouse GeneChip (Thermo Fisher) according to the manufacturers' recommendations. mRNA was transcribed into double stranded cDNA with the incorporation of a T7 polymerase binding site at the 3' end of the RNA molecule. Antisense RNA (aRNA) was generated with the T7 polymerase. 15ug of aRNA was used to generate a sense DNA strand, subsequently fragmented and end labelled with biotin, combined with a hybridization solution and incubated with the Mouse Clariome S microarray Gene Chip (Thermo Fisher). Following hybridisation, the chip was washed using the fluidics station and incubated with a streptavidin-phycoerythrein cocktail. The generated fluorescent signal was correlated with the hybridization level of the labelled material to the oligonucleotide probes on the chip. Gene Chips were scanned on the Gene Chip 7000G scanner, and the images collected as CEL files. This work was undertaken as a service by Dr Paul Heath (SiTRaN, University of Sheffield)

2.7.4 Bioinformatic analysis of Chip data

Microarray data in CEL files was analysed using Affymetrix Expression Console and raw data were processed using GeneSpring GX version 11.5 (Agilent Technologies). The initial raw data processing was performed by Miraj Kobad Chowdhury.

The Clariome S data sets were converted into a spreadsheet format using Transcriptome Analysis Console (TAC) tool version 4.02 and analysed in Microsoft Excel for gene specific intensity, fold change, and p-values. All calculations and statistical analyses such as average, standard deviation, p-value calculation and differential analyses were performed using Excel (Microsoft) and plotted using GraphPad Prism 9. Log2 intensity less than 5 was considered as negative for expression and Log2 intensity above 5 was considered as positive for expression of individual genes (Miller *et al.*, 2014). Microarray data were normalized and background-corrected using robust multi-array average-linear models for microarray data, and differentially expressed genes identified using limma voom (Ritchie *et al.*, 2015) with the criteria of absolute fold change > 1 and limma adjusted p-value < 0.05. p-values were adjusted using the Benjamini and Hochberg method (Gold *et al.*, 2009). Multidimensional scaling plot were generated using limma voom and gene ontology analyses were performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 (Dennis *et al.*, 2003).

Chapter 3:

Bioinformatic analysis and data mining of the human and murine *BPIFB1* gene and protein products

3.1 Introduction

BPIFB1 is the prototypic 2 domain containing family member of BPIF protein family. The full complement of BPIF genes has only been identified in placental mammals (Bingle and Craven, 2003; Bingle *et al.*, 2011a). It has been previously shown that birds do not have *BPIFB1, BPIFB5* or *BPIRB9* or any *BPIFA* genes (Bingle *et al.*, 2011c). As has been highlighted in the introduction, BPIF family members belong to the wider family BPI related proteins and BPIFB proteins share homology with both the N-terminal and C-terminal domains (Bingle and Craven, 2002; Krasity *et al.,* 2011). An interesting observation made on members of the wider BPIF-family is that they share relatively limited sequence similarity both within and between species (Bingle *et al.,* 2004, Bingle and Craven 2011). Therefore, in this chapter I sought to study the gene diversity and evolutionary development of *BPIFB1* more systematically.

BPIFB1 expression has been localised in mouse tissues to the tongue, more specifically the epithelium and minor glands of the dorsal tongue, ventral surface of the palate and lining of the upper pharynx and in the large airways (Musa *et al.*, 2012). This expression pattern is generally recapitulated human tissue with the gene being present in nasopharyngeal epithelial cells and submucosal glands and upper respiratory tract, with high abundance in nasopharyngeal and respiratory secretions, lavage fluid and sputum (Zhang *et al.*, 2003; Bingle *et al.*, 2010; Gao *et al.*, 2015). The more well-studied single domain containing paralogue BPIFA1, shows a partially overlapping expression pattern with BPIFB1, and has been found localised in the respiratory epithelium of the nasal passages, as well as in the large airways (Weston *et al.*, 1999; LeClair *et al.*, 2001; Genter *et al.*, 2003; LeClair *et al.*, 2004). Interestingly, it appears that the two proteins are produced by different cells in the respiratory tract. BPIFB1 is localised to a limited population of goblet cells in the upper airways whereas BPIFA1 found in non-ciliated/non-goblet cell population of the upper airway and the bronchial passages, (Bingle *et al.*, 2010; Musa *et al.*, 2012). As BPIFB1 shows partial overlapping expression patterns with BPIFA1, and as they share structural similarity, it could be speculated that BPIFB1 and BPIFA1 carry out similar biological roles in the upper respiratory tract.

A study investigating COPD patients suggested that BPIFB1 levels correlate with severity of COPD (Gao *et al.*, 2015; De Smet *et al.*, 2018) or severity of CF (Bingle *et al.*, 2012; Saferali *et al.*, 2015). Moreover, a number of studies have found that BPIFB1 appears to be

ectopically produced in multiple types of tumour tissue, suggesting that it may play a role in malignancies (Zhang *et al.*, 2003; González-Arriagada *et al.*, 2012; Wei *et al.*, 2018b; Li *et al.*, 2021b).

Therefore, constructing a solid understanding of BPIFB1 gene and protein expression and sequence and structural conservation, combined with temporal and spatial expression patterns, and information regarding dysregulation associated with respiratory diseases, may inform a better understanding of the to the biological function of BPIFB1 and allows for more targeted experimental approaches to be designed to uncover its biological function.

3.2 Aims

This chapter focuses on gathering and extraction of BPIFB1 expression data and disease association information based on current literature and published data sets, and by extrapolation of functional information of closely related PLUNC family members, such as BPIFA1. A specific focus was made on comparisons between human and mouse orthologues, as existing mouse model could provide important tools to further study BPIFB1 function in a more physiological relevant setting.

Using multiple online platforms in conjunction with published studies, the aim was to extract functional and expression data specific for BPIFB1 and related family members. In addition, the aim was to focus on human and murine orthologues establishing the following goals:

- Identification of gene diversity and evolutionary development of BPIFB1
- Exploration of BPIFB1 protein structure and conserved domains, and comparison between human and murine protein structures
- Study of BPIB1 expression on RNA and protein basis, from organ level down to specific cell type using published single cell RNA-Seq databases
- Investigation of genetic variations of BPIFB1 and potential links to disease states with specific focus on correlations to lung diseases

3.3 Results

3.3.1 Comparative analysis of BPIFB1 across multiple genomes

BPIF genes are known to be rapidly evolving and have been found to be highly divergent (Bingle *et al.,* 2004). To more fully investigate the molecular evolution of *BPIFB1* I identified orthologs of human BPIFB1 on Ensembl (https://www.ensembl.org/index.html).

Species set	Show details	With 1:1 orthologues
Primates (*23 species*) Humans and other primates	☐	23
Rodents and related species (*24 species*) Rodents, lagomorphs and tree shrews	☐	23
Laurasiatheria (*38 species*) Carnivores, ungulates and insectivores	☐	32
Placental Mammals (*90 species*) All placental mammals	☐	82
Sauropsida (*27 species*) Birds and Reptiles	☐	0
Fish (*65 species*) Ray-finned fishes	☐	0
All (*200 species*) All species, including invertebrates	☑	86

Figure 3.1: Species distribution of *BPIFB1* (ENSG00000125999) and respective orthologue numbers were identified using the ENSEMBL database.

The orthologue analysis carried out using the Ensembl database confirmed that there are no clear orthologs found in birds, reptiles, fish and invertebrates (Figure **3.1**). This is consistent with a previous study, showing the lack of the BPIFB1 orthologue in chickens, despite the presence BPIFB2, BPIFB3, BPIFB4 and BPIFB6 (Chiang *et al.*, 2011). However, it is important to bear in mind that not all species are represented in the Ensembl database and certain species appear not to have orthologs as the genome assembly may not be complete.

BPIFB1 is found in the majority of placental mammals (82 of 90 species) of which 32 out of 38 species belonging to the Laurasiatherian superorder encompassing the orders of moles, shrews, hedgehogs, bat, carnivores, ungulates and whales. Furthermore, rodents and related species (23 out of 24 species) and Primates (23 out of 23 species) share exclusively a single orthologue of BPIFB1 (Figure **3.1**). Overall, several BPIF genes are found in birds and reptiles, but BPIFB1 is not (Bingle *et al.*, 2011c).

Employing the ENSEMBL database, a phylogenetic tree was extracted using the human BPIFB1 protein sequence as query and covering all mammalian branches (Figure **3.2**). For simplicity some of the branches are compressed. The tree confirms that BPIFB1 orthologs are present in all classes of mammals. It also suggests that there may be more divergent orthologs present in marsupials, but not in species further away in evolutionary distance (Figure **3.2**).

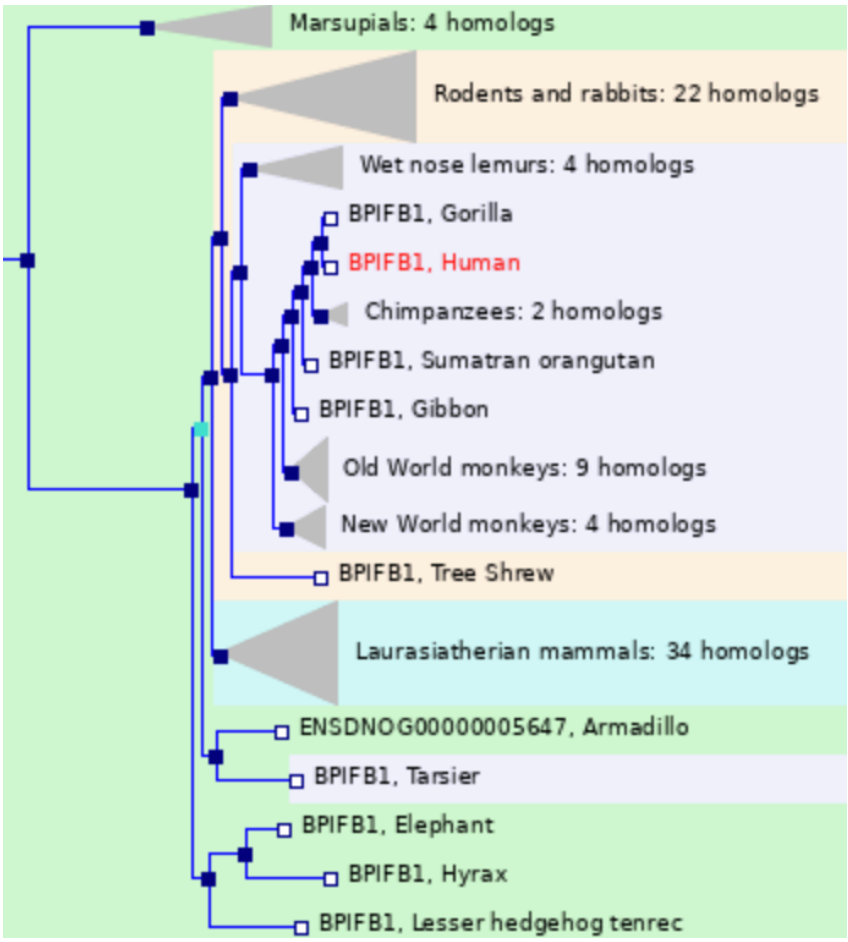


Figure 3.2: Phylogenetic tree of BPIF1 proteins. The phylogenetic tree was generated using the human BPIFB1 protein sequence as query (shown in red). Blue nodes contain at least one paralog of query. The size of a collapsed node (shown as triangle) is approximately relative to the number of homologs in that node (mentioned at each node). These are collapsed for clarity. The gene-tree was obtained from the ENSEMBL database (ENSGT01020000230460).

To examine the extent of conservation of BPIFB1 more specifically across species, a BLAST search was performed, analysing the reference sequence (RefSeq) protein database with the human BPIFB1 sequence. The output of this analysis is shown in Figure **3.3** as a multiple sequence alignment (MSA). The image is shaded to show differences in red (Figure **3.3**).

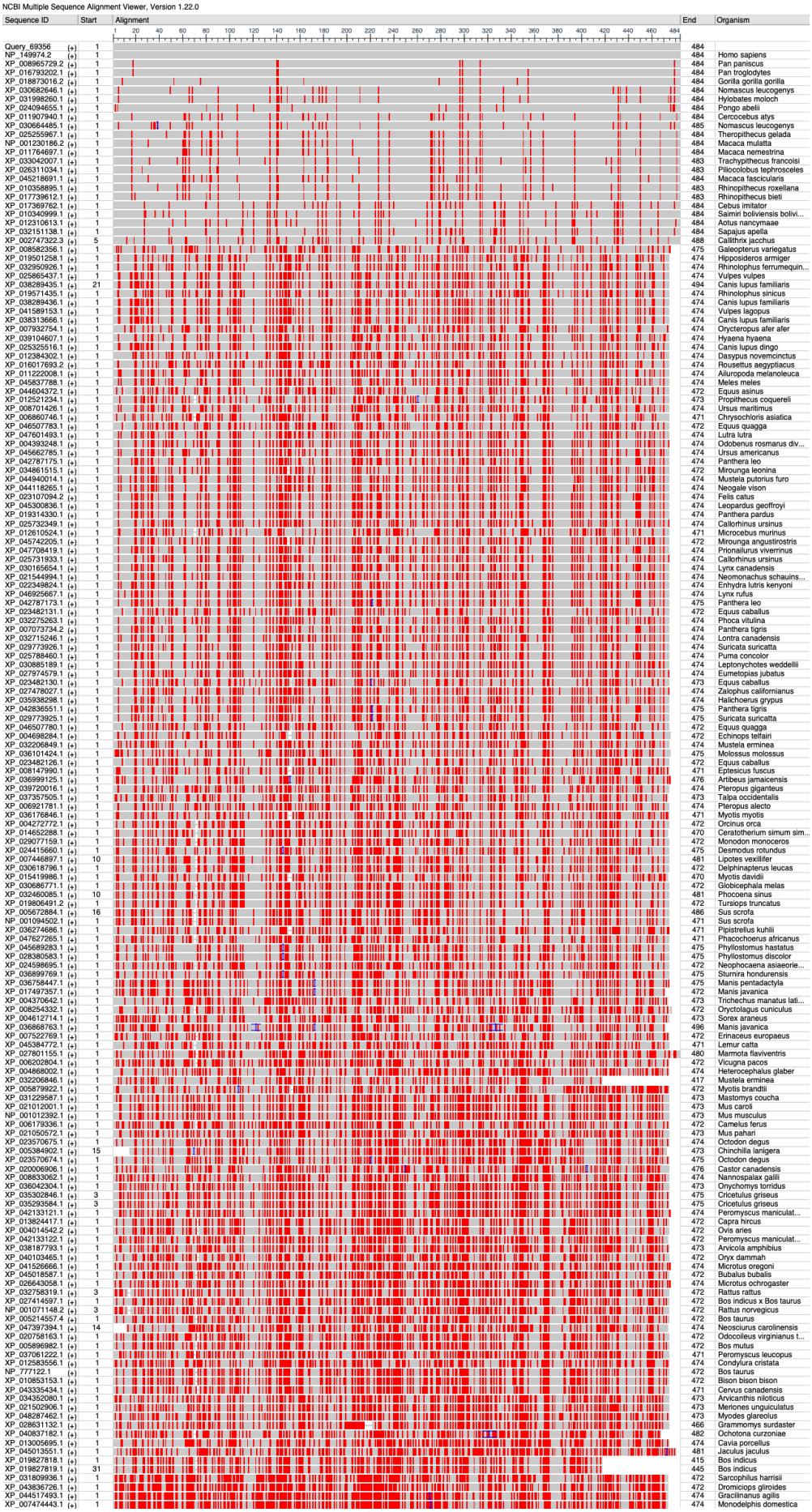

NCBI Multiple Sequence Alignment Viewer, Version 1.22.0
Sequence ID
Start
Alignment
End
Organism
Query_69356 (+) 1 484
NP_149974.2 (+) 1 484 Homo sapiens
XP_008965729.2 (+) 1 484 Pan paniscus
XP_016793202.1 (+) 1 484 Pan troglodytes
XP_018873016.2 (+) 1 484 Gorilla gorilla gorilla
XP_030682646.1 (+) 1 484 Nomascus leucogenys
XP_031998260.1 (+) 1 484 Hylobates moloch
XP_024094655.1 (+) 1 484 Pongo abelii
XP_011907940.1 (+) 1 484 Cercocebus atys
XP_030664485.1 (+) 1 485 Nomascus leucogenys
XP_025255967.1 (+) 1 484 Theropithecus gelada
NP_001230186.2 (+) 1 484 Macaca mulatta
XP_011764697.1 (+) 1 484 Macaca nemestrina
XP_033042007.1 (+) 1 483 Trachypithecus francoisi
XP_026311034.1 (+) 1 483 Piliocolobus tephrosceles
XP_045218691.1 (+) 1 484 Macaca fascicularis
XP_010358895.1 (+) 1 483 Rhinopithecus roxellana
XP_017739612.1 (+) 1 483 Rhinopithecus bieti
XP_017369762.1 (+) 1 484 Cebus imitator
XP_010340999.1 (+) 1 484 Saimiri boliviensis bolivi...
XP_012310613.1 (+) 1 484 Aotus nancymaae
XP_032151138.1 (+) 1 484 Sapajus apella
XP_002747322.3 (+) 5 488 Callithrix jacchus
XP_008582356.1 (+) 1 475 Galeopterus variegatus
XP_019501258.1 (+) 1 474 Hipposideros armiger
XP_032950926.1 (+) 1 474 Rhinolophus ferrumequin...
XP_025865437.1 (+) 1 474 Vulpes vulpes
XP_038289435.1 (+) 21 494 Canis lupus familiaris
XP_019571435.1 (+) 1 474 Rhinolophus sinicus
XP_038289436.1 (+) 1 474 Canis lupus familiaris
XP_041589153.1 (+) 1 474 Vulpes lagopus
XP_038313666.1 (+) 1 474 Canis lupus familiaris
XP_007932754.1 (+) 1 474 Orycteropus afer afer
XP_039104607.1 (+) 1 474 Hyaena hyaena
XP_025325516.1 (+) 1 474 Canis lupus dingo
XP_012384302.1 (+) 1 474 Dasypus novemcinctus
XP_016017693.2 (+) 1 474 Rousettus aegyptiacus
XP_011222008.1 (+) 1 474 Ailuropoda melanoleuca
XP_045837788.1 (+) 1 474 Meles meles
XP_044604372.1 (+) 1 472 Equus asinus
XP_012521234.1 (+) 1 473 Propithecus coquereli
XP_008701426.1 (+) 1 474 Ursus maritimus
XP_006860746.1 (+) 1 471 Chrysochloris asiatica
XP_046507783.1 (+) 1 472 Equus quagga
XP_047601493.1 (+) 1 474 Lutra lutra
XP_004393248.1 (+) 1 474 Odobenus rosmarus div...
XP_045662785.1 (+) 1 474 Ursus americanus
XP_042787175.1 (+) 1 474 Panthera leo
XP_034861515.1 (+) 1 472 Mirounga leonina
XP_044940014.1 (+) 1 474 Mustela putorius furo
XP_044118265.1 (+) 1 474 Neogale vison
XP_023107094.2 (+) 1 474 Felis catus
XP_045300836.1 (+) 1 474 Leopardus geoffroyi
XP_019314330.1 (+) 1 474 Panthera pardus
XP_025732349.1 (+) 1 474 Callorhinus ursinus
XP_012610524.1 (+) 1 471 Microcebus murinus
XP_045742205.1 (+) 1 472 Mirounga angustirostris
XP_047708419.1 (+) 1 474 Prionailurus viverrinus
XP_025731933.1 (+) 1 474 Callorhinus ursinus
XP_030165654.1 (+) 1 474 Lynx canadensis
XP_021544994.1 (+) 1 474 Neomonachus schauins...
XP_022349824.1 (+) 1 474 Enhydra lutris kenyoni
XP_046925667.1 (+) 1 474 Lynx rufus
XP_042787173.1 (+) 1 475 Panthera leo
XP_023482131.1 (+) 1 472 Equus caballus
XP_032275263.1 (+) 1 474 Phoca vitulina
XP_007073734.2 (+) 1 474 Panthera tigris
XP_032715246.1 (+) 1 474 Lontra canadensis
XP_029773926.1 (+) 1 474 Suricata suricatta
XP_025788460.1 (+) 1 474 Puma concolor
XP_030885189.1 (+) 1 474 Leptonychotes weddellii
XP_027974579.1 (+) 1 474 Eumetopias jubatus
XP_023482130.1 (+) 1 473 Equus caballus
XP_027478027.1 (+) 1 474 Zalophus californianus
XP_035938298.1 (+) 1 474 Halichoerus grypus
XP_042836551.1 (+) 1 475 Panthera tigris
XP_029773925.1 (+) 1 475 Suricata suricatta
XP_046507780.1 (+) 1 472 Equus quagga
XP_004698284.1 (+) 1 472 Echinops telfairi
XP_032206849.1 (+) 1 474 Mustela erminea
XP_036101424.1 (+) 1 475 Molossus molossus
XP_023482126.1 (+) 1 472 Equus caballus
XP_008147990.1 (+) 1 471 Eptesicus fuscus
XP_036999125.1 (+) 1 476 Artibeus jamaicensis
XP_039720016.1 (+) 1 474 Pteropus giganteus
XP_037357505.1 (+) 1 473 Talpa occidentalis
XP_006921781.1 (+) 5 474 Pteropus alecto
XP_036176846.1 (+) 1 471 Myotis myotis
XP_004272772.1 (+) 1 472 Orcinus orca
XP_014652288.1 (+) 1 470 Ceratotherium simum sim...
XP_029077159.1 (+) 1 472 Monodon monoceros
XP_024415660.1 (+) 1 475 Desmodus rotundus
XP_007446897.1 (+) 10 481 Lipotes vexillifer
XP_030618796.1 (+) 1 472 Delphinapterus leucas
XP_015419986.1 (+) 1 470 Myotis davidii
XP_030686771.1 (+) 1 472 Globicephala melas
XP_032460085.1 (+) 10 481 Phocoena sinus
XP_019806491.2 (+) 1 472 Tursiops truncatus
XP_005672884.1 (+) 16 486 Sus scrofa
NP_001094502.1 (+) 1 471 Sus scrofa
XP_036274686.1 (+) 1 471 Pipistrellus kuhlii
XP_047627265.1 (+) 1 471 Phacochoerus africanus
XP_045689283.1 (+) 1 475 Phyllostomus hastatus
XP_028380583.1 (+) 1 475 Phyllostomus discolor
XP_024598695.1 (+) 1 472 Neophocaena asiaeorie...
XP_036899769.1 (+) 1 475 Sturnira hondurensis
XP_036758447.1 (+) 1 475 Manis pentadactyla
XP_017497357.1 (+) 1 472 Manis javanica
XP_004370642.1 (+) 1 473 Trichechus manatus lati...
XP_008254332.1 (+) 1 472 Oryctolagus cuniculus
XP_004612714.1 (+) 1 473 Sorex araneus
XP_036868763.1 (+) 1 496 Manis javanica
XP_007522769.1 (+) 1 472 Erinaceus europaeus
XP_045384772.1 (+) 1 471 Lemur catta
XP_027801155.1 (+) 1 480 Marmota flaviventris
XP_006202804.1 (+) 1 472 Vicugna pacos
XP_004868002.1 (+) 1 474 Heterocephalus glaber
XP_032206846.1 (+) 1 417 Mustela erminea
XP_005879922.1 (+) 1 472 Myotis brandtii
XP_031229587.1 (+) 1 473 Mastomys coucha
XP_021012001.1 (+) 1 473 Mus caroli
NP_001012392.1 (+) 1 473 Mus musculus
XP_006179336.1 (+) 1 472 Camelus ferus
XP_021050572.1 (+) 1 473 Mus pahari
XP_023570675.1 (+) 1 474 Octodon degus
XP_005384902.1 (+) 15 473 Chinchilla lanigera
XP_023570674.1 (+) 1 475 Octodon degus
XP_020006906.1 (+) 1 476 Castor canadensis
XP_008833062.1 (+) 1 474 Nannospalax galili
XP_036042304.1 (+) 1 473 Onychomys torridus
XP_035302846.1 (+) 3 475 Cricetulus griseus
XP_035293584.1 (+) 3 475 Cricetulus griseus
XP_042133121.1 (+) 1 474 Peromyscus maniculat...
XP_013824417.1 (+) 1 472 Capra hircus
XP_004014542.2 (+) 1 472 Ovis aries
XP_042133122.1 (+) 1 472 Peromyscus maniculat...
XP_038187793.1 (+) 1 473 Arvicola amphibius
XP_040103465.1 (+) 1 472 Oryx dammah
XP_041526666.1 (+) 1 474 Microtus oregoni
XP_045018587.1 (+) 1 472 Bubalus bubalis
XP_026643058.1 (+) 1 474 Microtus ochrogaster
XP_032758319.1 (+) 3 472 Rattus rattus
XP_027414597.1 (+) 1 472 Bos indicus x Bos taurus
NP_001071148.2 (+) 3 472 Rattus norvegicus
XP_005214557.4 (+) 1 472 Bos taurus
XP_047397394.1 (+) 14 474 Neosciurus carolinensis
XP_020758163.1 (+) 1 472 Odocoileus virginianus t...
XP_005896982.1 (+) 1 472 Bos mutus
XP_037061222.1 (+) 1 471 Peromyscus leucopus
XP_012583556.1 (+) 1 474 Condylura cristata
NP_777122.1 (+) 1 472 Bos taurus
XP_010853153.1 (+) 1 472 Bison bison bison
XP_043335434.1 (+) 1 471 Cervus canadensis
XP_034352080.1 (+) 1 473 Arvicanthis niloticus
XP_021502906.1 (+) 1 473 Meriones unguiculatus
XP_048287462.1 (+) 1 473 Myodes glareolus
XP_028631132.1 (+) 1 466 Grammomys surdaster
XP_040837182.1 (+) 1 482 Ochotona curzoniae
XP_013005695.1 (+) 1 474 Cavia porcellus
XP_045013551.1 (+) 1 481 Jaculus jaculus
XP_019827818.1 (+) 1 415 Bos indicus
XP_019827819.1 (+) 31 445 Bos indicus
XP_031809936.1 (+) 1 472 Sarcophilus harrisii
XP_043836726.1 (+) 1 472 Dromiciops gliroides
XP_044517493.1 (+) 1 474 Gracilinanus agilis
XP_007474443.1 (+) 1 474 Monodelphis domestica

Figure 3.3: BPIFB1 multiple sequence alignment. Homologous mammals derived FASTA sequences of BPIFB1 were aligned using the online tool, BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). Conserved regions are highlighted in grey and amino acid differences are shown as red. Blue lines feature regions of sequence inserted in an individual protein sequence.

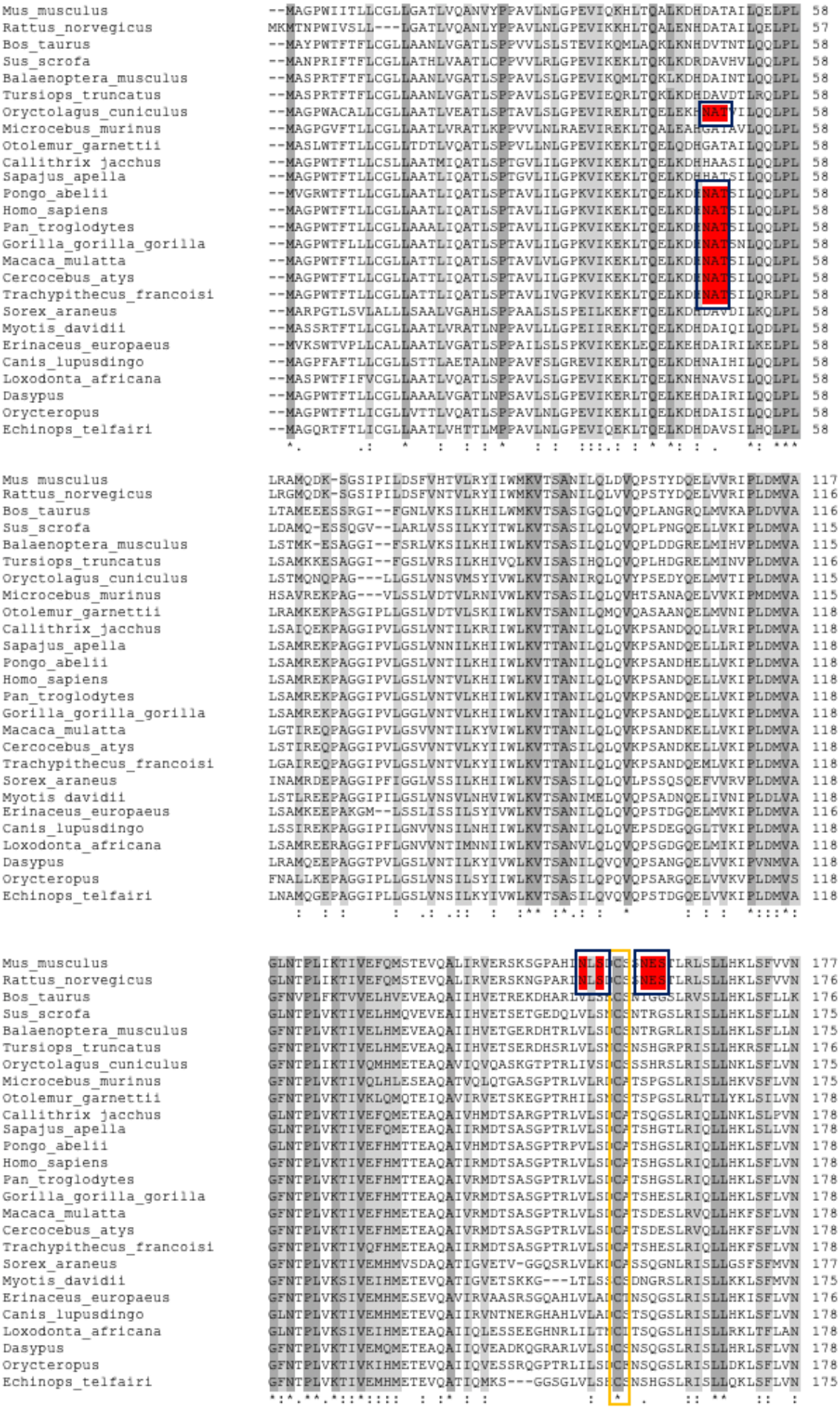
The most striking observation is the difference between the primate BPIFB1 sequences and those derived from other species. Not only can it be clearly seen that there is highest homology in this group, but another striking difference is the variation in protein lengths (Figure **3.3**) with those from non-primates being shorter.

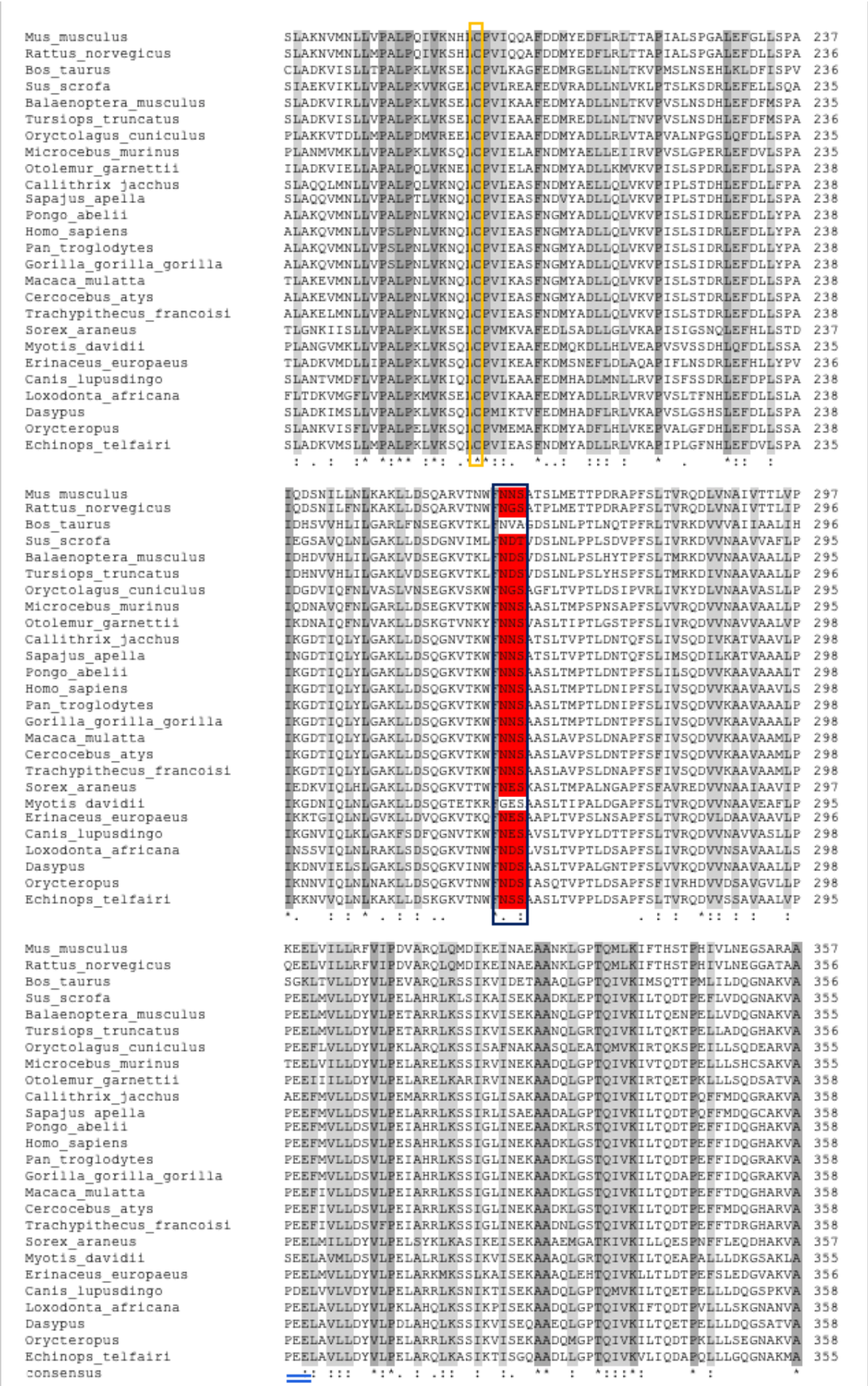
Next, I aimed to analyse the level conservation of BPIFB1 in more detail across a selected 26 mammalian species using the multiple sequence alignment and employing the Clustal Omega alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo) (Figure **3.4**) To achieve this, sequences were employed that have corresponding examples within the phylogenetic tree shown above and which cover all of mammalian phylogeny (Figure **3.2**). These sequences were extracted from Genbank as RefSeq sequences and were visually examined to ensure that they appeared to be complete predictions.

Overall, the most conserved (*) regions span amino acid residues 36 - 257 (the BPI domain 1) and 334 - 471 (the BPI domain 2/ LBP BPI CETP C). These contain predominantly glutamine (G), asparagine (N), proline (P), leucine (L), lysine (K), valine (V), alanine (A), cysteine (C) and phenylalanine (F) amino acid residues (Figure 3.4).

The C-terminal extension is found in higher land mammals such as primates and humans but is absent in lower species such as rodents (Figure **3.4**). This region, which exhibits conservation in primates, contains a ten amino acid sequence that is rich in serine (S) and prolines (P) residues, these are typical sites of (Manning *et al.*, 2002a; Manning *et al.*, 2002b) phosphorylation.

The MSA also illustrates the conservation pattern of the N-Glycosylation sites (**N**XT or **N**XS) among 26 species, which are confirmed through Uniprot, and are indicated by red colour within a blue box. N-Glycosylation sites **N**AT, **N**LS, **N**ES, **N**NS and **N**VS are identified in 8, 2, 2, 23 and 11 species respectively. Interestingly, the glycosylation site at position 264 (human) is found conserved across all species (with the exception of cattle and hedgehog), whereas the other glycosylation site positions vary. In addition, N-Glycosylation site at position 48 appears to be primate specific (with the exception of rabbits). Notably, rodents possess a fourth site, which also forms a cluster of two closely positioned sites (153 and 160), which are specific to mice and rats (Figure **3.4**).


Mus_musculus
Rattus_norvegicus
Bos_taurus
Sus_scrofa
Balaenoptera_musculus
Tursiops_truncatus
Oryctolagus_cuniculus
Microcebus_murinus
Otolemur_garnettii
Callithrix_jacchus
Sapajus_apella
Pongo_abelii
Homo_sapiens
Pan_troglodytes
Gorilla_gorilla_gorilla
Macaca_mulatta
Cercocebus_atys
Trachypithecus_francoisi
Sorex_araneus
Myotis_davidii
Erinaceus_europaeus
Canis_lupusdingo
Loxodonta_africana
Dasypus
Orycteropus
Echinops_telfairi


Mus_musculus
Rattus_norvegicus
Bos_taurus
Sus_scrofa
Balaenoptera_musculus
Tursiops_truncatus
Oryctolagus_cuniculus
Microcebus_murinus
Otolemur_garnettii
Callithrix_jacchus
Sapajus_apella
Pongo_abelii
Homo_sapiens
Pan_troglodytes
Gorilla_gorilla_gorilla
Macaca_mulatta
Cercocebus_atys
Trachypithecus_francoisi
Sorex_araneus
Myotis_davidii
Erinaceus_europaeus
Canis_lupusdingo
Loxodonta_africana
Dasypus
Orycteropus
Echinops_telfairi
consensus

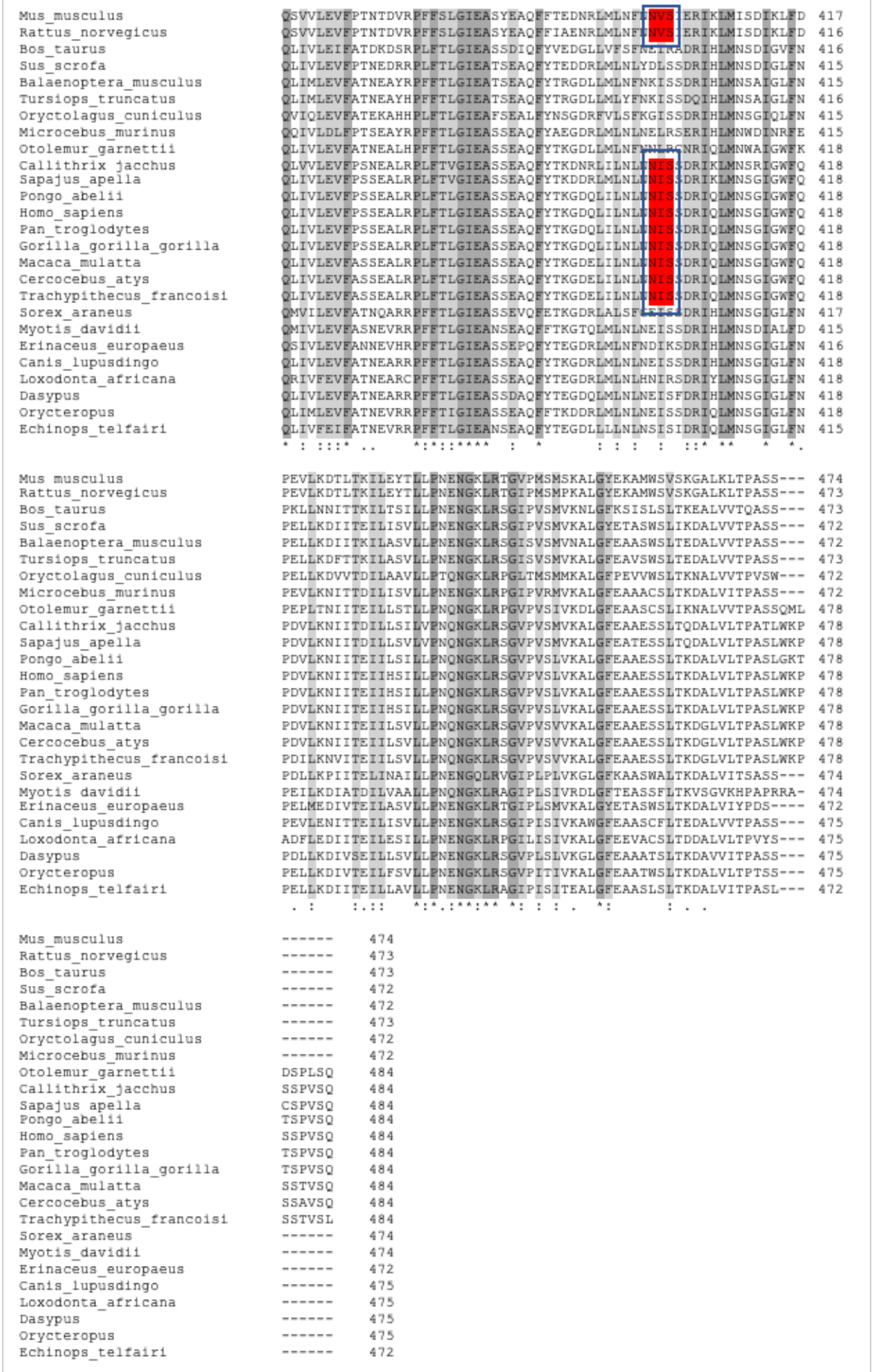

Mus_musculus
Rattus_norvegicus
Bos_taurus
Sus_scrofa
Balaenoptera_musculus
Tursiops_truncatus
Oryctolagus_cuniculus
Microcebus_murinus
Otolemur_garnettii
Callithrix_jacchus
Sapajus_apella
Pongo_abelii
Homo_sapiens
Pan_troglodytes
Gorilla_gorilla_gorilla
Macaca_mulatta
Cercocebus_atys
Trachypithecus_francoisi
Sorex_araneus
Myotis_davidii
Erinaceus_europaeus
Canis_lupusdingo
Loxodonta_africana
Dasypus
Orycteropus
Echinops_telfairi

Figure 3.4: Multiple sequence alignment of BPIFB1 protein across 26 mammalian species. Highly conserved amino acid residues are shown in dark grey colour, while semi-conserved are represented in light grey. N-Glycosylation sites NAT, NLS, NES, NNS and NVS are highlighted in red, whereas cysteine residues involved in formation of a disulphide bond are highlighted with a yellow box. A star (*) indicates completely conserved residues, a colon (:) residues which are semi conserved and share the same hydropathy and similar size, while period (.) indicates weakly similar residues, where the size or hydropathy has been conserved during the course of evolution.

It has previously been shown that BPIF proteins contain a two cystine residues that form a disulphide bond, which most likely carry an important function due to its conserved nature across the LBP/BPI family and related members (Beamer *et al.*, 1997; Beamer *et al.*, 1998; Stenvik *et al.*, 2004; Krasity *et al.*, 2011). The cystines involved in this potential disulphide bond formation are conserved among all 26 species in the MSA, present at position 137 and 174 (with -1 or -2 residues) in alignment, suggesting that they serve an important role in the tertiary structure and potentially biological activity (Figure **3.4**).

Next, it was of interest to investigate BPIFB1 on protein level with regards to genetic variations and potential clinical implications.

3.3.1.1 Genetic variations of the human BPIFB1

In the preceeding section I showed that BPIFB1 proteins from across mammalian evolution conserve a relatively limited number of amino acids. It seems lilely that these will be key structural or functional residues for the protein. Therefore, it was of interest to analyse potential genetic modification and/or protein modifications in a more systematic manner.

Studies of genetic variations seen with in *BPIFB1* might allow us to draw conclusions about potential hotspot regions or disease associations. Amino acid residues that are critical for function might be expected to have fewer variants. With regards to BPIFB1, this gene shows an indicated somatic mutation frequency of 0.9% (https://www.cbioportal.org) (Figure **3.5**). Overall, most identified alterations are missense mutations (green pins) or splice variant mutations (orange pins), which occur commonly between exon 11-12 and 13-14 (Figure **3.5**A).

With regards to missense mutations, these occur across the entire protein in an even manner and at a similar frequency, including the N-terminal signal peptide. Consequently, no hotspot regions could be identified (Figure **3.5**B). Taking a closer look at these recorded amino acid modifications, it appears that observed mutations occur across all amino acids with a similar frequency (Figure **3.5**C).

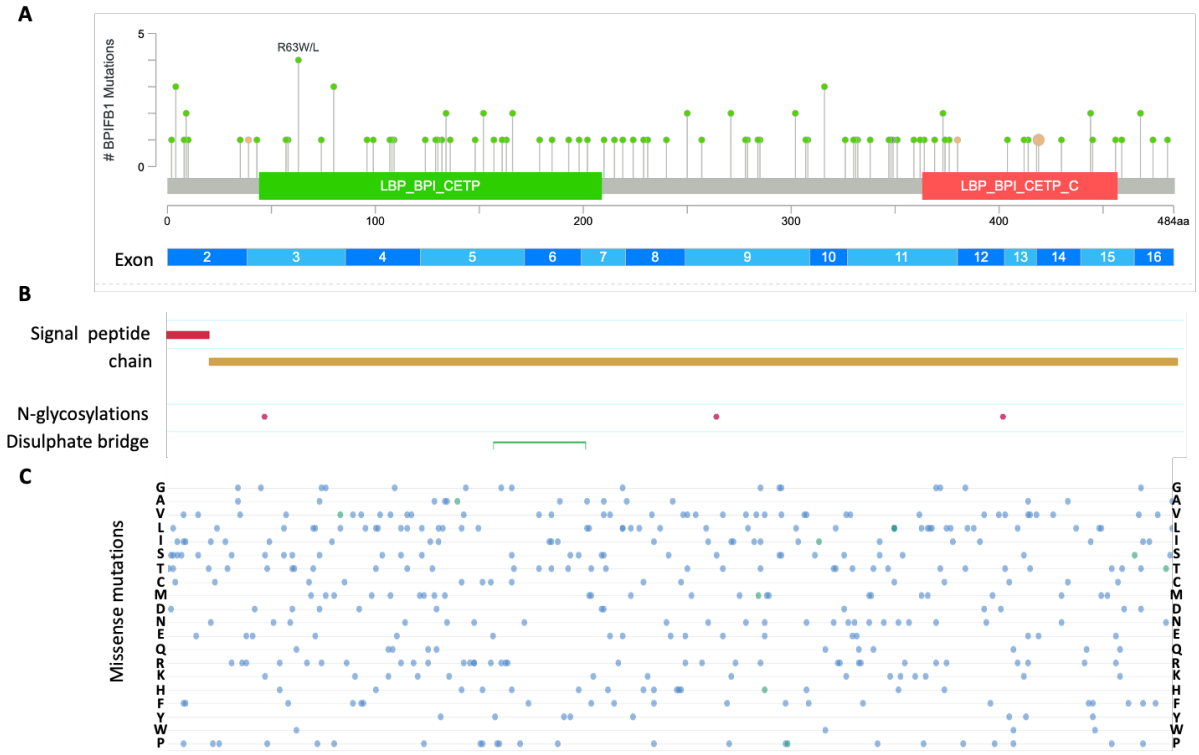
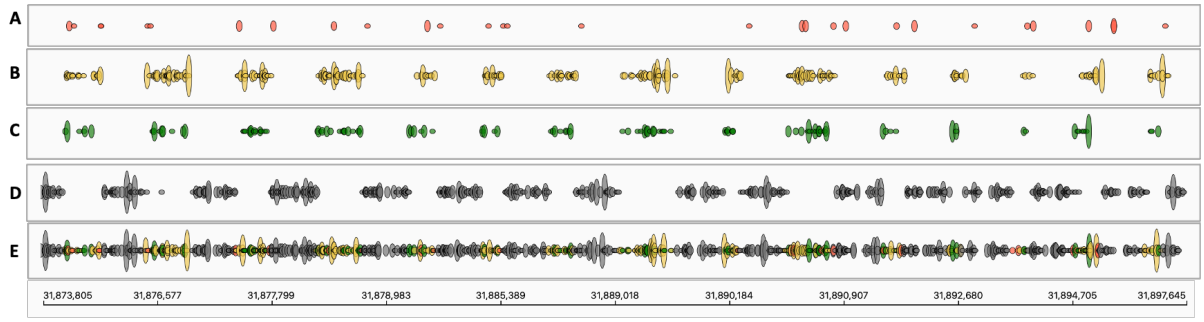


Figure 3.5: Human BPIFB1 protein domain structure and genetic variations. A) BPIFB1 identified mutations frequency has been illustrated by stick and ball symbols, highlighting the amino acid position and respective corresponding exon, where missense mutations are shown in green and splice variants in orange. This figure was generated by cBioPortal (https://www.cbioportal.org). **B)** Cartoon of BPIFB1 highlighting structural features such as the N-terminal signal peptide, position of glycosylation motifs and disulphide bridge formation. **C)** Identified missense mutation are obtained from the Pharos website (https://pharos.nih.gov) and illustrate which amino acids have been found altered at their respective position.

To take a closer look at these reported missense mutations I examined BPIFB1 gene modifications using the gnomAD v2.1.1. data base (https://gnomad.broadinstitute.org/). From this it became apparent that 938 variation events have been recorded in the gene (Figure **3.6**E). These modifications were classified as 30 cases of predicted loss of function mutations (Figure **3.6** and Figure **3.5**A), missense or in frame insertions or deletions in 271 events (Figure **3.6**B), and 121 synonymous mutations (Figure **3.6**C). Other genetic alterations such as splice or intronic modifications made up the remaining 516 events (Figure **3.6**D).

Using the Genomic variation as it relates to human health (ClinVar) data base, which holds genomic variation information and its relationship to human health (https://www.ncbi.nlm.nih.gov/clinvar/), four benign missense mutations have been recorded to date (accessed 17.04.2022), which contain the synonymous mutation P4P, and the three missense mutations K39R, P113S and T140A (Figure **3.6**F).



F

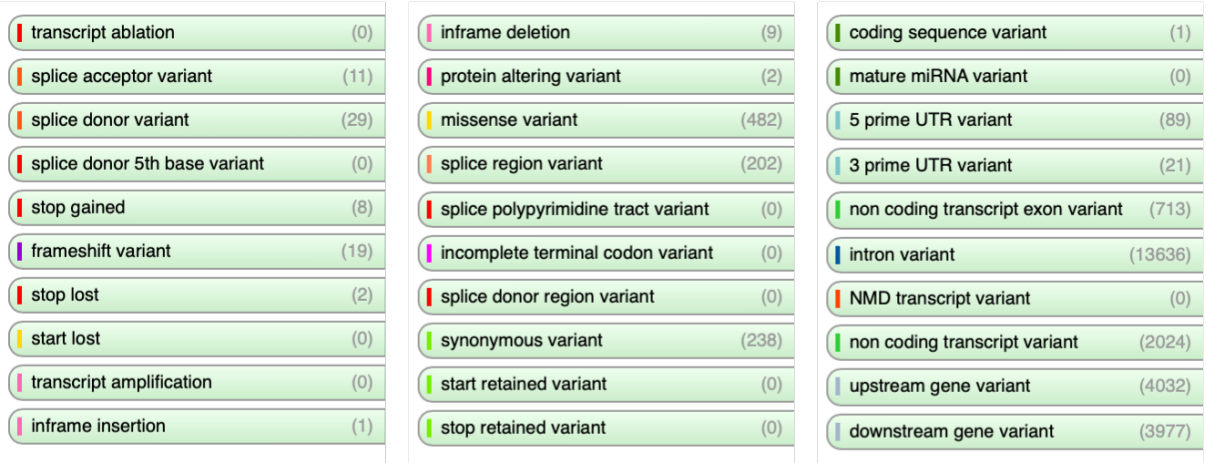
Variant ID	Consequence	Annotation	Clinical significance	Allele count	number of homozygoytes
20-31873891-G-A	p.Pro4Pro	synonymous	Benign	193	0
20-31876547-A-G	p.Lys39Arg	missense	Benign	738	13
20-31876616-T-C	p.Met62Thr	missense		144	0
20-31876681-A-G	p.Ile84Val	missense		84247	13737
20-31877713-A-T	p.Ile94Phe	missense		4293	43
20-31877770-C-T	p.Pro113Ser	missense	Benign	1015	20
20-31878815-A-G	p.Thr140Ala	missense	Benign	1367	14
20-31878894-G-A	p.Arg166His	missense		925	10
20-31879747-G-A	p.Val184Ile	missense		235	1
20-31885361-G-A	p.Gly210Ser	missense		220	0
20-31889141-G-A	p.Val284Met	missense		479	2
20-31889150-G-C	p.Asp287His	missense		5952	598
20-31889183-T-C	p.Ser298Pro	missense		11820	2174
20-31890175-G-T	p.Ser313Ile	missense		10371	1737
20-31891800-G-A	p.Asp393Asn	missense		108	0
20-31894766-C-T	p.Phe456Phe	synonymous		9474	231
20-31894789-C-G	p.Thr464Ser	missense		8597	178
20-31897554-G-C	p.Ser479Thr	missense		160318	47173

Figure 3.6: Overview of BPIFB1 protein variations. The gnomAD v2.1.1. data base was used (https://gnomad.broadinstitute.org/) to extract information regarding BPIFB1 recorded mutations and its respective location on the gene locus (bottom panel), highlighting loss of function mutations (30 events) **(A)**, missense or in frame insertions or deletions (271 events) **(B)**, synonymous (121 events) **(C)**, or other genetic alterations such as slice or intron modifications (516 events) **(D)**, leading to a total modification count of 938 events **(E)**. **F)** Selected most common variations have been extracted and assembled in a table, showing the frequency of missense or synonymous alterations by allele count and penetration to both alleles in number of homozygotes. The clinical significance has been annotated where available through ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), Genome build GRCh37 / hg19, used Ensembl gene ID ENSG00000125999.6, with the Ensembl canonical transcript ENST00000253354.1.

Furthermore, additional missense mutations were selected in this table, based on their high frequency of mutation (allele count) and the homozygote penetration (number of homozygotes). The variant ID also highlights the specific amino acid alteration (Figure **3.6**F). To show in a more detailed manner, all recorded variations of human BPIFB1, Table *3.1*

displays a summary including the frequency of events. The most common recorded alterations are intron variants (13636), as well as upstream (4032) or downstream (3977) gene variations at a similar rate or non-coding transcript variants (2024) versus non-coding transcript exon variants (713). On protein level, the highest number of variations occurred as missense (482) or as synonymous (238) modification. Frameshift changes (19), insertions (1) or deletions (9) occurred the least frequent (*Table 3.1*).

Table 3.1: Human BPIFB1 gene variation table. overview of occurring BPIFB1 variations. From http://www.ensembl.org/Homo_sapiens/



Taken together, I have shown that *BPIFB1* is found in mammals with no clear orthologs identified in birds, reptiles, fish and invertebrates. Although *BPIFB1* orthologs are present in all classes of mammals, more divergent orthologs are present in marsupials. As expected, primate *BPIFB1* sequences are closest related to the human counterpart. Moreover, a C-terminal extension is found in primates and humans, but is absent in other species, including rodents. This region may be functionally important. Finally, the post-translational N-Glycosylation modification sites show a variable conservation pattern, with only one position (aa 264 in human) found to be conserved across all species.

With respect to genetic modification or protein mutations, alterations were discovered at an even manner and at a similar frequency, indicating the absence of hot spot regions. The use of gnomAD and ensembl.org databases allowed to further study coding vs non-coding regions of the gene. However, these databases likely differ in sample number, sample background and sample bias.

After this detailed exploration of BPIFB1 sequence and protein variations in the human proteins, I next investigated BPIFB1 structure at the protein level and compared it to the murine counterpart.

3.3.2 BPIFB1 protein structure

BPIFB1 contains an N-terminal signal peptide of which are the first 21 amino acids (MAGPWTFTLLCGLLAATLIQA) are illustrated, followed by the remainder of 463 residues (Figure **3.5**B) (Bingle and Craven, 2002). As BPIFB1 is a secreted protein released into the extracellular space, the signal peptide is required to direct correct entry of the protein into the secretory pathway (Gao *et al.*, 2015). Comparing the human and murine N-terminal signal peptides, they show an 81% identity (17 out of 21 aa), or 85% positive overlap (18 out of 21 aa) (Figure **3.7**).

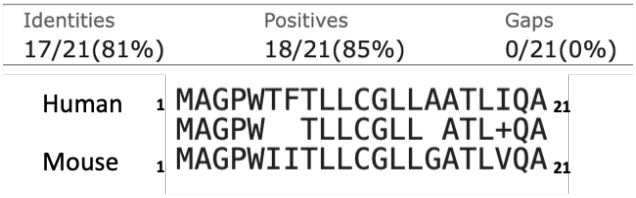


Figure 3.7: Comparison of the N-terminal BPIFB1 signal peptide between human and mouse. The initial 21 residues containing the signal peptides of both species have been aligned. Amino acids with the same chemical properties are indicated by a +. Lower case numbers denote the amino acid position of the respective protein.

3.3.2.1 Comparison of the mouse and human BPIFB1 protein

In order to investigate the conservation between murine and human BPIFB1 further, the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used. The sequence alignment shows a 59% conservation on protein level, which is lower than the average conservation between orthologues of about 70 % (Figure **3.8**A) (Waterston *et al.*, 2002).

Genomewide comparative analysis of sequence conservation in orthologues with structural domains has shown that such proteins generally share much greater levels of identity than it is observed in *BPIFB1 (*Waterston *et al*., 2002). An alternative measure of the evolutionary pressure on proteins is the Ka/Ks ratio, which compares the number of nonsynonymous

substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks) (Hurst, 2002).

It has been highlighted that higher Ka/Ks ratios are often associated with proteins involved in host defense (Emes *et al*., 2003). BPIFB1 has a Ka/Ks ration of 0.32 (Bingle *et al.*, 2004) and as a family BPIFs have Ks/Ks ratios higher than other rapidy evolving proteins (Waterston *et al.*, 2002; Bingle *et al.*, 2004) .

Overall, homology is a superficial indication. If structure and function is conserved, highly structured, domain-containing regions are often more compared to less-structured inter-domain regions based on secondary structure level rather than primary protein sequence. In other words, protein function might be conserved despite a divergent primary protein sequence. Conserved residues are highlighted in dark grey colour, while semi-conserved residues are shown in light grey (Figure **3.8**B). A star (*) indicates highly conserved residues, a colon (:) residues which are semi conserved and share the same hydropathy and similar size, while period (.) indicates less similar residues, where only the size or hydropathy has been conserved. An example of a strictly conserved feature amongst both proteins are the cysteine amino acid residues involved in disulphide bond formation, indicated by an orange box, connecting aa 158 - 201in the human protein and 157-200 in the murine counterpart (Figure **3.8**B).

Investigating the more conserved N-terminal part of the mouse protein in further detail, it was observed that it contains extra aromatic residues between positions aa76-83. Specifically, phenylalanine at position 76 (F76) and tyrosine at position 83 (Y83) (Figure **3.8**B. It is possible that these additional aromatic residues serve as a mouse-specific protein-protein interaction site, whereas analogues to that the extended C-terminus observed in higher mammals could support this function (Figure **3.8**B).

A

Species	Peptide length	% identity (Protein)	% coverage	Genomic location
Human (*Homo sapiens*)	484 aa	58 %	97 %	20:33273480-33309871
Mouse (*Mus musculus*)	474 aa	59 %	100 %	2:154032738-154062289

B

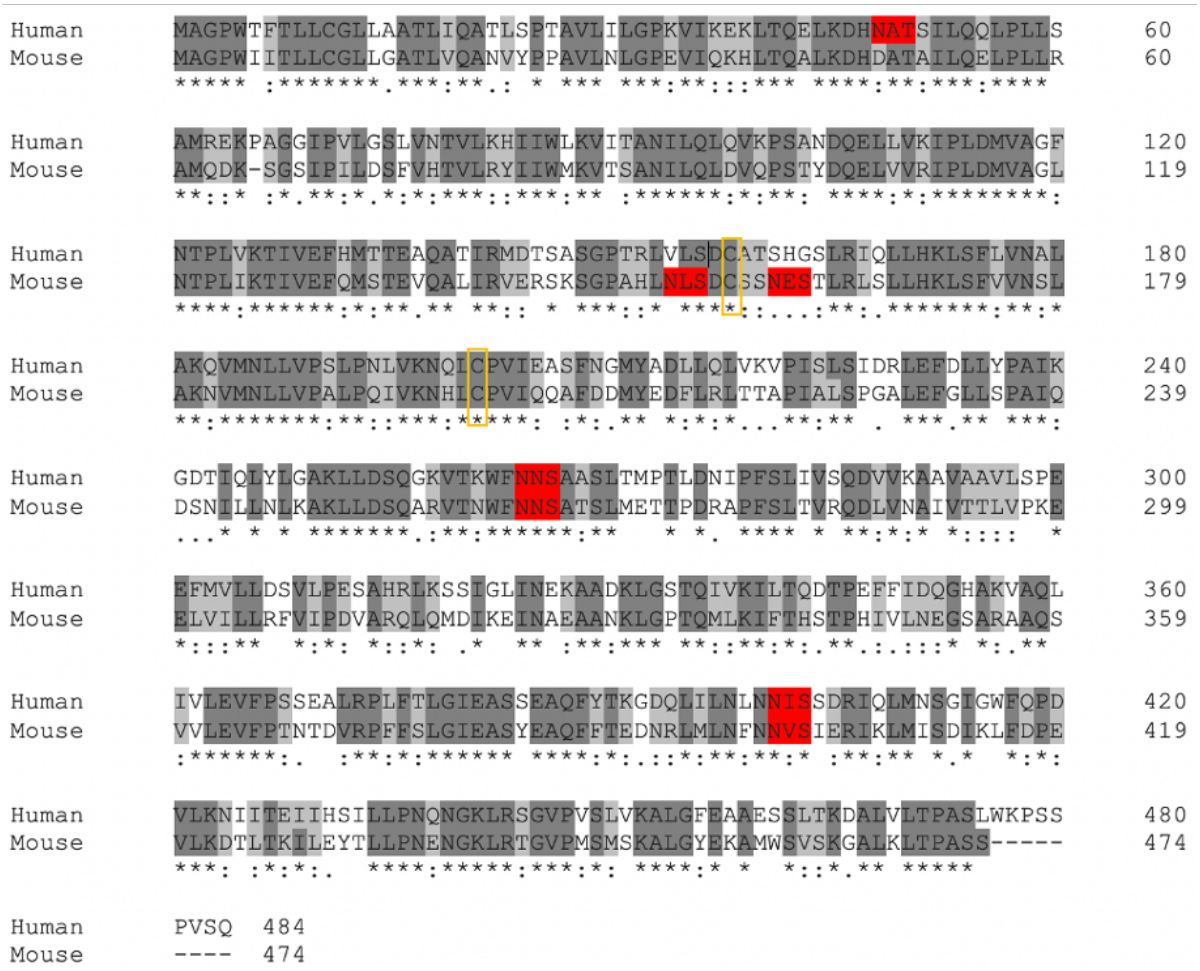


Figure 3.8: Sequence alignment between hBPIFB1 and mBPIFB1. A) Comparison of mouse (ENSMUSG00000027485) versus human BPIFB1 (ENSG00000125999) proteins regarding amino acid length, percent homology (% identity), coverage and genomic location. **B)** Conserved and semi-conserved residues are shown in dark grey and light grey colour, respectively. Cysteine amino acid residues involved in disulphide bond formation are surrounded by a yellow box, while N-glycosylation motifs are highlighted in red. A star (*) indicates highly conserved residues, a colon (:) residues which are semi conserved and share the same hydropathy and similar size, while period (.) indicates weakly similar residues, where only the size or hydropathy has been conserved. This sequence alignment was generated using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Another feature which is divergent between both orthologues is the number of putative N-glycosylation motifs, which are highlighted in red (Figure **3.8**B). Human BPIFB1 contains three N-glycosylation sites at aa48, aa264 and aa401, whereas in the murine protein aa153,

aa160, aa263 and aa400 have been identified. Interestingly, two sites are conserved between both species 264/263 (NNS) and 401/400 (NIS/NVS), whereas the other sites seemed to have evolved independently. In murine BPIFB1, these two sites centre around the cysteine responsible for forming the conserved disulphide bridge (Figure **3.8**, dark yellow box).

In order to study the protein sequences more in detail with respect to their secondary structures, an alignment between murine and human BPIFB1 proteins was generated using the UCSF Chimera online tool (https://www.cgl.ucsf.edu/chimera/download.html), aligning aa 84-471 for human and aa 83-470 for mouse (Pettersen *et al.*, 2004) and manually annotating alpha helices and beta sheets (Figure **3.9**). Of note, no secondary structures could be identified below aa 84 or 83 for human or mouse respectively.

Human BPIFB1 contains seven alpha helices and 18 beta sheets, compared to five alpha helices and 15 beta sheets in murine BPIFB1 (Figure **3.9**). Structural differences are observed at position V142-R144 (VER) in the fourth beta sheet in mouse and at position S155 converted into loop with respect to human BPIFB1. The sixth beta sheet is extended from position D156-L168 (DCSSNESTLRLSL). There is additional alpha helix in human BPIFB1 spanning from K261-F263 (KWF) and Q286-L297 (QDVVKAAVAAVL), while I283-S284 is formed by a beta sheet, A382-S384 (ASS) is converted into a loop in human BPIFB1.

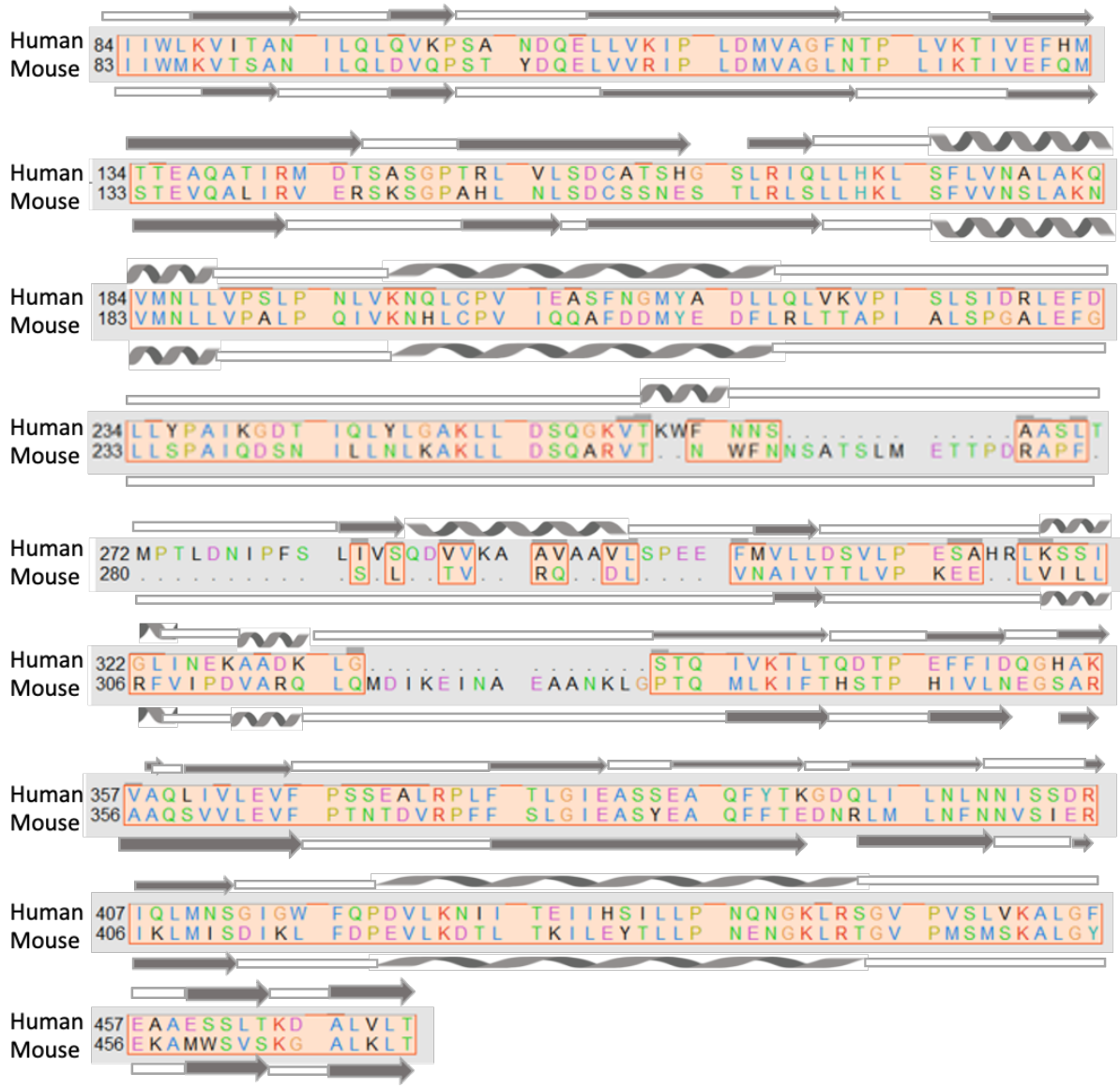


Figure 3.9: Structural conservation patterns between hBPIFB1 and mBPIFB1. Grey ribbons indicate alpha helices and beta sheets in structure. The UCSF Chimera online tool (https://www.cgl.ucsf.edu/chimera/download.html) was used to acquire structural information, which was manually assigned to the alignment. Proteins were aligned from aa 84 - 471 for human BPIFB1 and aa 83 - 470 for mouse BPIFB1, as the structure before aa 83/84 was too flexible for crystallography approaches.

Assembling this information in a more visual manner, the UCSF Chimera online tool 1.1.12 was used to model murine and human BPIFB1 secondary structures (Figure 3.20A).

Due to its' known similarity with other BPIF proteins, BPIFB1 exhibits a mixed secondary structure containing a combination of both alpha helices and beta sheets, specifically 22 % alpha helices and 32 % beta sheets. Highly conserved amino acid residues spanning from aa 119 - 133 represent block A, while aa 187 - 238 form block B. These residues include G119,

N121, L124, K126, V129 and are mostly hydrophobic in nature (Block A), while L188, V196, L200 demonstrate hydrophobicity in Block B, and approximately 36% of the residues are predicted to be solvent exposed. Putative protein binding sites were identified by Predict Protein Open software (https://predictprotein.org) between T274-D276 (TLD), I283-S285 (IVS), L306-P311 (LDSVLP), N424-I426 (NII), whereas potential DNA binding sites were predicted in the same manner between aa W5-L9 (WTFTL), G74 and L470-P478 (LTPASLWKP). No putative RNA binding sites have been identified (Figure **3.11**).

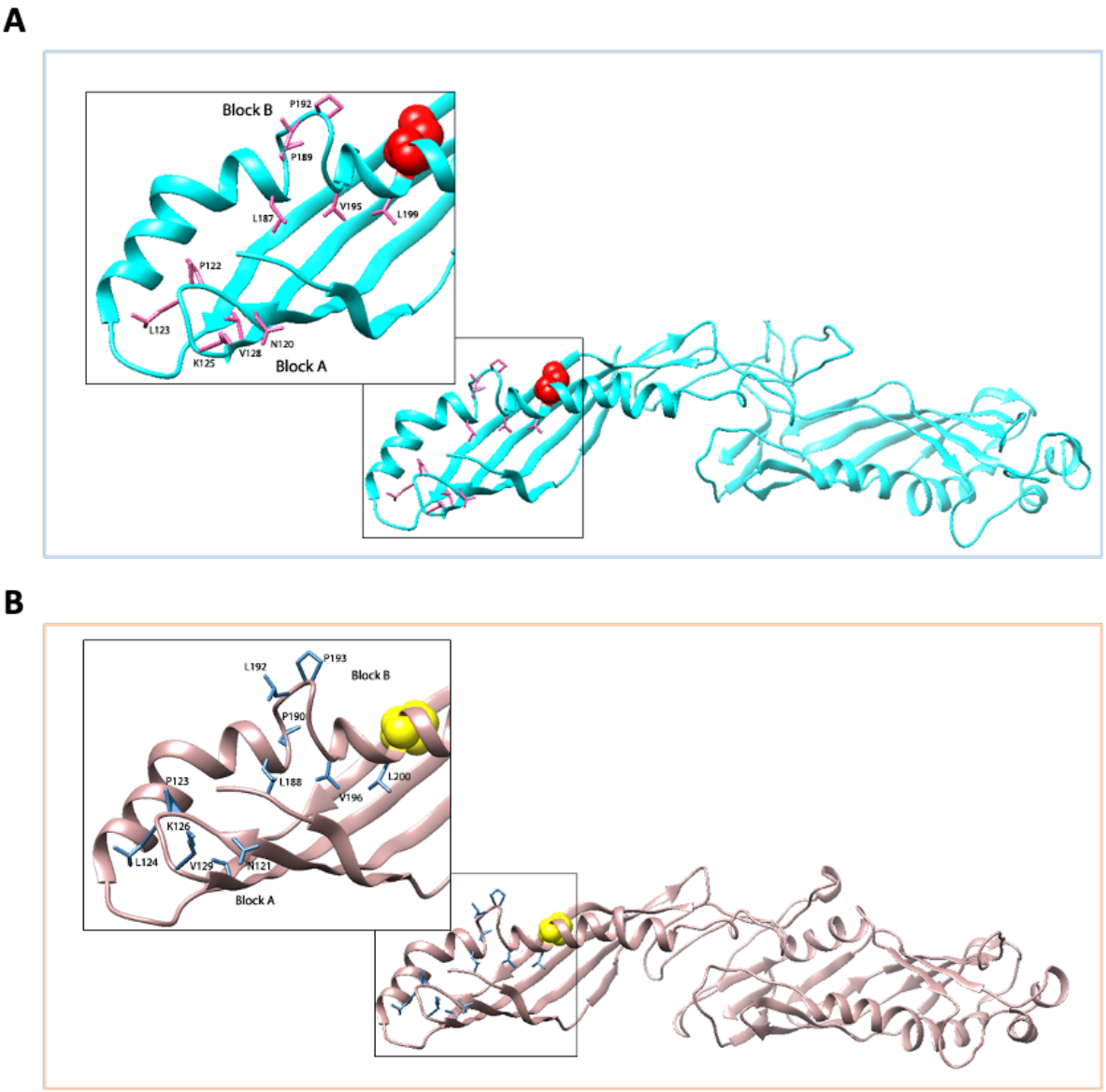


Figure 3.10: Structural conservation of hBPIFB1 versus mBPIFB1. Comparison of structural conservation between mouse **(A)** and human BPIFB1 **(B)** using Chimera 1.1.12 (https://chimera.software.informer.com/1.5/). Highly conserved residues are indicated with purple **(A)** or blue **(B)** sticks. Disulphide bonds are shown in red **(A)** or yellow **(B)**.

Murine BPIFB1 is structurally similar to its human counterpart, displayed in rose pink, illustrating highly conserved residues with blue sticks (Figure **3.10**B). Highly conserved residues span from aa residue 118 – 132 representing Block A, while aa 186-237 are located in Block B. These residues include L123, K125, V128, N120, G119 are mostly hydrophobic nature, while L187, V195, L199 exhibit hydrophobicity in Block B (Figure **3.10**). Thirty-seven percent of the residues are solvent exposed and putative protein binding sites were identified by Predict Protein Open software between residues L305-V308 (LRFV), L415-F416, Y431-T432, which are divergent to the regions predicted in the human protein (Figure **3.11**). Predicted DNA binding sites were identified by similar means for residues W5-L9 (WIITL), D73-S74 and T443-S448 (TGVPMS), of which T443-S448 is diverse between both species.

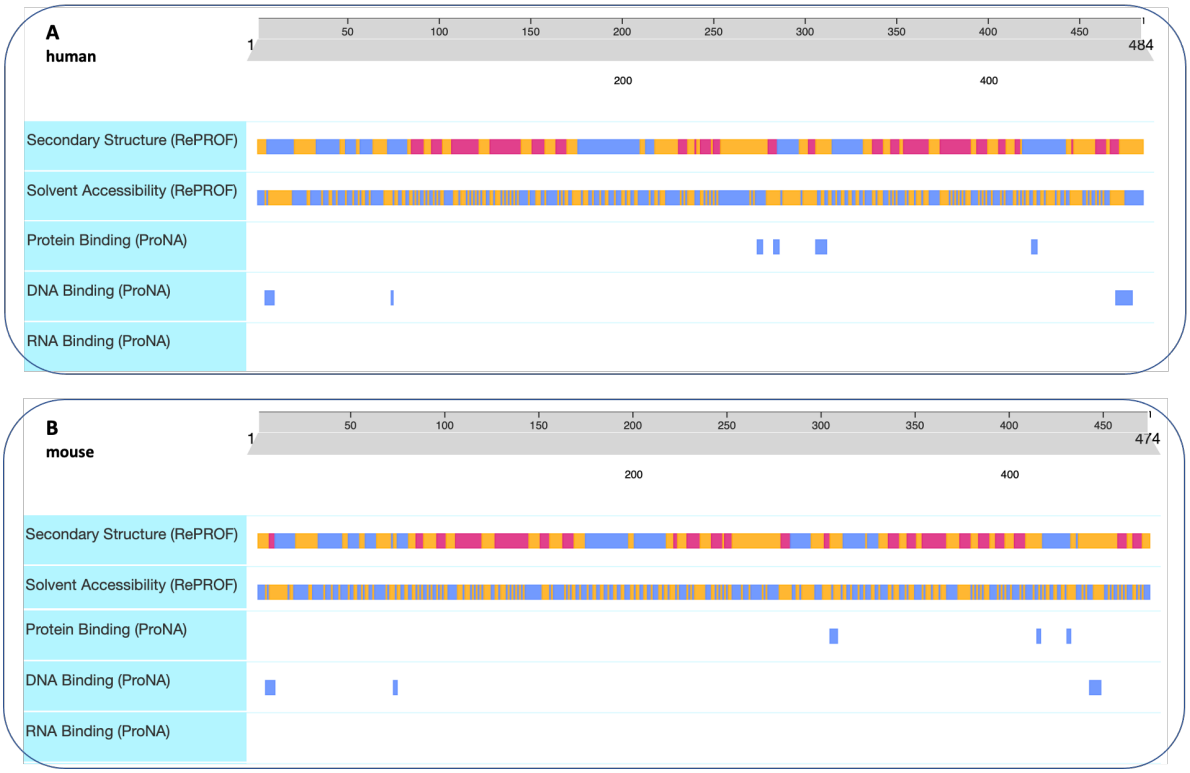


Figure 3.11: Predicted structure and interaction sites of BPIFB1 protein. The online tool Predict Protein Open software (https://predictprotein.org) was utilised to explore human **(A)** or mouse **(B)** BPIFB1 secondary structure (top panel), where purple sections depict alpha helices and red sections illustrate beta strands. Further, solvent accessibility has been predicted with blue stretches showing solvent exposure and orange areas highlights buried amino acids (second panel). Putative protein binding (third panel) sites, DNA binding sites (fourth panel) or RNA binding regions (bottom panel) are highlighted in blue. See text for detailed respective amino acid positions.

Finally, the online tool SWISS MODEL (https://swissmodel.expasy.org) was used to superimpose the human and murine BPIFB1 secondary structure to assess how similar the two proteins are (Figure **3.8**A).

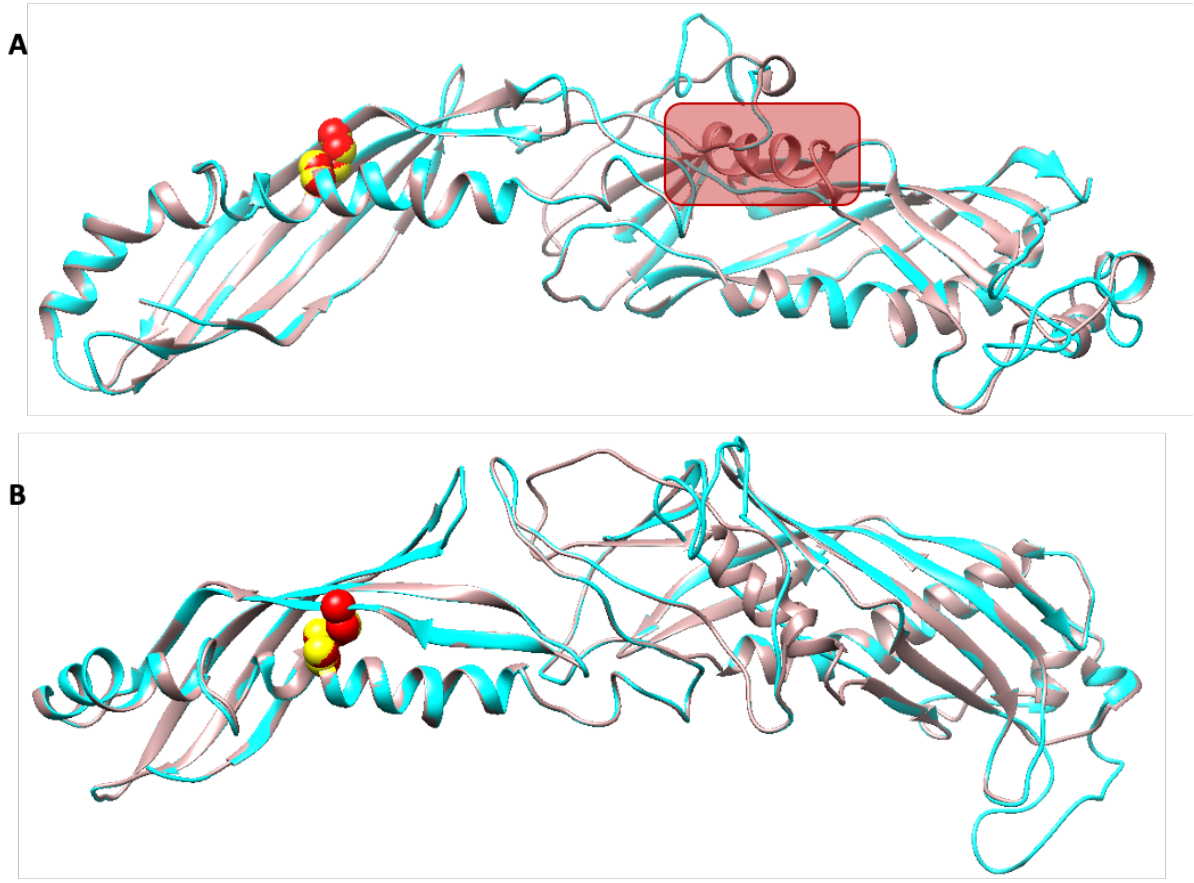


Figure 3.12: Superimposition of modelled hBPIFB1 and mBPIFB1. Mouse BPIFB1 (mBPIFB1, residues 84-470, cyan) was modelled after human BPIFB1 (hBPIFB1, residues 23-471, pink) using SWISS MODEL (https://swissmodel.expasy.org). Disulphide bonds are colored in red and yellow for mBPIFB1 and hBPIFB1, respectively. **A)** Superimposition of hBPIFB1 (Pink) and mBPIFB1 (cyan) with anterior orientation. **B)** Superimposition of hBPIFB1 (Pink) and mBPIFB1 (cyan) with posterior orientation. Red box indicates the alpha helix predicted to be only present in hBPIFB1.

Mouse BPIFB1 using residues 84-470 was modelled after human BPIFB1 employing amino acid residues 23-471, and the superimposition demonstrated strong structural similarities between the two proteins with a root-mean-square deviation (RMSD) of 0.171, suggesting a good alignment between these two proteins (Figure **3.12**). Despite noticeable differences in the primary amino acid sequence, both proteins demonstrate a high conservation of secondary structure elements or tertiary structure components such as the disulphide bond formation (Figure **3.12**). It is apparent that the N-terminus located to the left, containing the disulphide

bonds, shows a higher structural similarity between both species. Whereas the C-terminus on the right exhibit distinct species-specific loops, which do not contain alpha helix or beta sheet secondary structure elements and thus might be more flexible, interdomain regions (Figure **3.12** red box).

In summary, I explored in this section that BPIFB1 is a secreted protein released into the extracellular space, where the signal peptide is required to direct entry of the protein into the secretory pathway. The conservation on murine versus human on protein level is s lower than the average conservation between orthologues judged on primary sequence, with conserved N-glycosylation sites 264/263 (NNS) and 401/400 (NIS/NVS). Protein structure predictions revealed that the murine Bpifb is structurally similar to its human counterpart. Taken together, despite noticeable differences in the primary amino acid sequence, both proteins appear to demonstrate a high conservation of predicted secondary structure elements or tertiary structure components such as the disulphide bond formation. The N-terminus is predicted to gerater structural similarity between both species.

It is important to remember that the structural analysis is based on predictions as to date no crystal sturucture for BPIFB1 has been solved.

3.3.3 BPIFB1 expression and localisation

Next, it was important to assess where and potentially when this protein is expressed, to generate a more detailed expression profile of BPIFB1. This in turn would allow to draw more conclusions regarding its biological function or disease associations.

The majority of published data suggests that BPIFB1 protein is found most highly expressed in the trachea and lung, and weakly present in salivary glands, the duodenum and the stomach (Bingle *et al.*, 2010; Musa *et al.*, 2012; Bingle *et al.*, 2012; De Smet *et al.*, 2018; Li *et al.*, 2020). To investigate where BFIPB1 RNA and protein levels are found, I recovered data from the Human Protein Atlas project (HPA) (https://www.proteinatlas.org), an online database which aims to map the entire human proteome across cells, tissue and organs employing various omics technologies. Analysed tissues are divided into colour-coded groups according to functional features they have in common. Images of selected tissues give a visual summary of the protein expression profile furthest to the right (Figure **3.13**). The data shows that BPIFB1 exhibits a limited tissue distribution with no evidence for expression in many tissues.

As BPIFB1 has been implicated to potentially play a role in innate immunity in mouth, nose and lungs, as well being structurally related to BPI proteins, and able to bind bacterial LPS (Bingle and Craven, 2002; Bingle and Craven, 2003), it was surprising to note that RNA expression levels was not highest in respiratory tract tissues, and the highest amounts were detected in the cervix. Similar expression levels were detected for gastrointestinal tract tissues with a main tissue expression cluster for stomach and digestion (duodenum), and lower expression in proximal digestive tract tissues such as salivary gland and oesophagus. Very small amounts have been detected in the urinary bladder, but no expression was detected on protein level. Moreover, there are many tissues that do not express it at all (Figure **3.13** and Figure **3.14**). Overall, it seems that BPIFB1 is present in tissues with mucosal epithelia that have an exposure to the external environment and hence requires a higher immune protection. To assess BPIFB1 protein expression, the Human Gene Atlas data base generates protein expression scores based on a best estimate of the accurate protein expression from a knowledge-based annotation. Annotated protein expression profiles are performed using single antibodies as well as independent antibodies, directed against different, non-overlapping epitopes. For independent antibodies, the immunohistochemical data from all the different antibodies are taken into consideration.

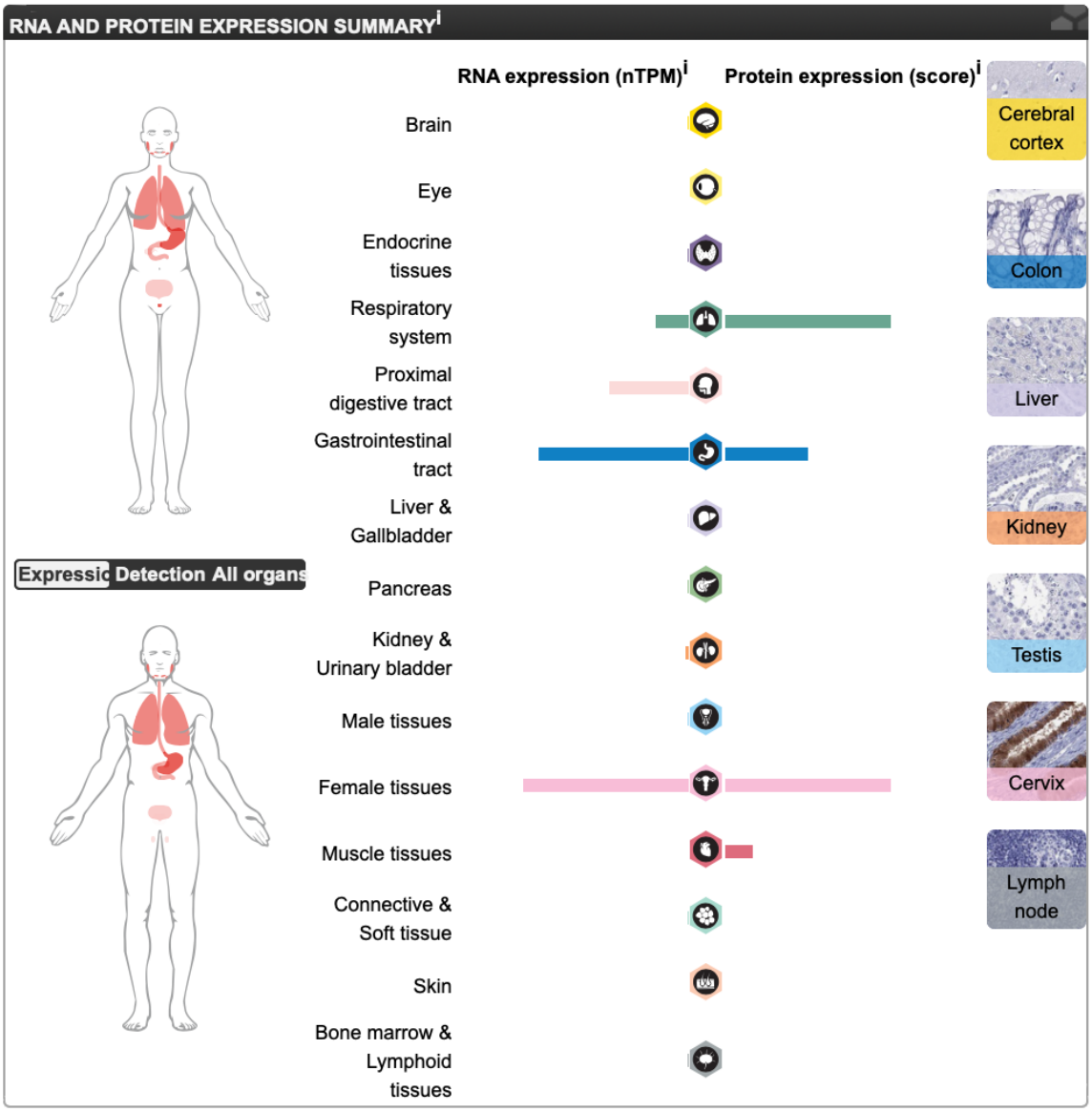


Figure 3.13: Graphical display of human BPIFB1 expression in tissue groups. The Human Gene Atlas data base (https://www.proteinatlas.org) was used to display the consensus RNA data (left hand side) based on normalized transcripts per million expression (nTPM) values. RNA-Seq data presented stems from the Human Protein Atlas (HPA) and the Genotype-Tissue Expression (GTEx) project. Colour-coding is categorised according to tissue groups with functional features in common. Protein expression (right hand side) is presented as and based on a best estimate of protein expression.

Cytoplasmic expression has been detected with the highest protein levels in mucosal epithelia cells in the endocervix, bronchus, nasopharynx, stomach, skeletal and smooth muscle. Notably, no protein expression was detected in lung tissue, salivary gland or oesophagus, despite showing significant RNA expression levels (Figure **3.13** and Figure **3.14**), which could be because the parts of the tissue samples do not have the right cells to express the gene. For example, BPIFB1 positive cells are only glandular cells within the stomach and

cervix, whereas it was completely absent from cervical squamous epithelial cells (https://www.proteinatlas.org).

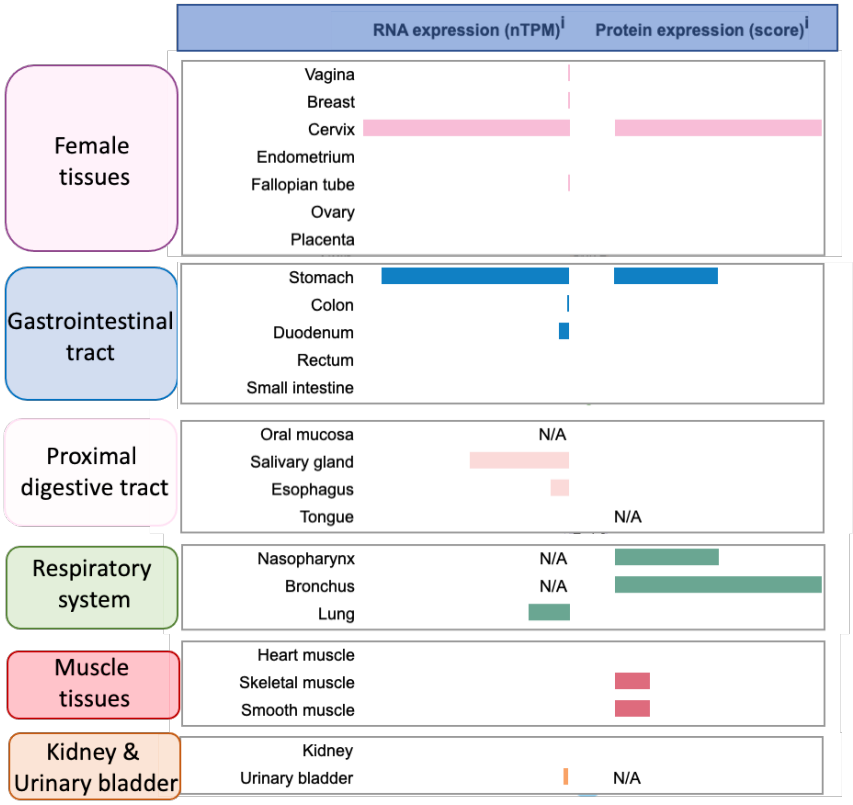


Figure 3.14: Expression data of human BPIFB1 in categorised tissues. The Human Gene Atlas data base (https://www.proteinatlas.org) was used to display the consensus RNA data (left hand side) based on normalized transcripts per million expression (nTPM) values. RNA-Seq data presented stems from the Human Protein Atlas (HPA) and the Genotype-Tissue Expression (GTEx) project. RNA (left hand side) or protein expression (right hand side) is presented as bars for indicated group of tissues. Only tissue groups were shown which show BPIFB1 expression on RNA and/or protein level.

Initial RNA expression data was further explored in a more quantitative manner, comparing RNA expression levels for 55 tissue types, normalized by expression (nTPM) levels, combining the HPA and GTEx transcriptomics data sets (Figure **3.15**A). Color-coding is based on tissue groups with functional features in common, is as indicated in Figure **3.13**. As mentioned above, the highest *BPIFB1* expression was detected in cervical tissue samples, and in stomach. Salivary gland expression showed moderate expression levels, with lowest levels in lung, duodenum, oesophagus and urinary bladder (Figure **3.15**A).

When BPIFB1 protein levels of this data set were estimated based on antibody staining, high levels were only detected in bronchus and cervix, whereas medium levels were found in

nasopharynx, and the digestive cluster of the stomach, duodenum and small intestine. Low BPIFB1 expression was identified in muscle tissues (Figure **3.15**B). However, it should be noted that to date no quantitative data on levels of BPIFB1 in secretions have been published. Most data were obtained by proteomic studies and our lab is working on establishing an ELISA setup for measuring BPIFB1 levels, which so far has only been used in patinent with COPD (Gao *et al.*, 2015).

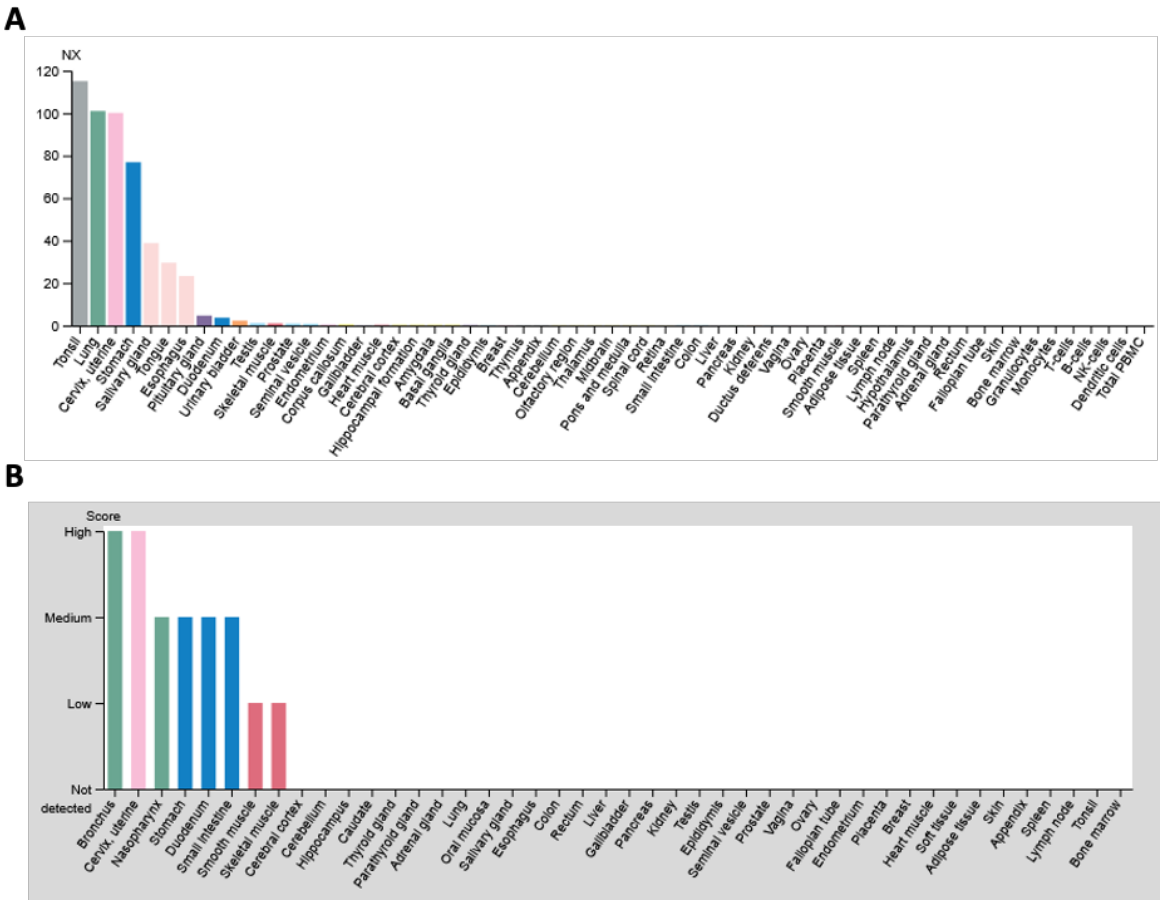


Figure 3.15: Expression data of human BPIFB1 in tissues. A) The Human Gene Atlas data base (https://www.proteinatlas.org) was used to display the consensus RNA data based on normalized transcripts per million expression values. RNA-Seq data presented stems from the Human Protein Atlas and the Genotype-Tissue Expression project. **B)** Protein expression is illustrated as bars in a particular group of tissues. Protein expression scores are based on a best estimate of protein expression and are divided in three expression categories (high, medium and low).

Comparing the expression findings with studies performed in the mouse, reported murine Bpifb1 expression across tissues compares moderately with human BPIFB1. Consistent are high levels of expression stomach and in lung, with lower levels reported for ileum and duodenum (Figure **3.16**). However, under these conditions no Bpifb1 protein has been

detected in female reproduction tissues such as the uterus or skeletal muscle (Figure **3.16**), which showed a clear expression for BPIFB1 in human derived samples (Figure **3.15**B).

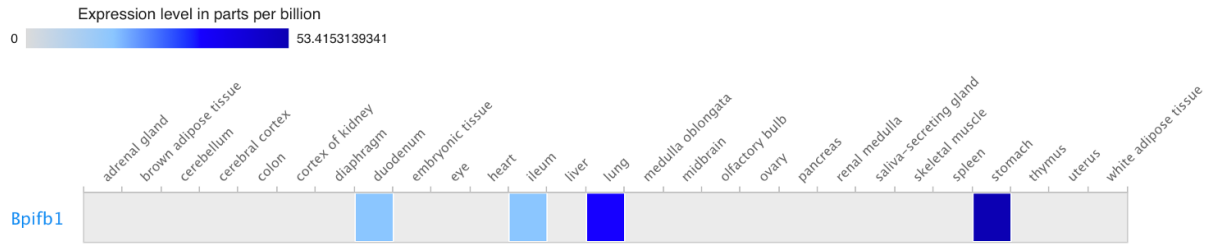


Figure 3.16: BPIFB1 protein levels across murine tissues. Tissue specific protein expression profile obtained from ENSEMBL database (ENSMUSG00000027485). Tissues are arranged in descending alphabetical order. Enhanced bpifb1 expression is found in organs related to respiratory and digestive systems. Proteomics baseline from from (Geiger *et al.*, 2013).

As BPIFB1 was highly expressed in tissues of the respiratory tract, and given that the focus of my study was on the airways, I further investigate in more detail cell specific expression of *BPIFB1* in respiratory tissue across human and mouse samples, using single cell RNA sequencing data from published literature.

3.3.3.1 Cell specific expression

Previous studies, predominantly using IHC methods, identified BPIFB1 and BPIFA1 to be localised to several epithelial cell types and minor mucosal glands residing in the nose, mouth and respiratory tract. BPIFB1 has been shown to be most strongly present in the minor glands of the proximal tongue and in a small population of goblet cells in the nasopharynx, whereas BPIFA1 is most strongly expressed in the nasal respiratory epithelium (Bingle *et al.*, 2010; Musa *et al.*, 2012).

Recent technical improvements have made single cell sequencing an important tool for understanding gene expression in individual cells (Svensson *et al.*, 2017; Hedlund and Deng, 2018; Greaney *et al.*, 2020). This is particularly important in tissues such as the lung and respiratory tract which contain multiple different cell types. Within the past five years multiple publications investigating respiratory samples have been published and made it possible to investigate cell type specific expression of genes of interest (Raredon *et al.*, 2019; Travaglini *et al.*, 2020; Han *et al.*, 2020; Deprez *et al.*, 2020). Therefore, using publicly

available data sets, it is feasible to explore which specific cell types are expressing *BPIFB1* in respiratory tissue specimen. Therefore, I used published single cell RNA sequencing data to profile the specific cellular expression of *BPIFB1* and its closest related paralogue *BPIFA1*. Investigation of this data enables a high resolution of cellular differences in gene expression and consequently a better understanding of the function of an individual cell in the context of its microenvironment.

Respiratory single cell RNA-Seq data

A recent single-cell RNA profiling study published by (Deprez *et al.*, 2020) explored the cellular heterogeneity of the human airway epithelium across ten healthy individuals. This study represents one of the most extensive ones performed on the human airway to date, collecting a total of 77,969 cells from 35 distinct locations, spanning from the nose to the 12th division of the airway tree (Figure **3.17**A). Samples were acquired by minimally invasive methods through biopsies and brushings, resulting in a large panel of epithelial cell subtypes of precise locations. Overall, cell populations in this study were categorised as epithelial (89.1%), immune (6.2%) and stromal (4.7%) cells. This extensive cell atlas was accessed through the online tool available at https://www.genomique.eu/cellbrowser/HCA.

Most cell cluster visualisation are displayed using the uniform manifold approximation and projection (UMAP) algorithm, which allows for high dimensional analysis to be viewed in 2D or 3D. It offers an advantage to the T-distributed Stochastic Neighbourhood Embedding (tSNE) methodology as tSNE clusters are distributed solely based on their similarities or differences of individual points rather than similarities of populations as performed by UMAP. Overall, both methods reduce the thousands of dimesions seen in the highly non-linear RNAseq data from each cell. A drop of dimensionality makes it possible to visualize the data on a 2D or 3D plot, allowing important insights to be gained by analysing these patterns in terms of clusters on a 2D plot.

Using the online visualisation tool of the Deprez *et al.,* 2020 cell atlas, I was able to explore specific expression of *BPIFB1* (Figure **3.17**B) and *BPIFA1* (Figure **3.17**D), as well as the goblet cell marker *MUC5B* (Figure **3.17**C) across a large panel of respiratory cell types in UMAP plots. This representation allows for a reduction of a vast amount of mRNA-seq data into four dimensions, namely the neighbour, distance, intensity/concentration and number/amount.

In this selection, *BPIFB1* is the most ubiquitous expressed member, being present in the majority of secretory, serous cells and mucous cells (Figure 3.17B). There is a distinct overlap of *BPIFA1* with *BPIFB1* in secretory and serous cell expression, however *BPIFA1* levels are significant lower. Furthermore, *BPIFA1* seems to be absent from mucous cell populations (Figure 3.17D). In contrast, *MUC5B* shows a good correlation with *BPIFB1*, being enriched in mucous and secretory cells and SMG goblet cells (Figure **3.17**D). However, the overall expression is significantly lower than observed for *BPIFB1*.

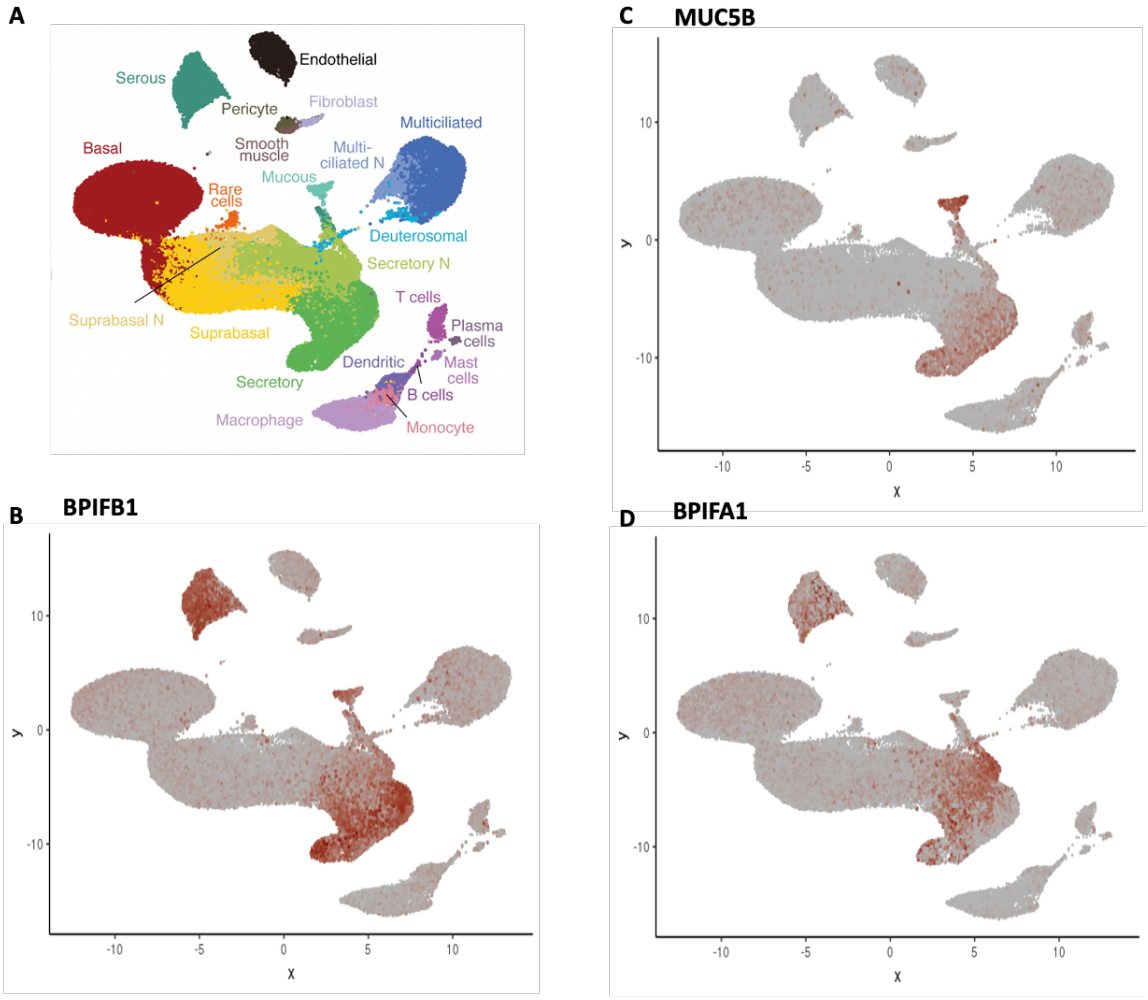


Figure 3.17: *BPIFB1, BPIFA1* and *MUC5B* expression across a healthy human airway. Single-cell RNA sequencing data comprising of a total of 77,969 cells, which were collected from 35 distinct locations, spanning from the nose to the 12th division of the airway tree. **A)** UMAP visualization of each distinct cell type by a specific colour. Expression of *BPIFB1* **(B)**, *MUC5B* **(C)** and *BPIFA1* **(D)** across the human airway data set. Images were generated using the Deprez et al., 2020 data set and were visualised through the interactive web tool (https://www.genomique.eu/cellbrowser/HCA/).

Further, a small subset of multiciliated cells show a significant expression of *BPIFB1*. It seems that expression positive cells are more mucous cells before they become ciliated. This would be consistent with the newer view that ciliated cells derive from goblet/mucus cells (Ruiz García *et al.*, 2019).

Overall, the green secretory portion of the plot includes nasal and airway cells that have slightly different expression profiles. It is observed that *BPIFB1* is more restricted to the cells with an airway origin compared to *BPIFA1*, which is seen more in the nasal cells. This is further illustrated in Appendix XIII. In addition, the secretory cell population seems to be more heterogenous based on transcriptional expression. It could be argued that it contains subsequent differentiation states. *MUC5B* appears to be 'sole'-heavy to a certain degree, however the overlap is not remarkable. Perhaps I should have also included *MUC5AC* in this context, and certainly there will be other genes enriched in ths region.

The Deprez *et al.,* 2020 data set was further used to extend the comparison to other BPIF members such as *BPIFB2, BPIFA2* and other mucins like *MUC5AC* (Figure **3.18**). It became apparent that *BPIFB1* and *BPIFA1* show a high cell type overlap, especially in secretory and serous cells. Whereas *BPIFB2* and *MUC5B* show only strong on overlapping expression in SMG goblet cells, which extends to a lower degree to *BPIFB1* and *MUC5AC* expression patterns. In contrast to *BPIFA1, BPIFA2* exhibits its most prominent expression in secretory cells, with a much lesser expression found in rare cells (Figure **3.18**). To conclude, *BPIFA1*, *BPIFB1* and *MUC5AC* show the highest overlap and diverse expression, whereas *MUC5B* seems to cluster specifically with *BPIFB2*.

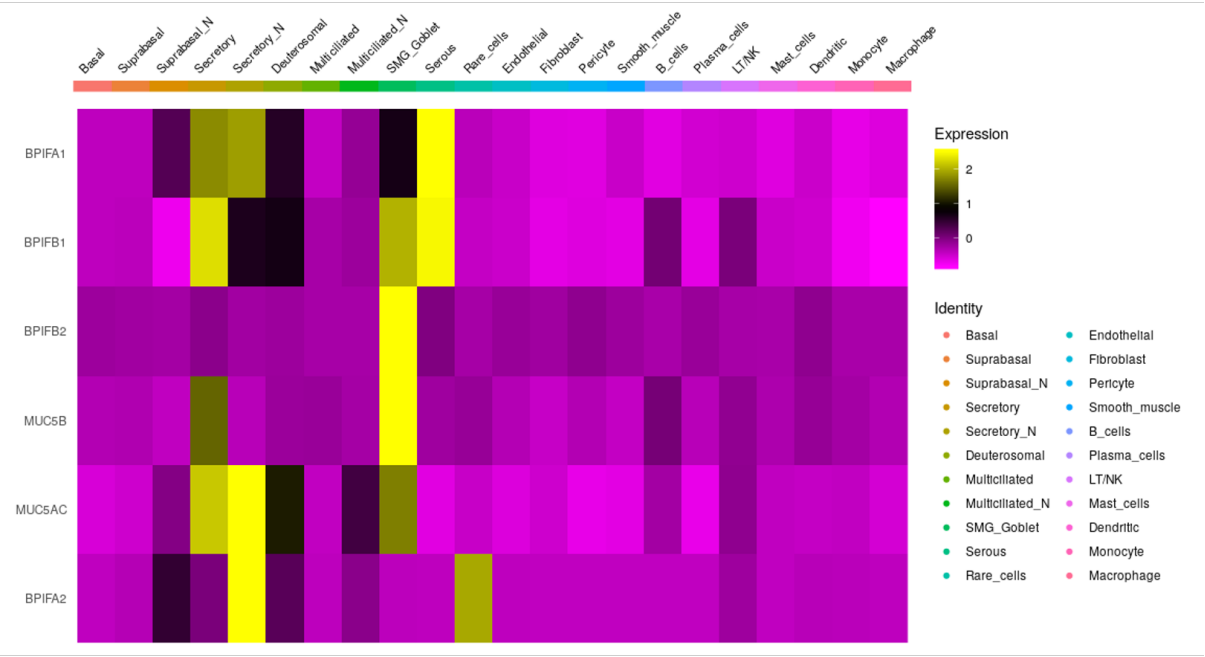


Figure 3.18: Heatmap comparing expression of selected BPIF members, *MUC5AC* and *MUC5B* in distinct airway cell populations. Generated from the Deprez *et al.,* 2020 data set by the Barbry and Zaragosi group.

I further extended the expression analysis to the majority of BPIF members using the Durante *et al.*, 2020 single-cell RNA-sequencing analysis of the human olfactory neuroepithelium, collected from 28,726 cells. This cell population has a significant respiratory component. Notably, *BPIFA1* and *BPIFB1* generate an almost identical distribution patterns, both in cell location and expression intensity, across a large panel of cell types (Figure **3.19**).

The highest expression levels are consistent with findings in other studies across respiratory samples, confirming presence of these two genes in respiratory epithelial and secretory cells. However, *BPIFB1*, and to a lesser extent *BPIFA1*, are found in the olfactory gland also termed Bowman's glands. Expression of *BPIFB1* and *BPIFA1* in this study in respiratory ciliated and columnar cells, as well as sustentacular cells, is found at a much lower expression and hence might require further investigation (Figure **3.19**). Overall, this analysis shows that *BPIFA1* and *BPIFB1* are the two most significantly co-expressed expressed *BPIF* genes in human nasal epithelial cells.

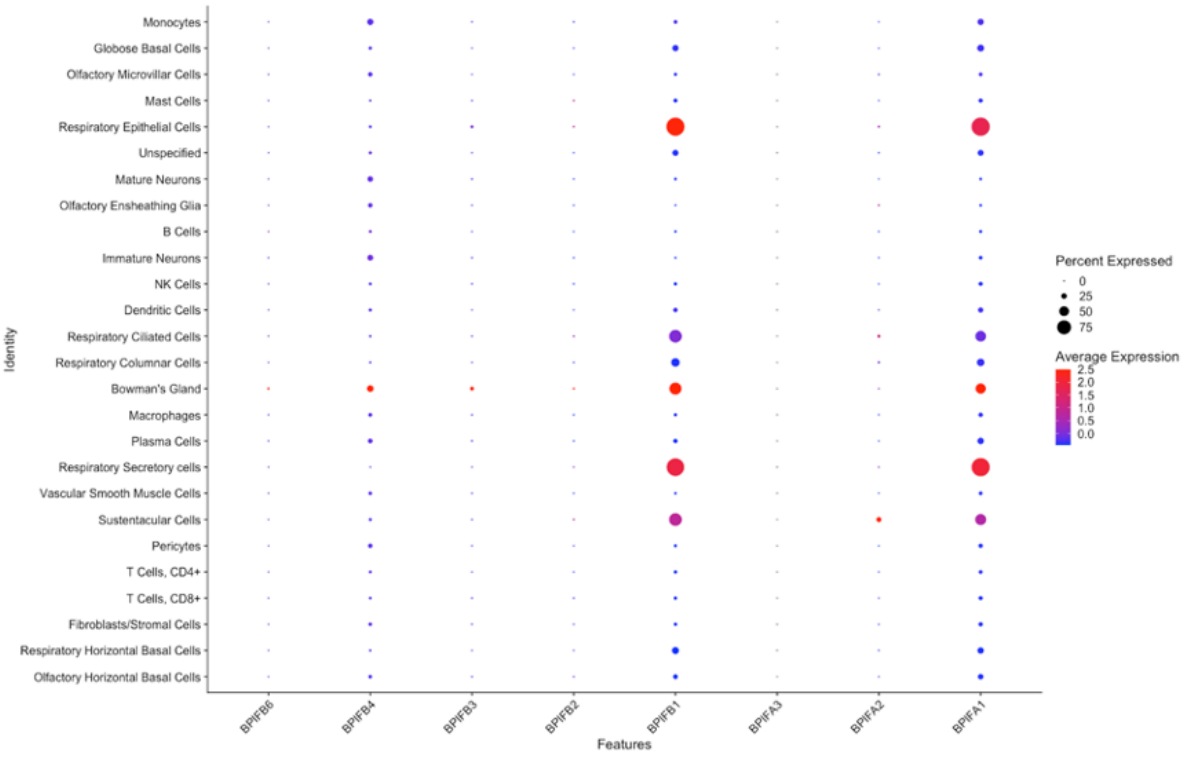


Figure 3.19: Dot plot visualization of *BPIF* gene expression across human olfactory and respiratory mucosal tissues. The plot depicts clusters from 28,726 combined olfactory and respiratory mucosal cells, collected from four patients. Cell identities are listed on the y-axis, plotted genes per cluster identified by log fold change. Genes are listed along the x-axis and the dot size reflects percentage of cells in a cluster, expressing each gene. Dot colour reflects expression level, as indicated on the legend on the right. Data was taken from Durante *et al.,* 2020.

As studies in my thesis focus on BPIFB1 in mouse and humans, next I directly compared expression *BPIFB1* between mouse and humans.

The Lung Connectome (http://lungconnectome.net/), is an interactive platform created by Raredon *et al.,* 2019, that allows transcriptional exploration of healthy adult lung across a number of species. Pre-clustered tSNE plots were generated to visualise the entire mouse and human single cell RNA sequencing data sets for these species (Figure **3.20**A and C). Each data set was then searched for *BPIFB1* expression and fold change mapped in the form of a tSNE plot and violin plots using this interactive platform (Figure **3.20**). Gene expression for each cell type in each species was further visualised with a heatmap (pkg. matplotlib) to illustrate cell type expression of *BPIFB1* in both human and mouse lung parenchyma. The data shows that human and mouse *BPIFB1* expression appears to be conserved across both species and is mainly restricted to secretory cell populations (Figure **3.20**B and D). A smaller amount of human *BPIFB1* was detected in mast, basal and ciliated cells, which was largely absent in the mouse data set (Figure **3.20**E versus F and G).

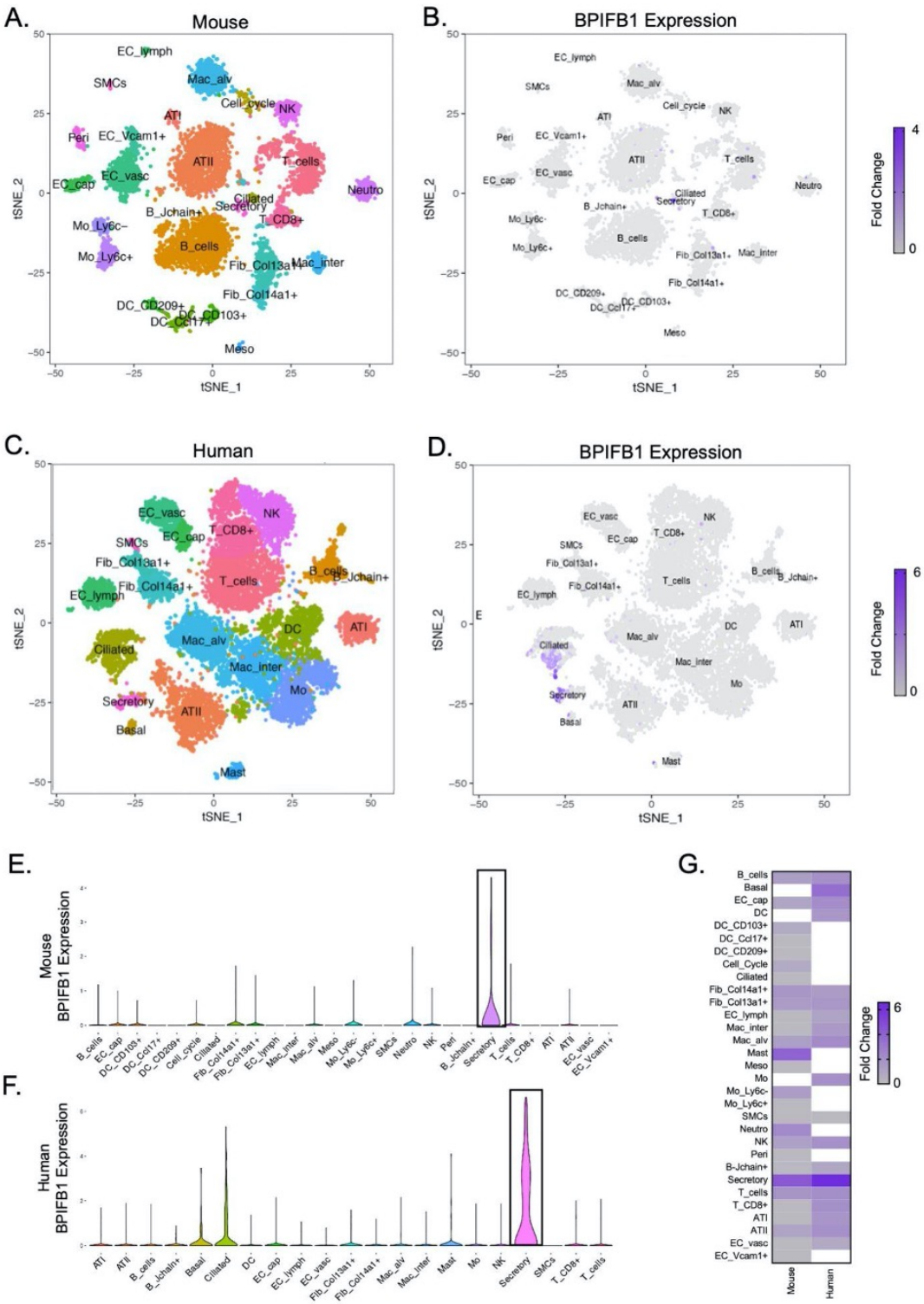


Figure 3.20: *BPIFB1* expression is conserved between human and mouse in lung tissue. **A)** Adult mouse lung single cell RNA sequencing (n=7744 cells). **B)** Expression of *BPIFB1* in the mouse lung scRNA-Seq data set. **C)** Healthy human lung parenchyma single cell RNA sequencing (n=17867 cells)**. D)** Expression of *BPIFB1* in the human lung scRNA-Seq data set. Violin plots displaying *BPIFB1* expression in both mouse **(E)** and human **(F)** data sets in each identified cluster. **G)** Heatmap exhibiting *BPIFB1* expression by cell type in each species. Data collection, clustering, and analysis was performed using data sets from (Raredon *et al.*, 2019) and visualized using their interactive platform Lung Connectome (http://lungconnectome.net/).

To explore which regions of the respiratory tract express *BPIFB1* in more detail, single cell RNA-Seq data sets were identified that contain nasal, bronchial, and parenchyma samples. Two data sets, one created by Deprez *et al.,* 2020 and the other generated by Vieira Braga et al. 2019, were found to contain cell populations from all three sections of the respiratory tract. The expression of *BPIFB1* was plotted using the cellxgene custom platform (https://www.covid19cellatlas.org/deprez19/) for the single cell RNA-Seq generated by Deprez *et al.,* 2019. The *BPIFB1* gene expression, breakdown of populations belonging to each area of the respiratory tract, and distribution of sample collection were visualized using generated UMAP plots on the interactive platform (Figure **3.21**A). The percentage of cells in each part of the respiratory tract expressing *BPIFB1*, as well as cells collected by sampling methods via brushing or biopsy, above a fold change > 1 was calculated and visualized using pie charts (Figure **3.21**D and F).

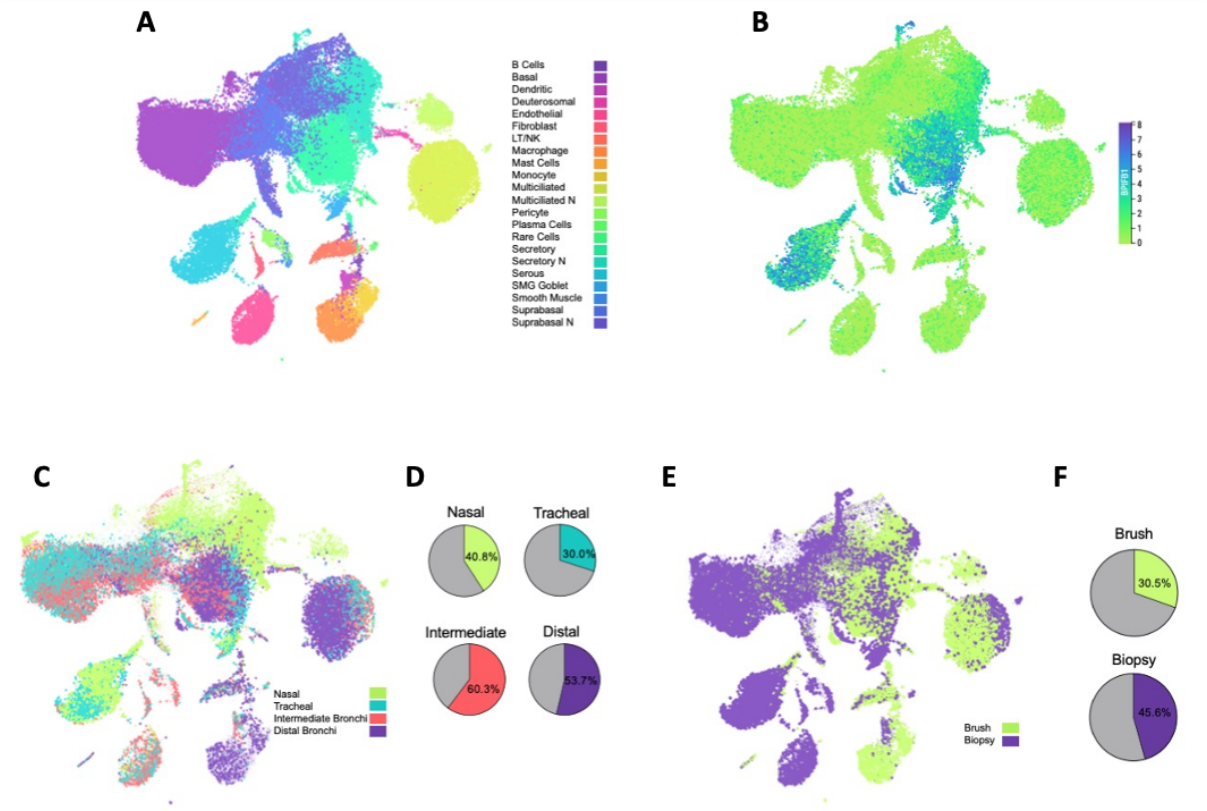


Figure 3.21: Expression of *BPIFB1* varies by location in the respiratory tract. A) UMAP depiction of healthy human airways reproduced from Deprez et al. 2020 data set (n=77969 cells). **B)** *BPIFB1* expression in healthy human airways. **C)** Locations of cells in the respiratory tract, extracted from the Deprez data set. **D)** Percentage of cells expressing *BPIFB1* fold change >1 in each section of the respiratory tract. **E)** Distribution of cells collected either by brushing or biopsy of the respiratory tract. **F)** Percentage of cells expressing *BPIFB1* fold change >1 as obtained by each sampling method. Data collection, clustering, and analysis was performed from the Deprez *et al.,* 2020 publication, using the interactive platform https://www.covid19cellatlas.org/deprez19.

Consistently, *BPIFB1* was shown to be found in secretory and serous cell populations of tracheal and nasal origin (Figure **3.21**B and C), and the overwhelming majority of *BPIFB1* cells were obtained by biopsy sampling (Figure **3.21**E). This suggests that the method of sampling may influences the type of cells collected. This could potentially be explained by the location of *BPIFB1* expressing cells such as goblet/mucous cells, some of which are located in SMGs, and therefore are less available to brushings and consequently found more represented in biopsies. However, this hypothesis needs further investigation as murine airway SMGs have been found negative for BPIFB1 (Bingle *et al.*, 2012).

To further investigate *BPIFB1* expression based on location in the respiratory tract, the Lung Cell Atlas (https://asthma.cellgeni.sanger.ac.uk/), an interactive platform generated by (Vieira Braga *et al.*, 2019) was used to examine cell populations throughout the respiratory tract. Data sets containing nasal, bronchial, and parenchyma samples were visualized using UMAP plots to analyse clusters and *BPIFB1* expression (Figure **3.22**). A heatmap was generated (via pkg matplotlib) to illustrate the expression of *BPIFB1* by cell type in different parts of the respiratory tract (Figure **3.22**D).

Nasal samples show the highest *BPIFB1* expression across goblet cell populations and a reduced level in ciliated cells (Figure **3.22**A), consistent with the findings of the Durante *et al.*, 2020 data set (Figure **3.19**). Whereas *BPIFB1* expression in bronchi is highest and exclusively found in club cells (Figure **3.22**B) and not detected in lung parenchyma (Figure **3.22**C).

Intriguingly, *BPIFB1* expression in nasal samples was detected in ionocytes and luminal macrophages at comparable levels to Club cell levels (Figure **3.22**D). This has not been documented before and requires further study. Pulmonary ionocytes in the respiratory airways have been a very recent discovery (Montoro *et al.*, 2018; Plasschaert *et al.*, 2018) and therefore not much is known about their function. On the other hand, luminal or alveolar macrophages are the most abundant innate immune cells present in the distal lung parenchyma, where they are located on the luminal surface of the alveolar space (Joshi *et al.*, 2018). In the latter case it could be speculated that *BPIFB1* could exert an immunomodulatory role in conjunction with the macrophage function. However, as *BPIFB1* expression in a macrophage background is a novel observation, and has not been noted in previous studies, it needs to be carefully assessed in future studies if this is a truthful result.

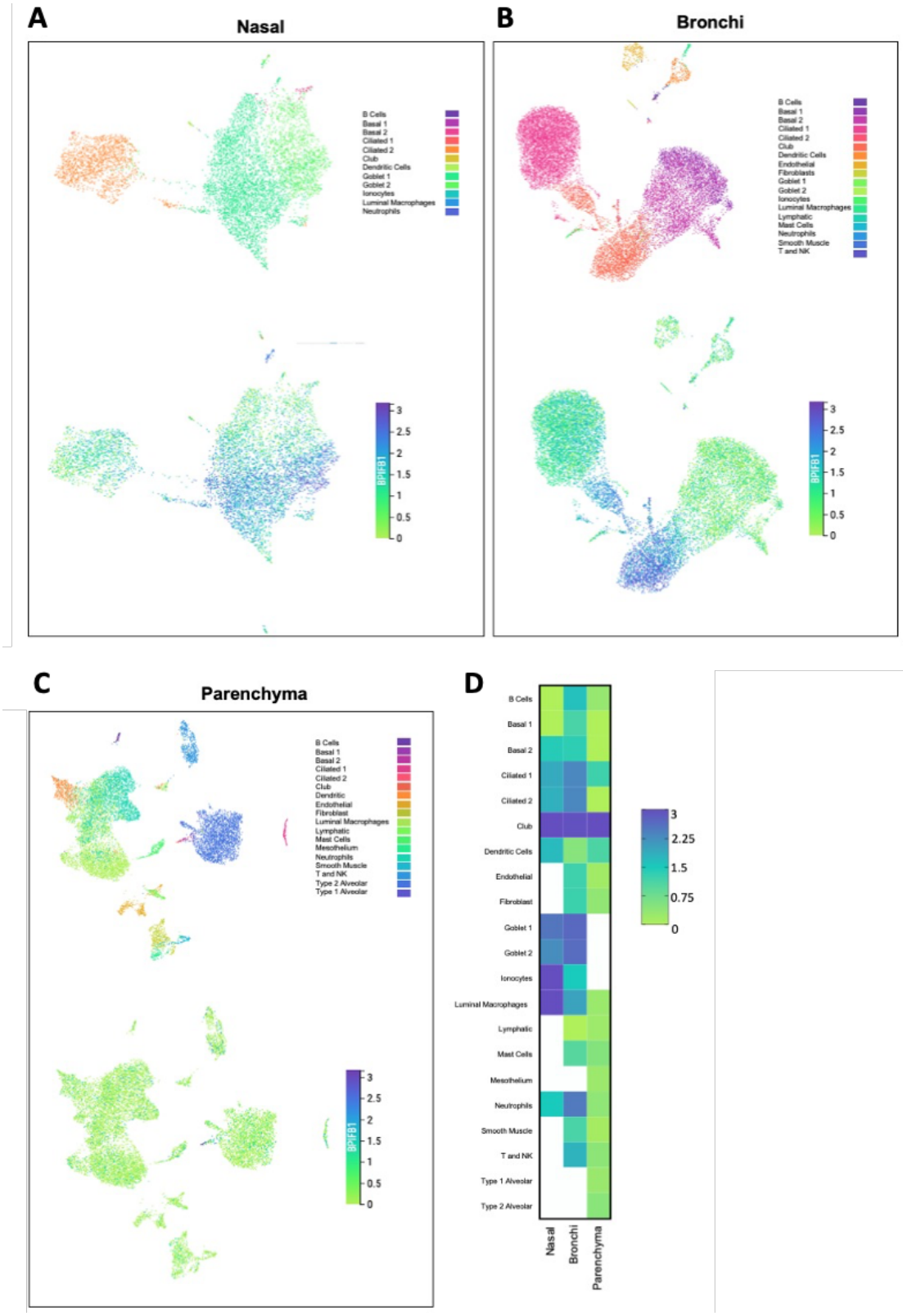


Figure 3.22: *BPIFB1* expression across different respiratory tissues. UMAP depictions of clustered healthy nasal (n=7087 cells) **(A)**, bronchial (n=16,873 cells) **(B)**, and parenchyma (n=12,971 cells) **(C)** cell populations with corresponding *BPIFB1* expression as measured by fold change from Vieira Braga et al. 2019 using the Lung Cell Atlas platform. **D)** Heatmap of *BPIFB1* expression throughout the respiratory tract separated by clustered cell types. Data collection, clustering, and analysis was performed from the Vieira Braga *et al.,* 2019 publication, using the interactive platform (https://asthma.cellgeni.sanger.ac.uk).

3.3.3.2 Temporal expression studies of *BPIFB1* in lung tissue

To explore temporal patterns of *BPIFB1* transcriptional activity during normal lung development, a single cell RNA-Seq data set was identified containing cells collected from foetal human lung tissue at week 11.5 to week 21 of development composed by (Miller *et al.*, 2020).

It should be emphasised that it is inherently very difficult to obtain human lung data, especially from late gestation and early post-natal life. This marks the time frame where the majority of essential lung development takes place, however it is at the same time the most difficult tissue to acquire for studies.

The expression of *BPIFB1* from this dataset was visualized using the cellxgene interactive platform (https://www.covid19cellatlas.org/miller20/) generated by Miller *et al.,* 2020. The percentage of cells expressing *BPIFB1* was visualised in a heatmap (via pkg matplotlib) by fold change >1 in each cell type, timepoint, and location (Figure **3.23**).

BPIFB1 expression is found in a very small and distinct cluster of Club-like secretory and Goblet-like secretory cells, but not in basal, secretory progenitor or multiciliated cells. As expected, immune cells were mostly negative for *BPIFB1* expression (Figure **3.24**A and B).

Regarding location and timeframe of *BPIFB 1*expression in the developing human lung, it appears to be specific for week 15 and restricted to a tracheal location (Figure **3.23**C). No detectable expression was observed in samples from weeks 11.5, 18 or 21, or at distal or airway locations. Of note, it seems that human fetal *BPIFB1* expression appears to be highly regulated spatially and temporally, as at week 18 in trachea no *BPIFB1* expression was observed (Figure **3.23**C). It could be speculated that low expression of *BPIFB1* could stem from a lack of fully differentiated cells due to the early developmental stage, in combination with a low cell number. Of note, it should be mentioned that the individual sample size of the study is very small and therefore great care must be taken not to over interpret the data.

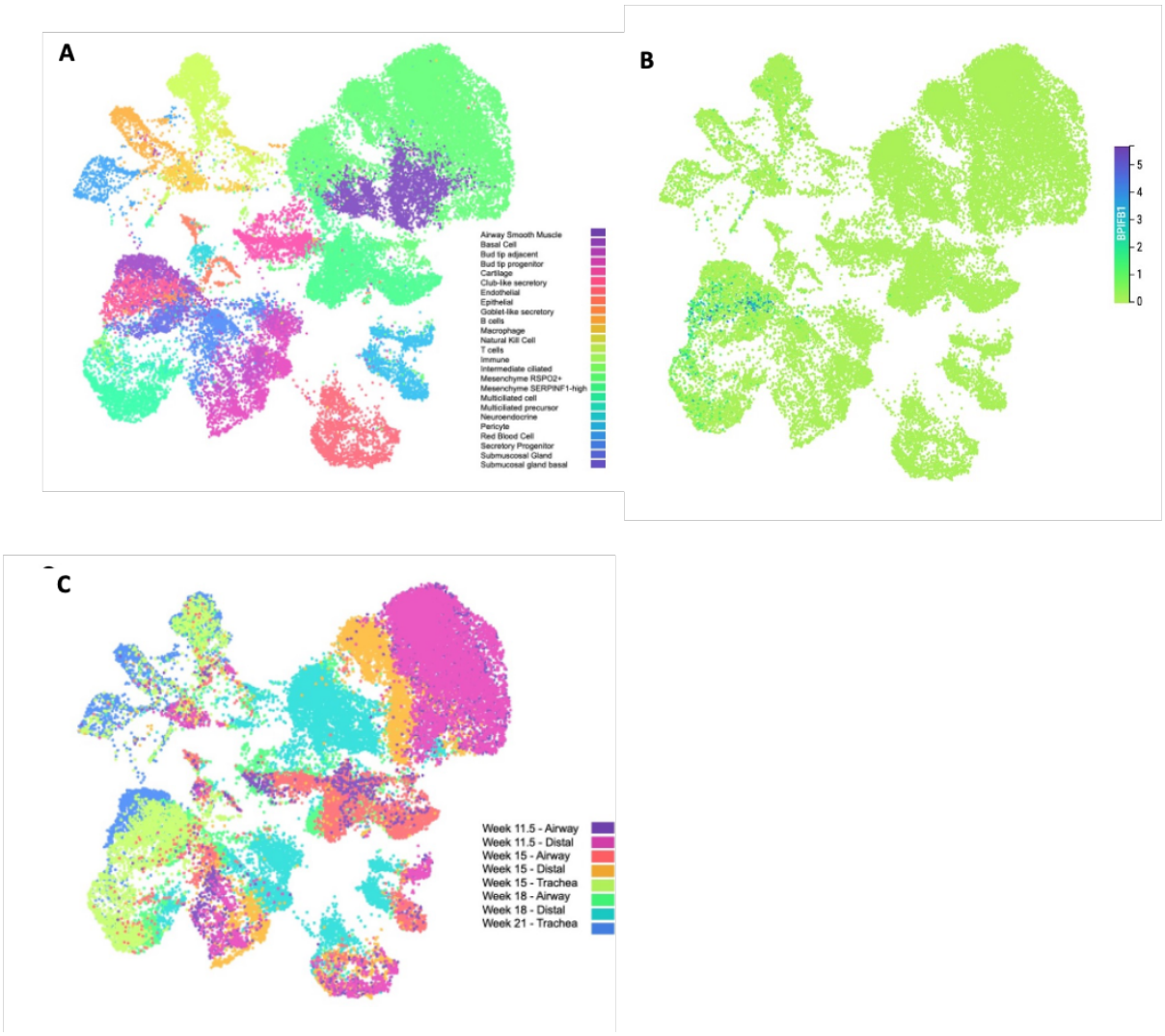


Figure 3.23: *BPIFB1* expression during fetal lung development in humans. A) UMAP depiction of human bronchial epithelial cells from fetal lungs during developmental stages covering week 11.5 to week 21 from Miller et al. 2020 (n=36807 cells). **B)** *BPIFB1* expression in developing fetal lungs as visualized by the interactive scRNA-Seq platform (https://www.covid19cellatlas.org/miller20/). **C)** UMAP displaying development timepoints by week (11.5, 15, 18 or 21) and location (Airway, Distal, Trachea) for human fetal lung cells.

As observed from the UMAP representation, *BPIFB1* expression in human fetal lung development is spatially and temporally restricted (Figure **3.23**). However, visualising the data in a heatmap format allows for more detailed observations (Figure **3.24**A). The highest expression of *BPIFB1* is found in goblet-like secretory cells and at a much-reduced level in club-like secretory cells. Expression is at initially detected at week 15 and maintained at a comparable level at week 18 and 21 in tracheal origin (Figure **3.24**A). However, due to the difficult nature of obtaining fetal human samples in addition to a very limited sample number, it is not possible to study *BPIFB1* expression at later developmental time points.

Next, the human data set was compared to murine data covering 26 pre- and post-natal time points in three common inbred strains of laboratory mice (C57BL/6J, A/J, and C3H/HeJ). To do this data from Beauhemin *et al.,* 2016 was acquired from the web interface (http://lungdevelopment.jax.org/) and visualized as a heatmap (via pkg matplotlib) (Figure **3.24**). This data set spans all five canonical stages of mammalian embryonic lung development: embryonic, pseudoglandular, canalicular, saccular and alveolar, with the latter one divided in four stages of alveolarization until establishment of the mature lung from post-natal day P21 to P56 (Figure **3.24** B and C).

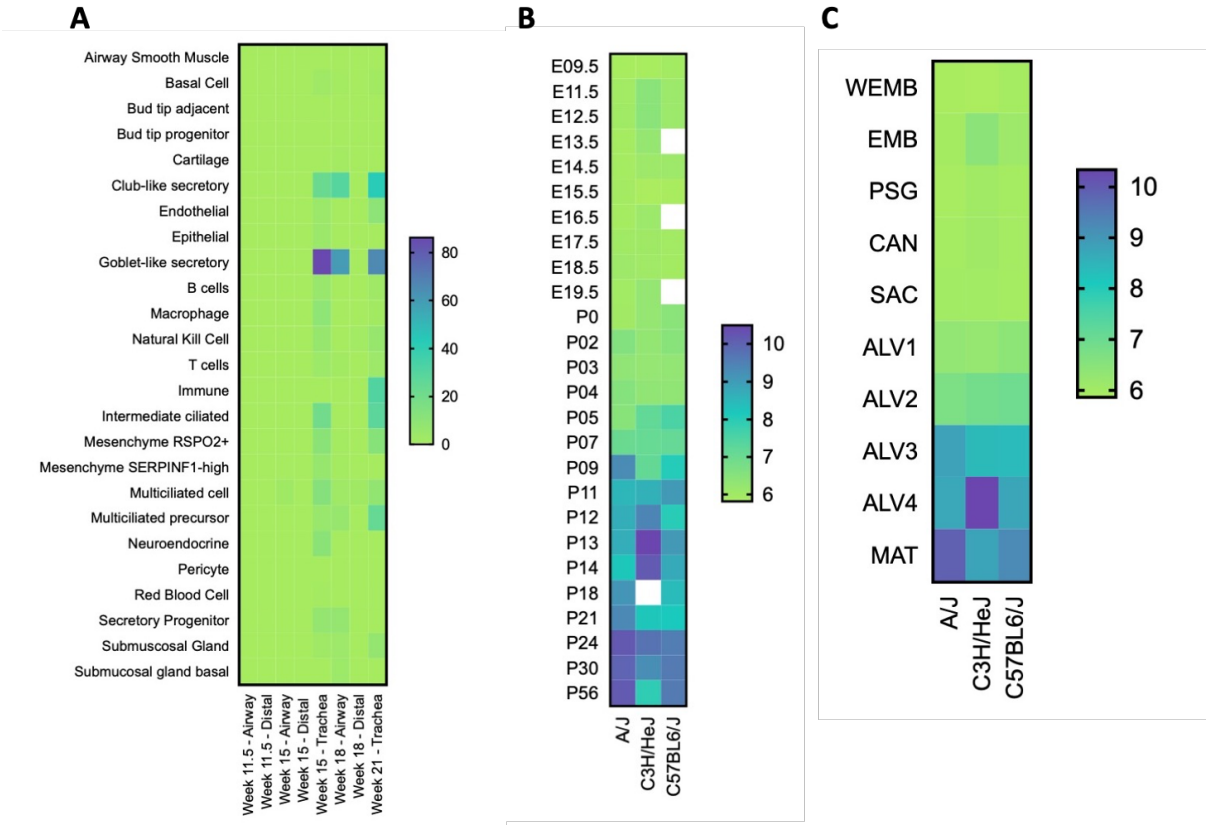


Figure 3.24: *BPIFB1* expression during lung development in human and mouse. **A)** Heatmap illustrating the expression of *BPIFB1* by fold change >1 in human fetal lung development for each cell type, developmental timepoint (week 11.5, 15, 18, 21, covering first and second trimester), and location (Airway, Distal or Trachea) from Miller et al. 2020. **B)** Heatmap demonstrating *Bpifb1* expression in three murine strains (A/J, C3H/HeJ, C57BL6/J) during lung development spanning developmental days E9.5 to E19.5 and post-natal day P0 to P56. **C)** Heatmap comparing Bpifb1 expression across three murine strains for each lung developmental stage. Data is from Beauchemin *et al*. 2016. WEMB - Whole embryonic embryo, EMB - embryonic, PSG - pseudoglandular, CAN -canalicular, SAC - saccular, ALV1-4 - alveolar stages 1-4, MAT - mature lung.

All three mouse strains show no *BPIFB1* expression until the post-natal phase and only after P7. There are differences in terms of peak expression time between strain C3H/HeJ (P13-

P14) and strain A/J and C57BL6/J, which seem to express significantly higher amounts of *BPIFB1* at the end of alveolarization during final lung maturation (Figure **3.24** B and C).

In this section I investigated BPIFB1 protein localisation and expression in databases and published RNAseq data sets. With respect to protein expression, BPIFB1 was detected in mucosal epithelia cells, in the endocervix, bronchus, nasopharynx, stomach, skeletal and smooth muscle. Notably, no protein expression was identified in peripheral lung tissue, salivary glands or oesophagus. This expression pattern was consistent with localisation reported for Bpifb1 in murine tissues.

Employing different published RNAseq studies, *BPIFB1* is the most ubiquitous expressed BIF member, being present in the majority of secretory, serous and mucous cells of tracheal and nasal origin, with expression partly overlapping with *BPIFA1*. Overall, *BPIFB1* expression appears to be conserved across both human and murine species and is mainly restricted to secretory cell populations.

Exploring temporal expression patterns of *BPIFB1* transcriptional activity during normal lung development revealed a very small and distinct cluster of club-like secretory and goblet-like secretory cells, but was not detected in basal, secretory progenitor or multiciliated cells. Although it appears that *BPIFB1* is expressed specifically around week 15 in the developing human lung, and is restricted to a tracheal location, this might not be representative as the study only contained a very small sample size. In contrast, murine developing lung tissue did not show *Bbifb1* expression until the post-natal phase and only after P7. This expression data suggests that the function of BPIFB1 is not required prior to birth.

Having investigated the cell specific and developmental expression of *BPIFB1*, I next investigated expression data sets from lung diseases as there are multiple reports of *BPIFB1* being differentially expressed in disease.

3.3.4 BPIFB1 implicated biological function and potential role in respiratory homeostasis

As mentioned in the introduction, functional analysis of BPIFB1 to date is still very limited. Therefore, I aimed to investigate *BPIFB1* gene associations by extracting data on its expression across a range of lung diseases, in order to widen the understanding of a potential role of BPIFB1 in respiratory homeostasis and disease.

Chronic obstructive pulmonary disease (COPD)

COPD is collective term for several lung conditions which cause breathing difficulties. These include emphysema, describing damage to the alveolar air sacs in the lungs, or chronic bronchitis, caused by long-term inflammation of the airways. COPD is commonly observed in middle-aged or older adults who smoke, with increasing breathing difficulty over time (Negewo *et al.*, 2015; Bagdonas *et al.*, 2015; Erhabor *et al.*, 2021).

To investigate the previous observation that *BPIFB1* levels correlate with severity of COPD (Gao *et al.*, 2015; De Smet *et al.*, 2018) further, the COPD Cell Atlas (https://p2med.shinyapps.io/copd-cell-atlas/_w_56264081/#tab-7237-3) (Sauler *et al.*, 2022) was used to examine *BPIFB1* expression in healthy and COPD lung tissue. It can be seen that the highest *BPIFB1* expression in healthy lung tissue is found in goblet cells and to a lesser extent in club cells. Much lower *BPIFB1* expression levels are detected in ciliated cells, with no significant expression in basal, PNEC or AT cells (Figure **3.25**A). Consistent with earlier observations (Gao *et al.*, 2015; De Smet *et al.*, 2018), *BPIFB1* levels are significantly upregulated in goblet cell and ciliated cell populations under COPD conditions, however at the same time there is a significant reduction of *BPIFB1* expression is in Club cells (Figure **3.25**A). *BPIFA1* is expressed overall at a lower extent compared to *BPIFB1*, although being found in the same cell population, being highest in goblet cells, low in Club cells and close to detection limits in ciliated cells. During COPD, a reduction of expression is observed in Club cells, however in contrast to *BPIFB1, BPIFA1* is found dowregulated in goblet cells (Figure **3.25**B).

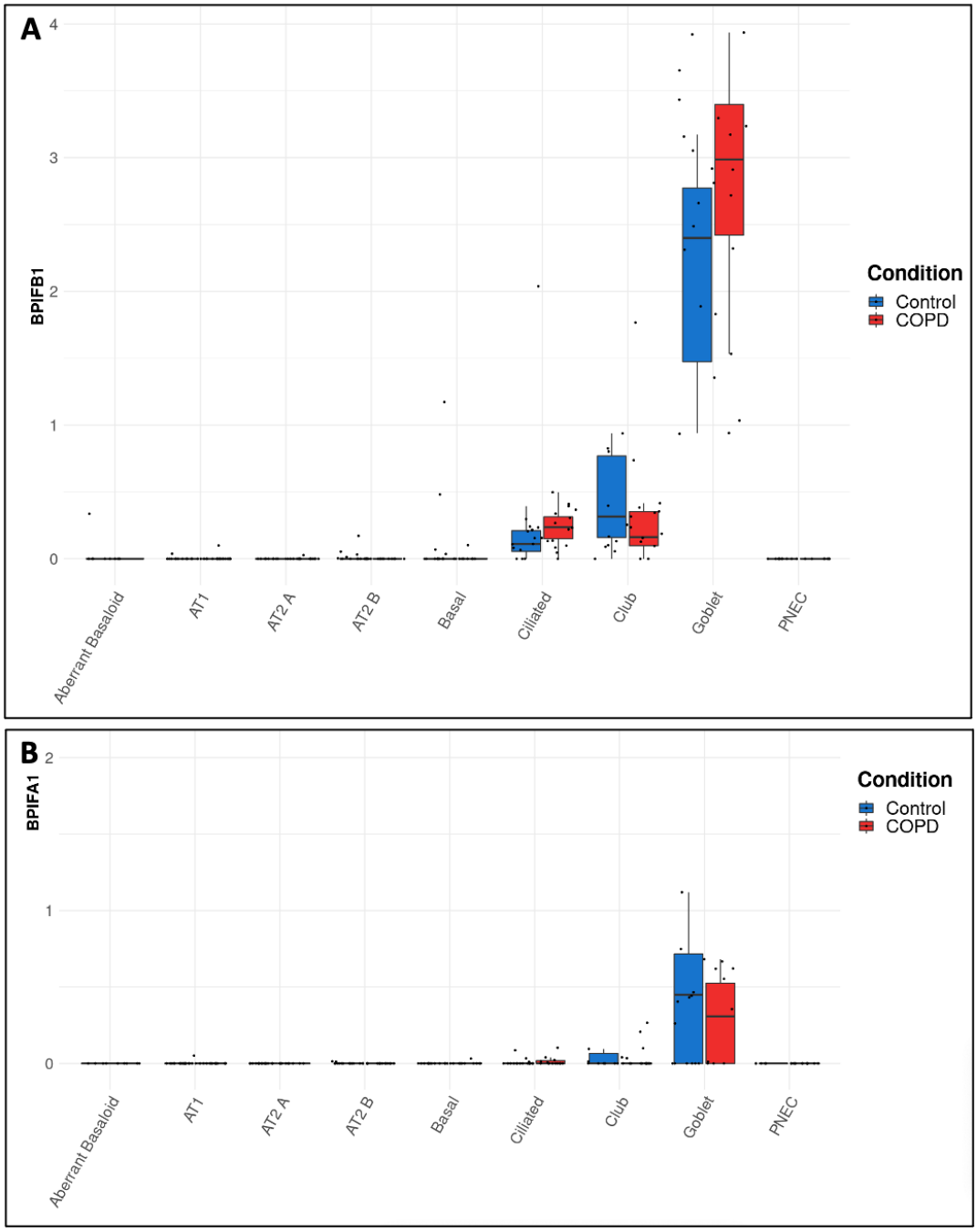


Figure 3.25: *BPIFB1* and *BPIFA1* expression in COPD. The data set combined lung tissue from 17 patients with advanced COPD and 15 age-matched controls (Sauler *et al.,* 2022), with 61,564 extracted cells from COPD lungs and 49,976 cells from control lungs. Overall, 37 distinct cell types were identified in both tissues as assessed by representative markers. **A)** Expression of *BPIFB1* and **B)** expression of *BPIFA1* across lung cell populations. Images were generated using the COPD Cell Atlas (http://www.copdcellatlas.com/)

Cystic fibrosis (CF)

The common and autosomal inherited disease cystic fibrosis (CF) results in thick mucus secretions in lungs, causing repeated bronchial infections and airway obstructions (Rafeeq and Murad, 2017; De Boeck, 2020) CF patients, as well as murine models of the disease,

exhibit markedly increased levels of BPIFB1. It is thought that increased production of the protein may be a response to the epithelial remodelling associated with this condition (Bingle *et al.*, 2012).

It was previously demonstrated that genetic variants in the *BPIFA1/BPIFB1* region are associated with decreased gene expression and increased lung disease severity in CF (Saferali *et al.*, 2015). This suggests that dysregulated *BPIFA1* and/or *BPIFB1* expression may be detrimental to CF lung function. For example, proteomic analysis of nasal epithelial cells from CF patients has demonstrated increased levels of *BPIFA1* (Roxo-Rosa *et al.*, 2006) and *BPIFA1* and *BPIFB1* are found elevated in sputum from patients with CF (McCray *et al.,* 2005), while epithelial cells from CF patients express abundant *BPIFA1* and *BPIFB1* (Scheetz *et al.*, 2004).

Carraro and colleagues published an extensive single cell study of the proximal airway epithelium of healthy and CF conditions. *BPIFB1* and *BPIFA1* expression data extracted from in this study showed that basal and ciliated subpopulations remain negative for any CF related differential gene expression (Figure **3.26**A) (Carraro *et al.*, 2021). However, a subset of secretory cells showed a modulated behaviour. Overall, *BPIFB1* exhibited a stronger modulated response in CF conditions compared to *BPIFA1*. The secretory cell population was subdivided into five categories: Secretory1-5 subsets as outlined in Figure **3.26**B. The Secretory 2 subset contains goblet cells based on the marker set containing mucins *MUC5B* and *MUC5AC*, *AGR2* and *SPDEF*. The Secretory 4 subset is defined by expression of *MUC5B*, *TFF1* and *TFF3*, which is representative for mucous-like cells, which are distinct from goblet cells (Widdicombe and Wine, 2015). The Secretory 5 subset contains a serous-like signature and represent glandular cell types of SMGs. Representative marker genes are *LYZ*), *Proline-Rich Proteins* (*PRBs,* and *PRRs*), and *Lactoferrin* (Carraro *et al.,* 2021). Interestingly, a *FOXN4*+ cell population was defined, which has been described as postmitotic subcell population in the developing proximal airway and this showed elevated *BPIFB1* levels during CF. It has been suggested that these cells may indirectly impact on alveologenesis during murine lung development (Li and Xiang, 2011).

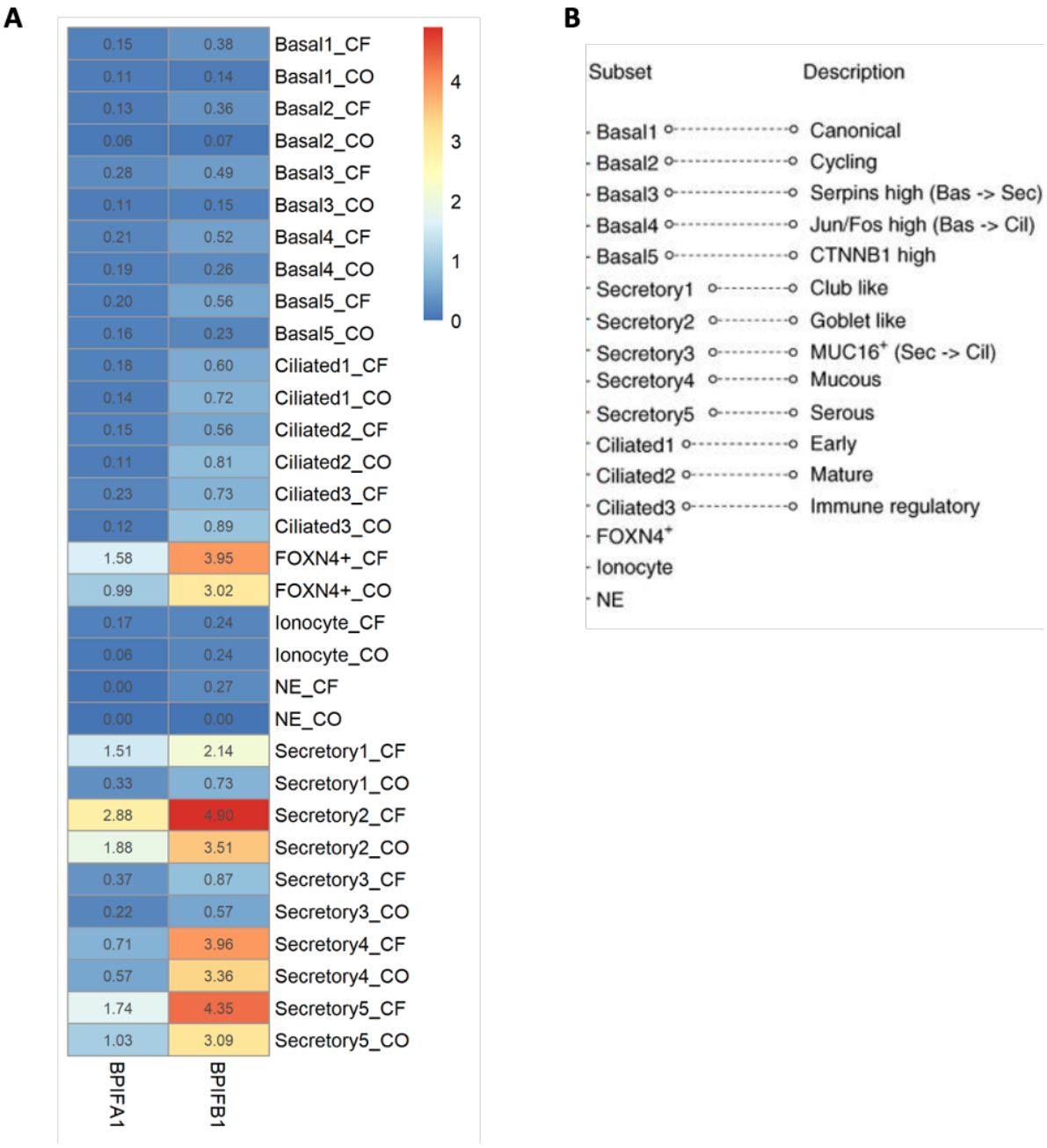


Figure 3.26: Comparison of *BPIFB1* with *BPIFA1* expression in Cystic Fibrosis lung tissue. A) Expression data for *BPIFA1* or *BPIFB1*. Expression table was provided by Brigitte Gomperts. **B)** Cell subset description from Carraro *et al.,* 2021.

In addition, Secretory 2 (goblet cell) and 5 (serous) cell populations show upregulated *BPIFB1* levels under CF conditions compared to control tissue, whereas Secretory 4 subcells do not demonstrate a modulated *BPIFB1* phenotype, rather express higher levels compared to *BPIFA1* (Figure **3.26**B). In contrast, *BPIFA1* shows across all populations a non-modulated phenotype, except for Secretory 2 cell populations (Figure **3.26**A).

Idiopathic pulmonary fibrosis (IPF)

Idiopathic pulmonary fibrosis (IPF) is a condition in which the lungs become scarred and thus lead to increasingly more difficult breathing. Till date reasons causing IPF are not well understood, and it usually affects people from 70 to 75 years of age, being rare in people under 50 years (Barratt *et al.*, 2018). A very recent genetic analysis investigating differentially expressed gene and pathways in IPF, identified *BPIFB1* as one of the three top DEGs showing the highest significant up-regulation. It was suggested that these DEGs play an important role in the manifestation of IPF through the mitogen-activated protein kinase (MAPK) (Li *et al.*, 2021a).

Additionally, BPIFB1 has been implicated in mucin regulation within the lung (Donoghue *et al.*, 2017) Donoghue *et al.*, 2020 on bioRxiv). *Bbifb1*-knockout mice appear to display abnormal mucin secretion, suggesting that the protein is associated with MUC5AC and MUC5B function (Donoghue *et al.*, 2017). Bingle *et al.,* (2013) reported that *BPIFB1* and *MUC5B* are upregulated in IPF, implying that increased *BPIFB1* expression might be linked to IPF.

The single-cell RNA-sequencing data set from Haberman and colleagues compared non-fibrotic control and pulmonary fibrosis lung tissue (Interstitial lung disease - ILD) contains over 114,000 cells from 20 pulmonary fibrosis and 10 corresponding control lungs and identified 31 distinct cell types (Habermann *et al.*, 2020). The online tool idiopathic pulmonary fibrosis (www.ipfcellatlas.com) was utilised to extract expression data for *BPIFB1* and to compare to BPIFA1 levels in the same data set (Figure **3.27**).

Overall, *BPIFB1* levels are higher (Figure **3.27**A) compared to *BPIFA1* (Figure **3.27**B). Notably, both genes are found in overlapping cell types, namely *MUC5B* and *MUC5AC* high cells as well as differentiating ciliated cells, and to a much lower level in *SCGB3A2* positive cells (Figure **3.27**D). This observation has been further visualised as violin blots (Figure **3.28**).

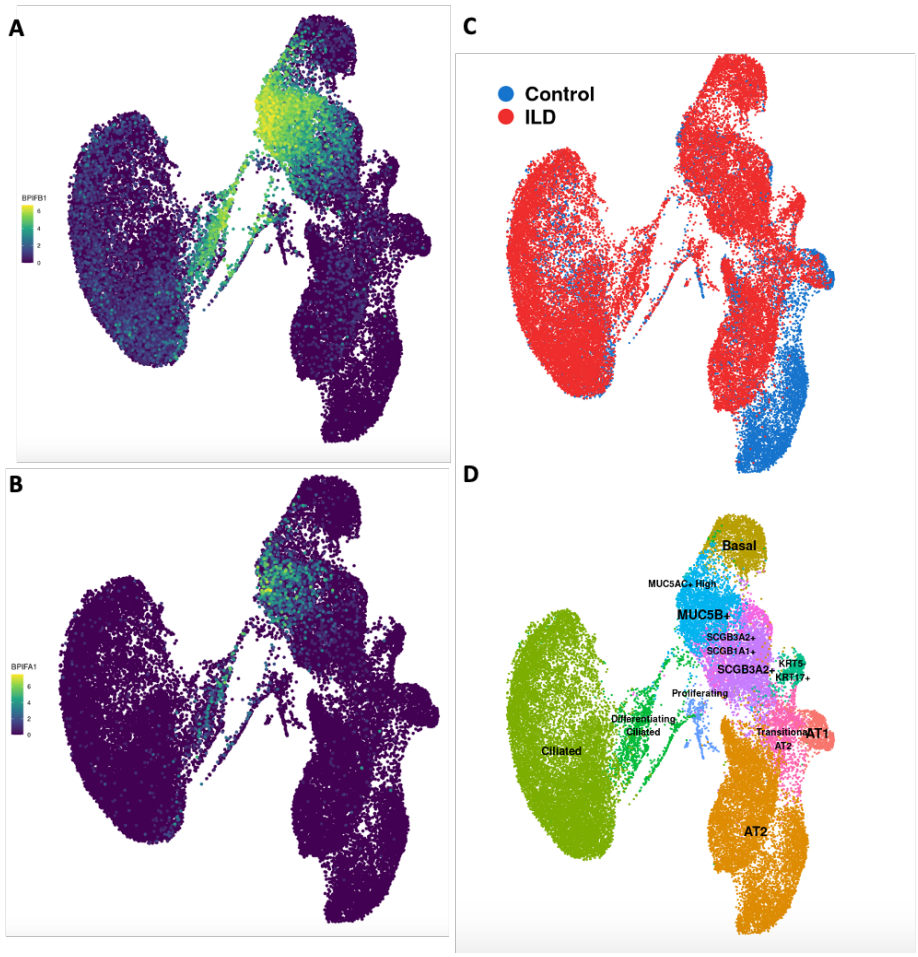


Figure 3.27: Expression *BPIFB1* and *BPIFA1* in control and ILD conditions is limited to mucin positive and differentiation ciliated cells. A) *BPIFB1* or **B)** *BPIFA1* UMAP expression representation. **C)** Demonstration of interstitial lung diseases (ILD) versus control cells (114,000 cells in total). D) Visualisation of distinct cell types across the data set. Data extracted from www.ipfcellatlas.com based on the Habermann *et al.,* 2020 study.

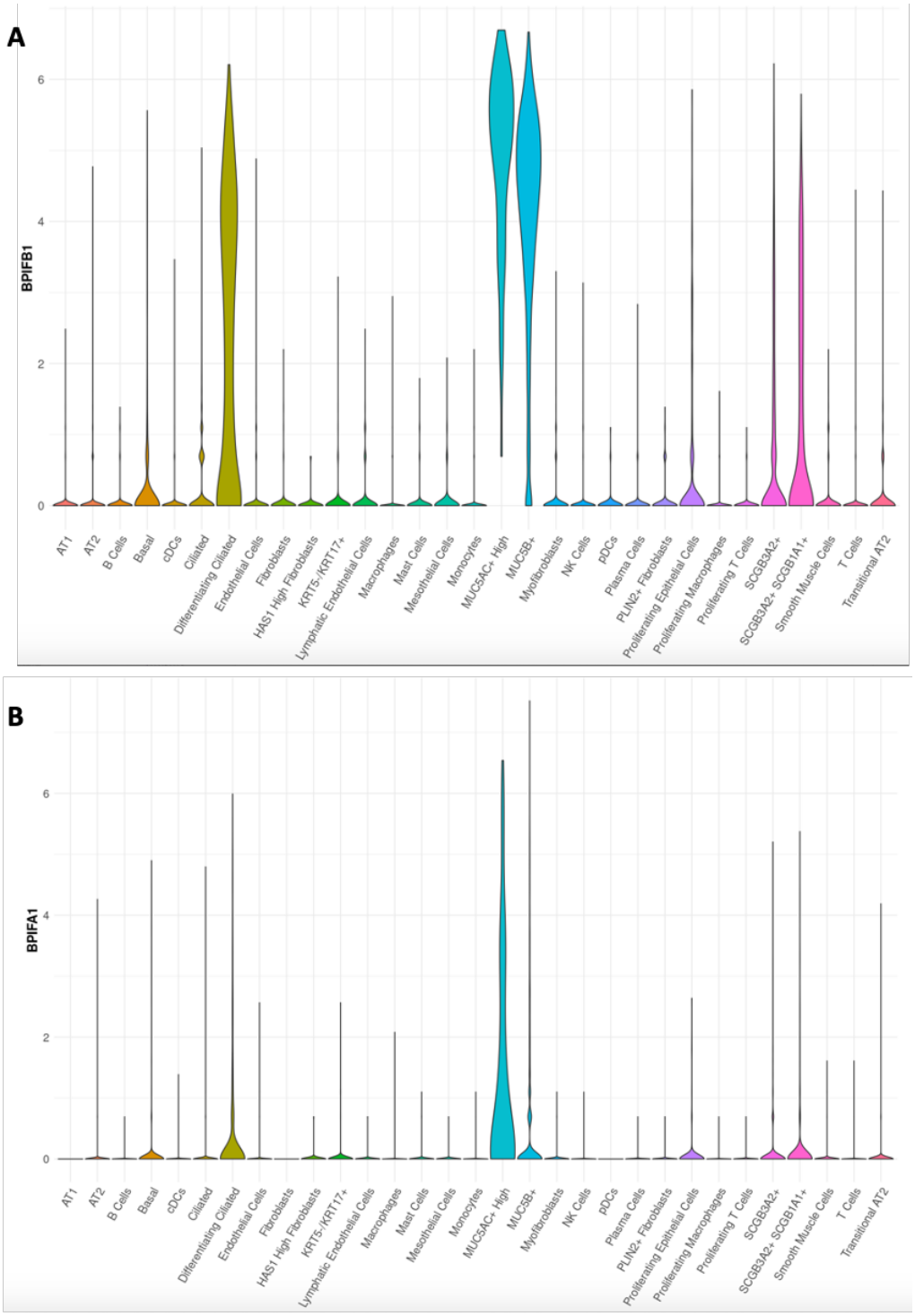


Figure 3.28: Violin plots of *BPIFB1* and *BPIFA1* expression. A) *BPIFB1* or **B)** *BPIFA1* violin plot representation across distinct cell types. Data extracted from www.ipfcellatlas.com based on the Habermann *et al.,* 2020 study, Banovich/Kropski gene explorer.

Asthma

Asthma is a common lung condition, which causes sporadic difficulty of breathing due to a narrowing and swelling of the airways, with potential additional mucus production. It is characterised by chronic airway inflammation and wheezing (Quirt *et al.*, 2018).

BPIFB1 has been found in sputum samples (Baines *et al.*, 2014), and BPIFB1 levels have found elevated upon mycoplasma infection (Fujii *et al.*, 2017), an infection which is sometimes associated with asthma (Hong, 2012; Wood *et al.*, 2013).

A recent scRNA-Seq study by Vieira Braga *et al.,* 2019 investigated the cellular landscape of the upper and lower airways in healthy lungs, as well bronchial tissue in asthmatic lungs. Asthmatic samples were obtained from bronchial biopsies from six volunteers demonstrating chronic, childhood-onset asthma and were compared to healthy sex- and age-matched volunteers. Key findings from the highlight that asthma induced mucous cell hyperplasia originates from a novel mucous ciliated cell state, as well as goblet cell hyperplasia, and the presence of pathogenic effector type 2 helper T cells (Vieira Braga *et al.*, 2019).

Employing the epithelial asthma airways atlas online tool provided by the Vieira Braga *et al.,* 2019 data set (https://asthma.cellgeni.sanger.ac.uk), it was possible to visualise differential expression of *BPIFB1* as well as *BPIFA1* across epithelia cells in asthma and healthy tissue (Figure **3.29**).

Similar to data from other studies, *BPIFB1* is expressed in club and goblet cells, submucosal and in a subset of mucus ciliated cells (Figure **3.29**C). Extending this view to the corresponding heatmap, it seems that asthma conditions mildly increase *BPIFB1* expression in ionocytes, goblet and ciliated cells as well as basal cycling cell populations (Figure **3.29**D). However, *BPIFA1* in contrast is less ubiquitously expressed and is found almost exclusively in club, submucosal and in a subset of mucus ciliated cell populations (Figure **3.29**E). In asthmatic conditions a significant elevation of *BPIFA1* is observed in club and ciliated cells, with a notable decrease of *BPIFA1* levels in mucous ciliated and basal activated cells, compared to healthy controls (Figure **3.29**F).

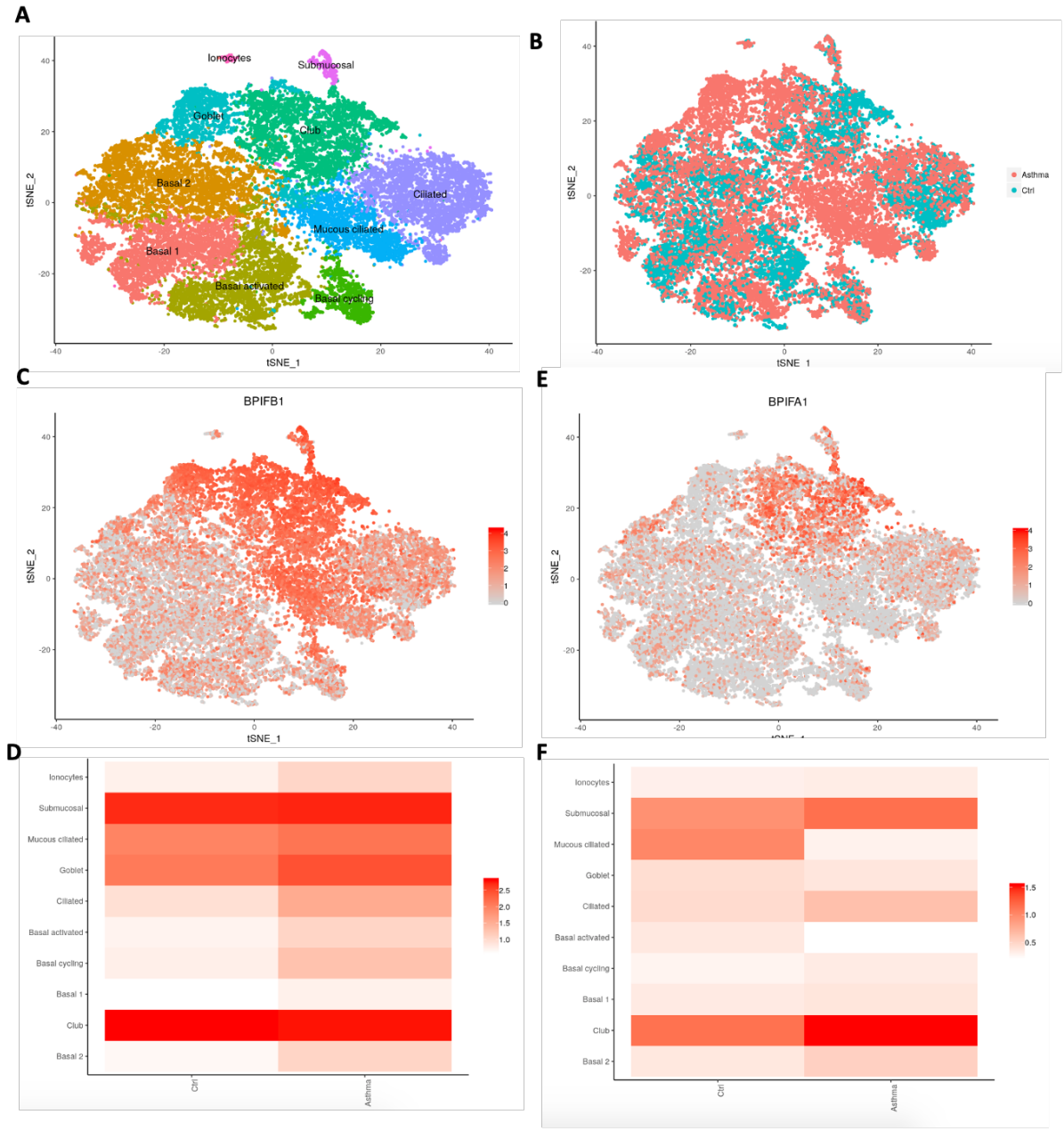


Figure 3.29: *BPIFB1* and *BPIFA1* expression in asthmatic and healthy control bronchial tissue. A) tSNE representation of cell clusters (n=25,146 cells) of bronchial tissues from the Vieira Braga *et al.,* 2019 data set. **B)** Distribution of cells derived from biopsies of the lower bronchi from asthmatic donors (red, n=15,033) or control counterparts (teal, n=10,113). **C)** Cell populations with corresponding *BPIFB1* expression and associated heatmap **(D)**, showing the control column on the left and asthmatic counterpart on the right. **E)** *BPIFA1* expression across bronchial tissue samples and corresponding heatmap **(F)**. Images were generated using the epithelial asthma airways atlas online tool at https://asthma.cellgeni.sanger.ac.uk.

This section aimed to investigate BPIFB1 gene expression across a range of lung diseases in order to widen the understanding of a potential role of BPIFB1 in respiratory homeostasis and disease.

BPIFB1 levels are upregulated in goblet cell and ciliated cell populations under COPD conditions, however at the same time a reduction of BPIFB1 expression is seen in Club cells. Another respiratory disorder, CF, showed an increased expression of BPIFB1 in the secretory cell pool containing goblet cells based on the marker set encompassing mucins MUC5B and MUC5AC, AGR2 and SPDEF. Additionally, BPIFB1 and MUC5B have been found upregulated in IPF, implying that increased BPIFB1 expression might be linked to IPF. In asthma mildly increase BPIFB1 expression in ionocytes, goblet and ciliated cells as well as basal cycling cell populations.

Overall, it is seen that *BPIFB1* levels are incrased in secretory cell populations in chronic lung diseases. However, this does not prove that increases in *BPIFB1* are causative.

3.4 Discussion

Among BPIF family members, the size and intron numbers are highly conserved, suggesting that they co-evolved during previous genetic events (Hubacek *et al.*, 1997; Bingle *et al.*, 2004; Bingle *et al.*, 2011a; Bingle *et al.*, 2011c). BPIF proteins which are found highly glycosylated, are expressed differently in different species or at different locations within the same species, which has been shown by different methods such as hybridization or multi-tissue cDNA microarray detection and thus demonstrated tissue specific expression of BPIF.

Using an alignments tool, it was demonstrated that BPIFB1 is found across mammalian species, although with a low level of similarity, and not in reptiles or fish. This perhaps indicates that evolving mammals with a lung based respiratory system are reliant on BPIFB1 function for homeostasis or immune defence. Furthermore, the alignment study revealed a specific primate C-terminus and conserved disulphide bridge, suggesting that these regions might be required for biological function as they have been preserved despite a rapidly evolving and divergent protein family (Bingle *et al.*, 2004).

Single cell studies confirmed that *BPIFB1* and *BPIFA1* are the two most abundantly expressed members of the BPIF protein family (Durante *et al.*, 2020; Deprez *et al.*, 2020). BPIFA1 and BPIFB1 proteins are largely confined to regions of the head and upper respiratory tract, where both proteins are found expressed in serous cells. Overall, *BPIFB1* expression has been found consistently higher, and more tracheo-bronchial and SMG goblet cell located. As it has previously been demonstrated that BPIFB1 and BPIFA1 were not co-localised (Vargas *et al.*, 2008; Musa *et al.*, 2012), this indicates a novel finding. This could be further verified using *in situ* hybridisation methods, which could confirm that both RNAs are expressed in the same cells (Haimovich and Gerst, 2018). In addition, RNA-Seq data largely confirmed previously published localisation of BPIFB1 and BPIFA1 proteins, which are expressed centrally in the nasopharynx, mouth, nasal cavity, respiratory tract, and digestive tract (Bingle and Bingle, 2011; Musa *et al.*, 2012; Alves *et al.*, 2017). However, RNA-Seq findings should be further validated on protein level, using immunofluorescence or double labelled IHC, where a dual-staining approach could verify co-expression of BPIFB1 and BPIFA1 (Chen *et al.*, 2010).

BPIFB1 expression has been detected in several tissues such as the respiratory or GI tract (Musa *et al.,* 2012, LeClair *et al.,* 2001; Hou *et al.,* 2004), however to my knowledge there are no publications which formally confirm BPIFB1 presence in the female reproductive tract or in muscle cells. With respect to the expression data of BPIFB1 in female tissues, more specifically the cervix, this data was derived from the GTex portal, which contained only a very small sample size of only 10 samples. In order to establish if BPIFB1 is truly expressed in these tissues, further studies need to be conducted, for example using an immunofluorescent approach to determine BPIFB1 protein levels.

An initial systematically protein expression analysis of BPIFB1 and BPIFA1 in mice demonstrated predominant expression in oral, nasopharyngeal, and respiratory tissues, however neither protein was present in kidney, heart, liver, pancreas, spleen, testes, or ovaries (Musa *et al.,* 2012). In addition, BPIFB1 is present in the serous glands in the proximal tongue where is co-localised with the salivary gland (Musa *et al.,* 2012). Studies using intestinal tract tissues, exhibited expression for Bpifb1 in the adult glandular stomach and neonatal intestinal tract(LeClair *et al.*, 2004; Hou *et al.*, 2004; Kallio *et al.*, 2010; Musa *et al.*, 2012) or BPIFB1 in respective human tissues (Sentani *et al.*, 2008; Shin *et al.*, 2011). However, as BPIFB1 is not found in other regions of the GI tract and located to Paneth cells only, this suggests that it is not a significant product of the gastrointestinal mucosa. In humans, this protein is present in a population of Paneth cells of the duodenum, and it has been shown to decrease pro-inflammatory responses to *V. cholerae* and *E. coli* LPS (Shin *et al.*, 2011). In contrast, BPIFA1 has not been detected in any gastrointestinal tissues (Musa *et al.,* 2012). Taken together, these proteins exhibit distinct and only partially overlapping localization based on protein expression studies, for example in the larynx. They exhibit limited expression outside of these regions (Musa *et al.*, 2012).

BPIFB1 is the most highly studied two-BPI domain containing member of the BPIF-family (Bingle and Bingle, 2011; Musa et al., 2012; Alves et al., 2017). Its distribution in the respiratory tract, as well as its structural relation to other innate immune proteins, has led to the proposal of the protein's role in the immune system. BPIFB1 may regulate the removal of debris and potential pathogens from the respiratory tract, and it has been implicated in respiratory diseases. However, no precise mechanistic role for BPIFB1 has been found yet and is has been suggested that BPIFB1 could act as a significant contributor to host immunity within the respiratory system. Therefore, the current working hypothesis is, that BPIFB1 may

be involved in the homeostasis of the immune system within the airways. In line with this, preliminary data from the host lab showed that *BPIFB1*$^{-/-}$ mice experience a higher viral load when challenged with influenza A virus. This data is consistent with other findings from the host lab implicating BPIFA1 as a regulator of IAV infection (Akram *et al.*, 2018).

BPIFA1, but not BPIFB1 or BPIFA2, have been demonstrated to inhibit the biochemical activation of ENaC sodium channels by preventing proteolytic processing and therefore activation of the channel (Garcia-Caballero *et al.*, 2009; Rollins *et al.*, 2010). Further, BPIFA1 is greatly increased in the small airways and plugged lumens in CF conditions (Bingle *et al.*, 2007). However, this was not observed in more recent single cell data sets (Carraro *et al.,* 2021). In contrast, *BPIFA1* has been found down-regulated in asthma models (Chu *et al.*, 2007; Wu *et al.*, 2017), which was only partly recapitulated the analysed data set from Vieira Braga *et al.,* 2019.

As mentioned above, BPIFB1 is most abundantly expressed in lung tissue and glandular cells of the stomach and thus might exert a protective role. This is consistent with findings discussing findings of an antiproliferative role of BPIFB1 in nasopharyngeal carcinoma (Wei *et al.*, 2018b).

Another interesting angle to test if BPIFB1 can exert a protective role in the lungs is the observation that BPIFB1 is found expressed in cells with high MUC5B or MUC5AC levels, and this has been associated with diseases such as COPD (Radicioni *et al.*, 2016) and *Helicobacter pylori* infection (Kato *et al.*, 2020; Li and Ye, 2020; Durazzo *et al.*, 2021). This could provide a link to observed BPIFB1 expression in stomach, where MUC5AC is expressed in the superficial epithelium and the upper part of the gastric pits. *H. pylori* positive patients exhibit a striking co-localisation of *H. pylori* and MUC5AC, where over 99% of the bacteria were associated with either extracellular MUC5AC or the apical domain of MUC5AC producing cells (Van den Brink *et al.*, 2000). Additional studies later confirmed that MUC5AC acts as a primary receptor for *H. pylori* (Van de Bovenkamp *et al.*, 2003).

Recent single cell RNA studies investigating respiratory tissues in healthy, or disease conditions led to a better understanding of *BPIFB1* expression, however different expression levels in various tissues or cell types potentially determine different biological activities. Overall, the efficacy of BPIFB1 in anti-bacterial activity, tumour inhibition, and respiratory disease is increasingly evident, but the detailed mechanisms behind its regulation are not yet clear and thus require further exploration and research.

Moreover, recent observations suggest that, although BPIFB1 and BPIFA1 do not seem to have an anti-bacterial role against *Pseudomonas aeruginosa*, they might be mediating an immunomodulatory function in CF airway epithelial cells (Saferali *et al.*, 2020).

Chapter 4:

Microarray data analysis of mTEC ALI culture during differentiation or IAV infection

4.1 Introduction

The closest related paralogue of BPIFB1 is BPIFA1, despite only possessing one BPIF domain, compared to two found in BPIFB1. From a respiratory focus, BPIFA1 localisation is restricted to the respiratory epithelium and the submucosal glands of the upper airways (Campos *et al.*, 2004; Bingle *et al.*, 2005), whereas BPIFB1 has been found expressed in the trachea, respiratory epithelium and submucosal glands of larger and some smaller airways. BPIFB1 has been identified in goblet cells of the respiratory system, whereas BPIFA1 derives from serous, nonciliated, nongoblet epithelial cells (Bingle *et al.*, 2005; Bingle *et al.*, 2010; Gao *et al.*, 2015). Notably, neither BPIFA1 nor BPIFB1 are detected in the alveolar regions of a healthy lung (Bingle *et al.*, 2010; Bingle *et al.*, 2012). However, under conditions of a compromised lung, BPIFB1 and BPIFA1 levels have been found dysregulated, compared to healthy tissue (Steiling *et al.*, 2009; Bingle *et al.*, 2012; Jiang *et al.*, 2013; Titz *et al.*, 2015; Gao *et al.*, 2015). Due to the fact that expression and secretion of both proteins is dependent on the differentiation status of the epithelial cell population, studies *in vitro* are relied on a robust experimental model which employs lung cell cultures grown at air–liquid interface (Campos *et al.*, 2004; Ross *et al.*, 2007; Bingle *et al.*, 2010; Martinez-Anton *et al.*, 2013).

Murine models are the most commonly employed model system to mimic respiratory infections and associated diseases, as the mouse trachea is very similar to the human trachea, bronchi, and bronchioles, and thus has been widely used for *in vivo* and *in vitro* studies (Cormier *et al.*, 2010; Bayes *et al.*, 2016; Pilloux *et al.*, 2016; Akram *et al.*, 2018). Therefore, establishing primary mouse tracheal epithelial cells (mTEC) provides a valuable device for studies of the composition and response of airway epithelial cells during host-pathogen interactions. In addition, utilizing genetically altered mice allows to determine the contribution of host genetics to the development and response of the airway epithelial cells.

Over the last two decades, the murine mTEC ALI culture has been further developed into a well-established model to study airway epithelial cells, which now allows for expansion of mTECs, while maintaining their differentiation capacity (Davidson *et al.*, 2000; You *et al.*, 2002; You and Brody, 2012; Mulay *et al.*, 2016; Schilders *et al.*, 2016). Consequently, the required numbers of animals required has been reduced, as has biological variation in the cultures. Furthermore, development of murine *in vitro* cell culture models, mimicking *in vivo* host airway epithelium and its microenvironment, provides unique possibilities to study immune responses of airway epithelium to viral stimuli.

The mTEC system relies on extraction of specific airway epithelial cells from murine tracheal tissue using a protease treatment, removal of contaminating cells such as fibroblasts and plating on transwell membranes for expansion of the basal, epithelial stem/progenitor cell population in a submerged fashion (Davidson *et al.*, 2000; You *et al.*, 2002). When confluence is reached, cells are lifted to the ALI and culture continued using specific culture media, only supplied from the basal side of the porous transmembrane, mimicking an airway epithelial environment. This stimulates the differentiation of basal cells and recapitulates the distinct cell types seen in mouse tracheal epithelium such as non-ciliated secretory cells and ciliated cells, as well as the formation of cellular tight-junctions (Ostrowski and Nettesheim 2009, Davidson *et al.,* 2000, Mulay *et al.,* 2016, You et al., 2002, You and Brody, 2013). Secretory cells produce mucins and antimicrobial peptides and metabolize toxins, whereas ciliated cells use their cilia to propel debris out of the lung (Jeffery and Li, 1997; Tata *et al.*, 2013; Tata and Rajagopal, 2017).

Airway epithelia present the first target of IAV infection, which are primary responders, thus contributing to immunity, pathology and recovery. The associated responses may depend, amongst other factors, on the host genetic background, which mediates susceptibility and protection. Using mTECs cultures in undifferentiated or differentiated conditions allows to focus on the contribution of epithelial cells in an infectious setting and to study the effect of host genetics on the response to IAV in addition to assessment of effectiveness of a potential therapeutic treatment against IAV. Thus, understanding the biological activity of BPIFB1 in host defence against viral and bacterial infections could provide important insights in the understanding the functions of the protein, and this would allow for an investigation to whether the loss of BPIFB1 in mTEC culture would lead to the greater susceptibility to IAV infection.

4.2 Aims

mTEC were harvested from C57BL/6J wt mice and isolated with the aim to expand and establish ALI cultures. My initial aim was to establish the culturing techniques so as to be able to undertake functional experiments on the role of BPIFB1 in antiviral host defence. Originally, the intention was to obtain *Bpifb1*$^{-/-}$ modified animals and draw direct conclusions between wild type (wt) and *Bpifb1*$^{-/-}$ mTEC cultures, but due to COVID-19 imposed breeding restrictions, only wt mice were available during the conduct of this work and thus these were used to setup the initial model system in a pilot study.

mTEC were differentiated in culture into upper airway-like epithelium and these were routinely validated by using a range of techniques. A genome wide microarray analysis was performed in order to analyse gene signatures associated with basal cell expansion and ALI cell differentiation and expression of BPIFB1 studied during this process. Finally, the genome wide responses of undifferentiated and differentiated mTECs to IAV infections were directly compared to focus on modulated genes during IAV challenge in both conditions.

4.3 Results

4.3.1 Validation of mTEC cultures and establishing infections with IAV

4.3.1.1 Cell harvest and differentiated cell model system

mTEC isolation is an established technique within the lab and has been used across several projects (Akram *et al.*, 2018). The current protocol was adapted from published methods (You *et al.*, 2002; You and Brody, 2012) and subculturing steps after extraction and employed were performed as illustrated in Figure **4.1**.

I initially sought to establish this technique for my own studies. Resected tracheas, cleaned of adhered tissue were placed in sterile mTEC basic media containing antibiotics. Tracheas were cut open lengthwise and incubated overnight in 0.15% pronase solution in mTEC basic media at 4ºC. Subsequently, pronase activity was neutralised with FBS and epithelial cells were gently dislodged from tracheas by gentle tube inversions.

The cell pellet collected by centrifugation was resuspended in cold DNase/BSA solution and incubated in a cell culture dish for 3-4 hours at 37°C, to allow for the fibroblast population to attach to the surface. The non-adherent cells, enriched for epithelium, were collected, and resuspended in mTEC plus media and a small population was used for RNA extraction, serving as an original or wt cell reference. The remainder of the isolated epithelial cell population was seeded at the density of 3-3.5x10^4 cells per transwell insert.

Cells were cultured in submerged conditions using mTEC plus media in both chambers until they reached confluency, which usually took seven to eight days. When confluent, this day was termed Day 0 (D0) and mTECs were differentiated into upper airway-like epithelium in ALI conditions for at least 14 days (D14) using mTEC ALI media in the basal chamber only (You *et al.*, 2002; You and Brody, 2012) (Figure **4.1**). D0 ALI cells were collected for RNA extraction to serve as undifferentiated cell reference compared to the D14 ALI differentiated sample set.

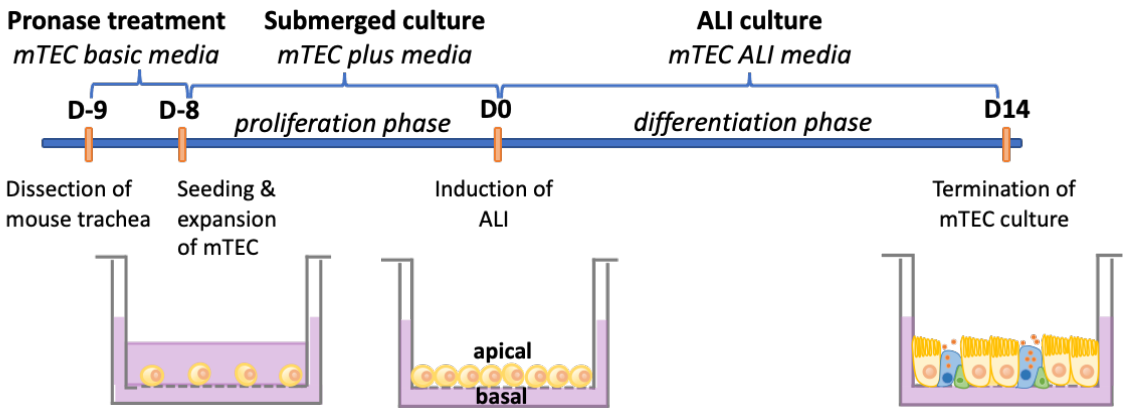


Figure 4.1: mTEC culture schematic and timeline. Cells were isolated from resected trachea approximately nine days prior (D-9) for induction of ALI and cultured in transwell inserts as submerged culture post pronase treatment (D-8) until reaching confluence (D0). Samples for validation and transcriptional analysis were collected at seeding (original cells, wt), day 0 and day 14. The submerged culture (D-8 to D0) was grown as proliferation and extension phase, and upon reaching confluence ALI was induced using a specific formulated media (mTEC ALI media) and only supplying it to the basal portion of the chamber, leading to the differentiation phase. Cells were cultivated on in the apical chamber on a semi-permeable membrane with a 0.4 µm pore size. The use and time frame of specific grow media is indicated. Yellow cells represent ciliated, blue secretory cells and green basal cell populations.

4.3.1.2 Validation of mTEC differentiation

The growth and morphology of mTECs seeded into transwells was always observed during submerged and ALI culture conditions. Epithelial cells appeared elongated, forming contacts with adjacent cells, leading to epithelial cell islands. Cells proliferated in the submerged culture condition and formed a confluent monolayer within seven to eight days of seeding.

Using this method, ALI culture was induced when the cells reached confluence on the wells, and within seven days of ALI culture cell morphology changed. At day 14 of ALI culture, two major types of mTECs, flat non-ciliated cells and ciliated cells are clearly observed and showed well-defined cell boundaries (Figure **4.2**).

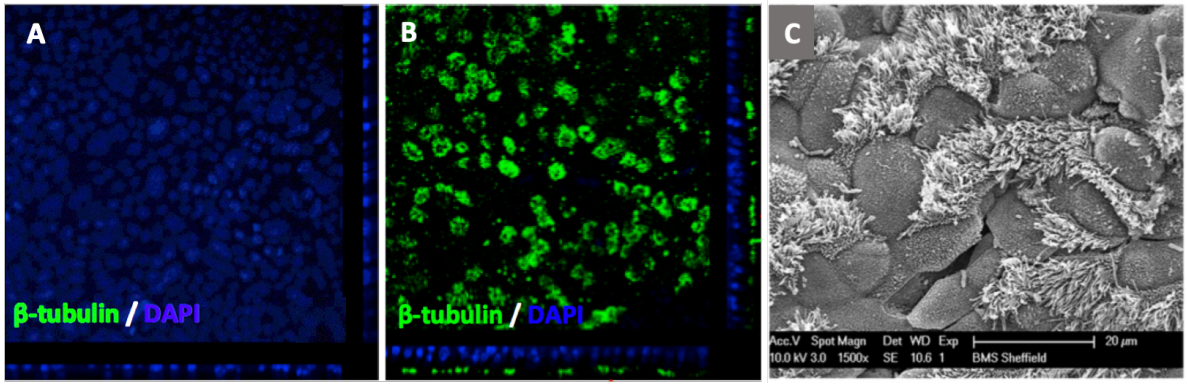


Figure 4.2: Visual characteristics of mTEC ALI culture. Representative confocal images of ALI D0 **(A)** and ALI D14 **(B)** mTEC cultures counterstained for cell nuclei (DAPI blue) and β-TUBULIN (green) identifying ciliated cell bodies. **(C)** Scanning electron microscope of a D14 ALI mTEC culture demonstrating ciliated and non-ciliated cell populations forming a confluent closed monolayer. Image was was generated by Dr Khondoker Akram.

The morphology of non-ciliated cells was of cobble-stone nature, characteristic of upper airway epithelium (Eenjes *et al.*, 2018). During the isolation procedure, mouse tracheal fibroblasts were also routinely cultured to obtain a pure population of fibroblasts to be used as negative control for validation studies.

mTECs grown in this manner were previously shown to exhibit staining of BPIFB1 and BPIFA1 (Figure **4.3**) in a non-ciliated cell population in mTEC cultures, consistent with the *in situ* localisation of the proteins in secretory cells of the native mouse upper airways (Musa et al 2012).

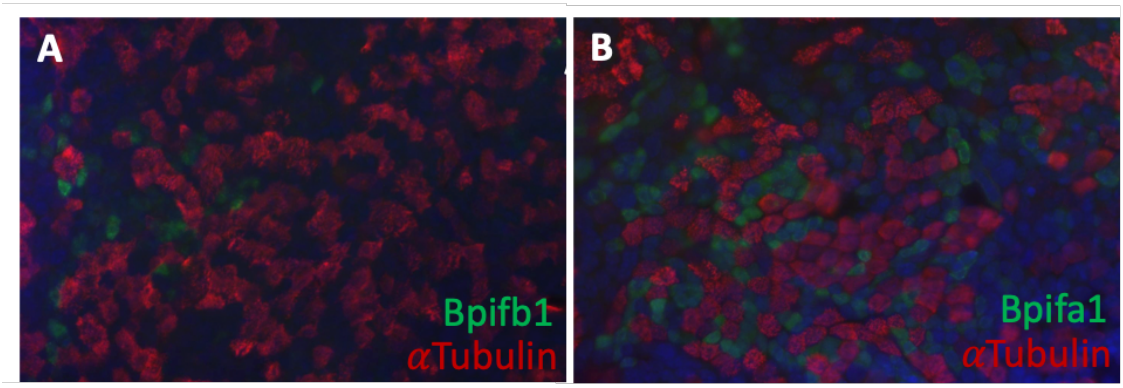


Figure 4.3: Immunofluorescent analysis of BPIFB1 and BPIFA1 in mTEC ALI at D14. BPIFB1 **(A)** or BPIFA1 **(B)** was visualised in the green-fluorescent channel, highlighting non-ciliated cell populations. Anti-acetylated α-TUBULIN (αTUBULIN red) has been employed as ciliated cell marker to identify ciliated cell populations and cell nuclei were counterstained with DAPI (blue). This figure was generated by Nick Yozamp and Steve Brody at Washington University, St Louis, USA.

Consistent with published localisation data, BPIFA1 and BPIFB1 were detected by immunofluorescence at D14 in a cell population with is strictly exclusive of a ciliated cell marker (Figure **4.3**). Notably, BPIFA1 staining intensity seem to be higher compared to BPIFB1. It may be that this is a primary antibody affinity related phenomenon. Nonetheless, this existing data shows that BPIFB1 is observed in differentiated mTECs in ALI cultures. This consistent with data from human HBECs produced in the lab,

ALI culture induced differentiation states were also routinely tested on a gene expression basis, by performing endpoint PCR investigating the level of expression of the epithelial cell markers *Tekt1*, specific for ciliated cell types, and *Bpifb1* and *Bpifa1* as secretory cell markers. *Oaz1* served as internal positive control (Figure **4.4**). As expected, none of these markers were expressed in fibroblast samples (Figure **4.4**).

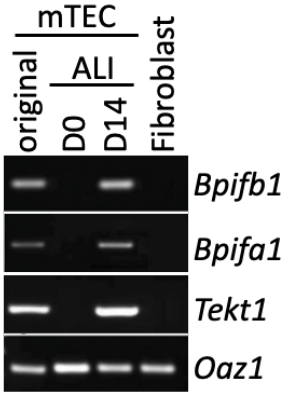


Figure 4.4: Representative endpoint PCR validation of epithelial cell markers. mTEC cells were isolated and cultured as outlined in the text, mRNA extracted from the original wt tissue after pronase treatment, and from cell cultures at 0 (D0) or 14 (D14) days post cultivation in a transwell ALI setup. RNA from a fibroblast culture served as negative control for epithelial cell markers. Gene expression analysis of *Bpifb1* (148 bp), *Bpifa1* (127 bp), *Tekt1* (373 bp) and *Oaz1* (274 bp) was carried out by amplifying equal amounts of cDNA by endpoint PCR. *Oaz1* served as an internal positive control.

In addition to the gene expression analysis, a protein-based approach was previously carried out on differentiating mTECs. The basal chamber media was changed every second day during ALI culture and apical washes performed with the aim to detect and analyse secreted BPIFB1 and BPIFA1 during differentiation. This data shows that cultures release BPIFB1 and BPIFA1 into the apical secretions in a time (and differentiation) dependant manner (Figure **4.5**). As expected, BPIFB1 was absent at the start of the ALI culture at day 0, but

showed low levels at day 3, which increased at day 7, and remained at a similar level at day 14 (Figure **4.5**). Comparably, secreted BPIFA1 was initially detected three days post ALI culturing, and was greatly increased at day 7 and then remained at a similar level at day 14 ALI (Figure **4.5**). Consistent with the IF data it appeared that BPIFA1 secretion was greater than BPIFB1.

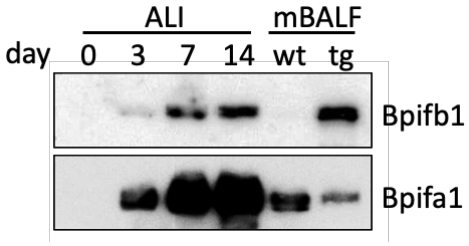


Figure 4.5: Western blot analysis of ALI cultured mTEC apical washes. mTEC cells were ALI differentiated as described in the text, and apical washes performed on day 0, 3, 7 and 14. Collected samples were separated by SDS-PAGE and western blotted with BPIFB1 and BPIFA1 antibodies to visualise secreted amounts of the respective proteins. BAL fluid from WT and *scgb1a1*^{Scnn1b} mice (tg) served as a positive control. Data from Dr Khondiker Akram.

Having established mTEC cultures as an experimental system recapitulating mTEC differentiation and that could be used for future studies, I was interested to study genome wide gene expression profiles of mTECs as they underwent development of a mucociliary phenotype of a pseudostratified epithelium. In the context of this study, my particularly interest was to investigate the expression of BPIF family members throughout that process.

4.3.2 Microarray data analysis

In order to analyse genome wide transcriptional changes, induced during cell cultivation after mTEC harvest and during ALI cultivation, a microarray study was designed and carried out by Priyanka Anujan and Fawaz Aljuhani. The initial array analysis was undertaken by the core facility run by Paul Heath at SITraN, University of Sheffield, and data extraction and normalisation was performed by Miraj Kobad Chowdhury and provide to me for analysis. This analysis used RNA from cell populations at the time of cell harvest (original wt cells) and mTECs at D0 and D14 of ALI culture (Figure **4.1**). In addition, the effects of IL-13in the differentiation model was also studied, as earlier observations reported that IL-13 causes

goblet cell metaplasia/hyperplasia (Kondo *et al.*, 2006). To mimic enhanced IL-13 levels, one set of ALI culture was continuously supplemented with 10 ng/ml IL-13 in the regular ALI culture media, which was replenished every second day (Figure **4.6**).

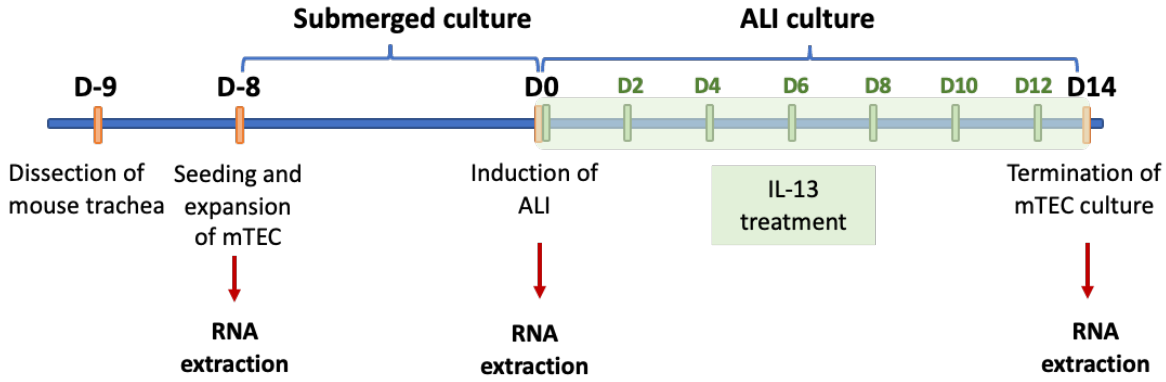


Figure 4.6: Sample collection time points and IL-13 treatment schedule. mTECs were isolated and expanded as described in the text. Triplicate sets of mTECs were continuously treated with ALI media supplemented with 10 ng/ml IL-13, starting at D0 until D12 as indicated.

A previous western blot analysis of apical washes had showed that treatment with IL-13 increases BPIFB1 levels in the apical secretions, observed from D3 onwards compared to IL-13 negative ALI samples (Figure **4.7**).

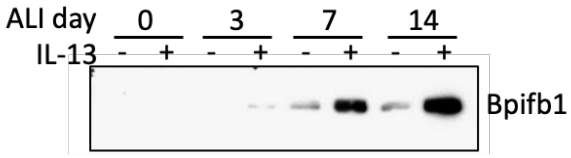


Figure 4.7: Western blot analysis of BPIFB1 in apical secretions during ALI differentiation in presence (+) or absence (-) of IL-13. Apical washes were collected on ALI day 0, 3, 7 and 14 and equal volumes of samples were resolved by SDS-PAGE and probed with a BPIFB1 antibody. Data from Dr Khondoker Akram.

The returned microarray data sets, presented as counts, were analysed for differentially expressed genes (DEGs) with limma-voom (Law *et al.*, 2014; Ritchie *et al.*, 2015; Costa-Silva *et al.*, 2017). Three biological replicates of each experiment were originally provided for micro array analysis but due to technical difficulties in the SITraN facility, the analysed datasets therefore only contain two replicates per condition.

4.3.2.1 Data quality control and validation

To assess which genes are involved in mTEC differentiation, three data sets were generated, comparing original cell populations after tracheal extraction (wt) with expanded cell populations grown as submerged cultures on transwell inserts before ALI culture day 0 (D0) and 14 day ALI cultured cells (D14) (Figure **4.1**). In addition, the IL-13 treated D14 ALI sample set was compared alongside the ALI D14 culture (Figure **4.6**). Before any DEG analysis were performed, the data sets were quality controlled and normalised as outlined below.

The initial step filtered for weakly expressed genes as they provide little to no evidence for differential expression and in addition interfere with statistical approximations. Furthermore, as they add to the testing burden with regards to estimating false discovery rates, this can reduce the power in detection of true differentially expressed genes. Therefore, poorly expressed genes were filtered out prior to further analysis, using minimum counts-per-million (CPM) threshold present in both samples per group, with genes retained expressed at a CPM above 0.5 (Law et al, 2016).

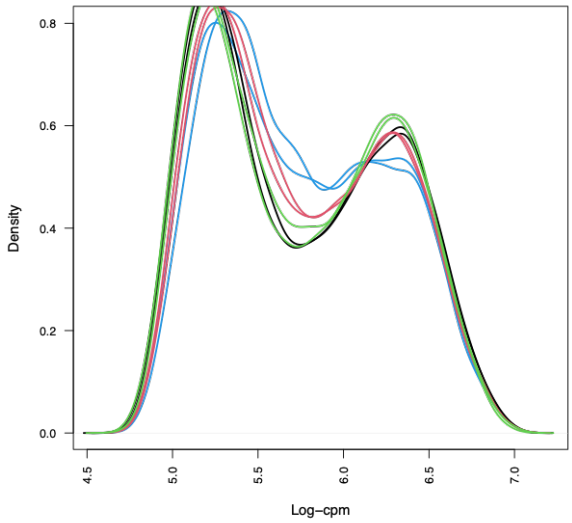


Figure 4.8: Density plot of differentiated sample groups. Counts distributions after filtering are displayed across all data groups (wt – blue, D0 – black, D14 – red, D14 IL13 – green). Genes without more than 0.5 CPM in at least two samples are determined insignificant and filtered out, which in this data set yielded 0 of 22206 (0%) genes were omitted due to low expression. Note: replicate samples are indicated as A or B.

The distribution of the raw counts is plotted in log2 scale, where a CPM value of 0.5 is equivalent to a log-CPM of -1. Examining the distribution before filtering in the density plot

(Figure **4.8**A) already reveals log-CPM values above 4.5 and thus, 0 of 22206 genes have been removed leaving the density plot shape unaltered (Figure **4.8**B). This indicates that this data has been normalised at the SITraN facility, as there were no poorly expressed genes present.

Data set counts were normalised with the aim to eliminate composition biases between samples. In other words, it has been observed that highly expressed genes could dominate samples, and consequently lead to lower read counts of other genes. For these data groups the default Trimmed Mean of M values (TMM) normalisation was applied. This normalisation algorithm applies a weighted trimmed mean of the log expression ratios to scale the counts for the samples (Robinson and Oshlack, 2010). To visualise the distributions of counts across the samples, box plots were generated, before and after TMM normalisation (Figure **4.9**). In general, it is expected that the samples are distributed close to the median horizontal line, which is the case for these differentiation data groups, and therefore they do not require further investigation or significant normalisation (Figure **4.9**).

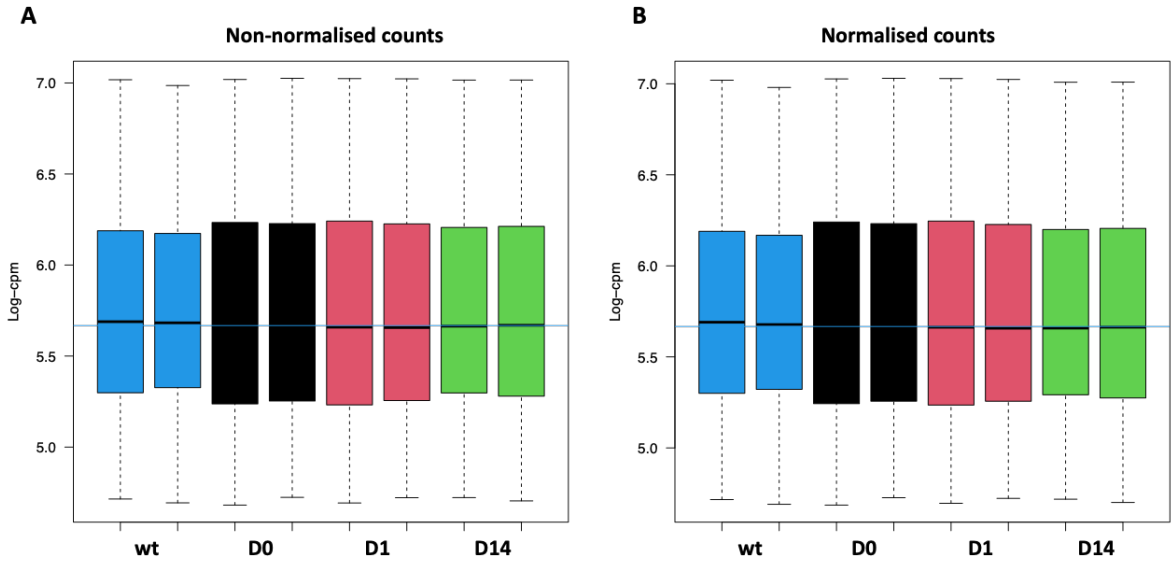


Figure 4.9: Box plots of mTEC differentiation data sets. Non-normalised **(A)** vs TMM normalised counts **(B)**. wt – original mTEC population, D0 – start of ALI culture, D14 – end of ALI culture. D0 is representative of the undifferentiated population, whereas D14 indicates the differentiated population. The D14 IL13 sample set has been IL-13 supplemented (10 ng/ml, every 48h) from D0 to D12. The blue line is indicating the data median.

Multidimensional scaling plots (MDS) are most appropriate to assess data quality, outliers, and that samples and replicate distribution behave as expected. They allow for visualisation

of a principal components analysis, and thus provide a tool to assess the greatest sources of variation in the data (Ritchie *et al.*, 2015). The analysis is carried out in an unsupervised manner and thus is sample group independent. Therefore, a well performed and controlled experiment should show the highest source of variation in the treatment group compared to the control samples.

The associated scree plot displays how much variation derives from each dimension, and therefore, for example batch variations, would cause high values for additional dimensions.

The MDS plot of the differentiation data set shows a very high degree of reproducibility between both data replicates, with a cluster of D14 and D14 IL13 treated samples, whereas wt and D0 data sets were found to be most distantly apart (Figure **4.10**).

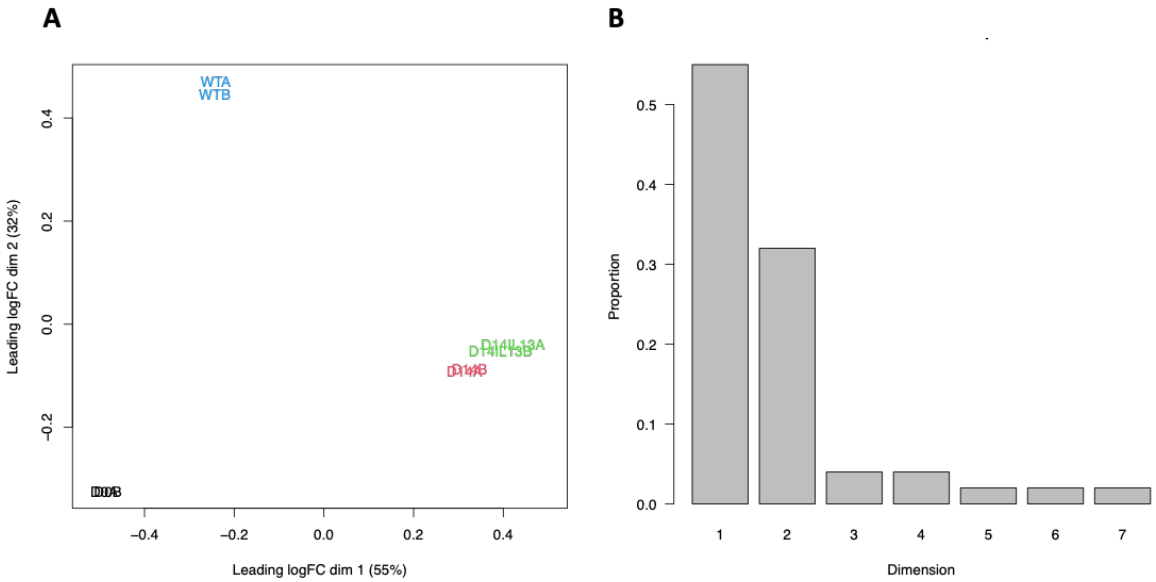


Figure 4.10: Quality control of variation within the mTEC differentiation data sets. (A) MDS plot dimension 1 versus 2 shows a high reproducibility between both replicates per sample group, with a distinct cluster of D14 (red) and D14 IL-13 (green) compared to D0 (black) or original mTEC (wt, blue). **(B)** The scree plot highlights variation derived from each dimension.

Voom variance plots represent all genes in the form of individual dots, and thus illustrate the mean-variance relationship of the genes in the data set (Ritchie *et al.*, 2015). This demonstration allows assessment of if low counts have been filtered satisfactorily and shows the level of data variation (Figure **4.11**A).

Another routine diagnostic tool is the scatterplot of residual-variances vs average log-expression (SA plot), which draws log2 residual standard deviations against mean log-CPM

values. This graph demonstrates removal of the dependency between mean and variance after fitting a linear model (eBayes) and voom weights are applied to the data (Figure **4.11**B) (Smyth, 2004; Phipson *et al.*, 2016).

The voom and SA variance plots investigating levels of variation in the differentiation data sets show a very low-level biological variation with very few outliers, highlighted in red (Figure **4.11**).

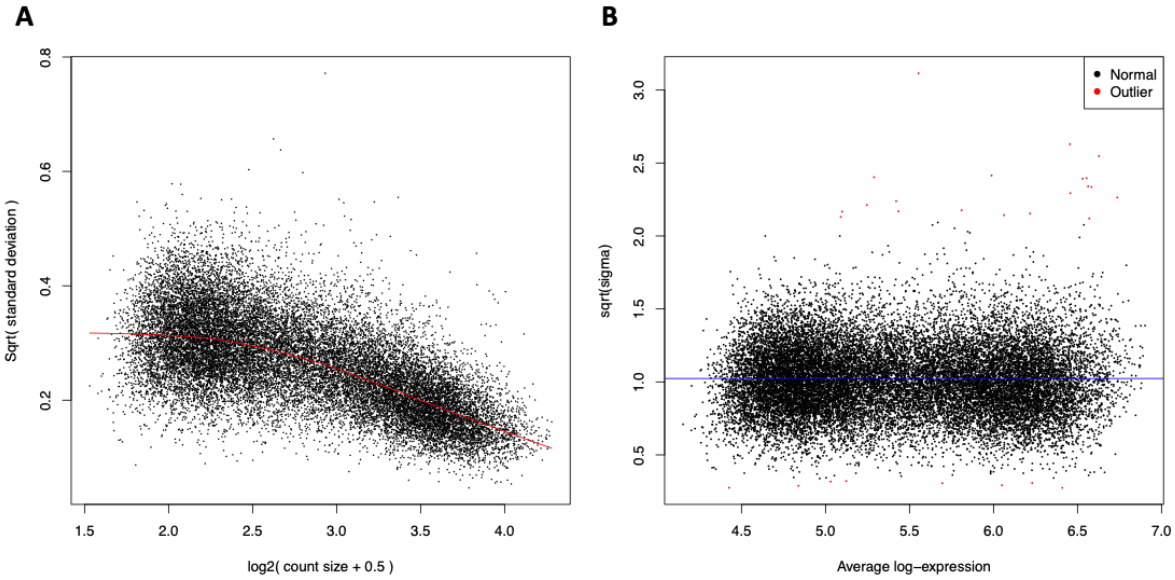


Figure 4.11: Quality control of mean-variance variation within the mTEC differentiation data groups. Voom mean-variance dot blot **(A)** versus scatterplot of residual-variances vs average log-expression (SA plot) **(B)**. The average log2 residual standard deviation is marked by a horizontal blue line.

The quality control of the differentiation data sets described above, revealed them to be of high quality with low sample variation but before the analysis of DEGs within the differentiation condition or IL-13 treatment was undertaken another level of data validation was carried out by comparing gene expression levels of selected epithelial cell markers, which should recapitulate expression differences observed with the PCR results (Figure **4.4**) shown earlier during validation of the mTEC model system (Figure **4.1**).

Extracting and plotting gene expression data for the secretory markers *Bpifb1* and *Bpifa1* alongside the cilia marker *Tekt1* across differentiation key time points showed the expected expression patterns (Figure **4.12**). The goblet cell marker *Bpifb1* (Figure **4.12**A), displays a high expression, which drops drastically at the start of ALI differentiation at D0. During ALI

culture, this level increased to about half of the original expression level at the point of mTEC extraction, and IL-13 treatment during the ALI differentiation phase rescued this expression to levels similar to that seem in the original state (Figure **4.12**A). Analogously, *Bpifa1* was detected at a high level in original mTEC cells, was significantly reduced at D0 and recovered completely after 14 days of ALI culture (Figure **4.12**B). In contrast to *Bpifb1*, IL-13 did not affect expression levels of *Bpifa1* (Figure **4.12**).

As observed in the initial validation studies, ALI cultures developed ciliated cells within 7-14 days (Figure **4.2**). Consistently, a reduction of the cilia cell marker *Tekt1* was detected at D0 compared to original wt mTEC and levels reached a higher than wt level at D14. This was further increased in culture were IL-13 treated (Figure **4.12**C). *Oaz1* was detected in all groups at similar expression levels, validating sample input and data processing (Figure **4.12**D).

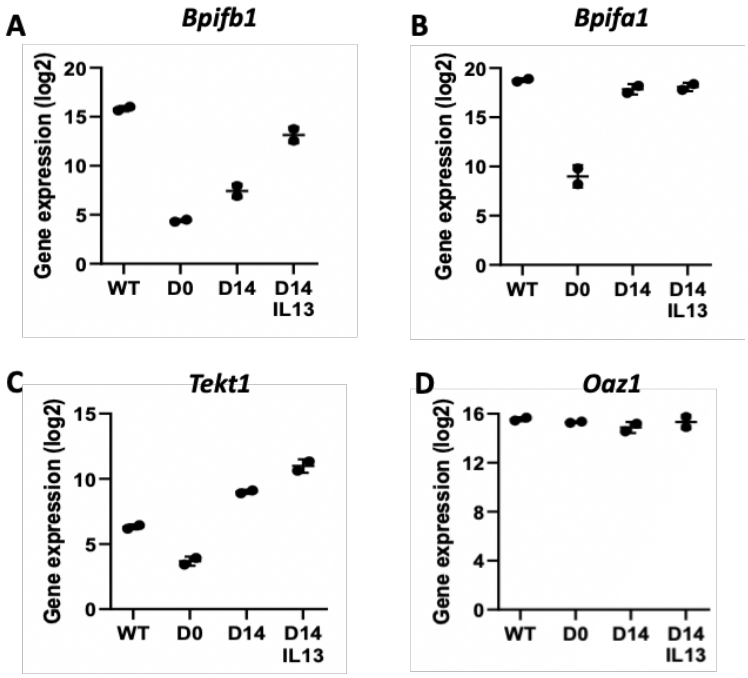


Figure 4.12: Gene expression levels of selected mTEC differentiation marker genes. Comparison of gene expression levels of the goblet cell marker *Bpifb1* **(A)**, secretory marker *Bpifa1* **(B)** and ciliated cell marker *Tekt1* **(C)** from original wt mTEC, undifferentiated (D0), ALI differentiated (D14) and ALI D14 IL-13 treated cell populations. *Oaz1* **(D)** served as internal control.

This expression data is consistent with that from an independent set of sample showing RNA expression profiles of mTEC at ALI D0 and D14 on secreted proteins of the Bpif family (Figure **4.13**) (Nemajerova *et al.*, 2016). For example, in this data set *Bpifb1* expression was

increased after 14 days of ALI culture (Figure **4.13**A) and *Bpifa1* is found dramatically elevated (Figure **4.13**B) to similar levels as observed with this study's data (Figure **4.12**). However, in this other experimental data set, the cilia marker *Tekt1* (Figure **4.13**C) was found to be more enhanced, between day 0 and day 14 compared to our data set. *Oaz1*used as internal marker and input control (Figure **4.13**D) was similar between the two time points in this study.

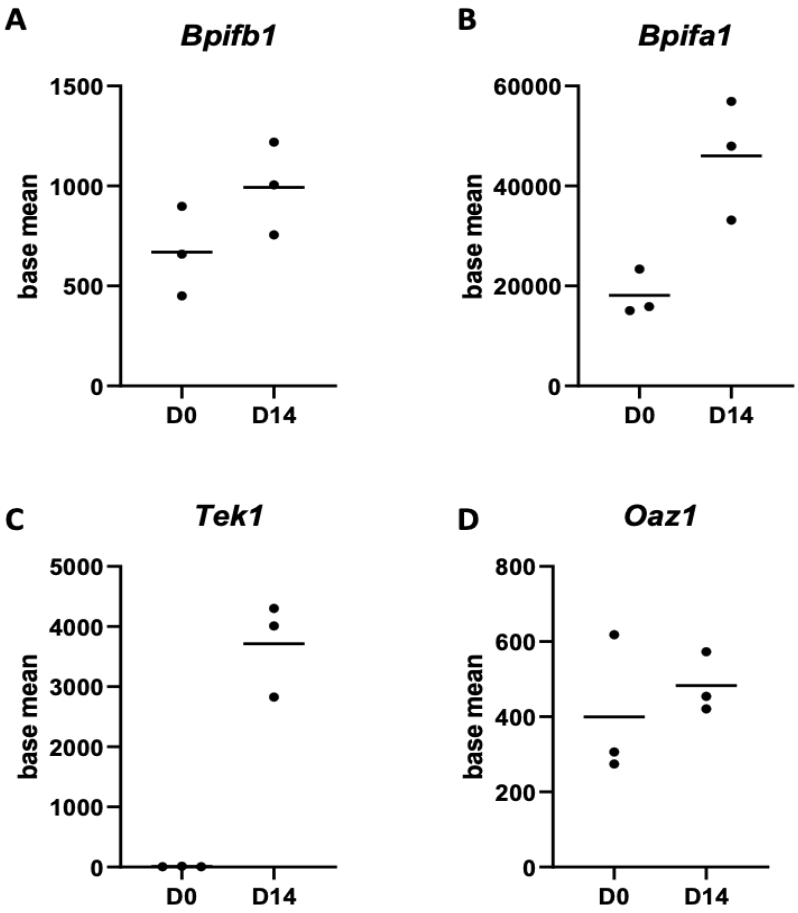


Figure 4.13: Gene expression levels of selected mTEC differentiation marker genes in data from Nemajerova *et al.* RNA expression profiles by RNA-Seq of mTEC at ALI D0 and D14 gene expression from the Nemajerova et al data set. Bar indicates mean of three samples. Differentially expressed genes were analysed using an adjusted p-value threshold of <=0.1 and log2-fold-change of 1 (Nemajerova *et al.,* 2016).

I next investigated the expression of all *BPIF* family members using our genome wide expression data. Notably, expression all *Bpifa* genes was around the level of detection of the assay with the exception of *Bpifa1*(Figure **4.14**A). A similar conclusion could be drawn for *Bpifb genes*, with *Bpifb1* being the only significantly expressed gene (Figure **4.14**B).

Therefore, it appears that *Bpifa* and *Bpifb1* are the only BIPF family members which are abundant in mTECs, as the other members are found expressed at detection threshold of log2 of 4-6 (Figure **4.14**) (Gonzalo Sanz and Sánchez-Pla, 2019).

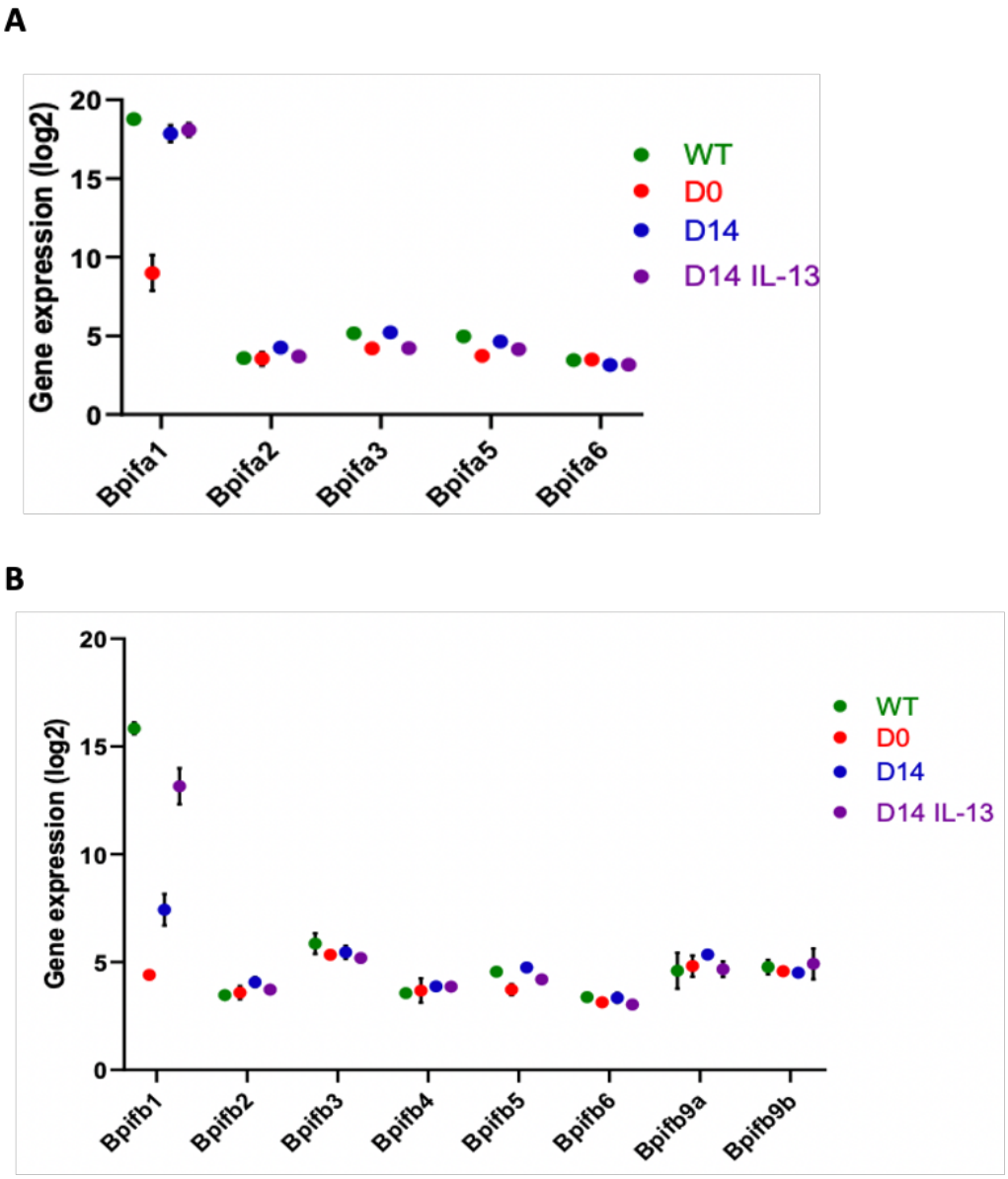


Figure 4.14: Expression of *Bpifa* (A) *and Bpifb* (B) family members during mTEC ALI differentiation. Comparision between original mTEC population (WT), undifferentiated (D0) start of ALI culture, differentiated cell population 14 days of ALI culture (D14) and IL-13 treatment throughout the differentation phase (D14 IL-13) displaying the mean of both replicates and associated standard deviation.

To verify that the mTEC isolation and culture conditions only expanded cells of epithelial origin, expression data points were extracted for genes of mesenchymal cell origin using common fibroblast markers vimentin (*vim*), CD34 antigen (*Cd34*) and lumican (*lum*) (Muhl

et al., 2020). Vimentins are intermediate filaments expressed in various non-epithelial cells, especially mesenchymal cells, and found attached to the nucleus, mitochondria or endoplasmic reticulum (Satelli and Li, 2011). The transmembrane phosphoglycoprotein CD34 has been initially described in hematopoietic stem and progenitor cells but nowadays has been established as a marker of several mesenchymal stem cells (MSC) alongside a multitude of other nonhematopoietic cell types such as vascular endothelial progenitors or embryonic fibroblasts (Fina *et al.*, 1990; Brown *et al.*, 1991; Sidney *et al.*, 2014). Lumican is a collagen binding proteoglycan initially discovered in the cornea, but it is found ubiquitously distributed in most mesenchymal tissues (Kao *et al.*, 2006; Krishnan *et al.*, 2012). It is required for maintenance of normal fibril architecture through facilitation of collagen fibril organization, corneal transparency, epithelial cell migration and tissue repair (Kao *et al.*, 2006).

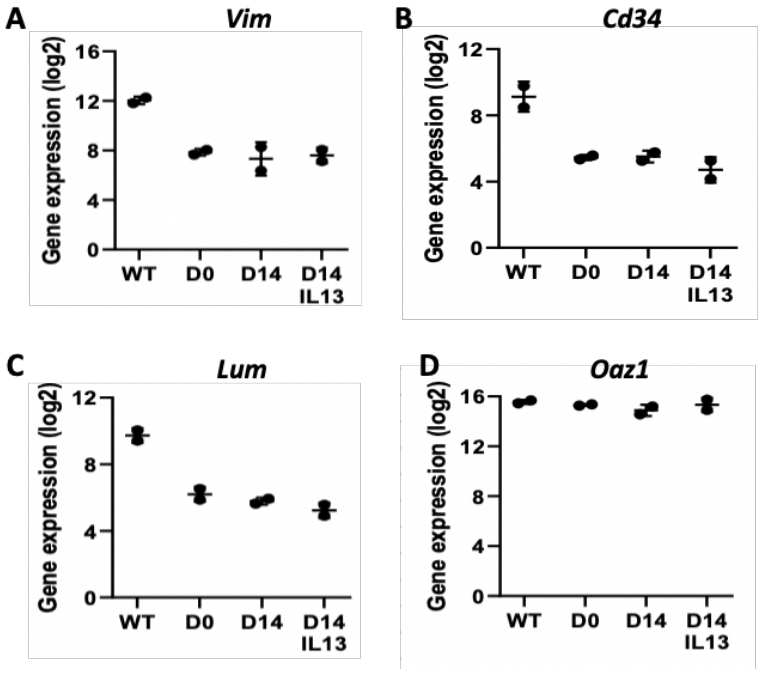


Figure 4.15: Fibroblast marker gene expression between mTEC extraction and ALI differentiation. Comparison of gene expression levels of common fibroblast markers such vimentin (*vim*) **(A)**, CD34 antigen (*Cd34*) **(B)** and lumican (*lum*) **(C)** from original wt mTEC, undifferentiated (D0), ALI differentiated (D14) and ALI D14 IL-13 treated cell populations. *Oaz1* **(D)** served as internal expression control.

Comparison of detected expression of fibroblast markers between wt, undifferentiated and differentiated conditions showed that in all cases highest levels were detected in the original tissue (wt) (Figure **4.15**A-C). At D0 and at the differentiated state, these levels were recorded at a decrease amount, remained stable and close to significant expression with a log2 value

around 5 for *Cd34* (Figure **4.15**B) and *Lum* (Figure **4.15**C). This indicates that the expanding population on the transwells are likely not to contain cells of mesenchymal origin.

The focus of the genome wide expression experiments was to understand the biological processes that occurred during epithelial proliferation until reaching of confluency in the submerged expansion phase, and what processes were active during mTEC differentiation in ALI conditions (Figure **4.1**). Therefore, my analysis was mainly focused on identification of upregulated genes during the mTEC expansion and differentiation phases. The gene expression profiles of extracted, uncultured cells, expanded and undifferentiated (D0), and 14 days ALI differentiated mTEC (D14) were compared (Figure **4.1**). In addition, I also investigated changes mediated by IL-13 treatment during ALI differentiation (D14 IL13) (Figure **4.6**).

4.3.2.2 Differentially expressed genes

By applying the criteria of an adjusted p-value of 0.05 and a log2 fold change (logFC) cut off of 1, 44 genes were show to be upregulated significantly between extracted mTEC (wt) and D0 cultures, compared to 106 genes induced during differentiation from D0 to D14 ALI (

Table 4.1). In contrast, 135 genes were downregulated during wt to D0 mTEC expansion, and 115 genes were expressed at lower levels during D0 to D14 ALI differentiation (

Table 4.1).

Consistent with the multidimensional scaling plots (Figure **4.10**A), only a small number of differentially expressed genes were expected between differentiated mTEC and the correspondent IL-13 treated sample group (Figure **4.6**). In fact, using the selected criteria of a logFC greater than 1 and adjusted p-value of 0.05, no genes were found downregulated and only a single gene was upregulated significantly amongst 22205 unchanged genes (

Table 4.1).

Table 4.1: *Differential expression counts upon mTEC expansion, differentiation or IL-13 treatment*

	Up	Flat	Down
WT-D0	44	22027	135
D0-D14	106	21985	115
D14-D14IL13	1	22205	0

The differential expression counts in

Table 4.1 correlated well with the principal component analysis plot (Figure **4.10**A) showing the wt sample set (blue) clustered the furthest away from the other three sample sets indicating the greatest variability, in contrast to D14 (red) and D14 IL13 (green) sample groups, which tightly clustered together demonstrating a much reduced variability.

The 25 most highly expressed genes between wt and D0 during mTEC expansion/proliferation (Figure **4.16**) and ALI differentiation between D0 and D14 (Figure **4.17**) were visualised as heatmap across all sample sets alongside respective MD and volcano plots.

Investigating DEGs for both stages, the proliferative/expansion and the differentiation phase, a subset of genes were found to be inversely correlated, meaning that they were found upregulated in wt original mTECs and subsequent proliferation and downregulated during ALI differentiation between D0 and D14 and vice versa (Figure **4.18**). Categorising these

genes by wide cast biological functions or protein properties using the DAVID Gene Functional Classification Tool (http://david.abcc.ncifcrf.gov) (Huang *et al.*, 2007), proliferation upregulated, and differentiation downregulated gene functions involve extracellular space localised or secreted molecules, hydrolases or glycoproteins (Figure **4.20**D). Genes which were downregulated during mTEC proliferation and upregulated during ALI differentiation are involved in cytoskeletal structure or cytoplasm residing proteins (Figure **4.20**E).

In addition to genes found inversely correlated, a larger portion of unique modulated genes were identified and have been highlighted in a more detailed volcano plot (Figure **4.19**). With regards to upregulated genes during mTEC expansion, the majority of these encode for cell membrane and extracellular region modulators. However, a much larger number of genes were downregulated, these function in immunity or belong to cell surface proteins and receptors, the hemoglobin complex, chemokine signalling pathway or lectins (Figure **4.20**A). During the mTEC expansion stage it was expected that the original isolated cells would be a mixed cell population consisting of epithelial cells and contaminating non-epithelial cell types such as macrophages, neutrophils, fibroblasts and smooth muscle cells and these would not be amplified during cluture, leading to a largely basal cell population at D0. Therefore, it was not surprising to observe a large proportion of genes which were down regulated and not associated with mucociliary function such as genes for the haemoglobin complex.

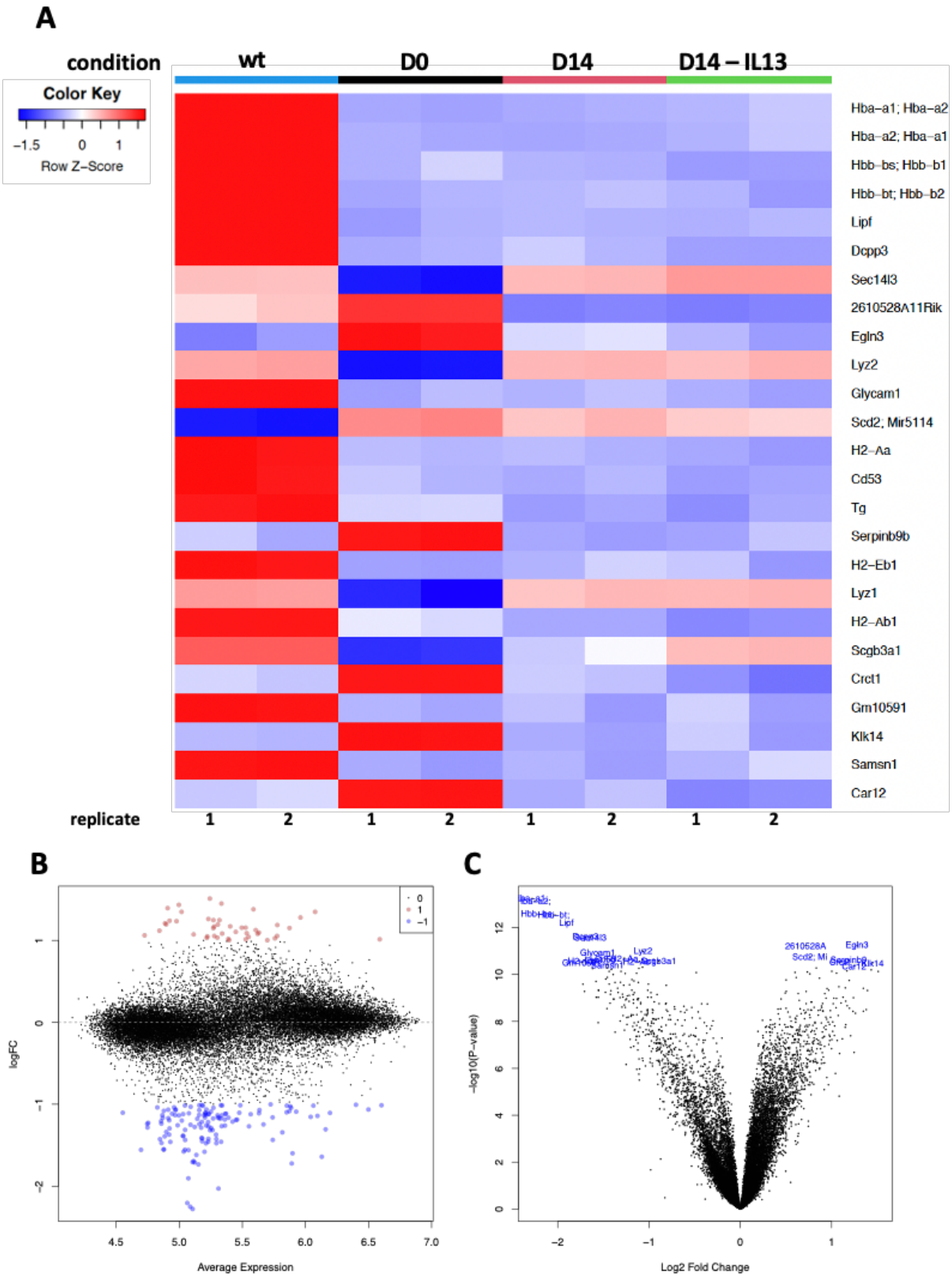


Figure 4.16: Differentially expressed genes between uncultured (wt), extracted mTEC and D0. (A) Heatmap of the top 25 genes. **(B)** Mean-Difference and **(C)** Volcano plot of differentially expressed genes with an adjusted p-value threshold of 0.05 and log2 fold change of 1.

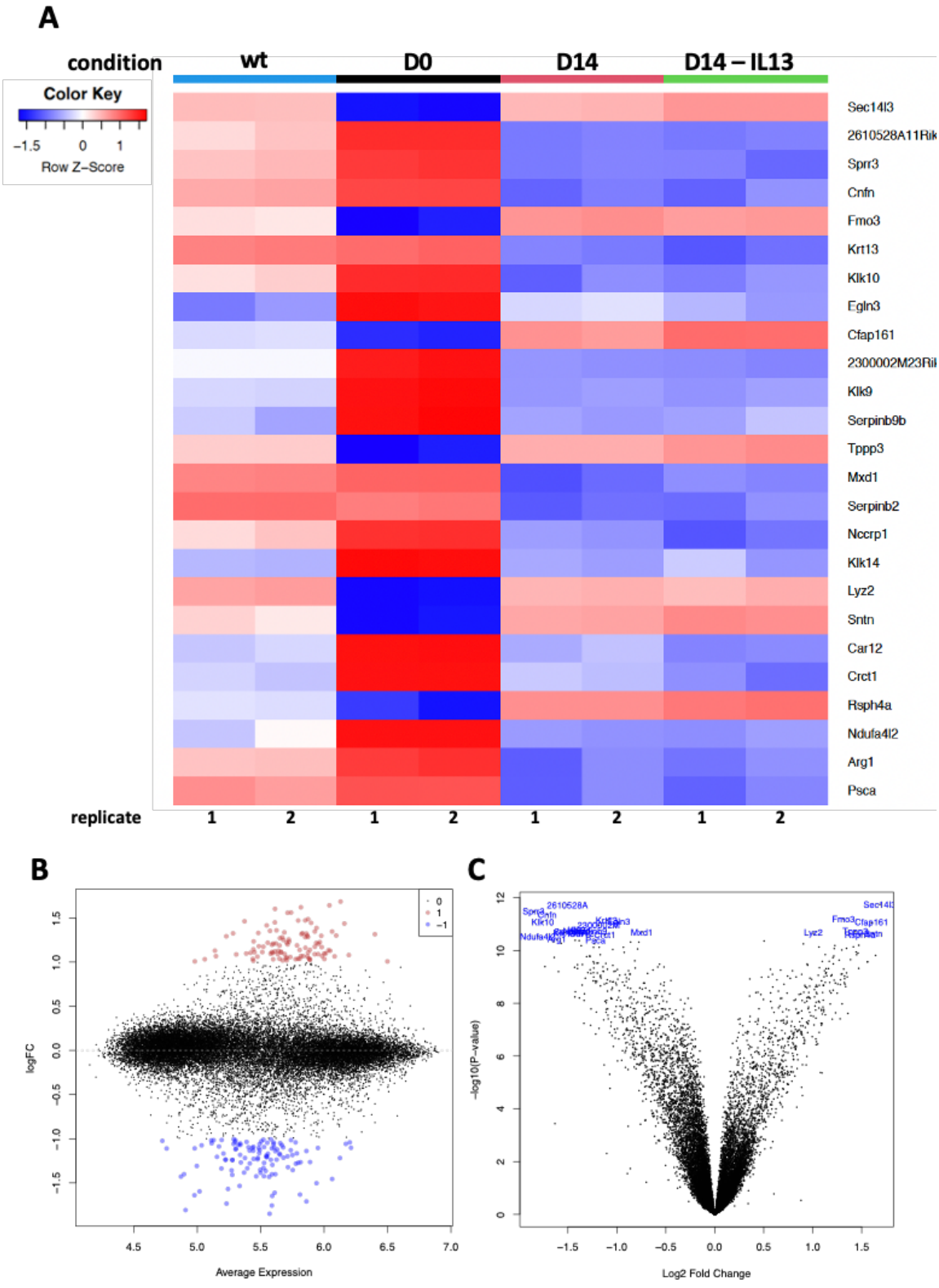
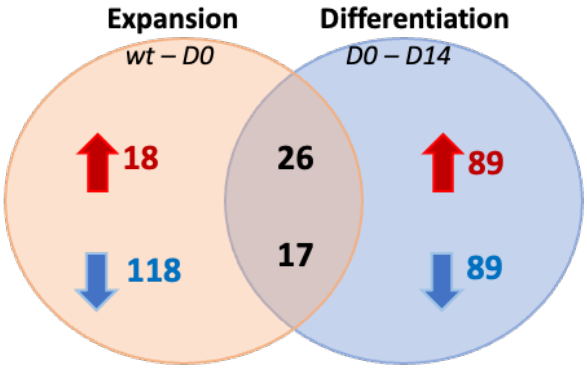


Figure 4.17: Differentially expressed genes between undifferentiated D0 and differentiated D14 ALI mTEC. (A) Heatmap of the top 25 genes. **(B)** Mean-Difference and **(C)** Volcano plot of differentially expressed genes with an adjusted p-value threshold of 0.05 and log2 fold change of 1.

A



B

Expansion (wt-D0)	
UP	DOWN
Kprp	Lrrc71
Klk14	Fmo3
Phgr1	Fmo6
Ndufa4l2	Rgs22
Olfm4	Rnf125
Gm2042	Fam183b
Klk8	Lyz2
Car12	Dynlrb2
Oog1; Gm2042	Spef2
Lce3f	Abi3bp
Ak4	5330417C22Rik
Serpinb9b	Cdhr3
Tmigd1	Odf3b
Klk9	Tppp3
Cda	Sntn
Crct1	Lyz1
U90926	Sec14l3
Gm6551	
Saxo1	
Bst1	
Serpinb7	
Upk3b	
Krt78	
Vsig10l	
Wfdc5	
Defb14	

Differentiation (D0-D14)	
UP	DOWN
Sec14l3	Saxo1
Cdhr3	Bst1
Sntn	Gm6551
Odf3b	U90926
Lyz1	Wfdc5
Tppp3	Tmigd1
Fmo6	Olfm4
Dynlrb2	Ak4
Fmo3	Krt78
5330417C22Rik	Crct1
Spef2	Upk3b
Lrrc71	Serpinb7
Rgs22	Serpinb9b
Abi3bp	Phgr1
Rnf125	Defb14
Fam183b	Car12
Lyz2	Lce3f
	Klk9
	Cda
	Klk14
	Kprp
	Gm2042
	Oog1; Gm2042
	Klk8
	Ndufa4l2
	Vsig10l

Figure 4.18: Inverse correlated differentially expressed genes between mTEC expansion and differentiation. **A)** Venn diagram outlining unique differentially expressed genes during expansion (orange circle) or differentiation (blue circle). Overlapping inverse gene number is found in the overlapping area of both circles. **B)** List of differentially expressed genes with an adjusted p-value threshold of 0.05 and log2 fold change (logFC) of 1. Note: the gene lists are organised from highest logFC to lowest increase (up condition) or lowest too highest decrease (down condition).

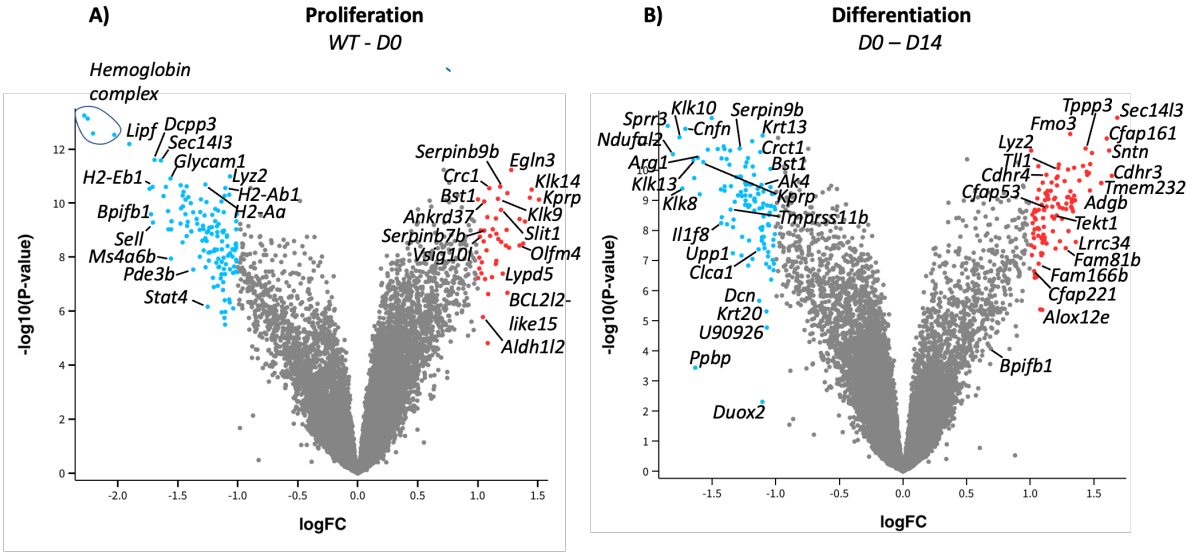
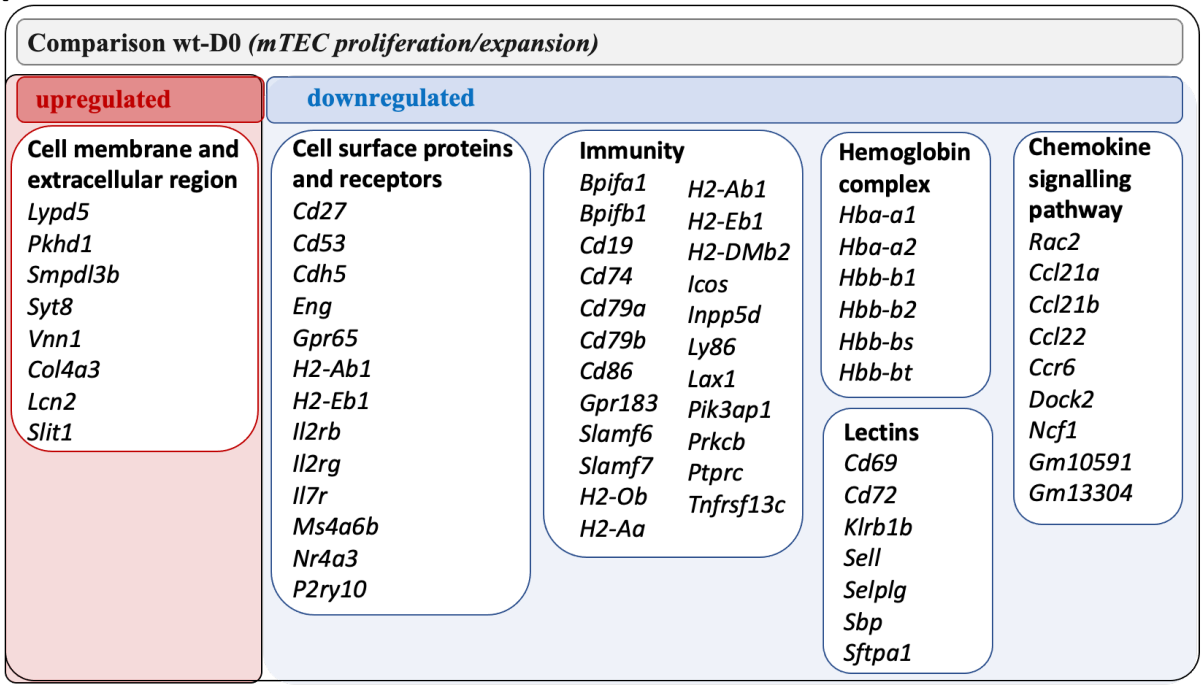


Figure 4.19: Annotated Volcano plot comparing proliferative versus differentiated phase. Selected differentially expressed genes have been highlighted with an adjusted p-value threshold of 0.05 and log2 fold change of 1.

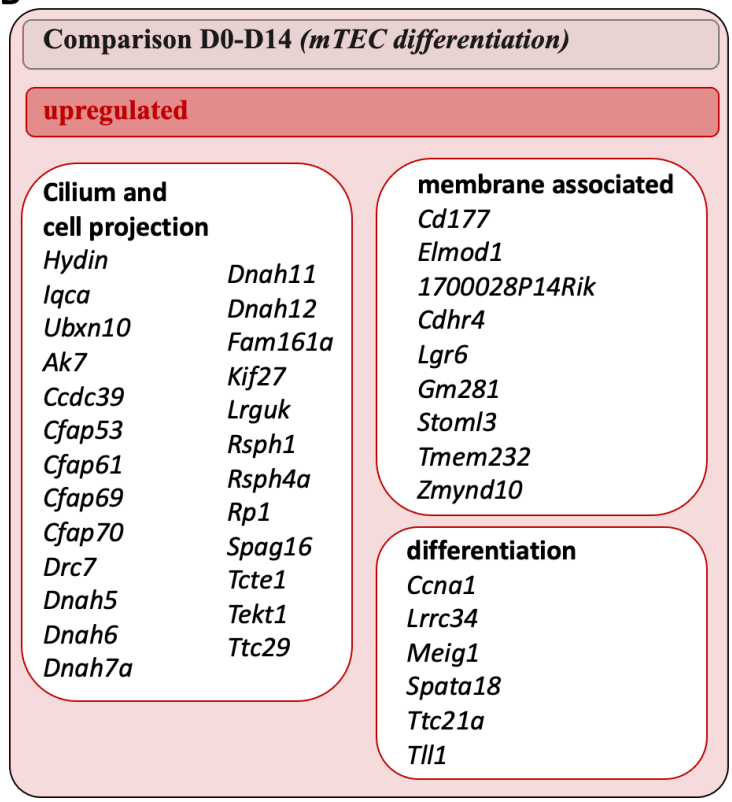
During ALI cultivation (D0-D14), cells undergo mucociliary differentiation, therefore upregulated genes were expected to belong to distinct differentiated cell types such as the ciliated, goblet and secretory cells. This is supported by the gene signatures observed in the analysis, where cilium and cell projection as well as differentiation genes are upregulated, alongside specific membrane associated proteins such as cadherin-related family members (Figure **4.20**B and C).

In contrast, genes encoding for extracellular space and secreted proteins, keratinization or keratinocyte differentiation or epidermis development were found to be downregulated during mTEC ALI differentiation. Functional categories also included peptidase inhibitor or hydrolase activity, membrane components or cytokine activity and negative regulation of cell proliferation (Figure **4.20**A).

A


Comparison wt-D0 (mTEC proliferation/expansion)
upregulated
downregulated
Cell membrane and extracellular region
Lypd5
Pkhd1
Smpdl3b
Syt8
Vnn1
Col4a3
Lcn2
Slit1
Cell surface proteins and receptors
Cd27
Cd53
Cdh5
Eng
Gpr65
H2-Ab1
H2-Eb1
Il2rb
Il2rg
Il7r
Ms4a6b
Nr4a3
P2ry10
Immunity
Bpifa1
Bpifb1
Cd19
Cd74
Cd79a
Cd79b
Cd86
Gpr183
Slamf6
Slamf7
H2-Ob
H2-Aa
H2-Ab1
H2-Eb1
H2-DMb2
Icos
Inpp5d
Ly86
Lax1
Pik3ap1
Prkcb
Ptprc
Tnfrsf13c
Hemoglobin complex
Hba-a1
Hba-a2
Hbb-b1
Hbb-b2
Hbb-bs
Hbb-bt
Lectins
Cd69
Cd72
Klrb1b
Sell
Selplg
Sbp
Sftpa1
Chemokine signalling pathway
Rac2
Ccl21a
Ccl21b
Ccl22
Ccr6
Dock2
Ncf1
Gm10591
Gm13304

B


Comparison D0-D14 (mTEC differentiation)
upregulated
Cilium and cell projection
Hydin
Iqca
Ubxn10
Ak7
Ccdc39
Cfap53
Cfap61
Cfap69
Cfap70
Drc7
Dnah5
Dnah6
Dnah7a
Dnah11
Dnah12
Fam161a
Kif27
Lrguk
Rsph1
Rsph4a
Rp1
Spag16
Tcte1
Tekt1
Ttc29
membrane associated
Cd177
Elmod1
1700028P14Rik
Cdhr4
Lgr6
Gm281
Stoml3
Tmem232
Zmynd10
differentiation
Ccna1
Lrrc34
Meig1
Spata18
Ttc21a
Tll1

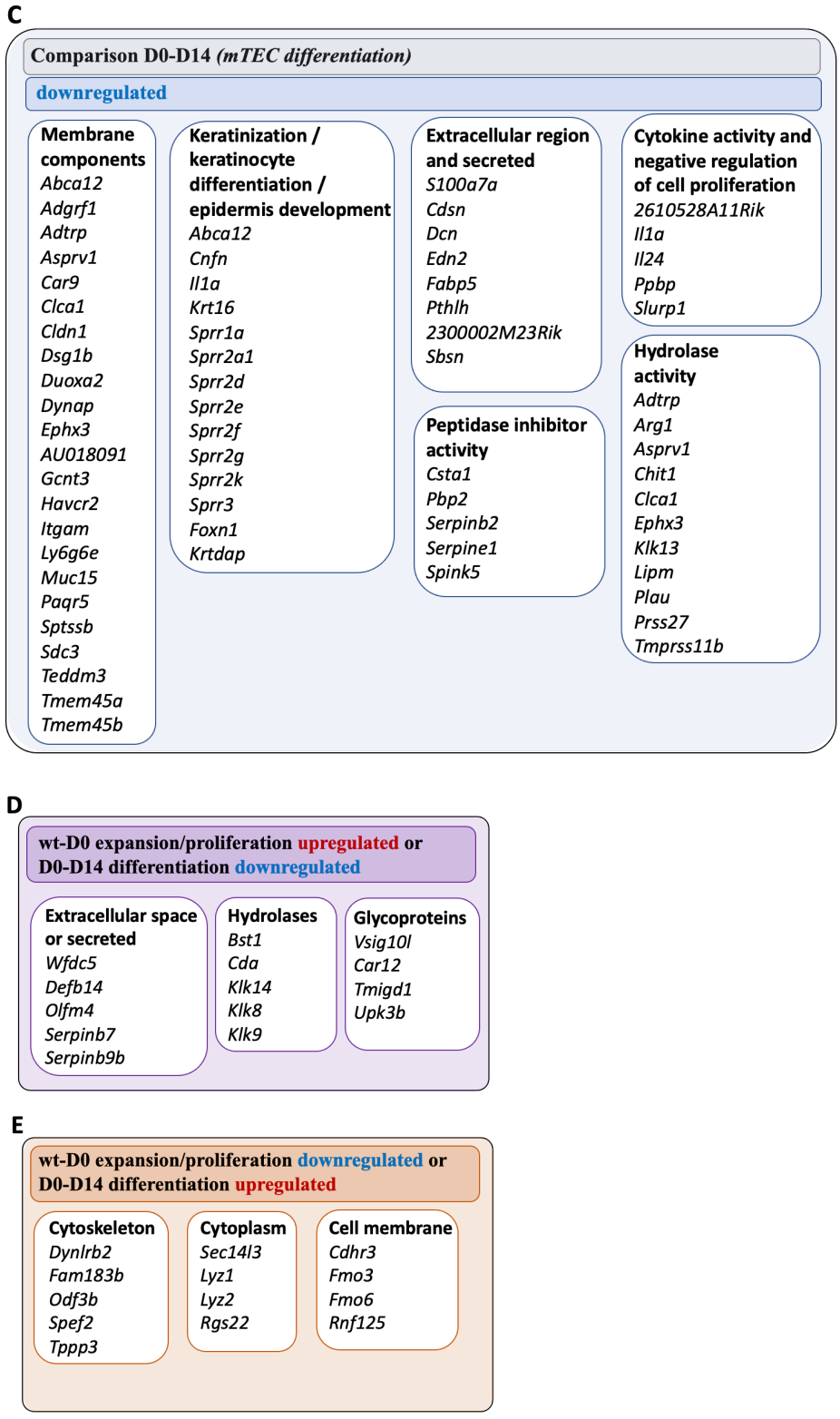


Figure 4.20: DEGs across expansion/proliferative and differentiation conditions. A) Comparison of mTEC expansion (wt-D0), **B)** mTEC differentiation (D0-D14), upregulated genes **C)** differentiation downregulated genes, **D, E)** show gene signatures inversely correlated between expansion and differentiation conditions. Genes were initially generally categorised using the DAVID Gene Functional Classification Tool (http://david.abcc.ncifcrf.gov) before manually refining them into categories listed as highlighted, as outlined in Table 9.1.

As well as looking across the differentiation process the other condition investigated was IL-13 exposure during mTEC differentiation. This treatment is reported to induce goblet cell (Atherton *et al.*, 2003; Kanoh *et al.*, 2011; Waddell *et al.*, 2019) differentiation and such culture could be used to test the role of goblet cells in infection. It was expected that under these culture conditions goblet cell markers such as *Bpifb1* would increase in expression and potentially ciliated cell marker genes would be decreased. As already mentioned earlier, D14 and D14 IL-13 treated samples showed very little variation during the microarray data validation (Figure **4.10**A), and consistently a much smaller numbers of genes were found to be modulated as shown in the heatmap and associated MD and volcano plot (Figure **4.21**) and the comparable volcano plot (Figure **4.22**). Investigating the volcano plot in more detail, shows there are a higher number of up-regulated genes, compared to downregulated ones, however the fold changes are small and lie almost exclusively below logFC of 0.6. The only gene identified to be significantly modulated at conditions of logFC greater 1, or logFC of -1, at an adjusted p-value of 0.05 is *Olfactomedin 4 (Olfm4)* reported to facilitate cell adhesion. Slightly below the criteria cut off, two other upregulated genes were identified, *Bpifb1*(logFC 0.79) and the tissue inhibitor of metalloproteinase 1 (*Timp1*), a glycoprotein inhibiting matrix metalloproteases, promoting cell proliferation and encompassing an anti-apoptotic function (logFC 0.87) (*Table **4.2***).

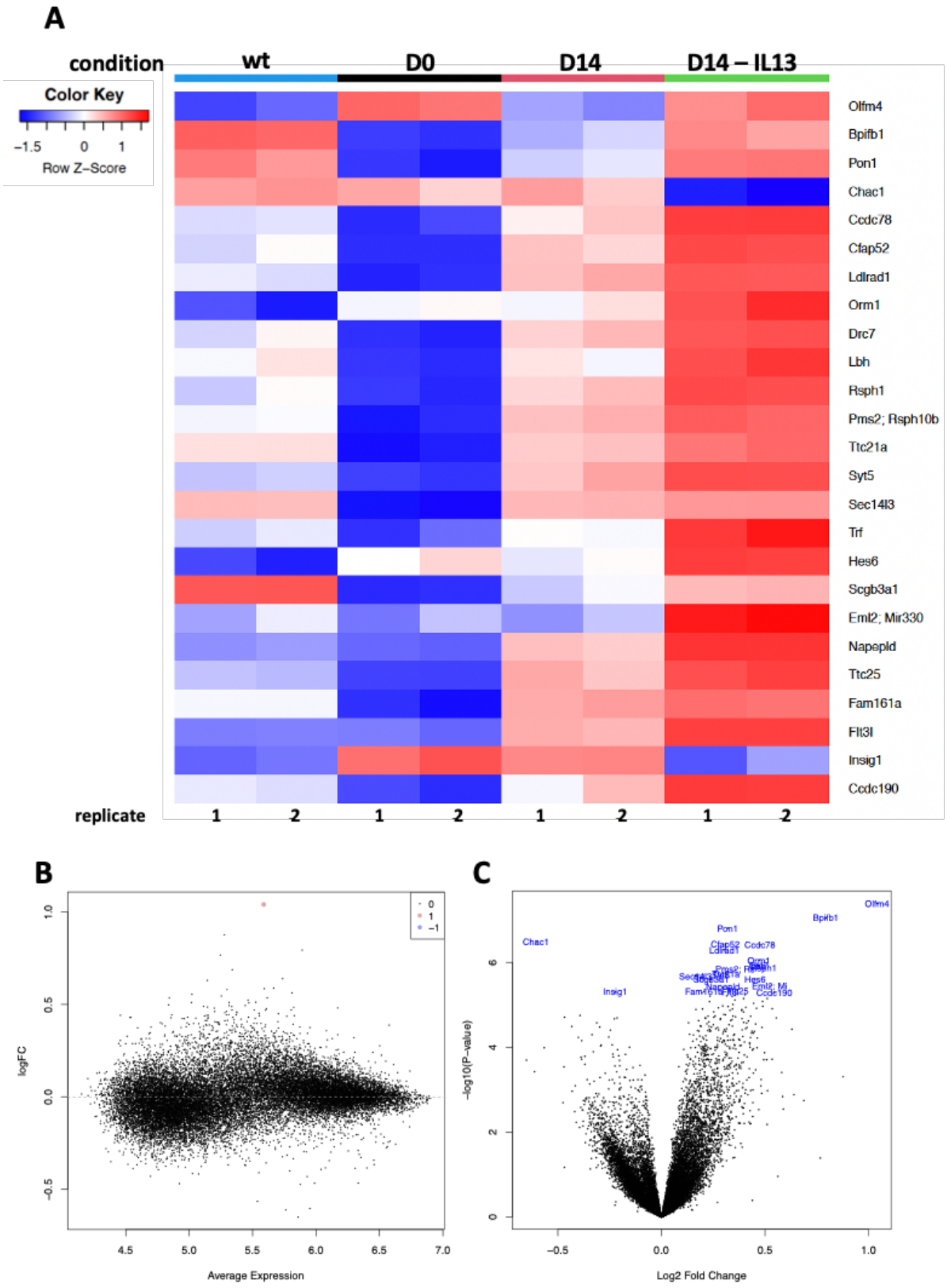


Figure 4.21: Differentially expressed genes between differentiated D14 ALI mTEC and IL-13 treated D14 ALI mTEC. (A) Heatmap of the top 25 genes. **(B)** Mean-Difference and **(C)** Volcano plot of differentially expressed genes with an adjusted p-value threshold of 0.05 and log2 fold change of 1.

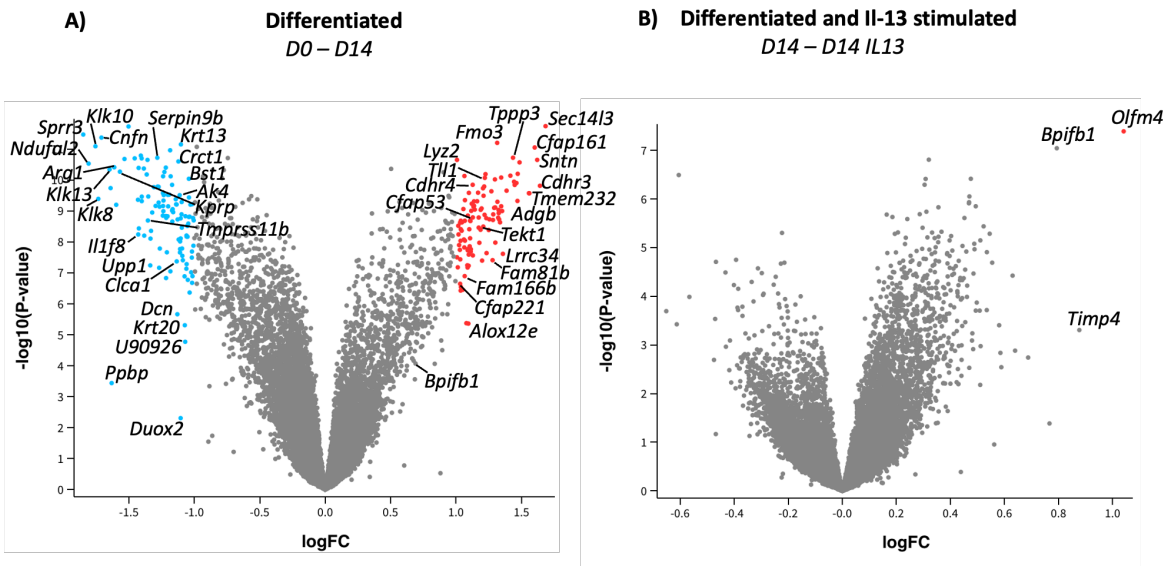


Figure 4.22: DEGs during mTEC IL13 treatment. mTEC were ALI differentiated with or without IL-13 (10ng/ml) for 14 days and compared with regard to gene modulation.

Table 4.2: DEGs during differentiation in presence of IL-13. *Gene entrez IDs are listed in front of each gene*

D0-D14 differentiation and IL-13 treated

upregulated

380924 *olfactomedin 4 (Olfm4)*
228801 *BPI fold containing family B, member 1 (Bpifb1)* – **logFC 0.794**
21857 *tissue inhibitor of metalloproteinase 1 (Timp1)* – **logFC 0.877**

4.3.3 Influenza A infection of mTEC populations

Previous experiments carried out in the lab had shown that mTECs provide a good model to study IAV infection. A published study from Akram and colleagues (Akram *et al.*, 2018) provided evidence that non-BPIFA1 positive expressing cell populations could be infected with IAV, and additional data from the lab suggested that IAV did not exclusively target ciliated cells (unpublished). Therefore, a study was designed investigate infection susceptibility in undifferentiated and differentiated mTEC populations with the overall aim to shed more light into molecules such receptors or proteases potentially required for viral uptake.

4.3.3.1 Comparison of X31 infection of undifferentiated and differentiated mTEC

To investigate the IAV infection capacity of undifferentiated versus ALI differentiated mTEC cells, initial experimental studies investigated viral infection by staining IAV nuclear protein (Figure **4.23**) and analysis of the antiviral cytokine response of *Interferon Lambda 2* (*Ifnl2)* at increasing MOIs of X31 IAV virus infection of undifferentiated cells (Figure **4.24**).

The X31 strain is an H3N2 virus, which has been demonstrated to only cause mild to moderate disease in murine models, compared to the PR8 strain, which is a mouse adapted H1N1 influenza virus, known to cause severe infection (Askovich *et al.*, 2013; Rutigliano *et al.*, 2014). In order to address the initial question if undifferentiated mTEC could be infected with IAV, mTEC populations were extracted and expanded as described and infection of undifferentiated (D0) cells were directly compared to ALI D14 differentiated cultures. These mTEC cell populations were imaged 24 hours post infection for expression of the viral nuclear protein (NP), counterstained for ciliated cells and overall cell nuclei with (DAPI) (Figure **4.23**). Surprisingly, viral uptake seemed to occur at a similar level between undifferentiated (Figure **4.23**A) and differentiated (Figure **4.23**B) mTEC populations, based on qualitive levels across representative images.

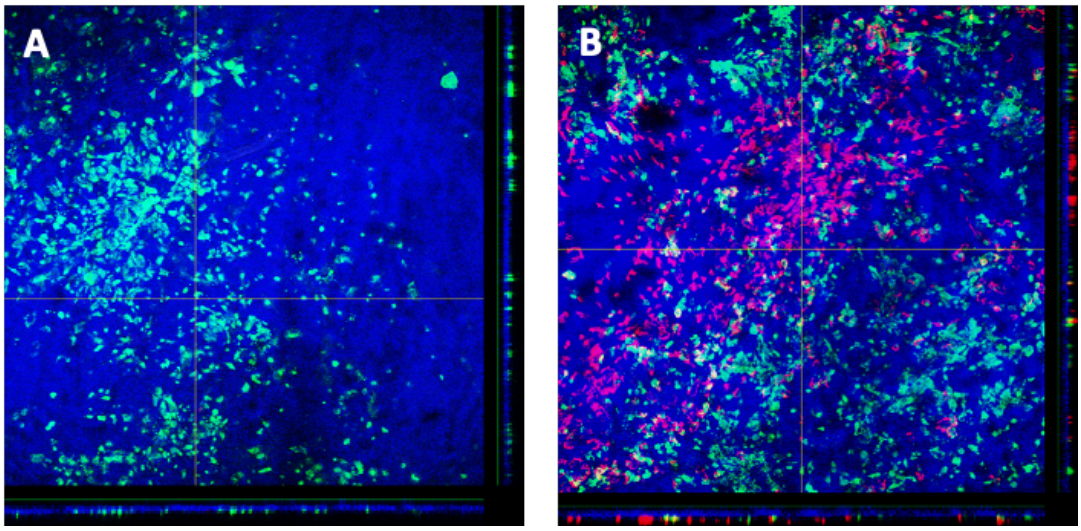


Figure 4.23: Confocal imaging of X31 infection of D0 and D14 mTEC. D0 (undifferentiated) **(A)** or D14 (differentiated) **(B)** cell populations were exposed to X31 virus with an MOI of 1 and imaged 24 hours later. X31 positive cells stained for IAV NP appear green, ciliated cells were stained with α-tubulin (red) and cell nuclei were counterstained with DAPI (blue). Image from Priyanka Anujan.

As expected, ciliated cells were only detected in differentiated mTEC (B) and the overwhelming majority of X31 infected cells were α-tubulin negative, providing evidence that this virus strain is able to infect non-ciliated cells.

To further analyse the observed X31 infection of undifferentiated mTEC, I used a titration approach using varying MOIs and examining expression of antiviral cytokine response of *Ifnl2* by endpoint PCR (Figure **4.24**).

As expected, the highest viral load at MOI 1 led to the highest *Ifnl2* amplification, further providing evidence that undifferentiated mTEC can be infected by IAV in a dose dependent manner, evoking expected biological responses (Figure **4.24**). Therefore, it appears that undifferentiated mTEC cultures could potentially provide a useful tool for assessment of recombinant BPIFB1 protective function and/or susceptibility modulation following X31 IAV infection.

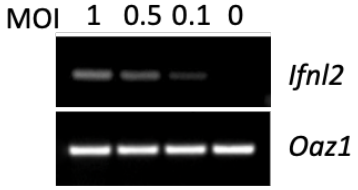


Figure 4.24: Endpoint PCR validation of infection in undifferentiated mTECs. mTEC cells were isolated and cultured as outlined in the text and mRNA extracted from undifferentiated cultures (D0) after 24h X31 virus challenge at indicated MOI ratios of 1, 0.5, 0.1 or from mock infected cells (0). Amplification of *Ifnl2* (200bp) was carried out by amplifying equal amounts of cDNA by endpoint PCR. *Oaz1* served as an internal positive control.

The discovery that undifferentiated mTECs seem to be susceptible of IAV infection at similar levels than the differentiated population, led to the design of a genomic experiment with the aim of investigating modulated gene signatures on a microarray format during the first 24 hours of X31 infection, comparing both differentiation stages (Figure **4.25**).

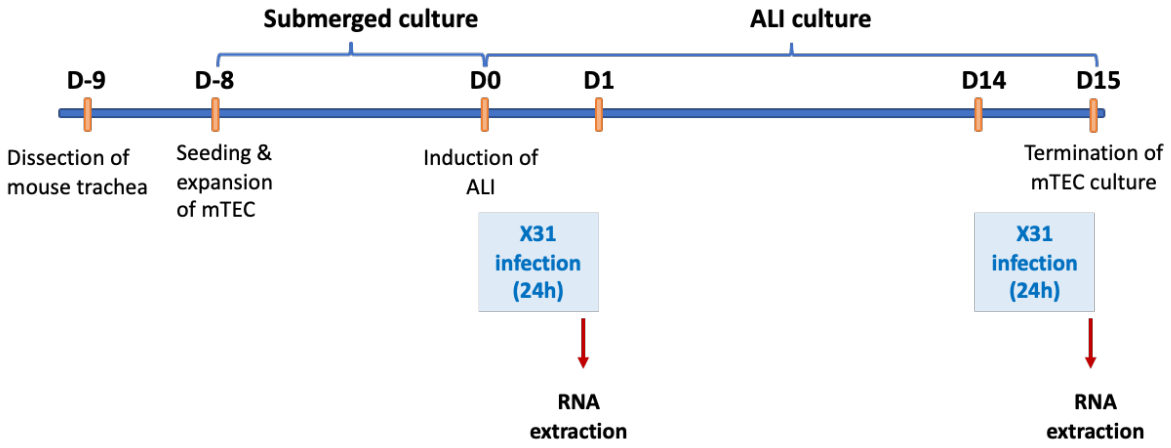


Figure 4.25: Schematic of experimental layout, timeline and sample harvest. mTEC isolated from resected trachea (D-9) were cultured in transwell inserts as submerged cultures (D-8) until reaching confluence at day 0 (D0, undifferentiated) before induction of ALI, or continued to culture for 14 days in ALI conditions (D14, differentiated). D0 or D14 cultures were either mock infected or infected with the IAV strain X31 at a MOI 1 for 24 hours, with subsequent media change and RNA harvest.

These experiments were undertaken by Dr Priyanka Anujan and Mr Fawaz Aljuhani. The microarray data sets, presented as counts, were generated at SITraN and analysed for differentially expressed genes (DEGs) with limma-voom (Law *et al.*, 2014; Ritchie *et al.*, 2015; Costa-Silva *et al.*, 2017). I undertook the analysis outlined below.

4.3.3.2 Quality control of X31 infection data sets

To investigate gene signatures, which are activated upon IAV infection, and to further explore potential differences between differentiated and undifferentiated mTEC population during X31 IAV exposure, mock infected and corresponding X31 infected sample groups were generated as outlined above and the provided data sets were quality controlled and normalised as outlined below.

The initial quality control step filtered for lowly expressed genes. Examining the distribution of the density plot before filtering (Figure **4.26**A) shows log-CPM values above 4.5 indicating absence of very low expressing genes at the cut of level of log-CPM -1. The analysis revealed, unsurprisingly, that 0 out of 22206 genes have been removed leaving the density plot shape unaltered (Figure **4.26**B), indicating that such a quality control step was already carried out on the raw data provided from the SITraN facility.

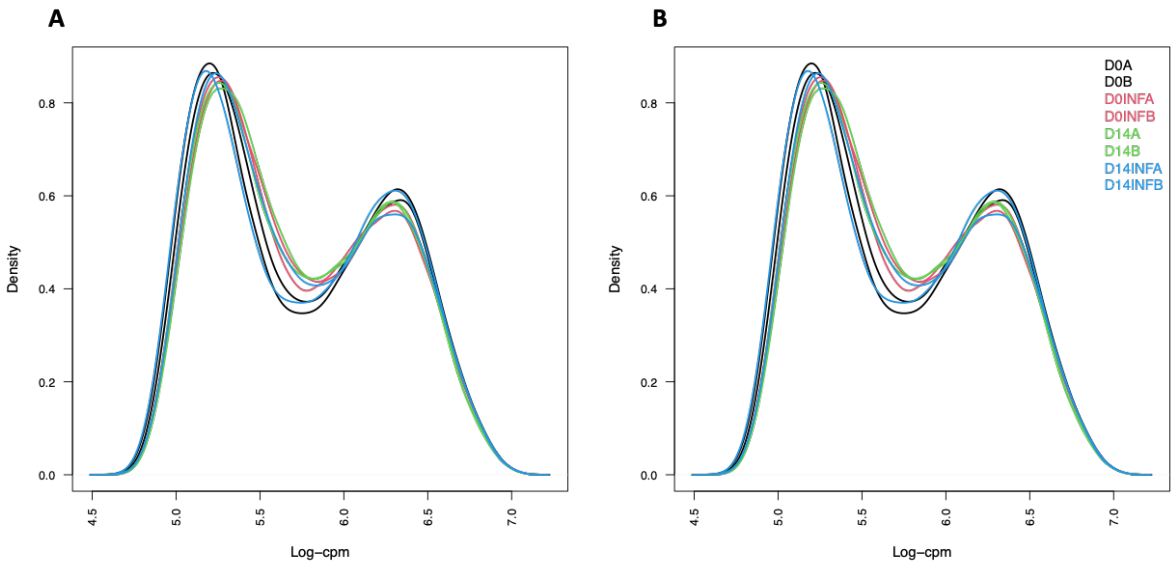


Figure 4.26: Density plot of infection sample groups. Comparison of the counts distributions before (raw) **(A)** and after filtering **(B)** are displayed across all data groups (D0 mock infection – black, D0 infection – red, D14 mock infection – green, D14 infection - blue). Genes without more than 0.5 CPM in at least two samples are determined insignificant and filtered out, which in this data set yielded 0 of 22206 (0%) genes were omitted due to low expression. Note: replicate samples are indicated as A or B.

Data set counts were TMM normalised in respect to differences between samples with the aim to eliminate composition biases between samples. Generated box plots exhibited that the infection sample groups are distributed close to the median horizontal line, and therefore they do not require further investigation or significant normalisation (Figure **4.27**).

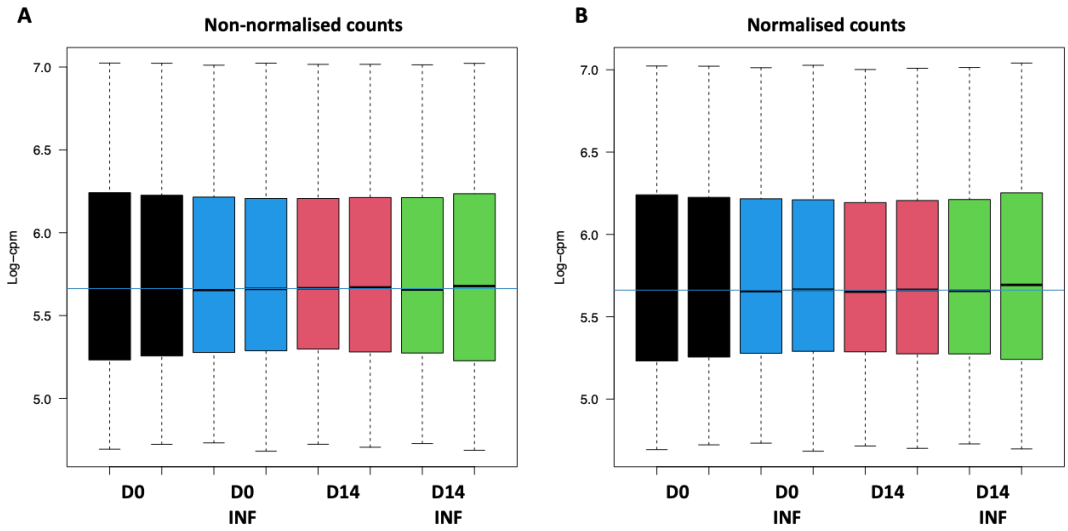


Figure 4.27: Box plots of sample normalisation of infection data sets. Non-normalised **(A)** vs TMM normalised counts **(B)**. D0 mock infection – black, D0 infection – blue, D14 mock infection – red, D14 infection - green). D0 is representative of the undifferentiated population, whereas D14 indicates the differentiated population. The blue line is indicating the data median.

The MDS plot of the infection data set is the best way to assess data variability between the data groups and across replicates. In this case, a very high reproducibility was observed in the D0 undifferentiated (black) and D14 ALI differentiated (green) samples, however the respective corresponding infection conditions displayed noticeable variation (Figure **4.28**).

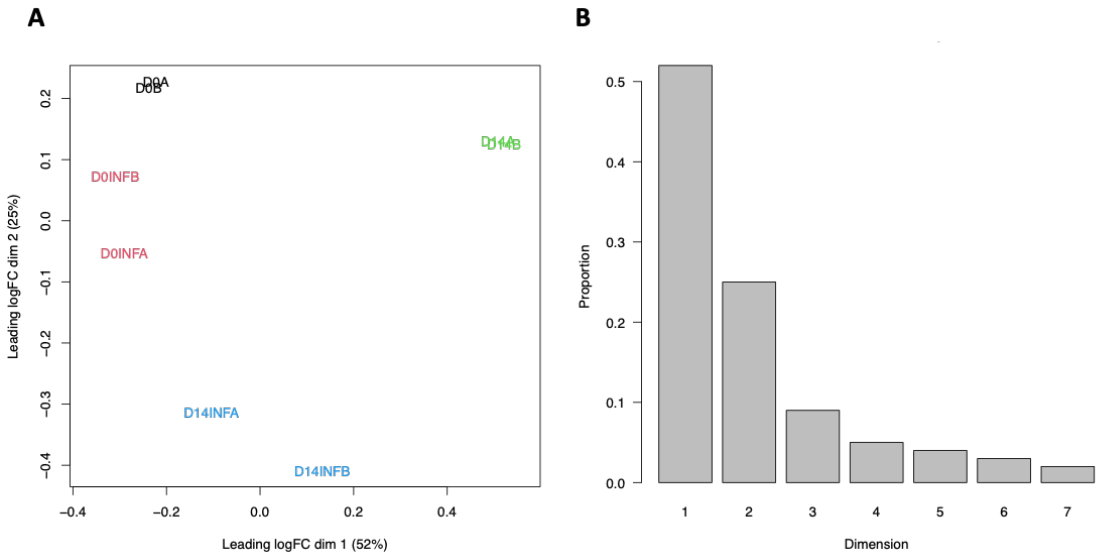


Figure 4.28: Quality control of variation within the infection data sets. (A) MDS plot dimension 1 versus 2 shows a high reproducibility between both replicates per sample set, with a distinct cluster of D0 undifferentiated (black) and D14 ALI (green) compared to undifferentiated D0 (red) or differentiated D14 (blue) X31 infected mTEC, which show a higher variability between sample sets, termed A or B. **(B)** The scree plot highlights variation derived from each dimension.

The voom (Figure **4.29**A) and SA variance (Figure **4.29**B) plots investigating levels of variation across the infection data groups showed a medium low-level biological variation with no visible outliers (Figure **4.29**B).

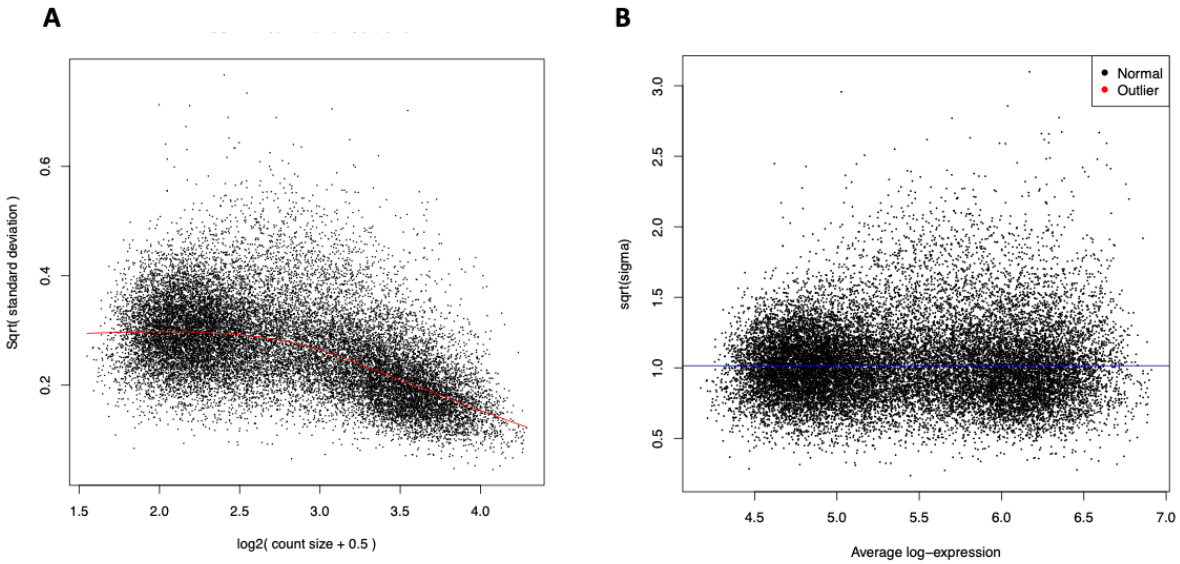


Figure 4.29: Quality control of mean-variance variation within the infection data groups. (A) Voom mean-variance dot blot versus **(B)** scatterplot of residual-variances vs average log-expression (SA plot). The average log2 residual standard deviation is marked by a horizontal blue line.

Before I conducted the full microarray data analysis and investigated differentially expressed genes, I extracted expression data for a number of genes expected show the modulation during viral infection (Figure **4.30**). For example *Ifnl2* (Figure **4.24**) showed an upregulated pattern, which is consistent in both differentiation stages, although it displayed a notable sample variation (Figure **4.30**A). Another infection marker, cytokine *C-X-C motif chemokine 10* (*Cxcl10*) (Liu *et al.*, 2011), showed a similar response pattern in X31 infection, with no distinction in intensity between undifferentiated (D0) or differentiated mTEC (D14) (Figure **4.30**B). *Oaz1* served as the internal control (Figure **4.30**C). Focussing, on *Bpifb1* (Figure **4.30**D) or *Bpifa1* (Figure **4.30**E) no significant response was observed. Expression of the cilia marker *Tekt1* did not change upon viral exposure (Figure **4.30**F). It should be noted that expression of some of these genes was different in the two samples.

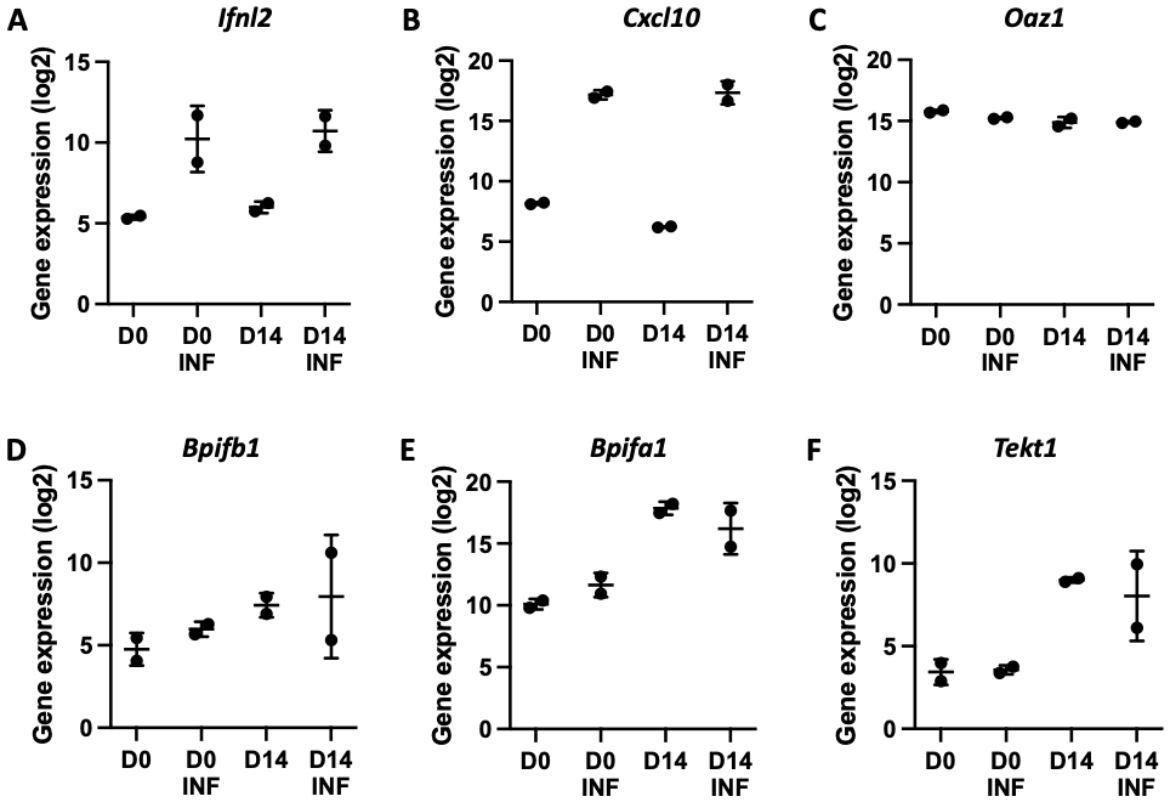


Figure 4.30: Gene expression levels of selected mTEC infection responsive and marker genes during X31 viral infection. Comparison of gene expression levels of cytokine *Ifnl2* **(A)** and chemokine *Cxcl10* **(B)** between undifferentiated (D0) and ALI differentiated (D14) cell populations, either mock treated or 24 h of X31 exposure (D0/D14 Inf). *Oaz1* **(C)** served as internal control. Analysis of secretory Bpif family members *Bpifb1* **(D)** or *Bpifa1* **(E)** or the cilia marker *Tekt1* **(F)**.

After completing data quality control, the limma-voom reported top 25 DEGs were investigated, to identify specific induced genes following X31 infection in both differentiation phases, meaning D0 versus D14, and to study which genes are unique or overlapping between both time points.

4.3.3.3 Differentially expressed genes upon X31 infection

Applying criteria of an adjusted p-value of 0.05 and a log2 fold change cut off of 1 resulted in 15 genes being significantly upregulated between undifferentiated mTEC D0 and X31 infected cells, compared to 116 genes being upregulated in differentiated D14 ALI mTEC 24h after X31 infection. Only a single gene was significantly down regulated under these conditions (Table ***4.3***) (Figure **4.25**).

Table 4.3: *Differential expression counts upon mTEC X31 infection*

	Up	Flat	Down
D0-D0INF	15	22191	0
D14-D14INF	116	22089	1

As mentioned earlier, the principal component analysis plot (Figure **4.10**A) showed high reproducibility across the D0 (black) and the D14 data set (green), which are distributed away from each other. However, the infection data sets for D0 (D0INF, red) or D14 (D14INF, blue) showed greater variation between the two replicates.

The 25 most significantly modulated genes following X31 challenge compared to mock treated undifferentiated (Figure **4.31**) or D14 ALI differentiated mTEC (Figure **4.32**) were visualised as heatmap across all sample sets alongside respective MD and volcano plots.

Analysing all the significantly upregulated genes of undifferentiated mTEC cells during infection with the corresponding differentiated counterpart revealed that these genes all overlap with the differentiated data group (Figure **4.33**). Moreover, no downregulated genes were identified in this condition, including my gene of interest *Bpifb1*. In order to have a more detailed look if the magnitude of overlapping upregulated genes found in both differentiation conditions was comparable, or were perhaps found amplified if the cells are differentiated, log2 expression values for the 15 genes were plotted to assess fold amplification (Figure **4.34**). It is apparent that these genes showed very similar basal expression levels, when compared to both differentiation states. In addition, 24h viral X31 challenge did not alter the magnitude of upregulated genes significantly compared between D0 and D14. Furthermore, the earlier observed replicate variation in the MDS plot (Figure **4.28**A) is observed in most of the plotted replicates (Figure **4.34**).

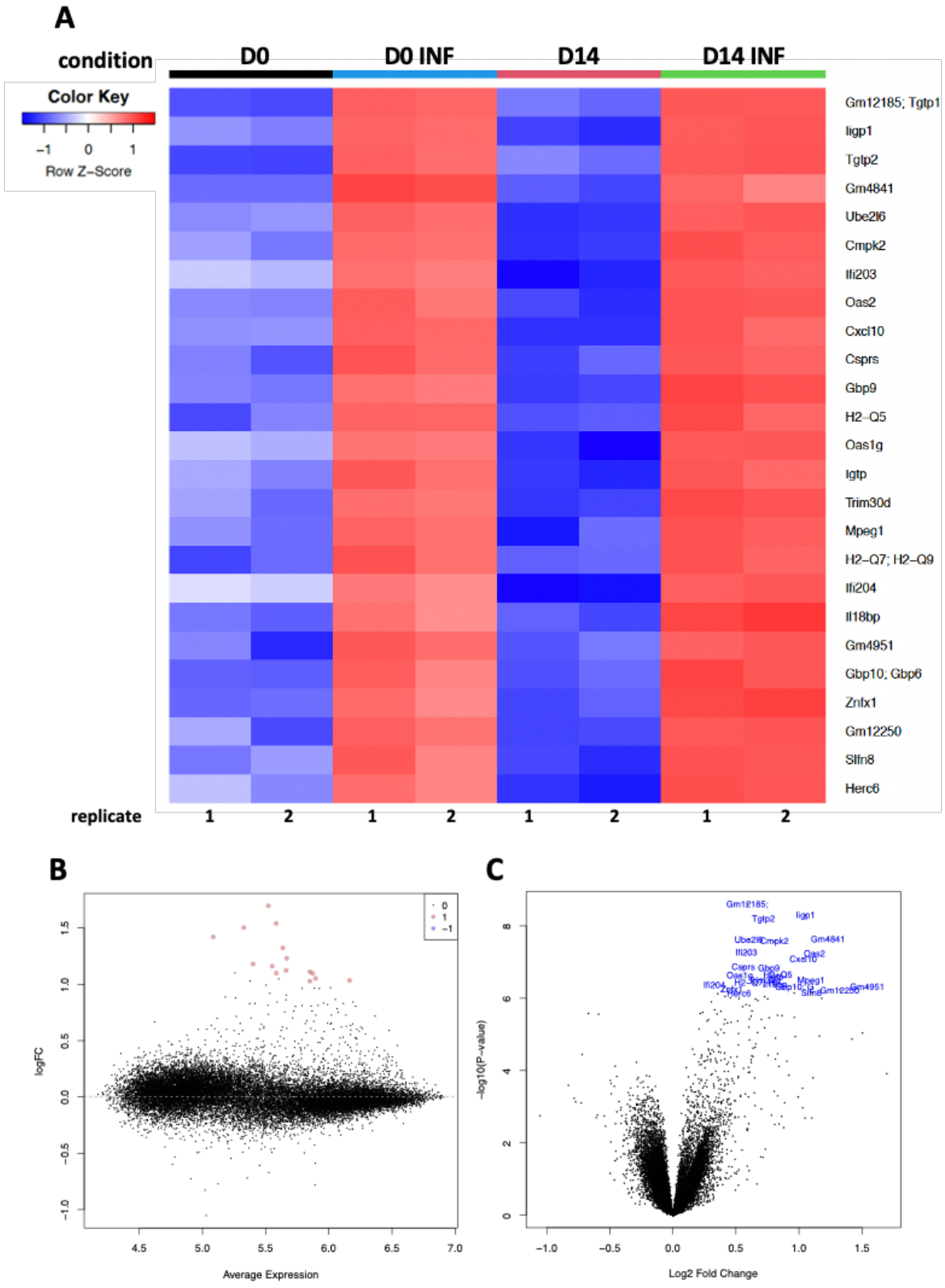


Figure 4.31: Differentially expressed genes between undifferentiated D0 ALI mock and X31 IAV exposed mTEC. Virus conditioned cell populations were exposed to virus for 1h and RNA was extracted 24h post infection. **(A)** Heatmap of the top 25 modulated genes. **(B)** Mean-Difference and **(C)** Volcano plot of differentially expressed genes with an adjusted p-value threshold of 0.05 and log2 fold change of 1.

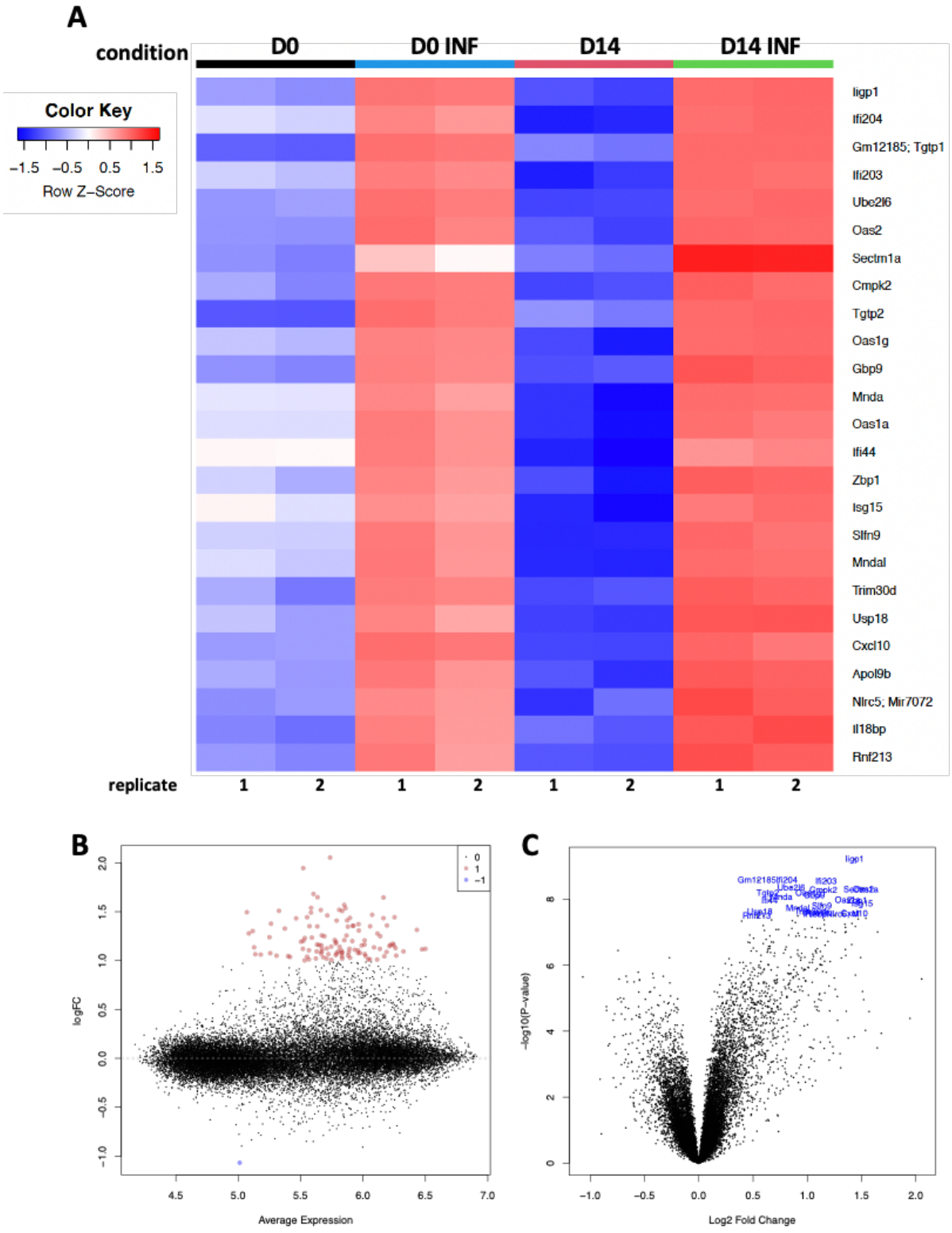


Figure 4.32: Differentially expressed genes between differentiated D14 ALI mock and X31 IAV exposed mTEC. Virus conditioned cell populations were exposed to virus for 1h and RNA was extracted 24h post infection. **(A)** Heatmap of the top 25 modulated genes. **(B)** Mean-Difference and **(C)** Volcano plot of differentially expressed genes with an adjusted p-value threshold of 0.05 and log2 fold change of 1.

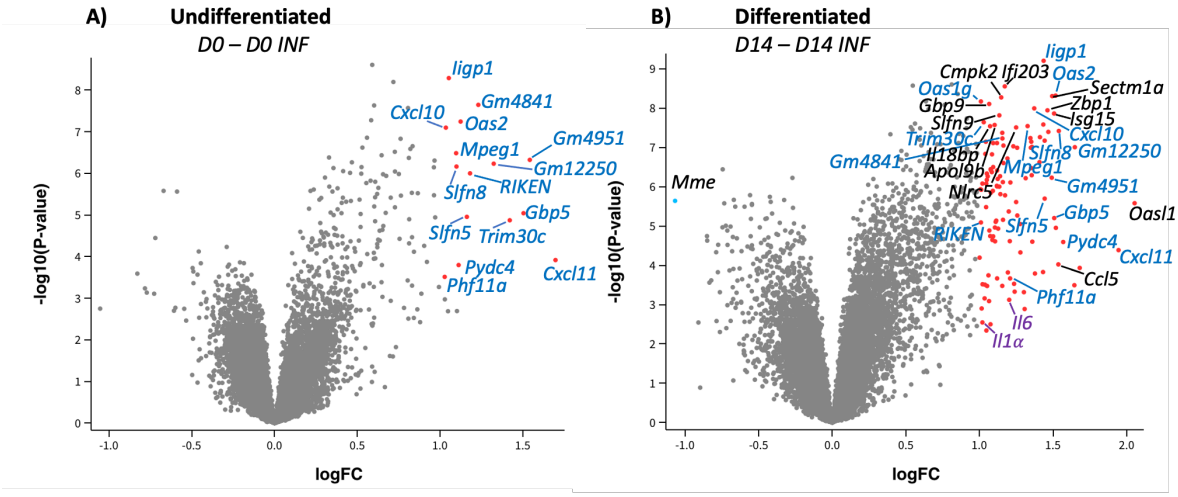


Figure 4.33: Volcano plots comparing modulated genes upon X31 infection. Undifferentiated D0 **(A)** or ALI D14 differentiated **(B)** mTEC were either mock or X31 IAV infected for 24h at MOI 0.1 and cells subsequently harvested for RNA extraction. Differentially expressed genes between the respective mock and virus challenged condition are plotted, applying an adjusted p-value threshold of 0.05 and log2 fold change of 1. Overlapping upregulated genes between both differentiation stages are annotated and highlighted in blue colour. Unique significantly upregulated genes are labelled in black. Interleukin related genes highlighted in purple indicate a response, which is absent in the undifferentiated condition.

The 15 genes found upregulated upon IAV challenge post 24h in the undifferentiated condition, which recapitulated in the set of the differentiated conditions, indicate that all cell types are able to be infected. Moreover, they induce a very similar gene set such as the ISGs.

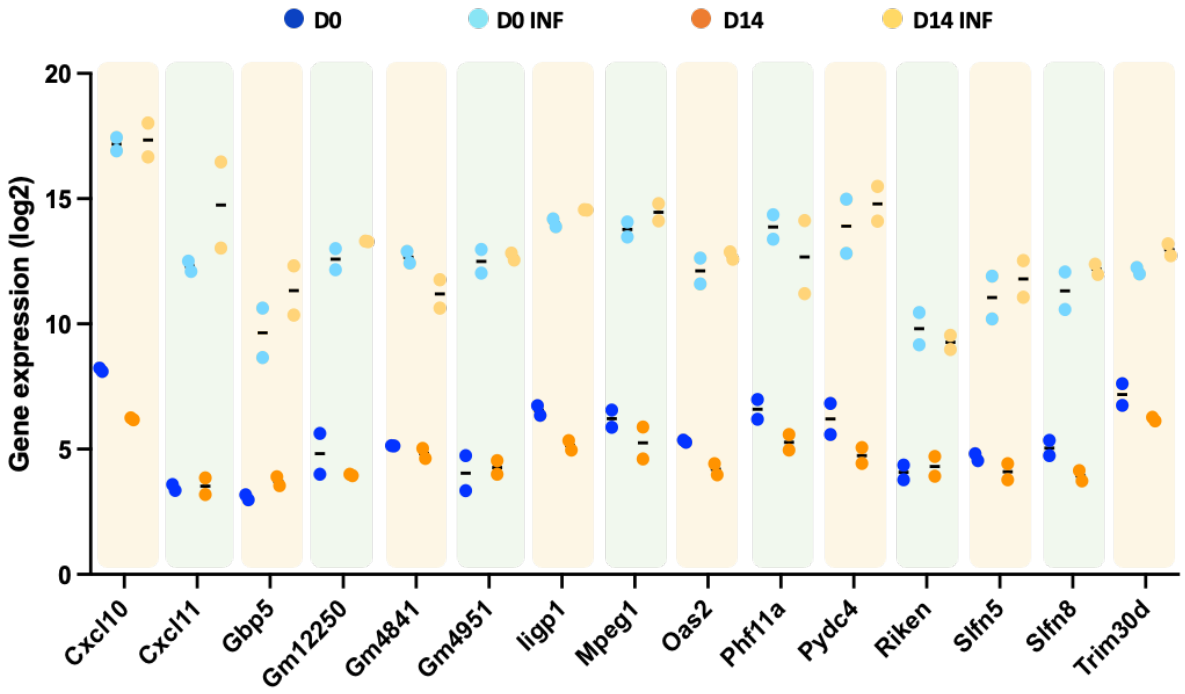


Figure 4.34: Comparison of expression induced genes shared between the undifferentiated and differentiated data sets. Log2 gene expressions are plotted by gene, comparing undifferentiated mock (D0, dark blue circle) with X31 viral infection (D0 INF, light blue circle) and differentiated mock (D14, orange circle) and the respective virus infected counterpart (D14 INF, yellow circle). Both replicates are plotted to indicate replicate specific variation, with the mean indicated by a black line. Genes are listed in alphabetical order, rather than logFC.

The data show that X31 infection of ALI D14 differentiated cells lead to a much higher response of upregulated genes compared to induction seen in undifferentiated cells. I next investigated if the highest upregulated genes in the differentiated condition were completely unmodulated in undifferentiated conditions or were just below the chosen cut off point (Figure **4.35**). The response observed differed from gene to gene, where in case for example of *Il1α* no modulation in undifferentiated cells at D0 was observed, however this became significantly upregulated in differentiated D14 cells. On the other hand, *Il6* was significantly upregulated in D14 with a lower response in D0 cells but exhibited a more variable expression across replicates. This pattern was observed for a large proportion of the significantly upregulated genes in differentiated D14 ALI cultures, where the fold increase between mock and infected counterpart was much higher than in the undifferentiated condition.

Interestingly, the only gene found downregulated in D14 infection conditions was the membrane metalloendopeptidase (*Mme*), which only showed modulation in the differentiated conditions.

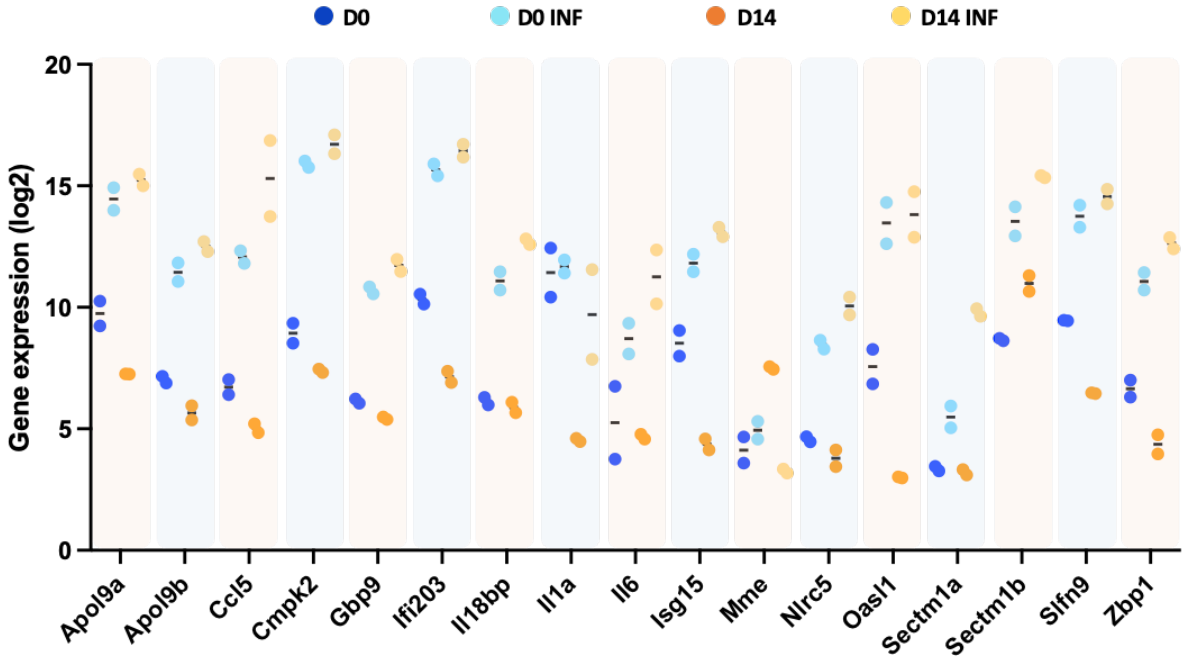


Figure 4.35: Comparison of upregulated genes in the differentiated data set. Log2 gene expressions are plotted by gene, comparing undifferentiated mock (D0, dark blue circle) with X31 viral infection (D0 INF, light blue circle) and differentiated mock (D14, orange circle) and the respective virus infected counterpart (D14 INF, yellow circle). Both replicates are plotted to indicate replicate specific variation, with the mean indicated by a black line. Genes are listed in alphabetical order, rather than logFC.

Categorising the identified genes into specific biological functions showed that antiviral response, innate immune response or interferon beta production were key functions amongst others as displayed on the enrichment grid (Figure **4.36**).

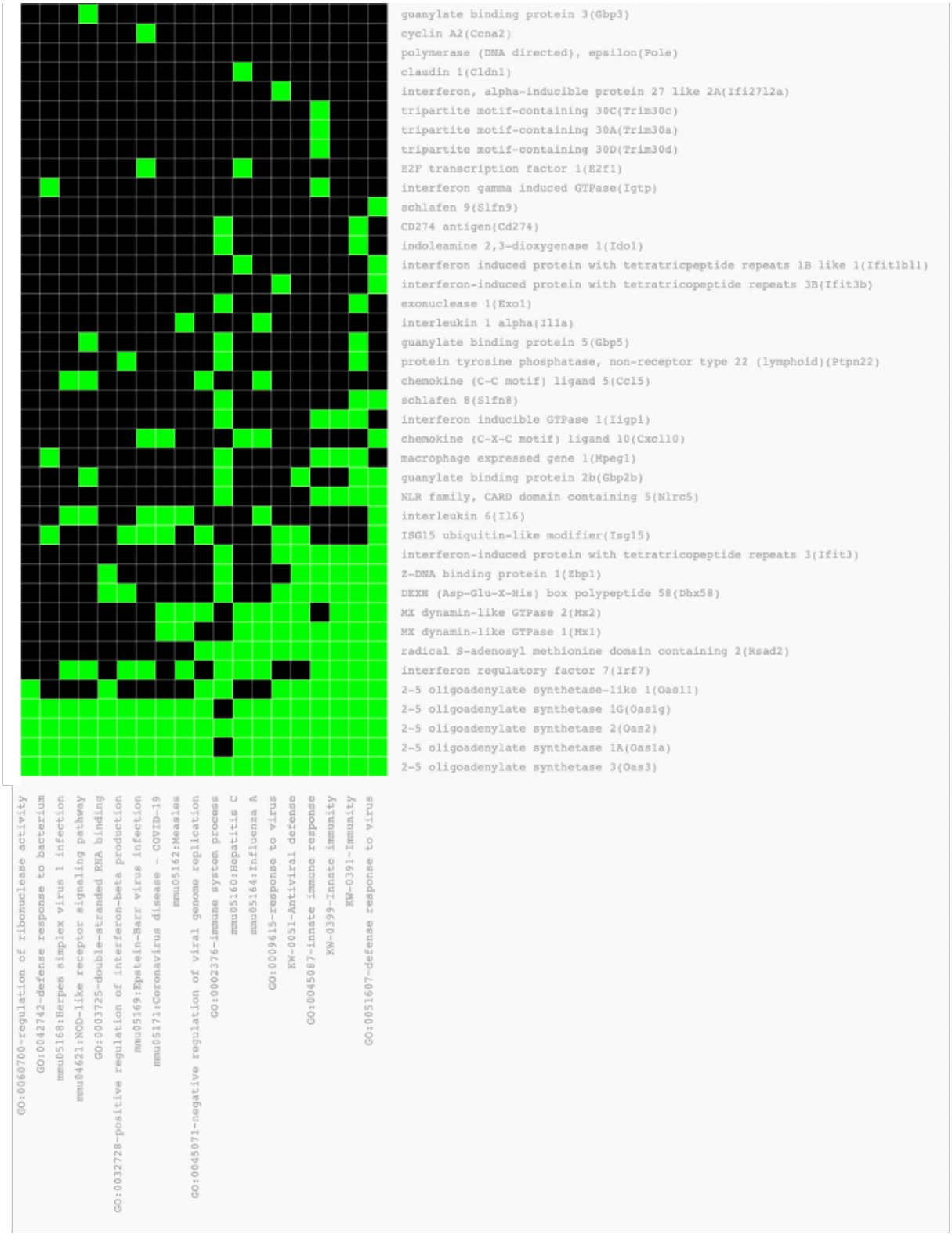


Figure 4.36: Enrichment grid displaying gene names allocated to biological functions or responses. Selected significantly upregulated genes from the D14 infection signature were chosen and displayed in an enrichment grid, which was generated using the DAVID Gene Functional Classification Tool (http://david.abcc.ncifcrf.gov) (Huang et al, 2007).

The data gathered so far from this initial pilot study does indicate that, surprisingly, undifferentiated cells can be infected, contrary to initial beliefs that IAV appears to target specific cell populations in a differentiated cell layer (Akram et al). How this is facilitated will remain a question which requires further investigation. Therefore, initial infection studies could be conducted in undifferentiated cell populations, which would be easier and faster to achieve.

4.4 Discussion

Air-liquid interface ALI cultures of human primary airway epithelial cells are a well-established *in vitro* model to investigate the role of airway epithelial cells in chronic lung disease (Nichols *et al.*, 2014; Schilders *et al.*, 2016). A large available variety of transgenic mouse lines offer the opportunity to investigate specific disease models through linked *in vitro* and *in vivo* experiments of mTEC cultures (Rock *et al.*, 2010). Multiple studies employed a mTEC model to investigate host-pathogen interactions of the mouse tracheal epithelium of viral (Cormier et al., 2010, Akram et al., 2018) or bacterial source (Bayes et al., 2016, Pilloux et al., 2016). A cell-based model system such as the mTEC system exhibits advantages such as reduced animal numbers and lower expenses for *in vitro* experiments requiring expansion and subsequent mTEC ALI differentiation, and accordingly results in a reduced biological variability. The disadvantage of using the mTEC system is the lack of the resident or recruited innate immune cells, which respond to infection by the release of cytokines, chemokines and other innate immune factors. And therefore, this limits the investigation to more simple infection models. Nonetheless, mTEC cultures provides novel insights in the dynamics of basal cell populations *in vitro*.

We did show that expanded mTEC cultured in ALI conditions retained the ability to develop a pseudostratified epithelial layer over a period between 14 to 28 days. However, the question remained if undifferenced cultures could be used for our IAV infection model. The advantage would be a simpler infection model of using undifferentiated cells rather than the more complicated ALI model, which requires additional reagents and time for development. Although it appears that IAV virus strains target specific cell populations in differentiated cell populations (Akram *et al.,* 2018), our observation showed that both differentiated and undifferentiated cell population are susceptible to IAV infection. However, how this is modulated and which receptor molecules and/or proteases are required to process HA prior to viral uptake has not been established yet. Nonetheless, our validation results show that mTECs can be a valuable model for IAV infection studies. They appear to readily infected even when undifferentiated and this provides the basis of a simple quantitative infection assay to uncover the role of specific proteins, in my case BPIFB1, in regulating IAV infection.

The fact that only a percentage of cells were infected in these assays reflects partly the complex nature of mTEC ALI cultures and where only a proportion of the cells express the α2,3- and α2,6 SA receptors (Akram *et al.,* 2018).

The X31 IAV strain used for the microarray study is an H3N2 influenza virus that has been reported to cause mild to moderate illness in mice, in contrast to the mouse adapted strain PR8, a H1N1 IAV strain that is known to cause severe infection in mice. Despite the fact that the X31 virus contains six internal genes of the PR8 strain (Kilbourne, 1969; Lamb and Krug, 1996; Rutigliano *et al.*, 2014). It has been recognised that a high-pathological acute influenza virus infection is associated with a dysregulated $CD8^+$ T cell response, which is likely caused by the more highly inflamed airway microenvironment during the early days of infection (Rutigliano *et al.*, 2014). Therefore, it might be of interest in future experiments to investigate the host response towards PR8 viral challenge in comparison in the mTEC system or a more physiological model.

It is well established that in the respiratory epithelium, interferon expression is a central cellular response to viral infection (Killip *et al.*, 2015), and it is essential in developing an antiviral state in both infected and neighbouring, non-infected cells to limit viral replication (Haye *et al.*, 2009). Consistent with this, our study identified various ISGs and immune modulatory genes as the most differentially expressed genes during influenza acute phase of infection.

Interferon inducible GTPase 1 *(Iigp1)* belongs to the immunity-related GTPases (IRGs) family, whose members are present at low resting levels in mouse cells but are strongly induced transcriptionally by interferons, being a mouse-specific ISG (Uthaiah *et al.*, 2003; Tian *et al.*, 2020). Iigp1 has been implicated in cell-autonomous resistance to intracellular pathogens by disrupting the parasite vacuole membrane and more recently demonstrated to restrict viral replication (Tian *et al.*, 2020). Other examples for IFNI stimulated response genes are 2′, 5′-Oligoadenylate Synthetase 2 (*Oas2*) (Liao 2020) or Schlafen (*Slfn*) genes (Katsoulidis *et al.*, 2009; Katsoulidis *et al.*, 2010; Li *et al.*, 2012), which interfere with viral replication (Liu *et al.*, 2012; Li *et al.*, 2012).

Other innate immune responsive genes which have been found overexpressed are for example the macrophage-expressed gene 1 (*Mpeg1*), which belongs to the membrane attack

complex/perforin (MACPF) branch of pore-forming immune effectors. This structurally diverse protein family introduces large transmembrane aqueous channels in target membranes, and is found upregulated in response to proinflammatory signals such TNFα and LPS (Bayly-Jones *et al.*, 2020). Consistently with our data, the membrane metalloendopeptidase (*Mme*) has been found in other studies as one of the highest downregulated genes in the acute phase of influenza virus infection (Zhai *et al.*, 2015).

Our expression data is consistent with other studies of host – pathogen gene response using airway epithelial cells, challenged with IAV. Interferon pathway, innate immunity endocytosis, and GTPase activity signatures are significantly upregulated during the acute phase of infection, while the expression levels of genes involved in translational elongation and protein biosynthesis are decreased (Ioannidis *et al.*, 2012; Zhai *et al.*, 2015).

Notably, a proteomic study investigating primary human bronchial airway epithelial cells after PR8 infection using 2D liquid chromatography and tandem mass spectrometry, identified a similar number of up-regulated 52 proteins above 2-fold, and 41 down-regulated proteins (Kroeker *et al.*, 2012). In contrast, we noticed a much higher number of overexpressed genes compared to the single downregulated gene *Mme* in the differentiated condition, applying criteria of an adjusted p-value of 0.05 and a log2 fold change cut off of 1. This observation could be possibly explained by the fact that identified modulated proteins on the array depend on viral strain (Scull *et al.*, 2009), cell type, location (Kroeker *et al.*, 2012) and differentiation state (Chan *et al.*, 2010), perhaps and the viral load measured as infected MOI level (Kroeker *et al.*, 2012). I would be interesting to test if the signature changes if a different MOI is used. Out study employed a MOI of 1, whereas studies using a MOI of 7 showed that 50% of cells demonstrated productive infection by 12 h post infection and over 95% of cells by 24 h (Kroeker et al., 2012). This was assessed by accumulation of viral NS-1 protein at 6, 12, and 24 h post infection protein of cell lysates assayed by Western blotting.

Bpifb1 (logFC0.2) and *Bpifa1* (logFC0.3) were not significantly modulated within the first 24h of acute viral response. This is consistent with a previous study, which only saw a significant loss of BPIFA1 protein levels in broncho-alveolar lavage seven days post infection (Akram *et al.*, 2018) and suggests that *BIPF* genes are not directly responsive to IAV.

Furthermore, *Bpifb1* expression is not particular high in mTECs, which might reflect the situation of the native tracheal epithelial cells. The higher expression observed in the original wt cells might stem from submucosal gland expression, however SMGs do not form in mTEC ALI cultures and therefore the restored differentiated *Bpifb1* levels at D14 reflect solely expression levels of the surface epithelium only and cannot rescue original levels.

Notably, *Bpifb1* and *Bpifa1* show the highest expression across their respective family members, and at are the same time the only genes which show a modulation during the ALI culture condition and compared to the wt mTEC cells after resection. The other BPIF family members might be only differentially expressed in certain conditions or specific locations.

IL-13 is a T helper 2 cytokine known to induce goblet cell metaplasia *in vivo* and *in vitro* using ALI cultures of airway epithelial cells (Kuperman *et al.*, 2002; Taube *et al.*, 2002; Fujisawa *et al.*, 2008; Alevy *et al.*, 2012; Mertens *et al.*, 2016). IL-13 exposure during differentiation induced MUC5AC positive goblet cells, which is likely to result from basal cells differentiating towards goblet cells rather than ciliated cells. Instead, it is also possible that IL-13 leads to an induced trans-differentiation of ciliated cells towards goblet cells, and thus consequentially results in fewer ciliated cells (Eenjes *et al.*, 2018). Moreover, IL-13 has been reported to induce a reduced barrier function *in vitro* (Ahdieh *et al.*, 2001; Saatian *et al.*, 2013; Eenjes *et al.*, 2018).

The rationale of IL-13 treatment was to generate cell populations with higher goblet cell populations with the expectation to generate more *Bpifb1* expression and potentially lead to other secretory molecules to be present, of which the effect on infection could be studied in the future (Seibold, 2018). However, only a very small number of DEGs were observed after IL-13 treatment for 14 days during ALI differentiation. The only significantly genes upregulated were Olfactomedin 4 (*Olfm4*) (logFC >1), Bpifb1 (logFC 0.794) and tissue inhibitor of metalloproteinase 1 (*Timp1*) (logFC 0.877). Timp1 proteins are natural inhibitors of the matrix metalloproteinases (MMPs). Interestingly, Olfm4 has been suggested to regulate innate immune response upon bacterial and viral infection (Liu and Rodgers, 2022), whereas Timp1 promotes the immune response after influenza PR8 infection (Allen *et al.*, 2018). Bpifb1 served as an internal control of goblet cell differentiation.

With respect to data processing and analysis, Limma-voom was used for identifying differentially expressed genes (Law *et al.*, 2014). Alternative frequently used microarray

analysis tools are edgeR (Robinson *et al.*, 2010) and DESeq2 (Love *et al.*, 2014). However, limma-voom has been reported to perform exceptional well with regards to precision, accuracy, and sensitivity, in addition to its speed, which allows handling large-scale data sets (Chen *et al.*, 2016).

Due to COVID-19 related animal work restrictions and later time constrains, it was not possible to carry out the original strategy to investigate the potential immune protective role of BPIFB1. Therefore, this should be followed up in future experiments, as the mTEC IAV infection model has been well established. Experimental plans would test the susceptibility to IAV infection of the two viral strains PR8 and X31 on differentiated mTECs derived from wild type *or* $Bpifb1^{-/-}$ mice. Whereas undifferentiated wild type mTEC cultures could be used to test the ability of recombinant BPIFB1 to modulate IAV infection. The next chapter describe studies designed to produce recombinant BPIFB1 for use in these type of experiments.

Chapter 5:

Generation and expression validation of murine BPIFB1 constructs

5.1 Introduction

BPIFB1 is a secretory protein produced predominantly by the epithelium of the respiratory tract. Abnormal expression of the protein has been linked to various pulmonary diseases including cystic fibrosis, idiopathic pulmonary fibrosis and COPD (Bingle *et al.*, 2012; Bingle *et al.*, 2013; De Smet *et al.*, 2018; Li *et al.*, 2021b). It has also been implicated in the progression of tumorigenesis, and has been suggested to have a protective role in viral or bacterial infections (Shin *et al.*, 2011; González-Arriagada *et al.*, 2012; Luo *et al.*, 2018; Wei *et al.*, 2018b; Zhou *et al.*, 2019; Li *et al.*, 2020). However, despite having been identified nearly twenty years ago its precise function remains unclear, and functional studies have been hampered due to its limited expression in differentiated primary airway cells grown at the ALI (Bingle *et al.*, 2012).

Based on known functionality of the closest paralogue of BPIFB1 -BPIFA1- in homology, and being part of the BPIF protein family, it has been speculated that BPIFB1 mediates antimicrobial properties. Consistent with this hypothesis, activity against infections have been reported: for example, modulating innate immune responses to bacterial infections (Shin *et al.*, 2011), viral infections (Akram *et al.*, 2018) or in disease conditions as reported for CF (Saferali *et al.*, 2020).

Human BPIFB1 has been reported to be N-glycosylated in glycoproteomic studies in saliva, and in biochemical experiments of airway cell secretions (Ramachandran *et al.*, 2008; Bingle *et al.*, 2010), containing three N-glycosylation positions identified by the motif N-X-S/T. As highlighted previously, the murine orthologue has four sites based on the motif N-X-S (section 3.3.2.1), and these have not been studied for glycosylation states at to date. It is well established that glycosylation modifications are important for both the structure (Imperiali and O'Connor, 1999) and function (Patterson, 2005) of proteins involved in immune defence (Dobrica *et al.*, 2020). The existence of glycosylation of the murine protein or in the function of BPIFB1 has not previously been addressed.

In addition, BPIFB1 is structurally related to BPI proteins, which have an established role in host defence, and have been reported to show activity against the Influenza A virus infections (Pinkenburg *et al.*, 2016). BPIFB1 is composed of two BPI-fold domains that exhibit pseudo-symmetry. The N-terminal domain is structurally related to the single BPIF-fold found in

BPIFA proteins. The two domains have been best studied in BPI where they have been shown to exert different functional effects, such as the N-terminal mediated antibiotic and endotoxin-neutralizing roles or the C-terminal facilitated BPI-dependent delivery of intact gram-negative bacteria and cell-fee endotoxin-rich particles to specific host cells (Ooi *et al.*, 1987; Elsbach and Weiss, 1993; Iovine *et al.*, 1997; Iovine *et al.*, 2002).

As outlined in the introduction, the central aim of this thesis was to address the question if BPIFB1 confers a protective role during IAV infections. I aimed to investigate this using recombinant BPIFB1 in the mTEC model system.

5.2 Aims

In order to explore the physiological role of BPIFB1 in respiratory host defence and carry out respective experiments, the generation of BPIFB1 expression constructs were required to enable the production of recombinant BPIFB1. Because my thesis was based a mouse model, the murine analogue of *BPIFB1* initially needed to be cloned from tracheal epithelial cells. This required cloning into a mammalian expression vector, and expression and secretion needed to be characterised. In addition, to be able to characterise the impact of the glycosylation sites or BPI fold domains on BPIFB1 function, the wild type full-length form of BPIFB1, a pan N-glycosylation defective mutant as well as N- and C-terminal truncated versions of BPIFB1 were generated and assessed by an enzymatic PNGase assay. To achieve this aim, the following objectives were performed:

- Generation of a wild-type murine *BPIFB1* (*mBPIFB1*) expression construct.
- Generation of a glycosylation deficient *mBPIFB1* and N- and C-terminal truncation expression constructs
- Validation of generated expression constructs by transfection into HEK293 cells
- Characterisation of secreted recombinant cell culture supernatants by western blotting
- Analysis of post-translational modification employing the PNGase enzymatic assay.

5.3 Results

5.3.1 PCR Amplification of cDNA for cloning

I have previously show that mTECs express abundant *Bpifb1* and so these cells were used to extract mRNA, to provide the template for cDNA generation that was subsequently utilised in a PCR reaction to produce amplification products of *mBPIFB1*. This product was initially cloned into the pCRII-TOPO vector allowing for a direct cloning of the PCR product (2.1.4).

For efficient expression of the exogenous *mBPIFB*1 open reading frame (ORF) in a mammalian background, the Kozak sequence (CGCCGCCACC) was incorporated upstream into the forward primer sequence alongside a *NotI* site and a *BamHI* restriction site within the reverse primer. The PCR product resolved on a agarose gel is depicted in Figure **5.1**.

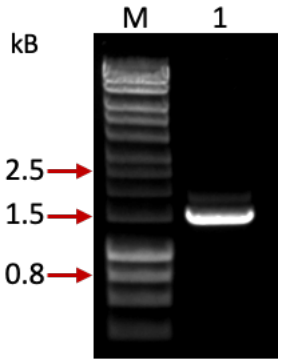


Figure 5.1: *mBPIFB1* amplification product from mTEC cDNA. mTEC mRNA was extracted and transcribed into cDNA and used to amplify the full length *mBPIFB1* ORF, including an upstream Kozak sequence and *NotI* (5') and *BamHI* (3') restriction sites. Lane M depicts a 1 kB molecular DNA marker allowing to extrapolate the amplicon size as ~1.5kB (lane 1). A negative control PCR (without cDNA) was carried out alongside and did not show any amplification products (data not shown).

5.3.1.1 Cloning of the PCR product into the pCRII-TOPO vector

The PCR product was directly cloned into the pCRII-TOPO vector without any purification steps, according to the cloning strategy performed as described in section 2.2.1. The recombinant construct was transformed into competent *E. coli* cells, and following the transformation, two individual white bacterial colonies were isolated from agar plates and cultured in LB. The recombinant plasmid DNA constructs were purified and subjected to *BamH1/XhoI* restriction enzyme digestion, to confirm successful cloning of the *mBPIFB1*

gene (Figure **5.2**). The *XhoI* restriction site is located within the multiple cloning site of the vector, downstream of the gene after PCR product insertion.

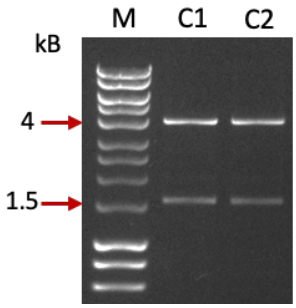


Figure 5.2: *BamH1/XhoI* restriction digest products of two individual pCRII-TOPO *mBPIFB1* clones. Lane M depicts the molecular marker, and two individual clones (C1, C2) show the vector backbone at around 4 kB and a released fragment of 1.5 kB, consistent with the expected gene size of *mBPIFB1*.

In addition to the release of the expected size of *mBPIFB1* by restriction digest, DNA sequencing was carried out to confirm the desired gene in the correct site and correct reading frame. Sequencing was performed as outlined in section 2.2.8, using T7 and Sp6 primers, which bind to sequences upstream and downstream within the pCRII-TOPO vector, and the sequence chromatogram for the start of the gene is shown in Figure **5.3** and the complete sequence is displayed in Appendix X.

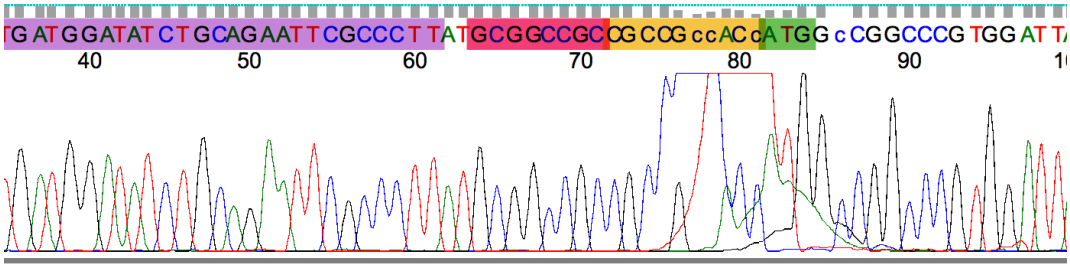


Figure 5.3: DNA sequence chromatogram of the pCRII-TOPO mBPIFB1 construct. The partial DNA sequence chromatogram of the pCRII-TOPO vector containing the *mBPIFB1* insert, demonstrating that the *mBPIFB1* PCR product had been cloned in the correct site. The sequence chromatogram was obtained using the T7 primer. The *Not1* restriction site is highlighted in red; the Kozak site in yellow, the start codon in green and the vector sequence in purple. The full-length sequence chromatogram is shown in Appendix X.

Following confirmation of the correct gene by DNA sequence analysis, the full-length insert was excised using *BamHI / NotI* restriction enzymes to subclone *mBPIFB1* into the mammalian VR1255 expression vector (Figure **5.4**).

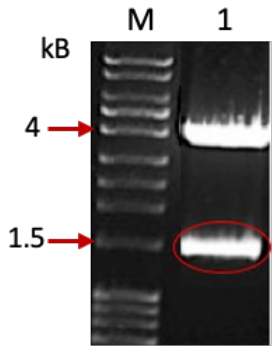


Figure 5.4: Restriction digestion of pCRII-TOPO mBPIFB1 construct with *BamHI* and *NotI.* Lane M depicts a 1 kB molecular DNA marker allowing to extrapolate the gene size as 1.5 kB (lane 1, red circle) alongside the pCRII-TOPO vector backbone at approximately 4 kB.

5.3.2 Subcloning of the full-length *mBPIFP1* insert into the pVR1255 expression vector

5.3.2.1 Digestion of the pVR1255 vector

Hartikka and colleagues initially developed and used the pVR1255 plasmid vector as a DNA vaccine delivery vehicle (Hartikka *et al.*, 1996), and in more recent studies this vector has been employed successfully to produce secreted proteins (Jayawardane *et al.*, 2008). The Bingle lab has previously used this vector for a number of protein expressions. Therefore, the *mBPIFB1* ORF was cloned into the mammalian pVR1255 expression vector, using *NotI* and *BamHI* restriction sites, with the aim of generating secreted mBPIFP1 protein from HEK293 cells.

The pVR1255 vector linearisation with restriction enzymes *NotI* and *BamHI* resulted in two fragments of ~ 1.5 kb and ~ 4.5 kb, confirming excision of the LUX sequence from the vector backbone (Figure **5.5**).

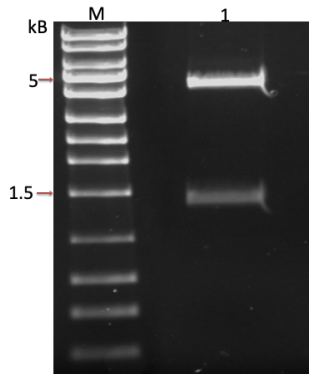


Figure 5.5: *NotI/BamHI* digestion of the pVR1255 plasmid. Lane 1 shows bands at ~1.5 kb and ~4.5 kB following the sequential digestion with *NotI* and *BamHI* enzymes, corresponding to the LUX gene sequence and vector backbone respectively. Lane M depicts the 1 kB DNA marker.

5.3.2.2 Diagnostic digest of the *mBPIFB1* insert cloned in the pVR1255 vector

The *mBPIFBI* ORF, which was excised from the pCRII-TOPO vector, was then subcloned into the pVR1255 vector. A diagnostic digest of the resulting construct was performed using *BglII*/*XhoI* restriction enzymes, to confirm that the cloning step was successful. As the *mBPIFB1* insert contains a unique internal *BglII* restriction sequence and the vector backbone contains a single *XhoI* restriction site, two product sizes of ~3.5 kb and ~2.5 kb were expected (Figure **5.6**).

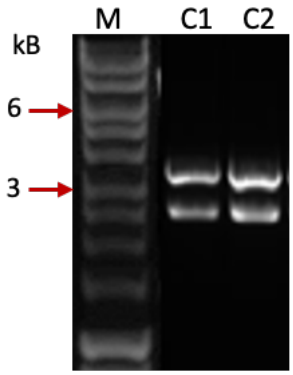


Figure 5.6: *BglII/XhoI* diagnostic digest of the pVR1255 mBPIFB1 construct. Restriction digestion with *BglII/XhoI* of two individual clones (C1, C2) confirmed existence of the mBPIFB1 gene in the pVR1255 vector, as demonstrated by presence of the two fragments at the expected size of ~3.5 kB and ~2.5 kB. Lane M shows the 1 kB DNA marker.

5.3.2.3 Sequencing of the pVR1255 construct containing the *mBPIFB1* insert

To confirm that the correct insert had been cloned into the pVR1255 vector, the construct was sequenced at the Genomics Core Facility at the University of Sheffield, using forward primer 'pVR1255', as shown in Table *2.3*, and as described previously (2.2.8). Sequencing reactions were performed using only the forward primer, since the insert DNA had been fully sequenced previously using both forward and reverse primers when it was within the pCRII-TOPO vector (Appendix XI).

5.3.3 Cloning of the glycosylation mutant *mBPIFB1* into the pVR1255 vector

In order to study mBPIFB1 post-translationally modified N-linked glycosylation events as reported for hBPIFB1, and to elucidate a potential biological role, the full-length *mBPIFB1* gene containing four glycosylation site mutations was synthesised by Biomatik. The *mBPIFB1* gene was modified in a way where all four proposed N-glycosylation sites were mutated from asparagine into amino acids, which no longer serve as an acceptor for N-glycosylation. The gene was provided in the pBSK(+) Simple AMP vector, and subcloned using *BamHI* / *NotI* restriction enzymes (Figure **5.7**). Following electrophoresis, the band at ~1.5 kb containing the mutant *mBPIFBI* DNA was gel excised, purified and subcloned into the pVR1255 expression vector.

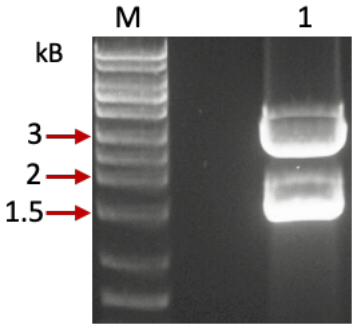


Figure 5.7: *NotI / BamHI* digestion of *the mBPIFB1* four glycosylation mutant from the pBSK(+) vector. Lane 1 shows digestion products at the expected sizes of ~3 kB (vector backbone) and the insert of ~1.5 kB. Lane M depicts the 1 kB molecular size marker. See Appendix IV for more information.

Following plasmid extraction of two clones (C1 and C2), the construct was diagnostically digested with restriction enzymes *BglII/XhoI* in order to confirm presence of the glycosylation mutant mBPIFB1 gene (Figure **5.8**).

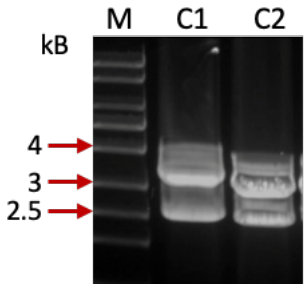


Figure 5.8: *BglII/XhoI* diagnostic digest of pVR1255 mBPIFB1 glycosylation mutant. *BglII/XhoI* restriction enzyme digestion of the glycosylation mutant mBPIFB1 gene cloned into the pVR1255 vector from two individual clones (C1, C2) resulted in two fragments at a size of ~2.5 kB and ~3.5 kB. Lane M depicts the 1 kb molecular size marker.

To confirm that the mutant insert had been cloned successfully into the pVR1255 vector, and in the correct orientation, the construct was sequenced in both the forward and reverse direction at the Genomics Core Facility at the University of Sheffield. The sequencing data spanning the four N-glycosylation sites of mutant *mBPIFB1* are shown below in Figure **5.9** to highlight successful abrogation of the asparagine (Asn) amino acid, responsible for N-glycosylation modification.

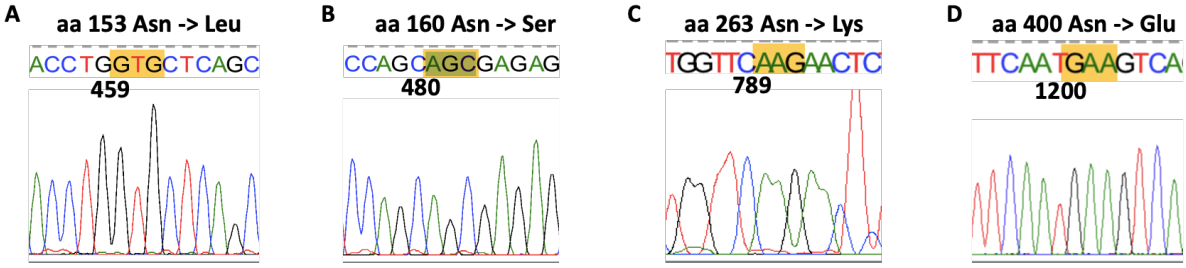


Figure 5.9: Sequence chromatograms of glycosylation deficient mutant sites within *mBPIFB1*. Sequence chromatograms showing mutations encoding amino-acids 153 (N to L (Leucine)) **(A)**, 160 (N to S (Serine)) **(B)**, 263 (N to K (Lysine)) **(C)** and 400 (N to E (Glutamate)) **(D)**, which removed the four N-glycosylation sites from *mBPIFB1*, therefore potentially abrogating post-translational glycosylation events within the protein. Wild-type nucleotide sequences at these sites were: AAT, AAT, AAC and AAC, respectively and are indicated by the respective nucleotide number in bold. The full-length sequence chromatogram is outlined in Appendix V.

5.3.4 Cloning of the N- and C-terminal truncation constructs of *mBPIFB1* into the pVR1255 vector

As outlined in the introduction, BPIF protein family members are divided into two subgroups according to their structural features (Bingle and Craven 2002, Bingle et al 2011). BPIFA proteins are homologous only with the N-terminal part of BPI proteins, containing only a single BPI-domain, whilst the long BPIFB proteins are homologous with both the C- and N-terminal of BPI, thus encompassing two BPI-domains (Bingle et al 2011). Both domains are known to be key mediators in the immune response of BPI (Elsbach and Weiss, 1998, Beamer et al. 1997, Gray et al. 1993). Therefore, one of my aims was to generate either N- or C-terminal domain containing versions of mBPIFB1 in order to study a potential antiviral function of the BPI protein fold domains.

Truncation constructs for N- or C-terminal abrogated mBPIFB1 were synthesised by Biomatik, and contained in both cases the original 21 amino acid N-terminal signal sequence including the start codon ATG (MAGPWIITLLCGLLGATLVQA) and a C-terminally located Flag epitope for detection by western blotting (Figure **5.14**A) (Appendix V and VI).

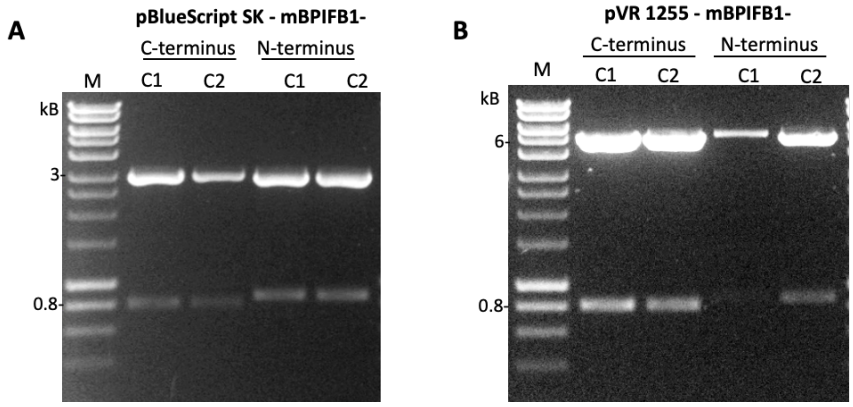


Figure 5.10: Excision of N- and C-terminal truncations of mBPIFB1. A) *NotI / BamHI* digestion of the pBlueScript II SK vector containing N-terminal and C-terminal *mBPIFB1* truncation sequences. C-terminal (772 bp) and N-terminal (856 bp) *mBPIFB1* truncation constructs were excised using *NotI/BamHI* restriction enzymes from two individual clones (C1, C2) per construct, releasing a vector backbone fragment of ~3 kB and gene insert size of either 772 bp (C-terminal C1, C2) or 856 bp (N-terminal C1, C2). **B)** Diagnostic digest of two N- and C-terminal construct clones of pVR1255 *mBPIFB1* using *NotI/BamHI* to confirm correct subcloning. The restriction digest released expected fragments of either 772 bp (C-terminal) or 856 bp (N-terminal) *mBPIFB1* and the pVR1255 vector backbone of 5 kB.

These truncation sequences consist of either of the N-terminal (856 bp, spanning amino acid 1-273) or the C-terminal (772 bp, containing the signal peptide and spanning amino acids 239-474) parts of the mBPIFB1 gene, and were provided in the pBlueScript II SK vector (Figure **5.14**A. Subsequently, the constructs were mini-prepped and restriction digested with *NotI / BamHI* enzymes (Figure **5.10**A) before cloning into the expression vector pVR1255. The resulting pVR1255 clones were diagnostically digested (Figure **5.10**B) and sequenced to confirm the correct and error free base pair sequence.

5.3.5 Expression validation of mBPIFB1 constructs by transfection and Western blotting

The expression vector pVR1255 was used to generate secreted recombinant mBPIFB1 in HEK293 using the calcium phosphate nucleic acid delivery. Initially, midi preparations of required expression plasmids were prepared, and the corresponding plasmid DNA extracted. HEK293 cells were used for expression and subsequent protein extraction experiments as this cell line has been widely used in biochemical applications due to its fast and steady growth and propensity of DNA delivery by transfection (Kavsan *et al.*, 2011).

Initial transfection trials took place at small scale in a 24 well format to assess delivery efficiency by green fluorescent protein (GFP) transfection, alongside the wild type and respective pan glycosylation deficient mutant of mBPIFB1 (Figure **5.11**).

As the calcium phosphate transfection method is economic compared to commercial products such as FuGENE transfection reagent and resulted in a highly efficient GFP transfection judged by expression in > 95% of cells (Figure **5.11**), this method was used for all subsequent transfection protocols.

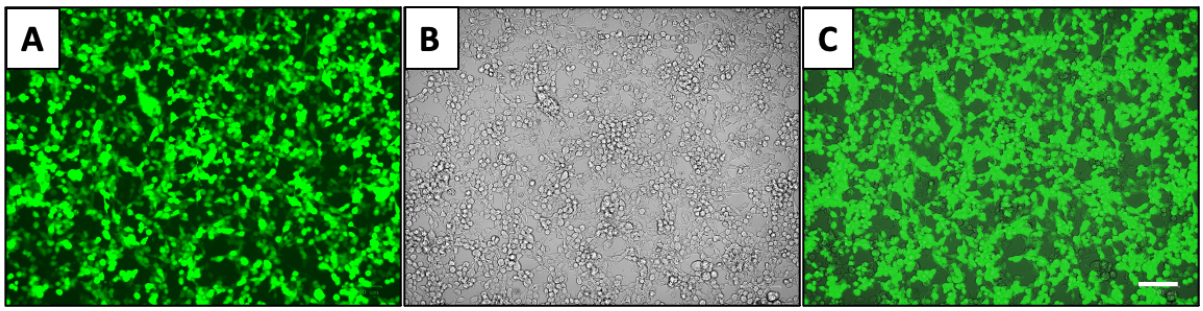


Figure 5.11: Representative images of GFP transfected HEK293 cells. Cells were transfected with a GFP construct using the calcium phosphate method for visualisation of transfection efficiency and imaged 48h post transfection. Cells were recorded using the green fluorescence channel **(A)**, brightfield **(B)** or both images overlaid **(C)**. Bar indicates 100 μm.

To assess expression and successful secretion of recombinant mBPIFB1 in the cell culture media, western blots samples were prepared from serum free conditioned media harvested at 24h intervals starting at 48h post transfection, and the first harvest was subjected to gel electrophoresis using the same sample volumes across all conditions. Recombinant proteins were detected using the C-terminal Flag epitope (Figure **5.12**). Transfection of the wild type mBPIFB1 construct resulted in detection of two distinct protein bands spanning a range from 50-54 kDa in size by western blot compared to the glycosylation deficient mutant at around the predicted size of 48 kDa, suggesting that the wild type protein is post-translationally modified by glycosylation events prior secretion (Figure **5.12**A and B). To assess if this is only restricted to a time point around 48h post transfection and to test expression levels across a broader time scale, serum free condition media harvests were performed every 24h for three sequential days (H1-H3, 48h – 96h post transfection), starting 48 h post calcium phosphate transfection (Figure **5.12**C).

The previously observed pattern of a slower migrating wt mBPIFB1 protein at around 54 kDa was shown in harvest 2 (H2, 72h post transfection) and harvest 3 (H3, 96h post transfection) compared to the glycosylation deficient protein (mut) at around 48 kDa. Further, it was noted that transfection of comparable amounts of wt versus mutant *mBPIFB1* plasmid DNA appeared to yield higher amounts of secreted mutant protein compared to wild type in harvest 1 and 2 but not later points (harvest H3) (Figure **5.12**C and D). However, as the objective was to harvest the serum free supernatant for protein purification of both mBPIFB1 variants, these results suggest that all three harvests will contain sufficient amounts for subsequent purification approaches by anti-flag resin.

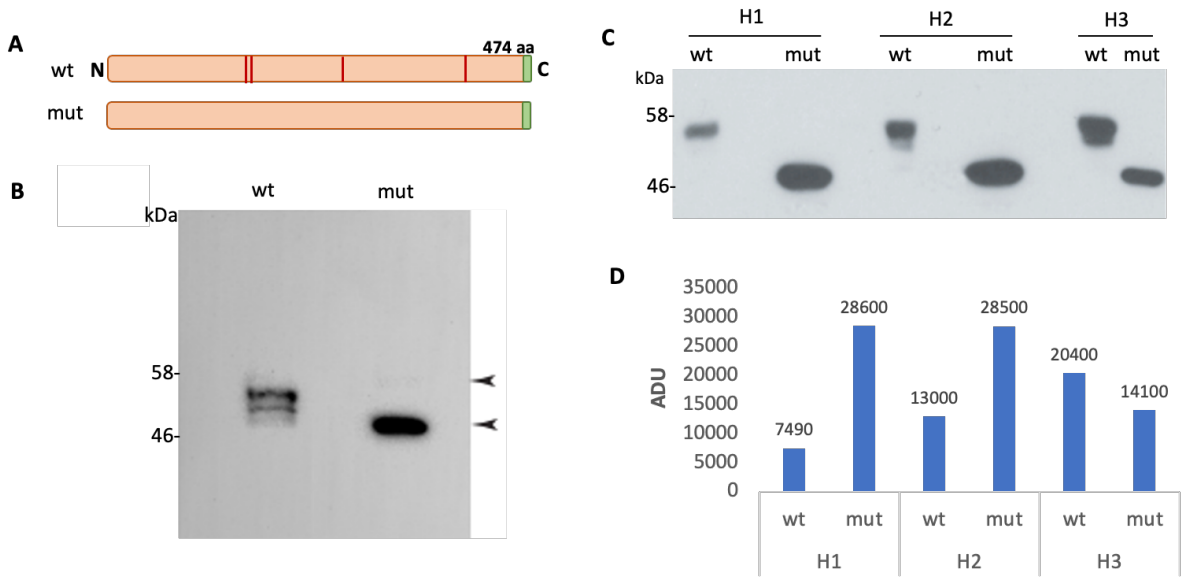


Figure 5.12: Western blot detection of secreted mBPIFB1 proteins. Transfection of wt and N-glycosylation mutant (mut) pVR1255 *mBPIFB1* constructs using the calcium phosphate method into HEK293 cells and detection of secreted proteins using a Flag antibody. **A)** Depicts the wt or mut mBPIFB1 proteins indicating the location of the four N-glycosylation sites (red lines) and the C-terminal Flag epitope (green). **B)** The full-length protein is predicted to migrate around 52 kDa by western blotting, and the wt transfected cells (48h post tranfection) showed a distinct double protein band around 50-54 kDa, indicative of glycosylation events. However, transfection of the glycosylation mutant construct only showed a single band at the expected size around 48 kDa, suggesting absence of post-translational modifications. **C)** Assessment of mBPIFB1 expression levels in serum free media was performed by western blotting of conditioned media 48h (harvest 1 -H1), 72h (H2) or 96h post (H3) transfection, and detection of the recombinant protein by flag antibody. **D)** Shows the densitometry estimation of protein levels respective to each other in arbitrary densitometry units (ADU).

In order to confirm that the difference in migration between the wt and mut proteins is due to absent N-linked glycosylation modification, and not caused by other post-translational events, both proteins were subjected to treatment with the amidase enzyme, Peptide-N-Glycosidase F (PNGase F), which cleaves between the innermost N-Acetylglucosamine (GlcNAc) and asparagine residues of high mannose (Maley *et al.*, 1989). Therefore, it effectively removes high mannose N-glycans from glycoproteins. Exposing both proteins, wt or mut mBPIFB1 conditioned media samples to PGNase F and assessing the migration pattern by western blotting, showed that the previously observed doublet bands in the wt sample around 54 kDa mimicked the pattern of the mut protein, resulting in a single band around 48 kDa (Figure **5.13**). Treatment of the mut protein with PNGase F did not alter the migration pattern confirming that this protein was indeed not N-glycosylated. Therefore, it seems that the wt

mBPIFB1 is glycosylated on at least two distinct sites, resulting in two prominent bands, which differ from the migration pattern of a non-glycosylated protein (Figure **5.13**). This observation is consistent with *in silico* N-glycosylation analysis using the online tool NetNGly (http://www.cbs.dtu.dk/services/NetNGlyc/), which predicts two out of the four potential sites (aa 153 and aa 400) to be N-glycosylated (Appendix XII) (Gupta and Brunak, 2001).

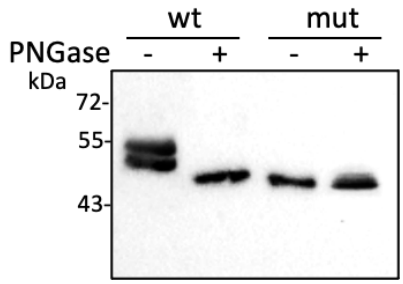


Figure 5.13: mBPIFB1 is post-translationally modified through N-glycosylation events. Secreted wt or mut mBPIFB1 protein from harvest 1 (48h post transfection) was incubated in the presence (+, lane 2 and 4) or absence (-, lane 1 and 3) of the PNGase, the reaction resolved by western blotting and the proteins detected using a flag-antibody.

The N-terminal mBPIFB1 construct is predicted to have an expected size of 30.2 kDa, according to the Expasy tool (https://web.expasy.org/compute_pi/), and contains three potential glycosylation sites (aa 153, 160, 263), whereas the C-terminal construct should yield a slightly smaller protein of a predicted size of 28.4 kDa, which contains two potential glycosylation sites, namely aa 263 and aa 400 (Figure **5.14**A). Western blotting of harvested serum free supernatant of N- or C-terminal construct transfected HEK293 cells showed abundant secretion of the C-terminal construct, across a range from 33 kDa to 44 kDa, displaying a range of bands (Figure **5.14**B). In contrast, the N-terminal construct appeared to be secreted at a much lower level, despite maintaining the same transfection conditions and no prominent higher migrating band is observed besides the expected size around 30 kDa (Figure **5.14**B). Reasons for this observation could be due to potential protein instability to less efficient secretion. In addition, the observation of a single prominent band only at the expected size suggests that there is no N-glycosylation modification taking place (Figure **5.14**B).

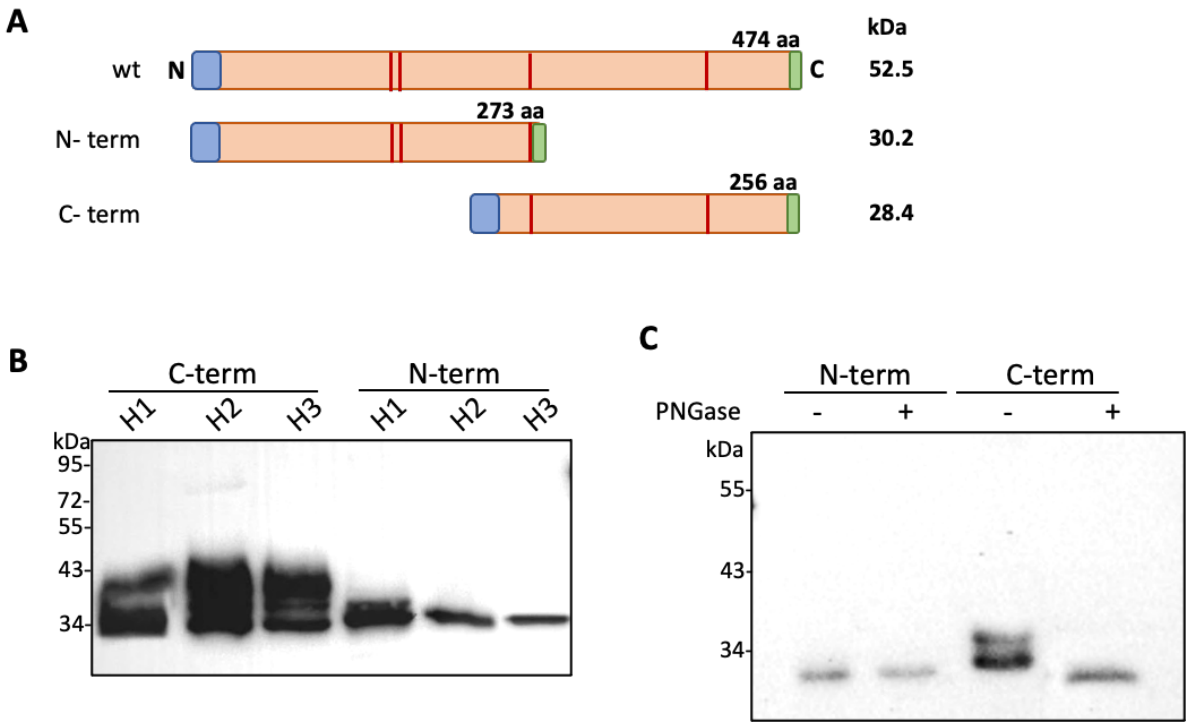


Figure 5.14: Western blot detection of secreted N- or C-terminal truncated mBPIFB1 proteins. A) Generated constructs contained in both cases the original 21 amino acid N-terminal sequence including the start codon ATG (blue) and a C-terminal Flag epitope (green) for detection by western blotting. The position of predicted glycosylation sites is highlighted as red lines and expected protein sizes in kDa are indicated on the right. **B)** Assessment of secreted recombinant N-terminal (N-term) or C-terminal (C-term) mBPIFB1 protein expression levels in serum free conditioned media of HEK293 cells was performed by western blotting to detect the flag epitope 48h post (harvest 1 -H1), 72h post (H2) or 96h post (H3) transfection using an anti-flag antibody. **C)** PNGase F enzymatic reaction (500 units +, or absence -) were incubated with N-terminal and C-terminal BPIFB1 proteins as described, and samples resolved by western blotting and proteins detected with anti-flag antibodies.

Assessment of glycosylation of the N- and C-terminal BPIFB1 secreted recombinant proteins using PNGase showed that the prominent bands observed in the C-terminal construct at 32 kDa and 35 kDa, are reduced to a protein size of around 30kDa after treatment (Figure **5.14**C). This is consistent with earlier observations that there might be two distinct glycosylation events taking place, at sites aa 263 and aa 400. However, the N-terminal construct, containing three potential N-glycosylation sites (including one shared with the C-terminal domain construct) does not show any change in size of the observed protein band after PNGase treatment, suggesting that this modification does not take place in this construct (Figure **5.14**C) despite it being predicted to be glycosylated at aa 153 (Appendix XII).

5.4 Discussion

The main aim of this chapter was to generate and validate expression constructs for mBPIFB1 for use in functional studies in the mTEC system. The wild-type and glycosylation deficient mutant *mBPIFB1* expression constructs (in which each of the putative glycosylation sites had been exchanged) were successfully cloned into the mammalian pVR1255 vector and expression of the recombinant protein was validated by western blotting of secreted mBPIFB1 protein following transfection into HEK293 cells. The recombinant wt protein showed two distinct protein bands suggesting that at least two glycosylation events take place across the four sites predicted to exist within the protein, compared to the lower migrating glycosylation deficient mutation. Both recombinant proteins were readily secreted into the cell culture medium and could be detected in abundance and at generally comparable level up to 96h post transfection. Secretion over longer time points were not investigated. These validated constructs could subsequently be used to generate significant quantities of both proteins for functional studies. This could be achieved by using flag-affinity resin to capture secreted recombinant mBPIFB1 in serum free supernatant of transfected HEK293 cells. After washes of the resin to remove unbound proteins, the recombinant BPIFB1 is recovered using flag-peptide to elute in a competitive manner (Gerace and Moazed, 2015). In order to estimate the potential protein yield for subsequent protein purification, protein quantification assays such as Bradford or BCA should be employed in the future. Alternatively, an enzyme-linked immunosorbent assay (ELISA) could be performed using the flag tag as epitope in order to quantify protein amounts.

Although a number of studies have suggested that the protein plays a role in immunity within the respiratory tract, the precise role of BPIFB1 is not well defined. Human BPIFB1 has been shown to be secreted as an N-glycosylated protein in human airways and saliva (Ramachandran *et al.*, 2008; Bingle *et al.*, 2010). Glycosylation is a post-translational modification of proteins mediated by attachment of sugar moieties to proteins. This process is critical for a wide range of biological processes, such as cell attachment to the extracellular matrix and protein–ligand interactions in the cell. Moreover, this modification is vital in determining protein structure, folding, function and stability (Dennis *et al.*, 1999; Trombetta, 2003; Aebi, 2013; Jayaprakash and Surolia, 2017). While it has been suggested that post-translational glycosylation is important for the function of BPIFB1 (Gao *et al.*, 2015), there have been no studies investigating the role of this glycosylation. Moreover, there are no existing studies on the mouse protein regarding this. Therefore, a glycosylation defective

mBPIFB1 construct was produced by gene synthesis and was subcloned into a mammalian expression vector. After validation and characterisation of the expression construct it was aimed to investigate if the N-glycosylation events mediate biological effects like Influenza A infection efficiency. The design was to use murine mTEC cells from a *Bpifb1* $^{-/-}$ mouse strain as employed in a published study from Donoghue and colleagues (Donoghue *et al.*, 2017), and to test the glycosylation deficient as well as the generated truncation mBPIFB1 mutants in order to identify structural regions of BPIFB1 associated biological activity.

The rationale behind the expression of the N- or C-terminal domain of BPIFB1 was to investigate an immune protective role of BPIFB1 in more detail. For example, the N-terminal section is structurally related to BPIFA1, which only possesses one BPI fold X-ray crystallography studies of BPI discovered a unique boomerang-like shape, which encompasses two domains spanning from residues 1–230 and 250–456, whereas limited proteolysis experiments generated two BPI fragments comprising residues 1–199 and 200–456. Particularly, only the N-terminal domain is essential for its biological activity as it mediates the antibiotic and endotoxin-neutralizing functions of BPI (Ooi *et al.*, 1987; Beamer *et al.*, 1997; Iovine *et al.*, 1997; Iovine *et al.*, 2002).

Expression analysis of the N-terminal construct in 293HEK cells showed a significant decrease compared to the C-terminal counterpart by over ten-fold. This might imply, at comparable transfection and sample conditions, that the N-terminal truncation protein is less stable (Deller *et al.*, 2016) and hence a significant lower amount is secreted. To test this hypothesis, transfected cells could have been lysed 48h post transfection and the lysate assessed for expression of the N-terminal versus C-terminal recombinant protein, ideally in presence or absence of a proteasomal inhibitor such as MG132. It would also be helpful to quantitate secretion levels using an ELISA against the epitope tag.

Moreover, the N-terminal truncation protein, despite containing three of the four glycosylation sites, migrated at the expected size and did not show any change in size after PNGase treatment, suggesting that no N-glycosylation events are taking place. In contrast, the C-terminal secreted protein demonstrates two distinct bands which run higher than the PNGase treated sample, suggesting that more than one site is glycosylated, which is consistent with observations of the wt protein. Interestingly, the C-terminal protein only encompasses two glycosylation sites, of which one is overlapping with the N-terminal protein (aa 263). However, no evidence has been observed that this site is glycosylated within the N-

terminal expressed protein, suggesting that potential protein stability issues preventing post-translational modifications.

We observed that the C-terminus containing secreted recombinant BPIFB1 protein showed multiple bands, which were undistinguishable, despite only containing two glycosylation sites within this construct namely N263 and N400. An explanation for this observation could be that glycans are produced by multiple competing glycosyl transferases, therefore if both sites are decorated with different glycans, this would lead to a large structural heterogeneity, and therefore to manyfold bands when analysed by gel electrophoresis (An *et al.*, 2009). In case it is of interest which sites and what type of N-glycans are attached, mass spectrometry using specific protocols for detection and analysis of glycopeptides would be a great tool to pursue (Harvey, 2005; Geyer and Geyer, 2006; Suttapitugsakul *et al.*, 2021).

Notably, as highlighted in chapter 3, BPIFB1 glycosylation sites are found across all species, however the only glycosylation position which is found conserved across all species is aa 264 in human or 263 in mice, with the exception of cattle and hedgehog. N-glycosylation site aa 48 (human) appears to be primate specific, with the exception for rabbits. Remarkably, solely rodents possess a fourth site, which also forms a cluster of two closely positioned sites (aa 153 and aa 160), which are specific to mice and rats, and the most terminal site (aa 401 in humans) is only found in primates and rodents.

BPIFB1 glycosylation occurs through attachment of an GlcNAc residue to an asparagine residue belonging to a consensus sequence NX(S/T), where X can be any amino acid except proline. The presence of the consensus sequence is required for N-linked glycosylation; however the occupation of a potential site is not obligatory. With regards to human BPIFB1 N-glycosylation, only the N- and C-terminal sites (N48 and N401, respectively) have been found to be glycosylated, not position N264 (Boersema *et al.*, 2013; Hu *et al.*, 2018).

It has been reported that in case of human hyaluronidase 1, N-glycosylation is required for secretion and enzymatic activity (Goto *et al.*, 2014). Thus, it could be speculated that the N-terminal construct of mBPIFB1 lacking N-glycosylation modification and therefore these proteins are not secreted or to a much lower extend. This hypothesis could be tested by using Tunicamycin, an antibiotic, which inhibits N-glycosylation in eukaryotes by blocking the transfer of *N*-acetylglucosamine-1-phosphate. This inhibitor has been used extensively to study N-glycans in glycoprotein maturation, secretion, and function (Kim *et al.*, 2013) and

could be used to investigate if N-glycosylation modification is required for the secretion of the C-terminal construct.

In case of an observed biological effect due to N-glycosylation, a site-directed mutagenesis approach was planned to investigate in detail which sites are required and essential for observed biological effects. However, due to the impact and associated restrictions on research imposed during the COVID-19 pandemic and related inaccessibility of experimental mice, it was not possible to address a biological role of the generated glycosylation defective mutant nor the N- and C-terminal truncation and therefore contribution of individual N-glycosylation sites to the physiological role of mBPIFB1 was not explored. In addition, due to the limited time and circumstantial constrains, the aim to further purify these recombinant proteins for biological assays could not be achieved, and no additional experimental efforts were made to investigate these observations further. I have however developed the tools to undertake these studies in the future. Because of these limitations I sought to address the role of BPIFB1 in IAV infection in a different, technically tractable model in the laboratory and these experiments are outlined in the next chapter.

Chapter 6:

Generation of a CRISPR based BPIFB1 knock out cell line and establishment/validation of an influenza A infection assay

6.1 Introduction

Although several studies have suggested that BPIFB1 plays a role in immunity within the respiratory tract, the precise physiological role of this protein is not known. Human BPIFB1 has been shown to be secreted as an N-glycosylated protein and its expression is limited to differentiated primary airway cells, gastrointestinal tract and female organ tissue (Ramachandran *et al.*, 2008; Bingle *et al.*, 2010) (Figure **3.13**).

IAV virus is a significant respiratory pathogen and previous reports showed that BPIFA1, besides demonstrating a protective role against bacterial infections, is also able to exhibit an antiviral effect towards IAV (X31) infections (Chu *et al.*, 2007; McGillivary and Bakaletz, 2010; Liu *et al.*, 2013; Krammer *et al.*, 2018a; Akram *et al.*, 2018). Therefore, the central hypothesis underlying this thesis was that BFIFB1 could contribute to the defence mechanism against IAV infections and could play a role in maintaining homeostasis within the airway epithelium. The experimental strategy to elucidate a potential role of BPIFB1 in the host defence against IAV and discovery of the underlying protective mechanism, initially involved a *Bpifb1*$^{-/-}$ knockout mouse model in combination with a primary lung cell culture, mTEC grown at air–liquid interface to allow for formation of a pseudostratified mucociliary epithelium (Ross *et al.*, 2007; Martinez-Anton *et al.*, 2013).

Unfortunately, due to COVID imposed restrictions, we were not able to obtain thew *Bpifb1*$^{-/-}$ transgenic mouse model from our collaborator at the University of Liverpool and therefore I had to abort the initial plan of using mTEC cells and the generated murine BPIFB1 mutation and truncation constructs (Chapter 5) for the rest of my thesis.

I needed to utilize an alternative model system to continue my study. For this I chose to employ human bronchial epithelial cells immortalized with CDK4 and human telomerase hTERT, commonly known as HBEC3-KT cells. These cells have been used in previous studies of airway biologu and are well established in my host laboratory (Ramirez *et al.*, 2004; Nakauchi *et al.*, 2019; Matsuya *et al.*, 2022). A key feature of these cells that is not shared with primary mTEC or HBECS is that they can be transfected and selected for clonal growth.

6.2 Aims

In order to generate an alternative readout for BPIFB1 biological function within the airway epithelium, the bronchiolar cell line HBEC3-KT was used as human cell model to study modulation of IAV susceptibility by BPIFB1. The strategy was to ablate *BPIFB1* expression by a CRISPR gene approach, and subsequently analyse if this would modulate IAV infection efficiency of HBEC3-KT cells in a cell-based assay format. This technique involved transient transfection using a double guide RNA approach, selection and clonal cell line generation by limiting dilution and expansion. The resultant cells could then be used for functional studies.

Specifically, I aimed to

- Generate a CRISPR mediated HBEC3-KT BPIFB1 knock out cell lines
- Validate generated clones by genomic, proteomic, phenotypical and functional means
- Undertake IAV PR8 infection time course in undifferentiated HBEC3-KT cells
- Analyse cytokine response upon IAV PR8 exposure in HBEC3-KT cells
- Study the effect of IAV PR8 infection levels in the presence of CRISPR clone derived lavage or recombinant BPIFB1 protein upon PR8 infection

6.3 Results

The CRISPR/Cas9 gene editing technology was employed to generate a knock-out cell line for the *BPIFB1* gene using the immortalised human bronchial epithelial cell line HBEC3-KT (Ramirez *et al.*, 2004; Nakauchi *et al.*, 2019). These cells are used extensively in my host laboratory has they have the capacity to produce mucociliary airway phenotypically similar to that of the native human airway when grown in ALI culture conditions. Initial experiments undertaken by Zeyad Alharbi confirmed that BPIFB1 expression was induced after 21 days of ALI culture (Figure **6.1**A) and label free proteomic analysis, performed by using a Thermo Fisher Orbitrap Elite nanoflow liquid chromatography tandem mass spectrometry (LC-MS/MS), showed that BPIFB1 was one of the proteins secreted by differentiated cells (Figure **6.1**B).

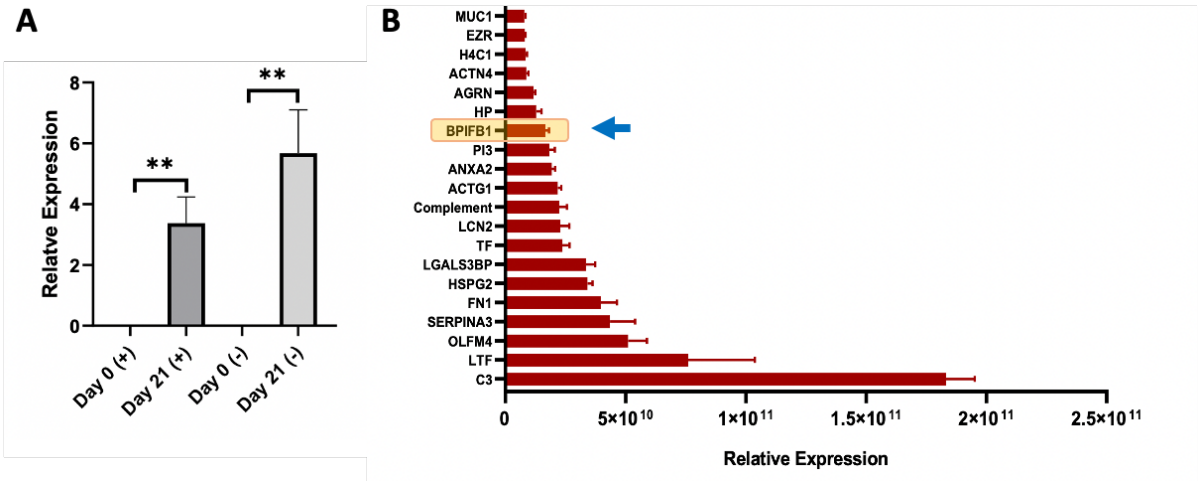


Figure 6.1: Relative protein expression of BPIFB1 in differentiated HBEC3-KT and proteomic secretome. A) *BPIFB1* expression (blue arrow) after ALI differentiation of 21 days in the presence (+) and absence (-) of collagen. **B)** The top 20 most abundant proteins secreted by HBEC3-KT 21 days post ALI culture. Proteins are designated by their corresponding gene. The error bar represents standard error of the mean. Data provided by Zeyad Alharbi.

6.3.1 CRISPR/Cas9 gene editing of *BPIFB1* in the HBEC3-KT cell line

The editing strategy employed involves plasmids containing gene specific guide RNAs (gRNA) of 20 nucleotides and co-expression of the endonuclease Cas9 to facilitate precise gene recognition and excision (Adli, 2018). In addition, the chosen vector for this study expresses a GFP protein, the co-expression of which facilitates distinction of transfected cells

and consequently successful delivery of the CRISPR plasmid (Ran *et al.*, 2013c). Fluorescence positive cells are isolated and clonal population generated using fluorescence-activated cell sorting (FACS) and selected clones were validated for successful gene editing.

6.3.1.1 CRISPR/Cas9 gene editing design and genomic target sites

In order to direct the Cas9 endonuclease successfully to the genomic location of the *BPIFB1* gene, three single guide RNAs (sgRNAs) of 20 nucleotides in lengths were designed and synthesized, targeting exon 2 of *BPIFB1* that contains the start codon (*Table 2.7*) (Figure **6.2**). The approach included two sgRNAs which were delivered in a co-transfection approach, gRNA1 in combination with gRNA3 (gRNA1+3), or gRNA2 with gRNA3 (gRNA 2+3), resulting in an excised region of over 200 bp including the entire exon 2 containing the gene start site (Figure **6.2**). Both approaches were carried out simultaneously and yielded successful knock out clones. To avoid duplication, data is only shown for generation and validation of cell lines derived from the gRNA 2+3 approach.

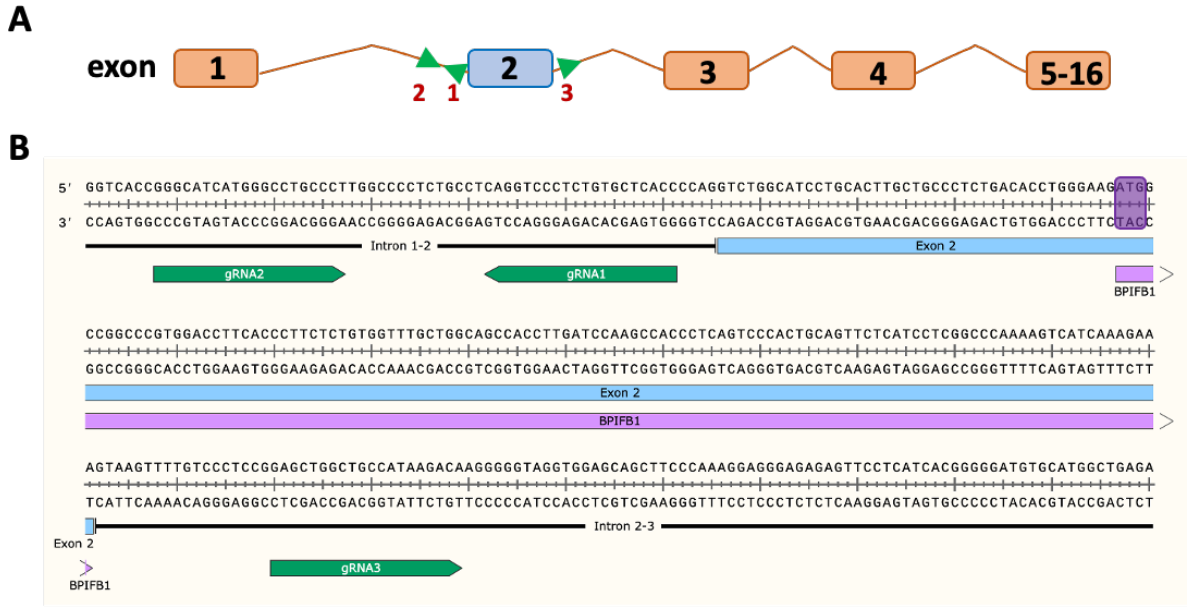


Figure 6.2: CRISPR sgRNA design and target site. A) Schematic representation of the CRISPR target sites surrounding exon 2 of the human *BPIFB1* gene containing the translation start site. **B)** Sequence of exon 2 and adjacent intron nucleotides, gRNA target sites are labelled in green, the exon 2 region in blue and the *BPIFB1* coding sequence is coloured in pink, with the gene start codon highlighted in purple.

Target sgRNAs were *BbsI* cloned into the delivery vector pX458 (Appendix VII), designed to co-express the Cas9 and GFP protein, and the plasmids amplified using *E. coli* bacteria (Ran

et al., 2013c). Extracted plasmid DNA was subjected to restriction digestion to confirm presence of the delivery vector (data not shown) and the vectors were sequenced to confirm they contain the sequences and were subsequently used for delivery into HBEC3-KT by transfection.

6.3.1.2 Plasmid delivery by transfection, selection and clonal expansion

HBEC3-KT cells were seeded in a 6 well format and transfected using the FugeneHD transfection reagent with either the empty vector pX458 serving as background control, or with a combination of vectors expressing the individual sgRNAs (pX458-sgRNA). Cells were microscopically observed at 48 hours post transfection for green fluorescence protein expression, indicating successful delivery of the vector and expression of the Cas9 protein alongside sgRNAs (Figure **6.3**).

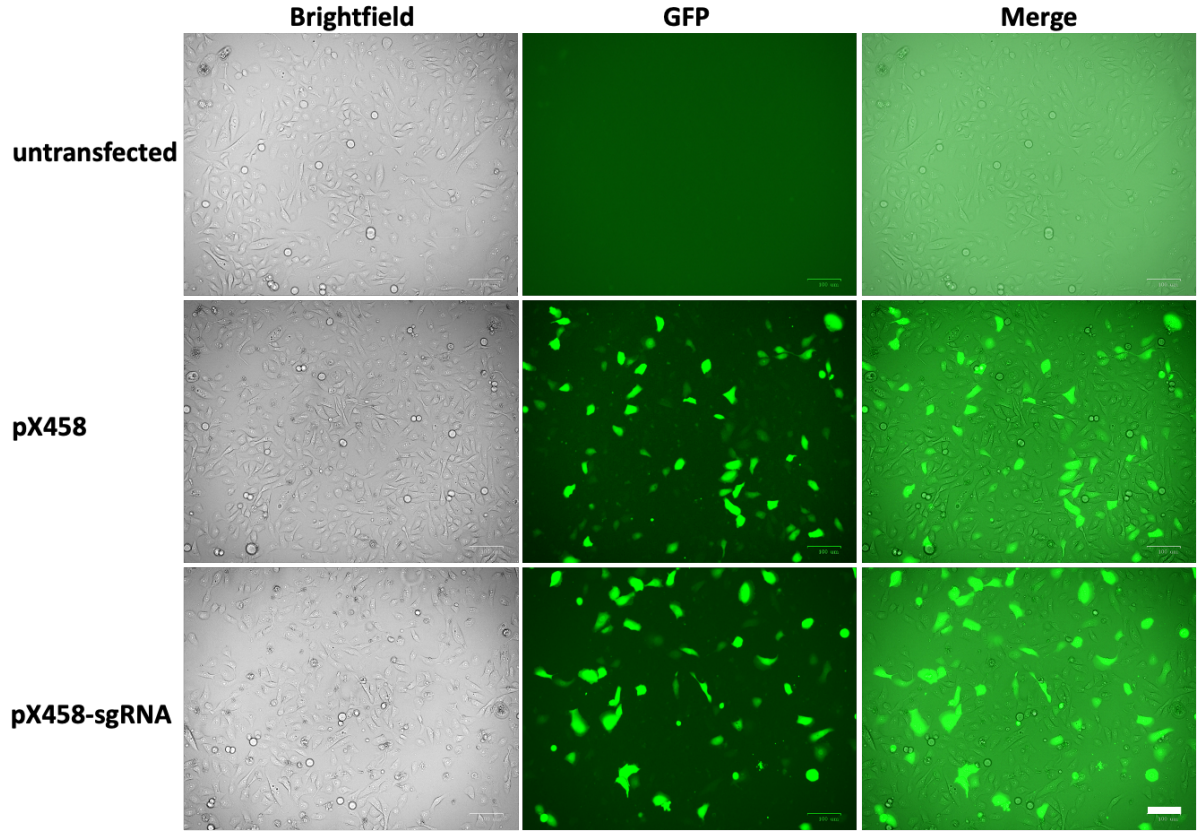


Figure 6.3: GFP expression following pX458 vector delivery into HBEC3-KT cells. HBEC3-KT parental cells were transfected with empty pX458 (pX458) or sgRNA containing (pX458-sgRNA) vectors using FugeneHD transfection reagent or left untreated (untransfected). Transfection efficiency was observed as fluorescent readout and imaged 48h post transfection. Cells were recorded using the green fluorescence channel brightfield or both images overlayed as indicated. Bar indicates 100 μm.

As expected, the empty vector transfection yielded a similar transfection rate compared to the sgRNA vector equivalent (Figure **6.3**) of an average about 15% (Figure **6.4**). In order to establish and subsequently test clonal populations for successful gene editing, GFP expressing cells were FACS sorted (Figure **6.4**).

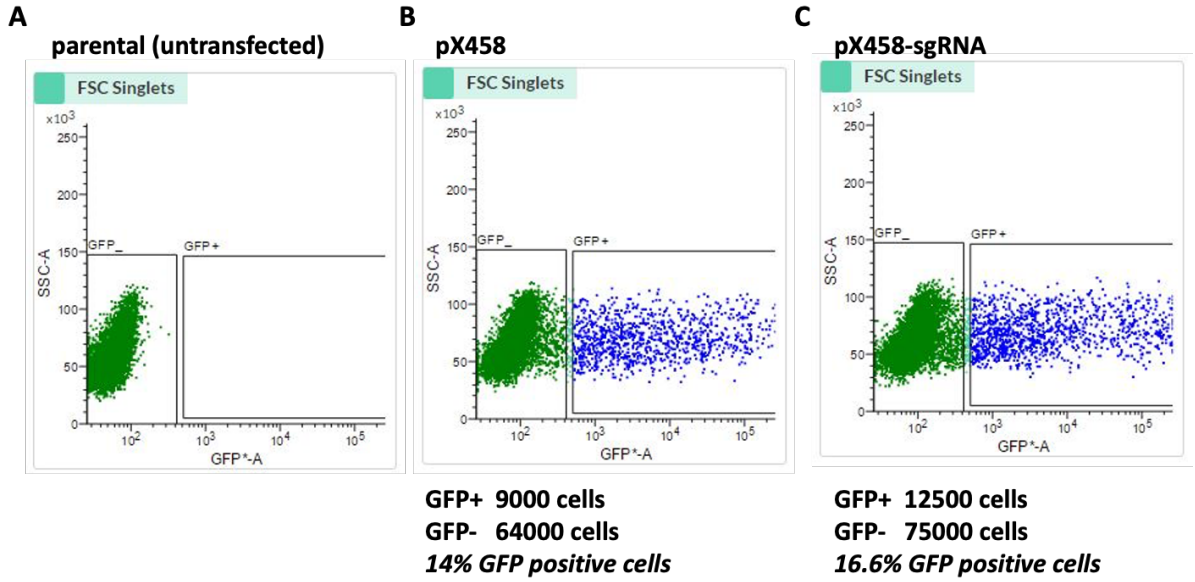


Figure 6.4: FACS profiles of transfection pools. Parental HBEC3-KT cells (untransfected) **(A)**, empty vector control pX458 **(B)** or pX458-sgRNA **(C)** were trypsinised and sorted by GFP fluorescence, using parental cells as a background level. Fluorescence positive (GFP+) and negative (GFP-) cell counts are indicated below the FACS profiles, and percentage of positive cells across the population was calculated.

As evident from GFP visualisation after transfection, GFP intensity varied widely across the transfected cell population (Figure **6.3**), which was consistently observed in the FACS profile, where GFP fluorescence intensity ranged across three magnitudes (Figure **6.4**).

Single cells were immediately seeded into a 96 well plate for clonal expansion to achieve a clonal cell population deriving from a single progenitor cell and cultured for 1-2 weeks for expansion.

6.3.1.3 CRISPR/Cas9 gene editing selection and validation of clonal populations

Clones originating from a single colony were subjected to genomic DNA isolation (from a portion of the colony) and end point PCR amplification. Successful CRISPR gene editing was

assessed based on generated amplicon size (Figure **6.5**A). PCR primers flanking regions upstream and downstream of both gRNA target sites generated an amplicon size of 551 bp. In case of successful gene editing an excision of around 232 bp should be observed, generating a shorter product size of only 319 bp. 18 clonal cell lines were analysed by PCR alongside a background wt control extracted from the untransfected parental cell line (Figure **6.5**). Among the 18 tested clonal cell lines, three (16 %) were homozygous for the shorter amplicon, indicating generation of a potentially full knock-out cell line. In contrast, 13 (72 %) were partial, heterozygous knock-out cell lines, where one of the two alleles was only modified. And two (11 %) remained unmodified and could be used as wt control cell lines for downstream analysis (Figure **6.5**). In addition, an empty vector transfected control cell population was expanded and included in subsequent functional assays.

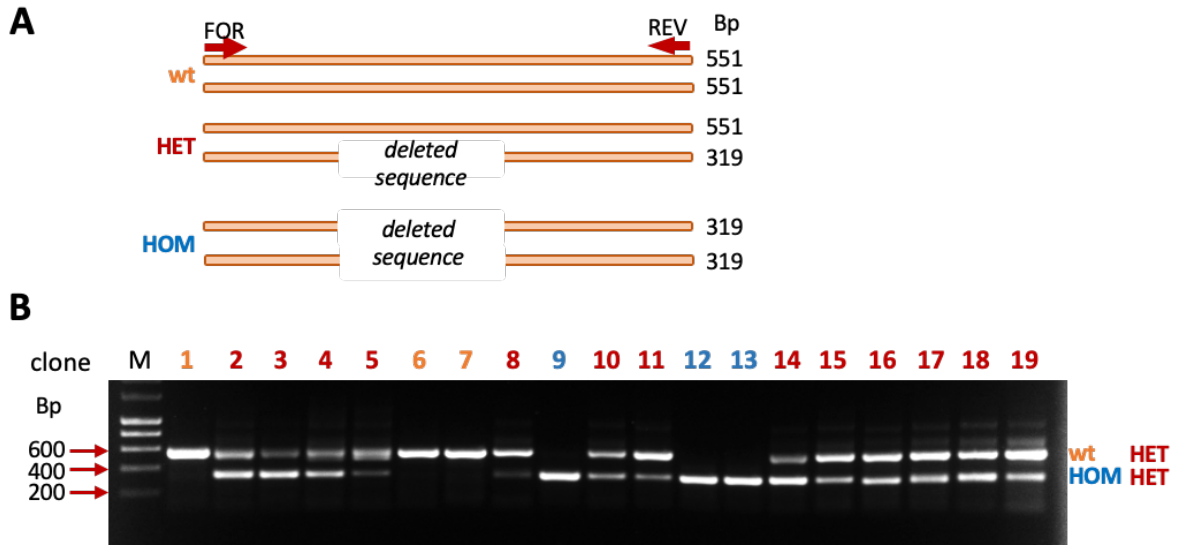


Figure 6.5: Genotyping of CRISPR edited clonal cell populations. A) Schematic of genotyping approach. A forward (FOR) and reverse (REV) primer pair was designed to amplify a product of 551 bp in length (wt, orange), whereas homozygous knock-out clones should result in a shorter amplicon size of 319 bp (HOM, blue). Both amplicon products are present in heterozygous edited clones (HET, red). **B)** Primer pairs flanking the sgRNA target region were used in a PCR reaction of genomic DNA isolated from the parental HBEC3-KT cells (1) or clonal populations of pX458-sgRNA derived pools (clone 2-19).

From the genotyping analysis, a homozygous knock-out clone (clone 13) and a corresponding wt clone (clone 7) was chosen for further validation steps. The PCR product of HOM clone 13 and wt clone 7 reactions were cloned into the pCRII-TOPO vector, amplified, extracted and sequenced (Figure **6.6**) to validate correct deletion of the targeted exon 2 without any acquisition of additional unwanted genomic alterations.

Sequencing data confirmed deletion of 232 bp including the entire exon 2 of clone 13, which was intact in clone 7 (Figure **6.6**).

Figure 6.6: Sequence chromatograms and alignment of wt clone 7 and homozygous *BPIFB1*$^{-/-}$ clone 13. CRISPR editing results in an excision of 232 bases (clone 13) compared to the original sequence (clone 7) **(A)**. Partial sequence chromatograms of the wt sequence (start indicated in orange) **(B)** or homozygous knock-out clone (HOM) 13 **(C)**, highlighted in pink shading.

In order confirm successful disruption of *BPIFB1* gene expression on protein level, western blot analysis of secreted BPIFB1 from ALI cultured cells alongside a control protein was carried out using wt (clone 7) and KO (clone 13) cell lines (Figure **6.7**). Employing antibodies targeting endogenous BPIFB1, a single protein band is detected around 55 kDa in the wildtype apical secretion. In contrast, the KO apical secretion did not yield any detectable protein band. To rule out that variation of total protein amounts loaded led to the absent band in the KO sample, Olfactomedin-4 (OLMF4) protein served as loading control and showed

comparable protein amounts across samples (Figure **6.7**A). OLFM4 was shown by proteomics to be abundantly secreted by HBEC3-KT cells (Figure **6.1**B). To confirm that the secreted wt BPIFB1 protein is glycosylated, PNGase treatment of the wt sample was carried out and showed that BPIFB1 is detected at the expected size of 52 kDa, rather than at 55 kDa in the sample left untreated (Figure **6.7**B).

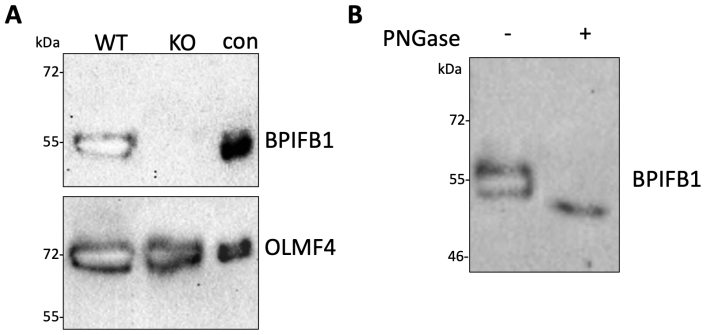


Figure 6.7: Western blot validation of BPIFB1 knockout cells. A) Secreted supernatant of HBEC3-KT clonal population 7 (WT) or knock-out 13 (KO) cells were resolved by SDS-PAGE alongside control supernatant from an unrelated apical secretion, and probed for presence of BPIFB1 or Olfactomedin-4 (OLMF4) using respective protein specific antibodies. OLMF4 served as an unrelated protein expression control indicating similar protein loading across samples. **B)** Clonal population 7 (WT) supernatant was incubated in the presence (+) or absence (-) of the Peptide-N-Glycosidase F (PNGase) enzyme and the endogenous protein detected using a BPIFB1 specific antibody.

Additional validation experiments were carried out to characterise the behaviour of the CRISPR clonal cell lines in an undifferentiated state to assess basal expression levels of selected genes in comparison with levels 14 and 21 days post ALI cultivation in differentiation media, to show that these cells still maintain the capability to differentiate into different cells types as expected for HBEC cultures (Vaughan *et al.*, 2006).

Maintaining cells in basal growth media showed that this does not induce *BPIFB1* expression (Figure **6.8**) and therefore the wt and KO cell line for *BPIFB1* should behave biologically comparable to the empty vector control clonal population when assessed in proliferation assays.

To assess basal expression patterns and to monitor stimulation of selected genes, cells were maintained a three-dimensional transwell setup at an ALI to initiate differentiation into various cell types and harvested after 14 and 21 days. Differentiation was assessed by endpoint PCR (Figure **6.8**). ALI initiated differentiation induced *BPIFB1* after 14 days, which was maintained at 21 days. *BPIFA1* served as an additional secretory cell marker and

showed a similar temporal pattern of expression. In contrast, the ciliated cell marker *TEKT1* was only be detected after 21 days at the ALI. *OAZ1* as a loading control was present at all time points at comparable levels. In summary, the observed gene expression patterns performed as expected, with no noticeable difference between the wt and KO CRISPR clonal cell line for the selected genes (Figure **6.8**). Notably, despite multiple optimisation attempts, it was not possible to design a primer pair for *BPIFB1,* which would amplify a product across exon 2 and thus would be able to distinguish the wt from the KO allele (Figure **6.2**). The primer pair employed in this study amplified a region of 207 bp across exon 5 to exon 8, thus wildtype and knock out CRIPSR clonal population result in a same sized amplicon of 207 bp (Figure **6.8**), leading to an RNA molecule but not resulting in protein expression.

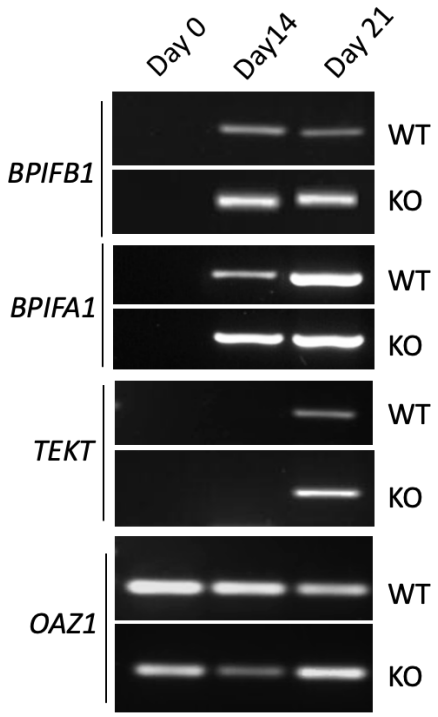


Figure 6.8: Analysis of selected gene expression upon HBEC3-KT differentiation. CRISPR generated wild type (clone 7 - wt) and knock out (clone 13 - KO) *BPIFB1* clonal cell lines were differentiated in ALI culture conditions simultaneously. Gene expression analysis of *BPIFB1* (207 bp), *BPIFA1* (466 bp), *TEKT1* (198 bp) and *OAZ1* (164 bp) was carried out by amplifying equal amounts of cDNA by endpoint PCR from cell cultures 0, 14, or 21 days post cultivation in a transwell ALI setup. *OAZ1* served as an internal control.

Assessing the CRISPR generated HBEC3-KT cells phenotypically, there was no noticeable difference between the parental cell population, the CRISPR modified cell lines, presented as empty vector control, *BPIFB1* wt (clone 7) or *BPIFB1* KO (clone 13) (Figure **6.9**). Of note,

in subsequent experiments analysing cell cycle progression or proliferative properties of these clonal populations for validation purposes, the empty vector control was used as a reference for comparisons to the wt or KO cell line.

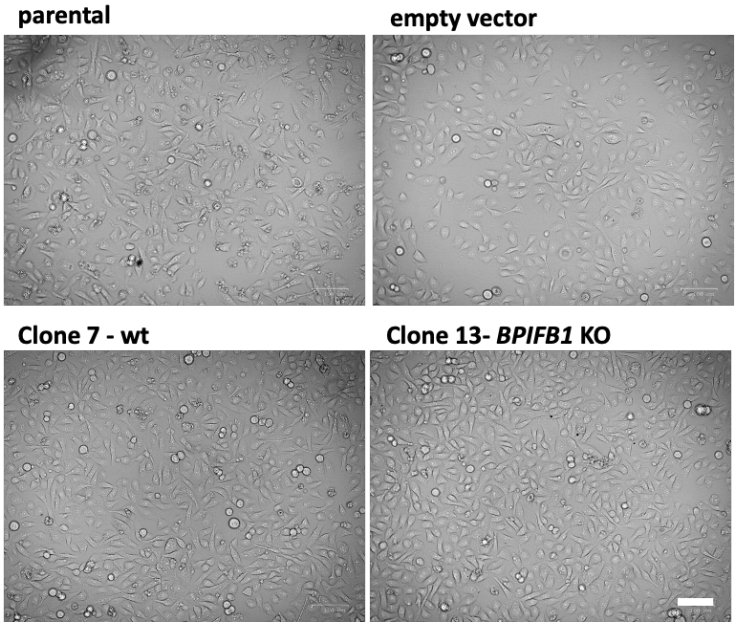


Figure 6.9: Morphology of HBEC3-KT parental and CRISPR generated clones. Parental, empty vector transfected (empty vector), wild type clone 7 (wt) or *BPIFB1* knock out clone 13 (KO) undifferentiated HBEC3-KT cells were maintained submerged in KSFM for three days prior brightfield imaging. Representative images are shown. The bar indicates 100 μm.

To validate the CRISPR generated clonal cell populations with regard to off target effects, which could hamper subsequent assays for BPIFB1 function, cell cycle analysis and an MTS based proliferation assay was conducted. It was expected that the empty vector control and the wt clone 7, as well as the KO clone 13, would behave comparably.

6.3.1.4 Cell cycle analysis of CRISPR clones

Based on *BPIFB1* expression analysis, undifferentiated HBEC3-KT cells do not express the *BPIFB1* gene (Figure **6.8**) and therefore no physiological effect of BPIFB1 loss in these cells were anticipated.

HBEC3-KT cell lines were cultured in Keratinocyte Serum-Free media (KSFM) to maintain an undifferentiated phenotype and were subjected to propidium iodide (PI) staining for cell cycle profiling using FACS. In order to quantitate the percentage of cells in each cell cycle

phase, a single cell population was identified and gated by plotting pulse width against pulse area. This population was then presented as scatter plot to remove visible debris and then represented as a histogram plot over PI intensity (Figure **6.10**A) (Pozarowski and Darzynkiewicz, 2004).

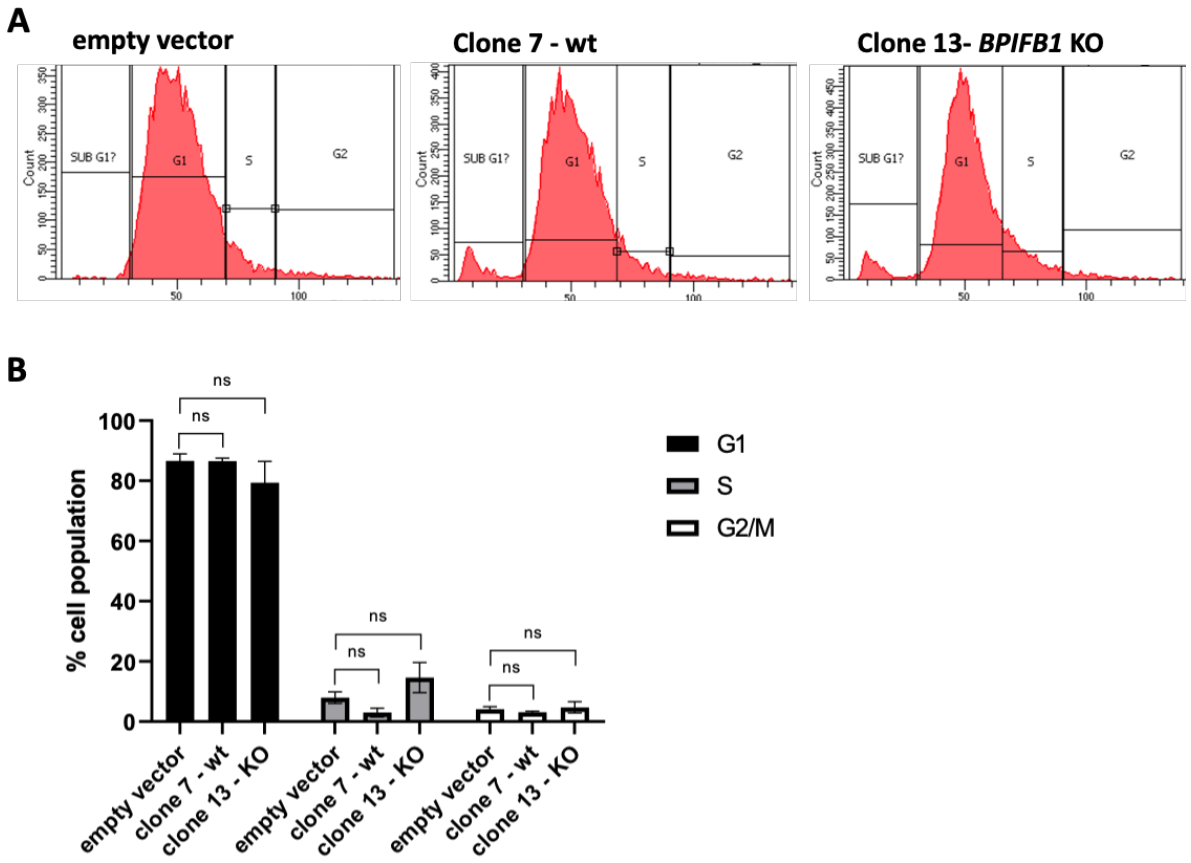


Figure 6.10: Cell cycle analysis of CRISPR clones. Undifferentiated HBEC3-KT empty vector transfected (empty vector), wild type clone 7 (wt) or *BPIFB1* knock out clone 13 (KO) cells were passaged in KSFM to maintain an undifferentiated phenotype and cell cycle phases analysed by means of flow cytometry. **A)** Representative gated FACS histogram profiles displaying cell count over propidium iodide intensity for quantification of cell cycle phase distribution. **B)** Quantification of cell populations in G1, S Phase or G2/M, represented as mean ± SD of three technical replicates. ns – non-significant.

In summary, all three clonal cell lines performed comparably with no significant change in cell cycle phase populations (Figure **6.10**B). Notably, it was observed that the G1 population was lower in the KO cells compared to empty vector control or wt conditions, and increased in the S phase, however this result was not significant due to a higher data variability.

6.3.1.5 Cell proliferation analysis of CRISPR clones

To validate cell proliferation and viability of the CRISPR generated, undifferentiated HBEC3-KT cell lines, an MTS assay was employed. This assay is based on the reduction of the MTS tetrazolium compound through viable mammalian cells, which generates the soluble, coloured formazan dye, and can be recorded by means of a standard colorimetric read out.

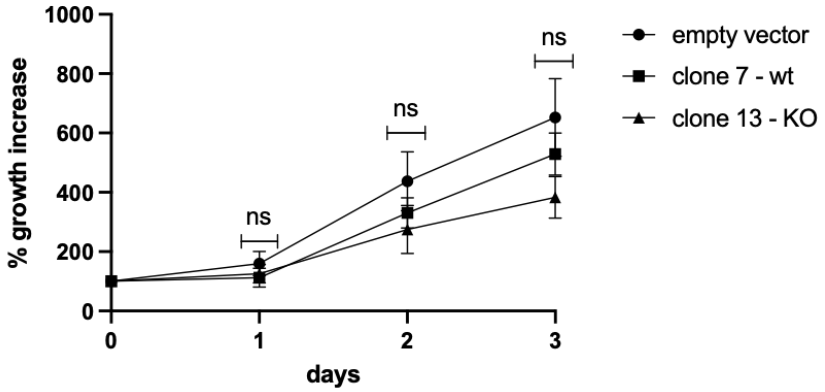


Figure 6.11: Proliferation analysis of CRISPR clones. Undifferentiated HBEC3-KT empty vector transfected (empty vector), wild type clone 7 (wt) or *BPIFB1* knock out clone 13 (KO) cells were passaged in KSFM to maintain an undifferentiated phenotype and cell viability and proliferation assessed over three days using an MTS colorimetric readout in 24h intervals. Data is represented as mean ± SD of three independent experiments. ns – non-significant.

It was observed that all three clonal cell lines behave comparably with no significant change in proliferation across an observed time frame of 72h (Figure **6.11**). However, it might should be investigated if the KO cell line leads to a delayed proliferative phenotype over longer periods of culturing, and thus further experimental work will be required.

6.3.2 Functional assays

After extensive quality control and validation studies of the CRISPR clones, experiments were set up to analyse the protective effect of BPIFB1 following IAV infection. To establish an appropriate experimental setup, undifferentiated HBEC3-KT cells were infected with the influenza A strain A/Puerto Rico/8/1934, referred to as PR8, co-expressing GFP for visual confirmation of successful viral entry. Studies were carried out at different MOIs and observed over time for infection efficiency. In addition, two culture media were investigated for suitability to maintain HBEC3-KT cells in an undifferentiated state.

6.3.2.1 PR8 infection time course in undifferentiated HBEC3-KT cells

HBEC3-KT cells are usually maintained and expanded in KSFM, supplemented with 50 mg/L prequalified human recombinant Epidermal Growth Factor 1-53 (EGF 1-53, 50 µg/ml) and Bovine Pituitary Extract (BPE) at 5 ng/ml.

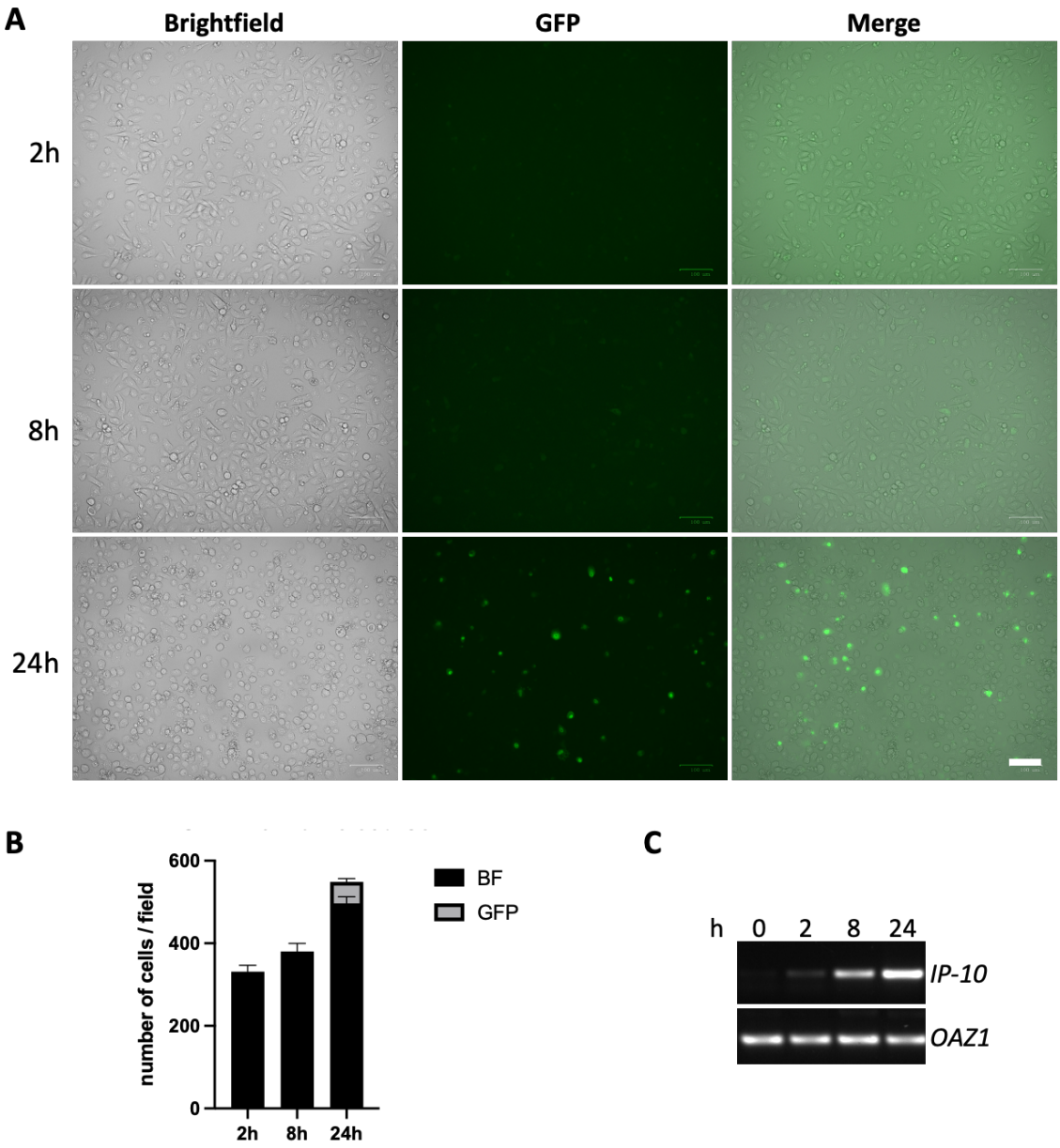


Figure 6.12: Time course of *IP-10* expression upon PR8 viral infection in KSFM maintained cells. Undifferentiated parental HBEC3-KT were cultured in KSFM prior PR8 viral infection at a MOI ratio of 1. **A)** Cell morphology and virus induced GFP co-expression was monitored by brightfield microscopy or the green-fluorescent channel (GFP), respectively, 2, 8 and 24 hours post infection. **B)** Bar graph indicates detected cells per brightfield area (BF) in A and corresponding green-fluorescent positive cells (GFP). **C)** End point PCR analysis of isolated RNA, 2, 8 and 24 hours post viral infection was used to assess *IP-10* (190 bp) levels, an indicator of viral infection. *OAZ1* (164 bp) served as internal loading control. Bar indicates 100 µm.

To test how well undifferentiated HBEC3-KT cells could be infected with the PR8 strain, an MOI of 0.1 and 1 was chosen, and cells monitored over the next 24 hours phenotypically for green-fluorescent protein expression, indicative of viral entry and replication. Further, RNA was extracted and expression of IFN-gamma-inducible protein 10 (*IP-10)* at 2h, 8h and 24h post viral infection (pvi) assessed by end point PCR (Figure **6.12**).

GFP expression was not visible at 2h and 8h pvi but could be detected 24h pvi using an MOI of 1 (Figure **6.12**A). Notably, only a small subpopulation of about 10% of cells showed GFP expression (Figure **6.12**A and B). This might be also attributed to the fact that master PR8 viral aliquots were thawed, aliquoted and refrozen prior use for experimental conditions. This might have led to reduction of the actual viral titre compared to the established one upon virus production, which was used for calculation of required viral volumes.

Infection conditions using an MOI of 0.1 resulted in about 1% fluorescent cells 24 h pvi, consistent with the 10-fold higher number associated with MOI of 1 (data not shown). As this is a negligible number of cells and a biological effect might be masked by the overriding number of non-fluorescent, uninfected cells, the data acquired with MOI of 0.1 therefore was not considered in the analysis and in further experiments.

It has been established that *IP-10* expression is stimulated in airway epithelial cells upon infection with respiratory viruses (Ramos *et al.*, 2019). Endpoint PCR analysis for *IP-10* expression showed that elevated levels could be detected as little as 2h pvi, and the highest level was detected at 24h pvi (Figure **6.12**C).

The other media tested was the defined, commercially available PneumaCult™-Ex Plus (referred subsequently as ExPlus) expansion media, which in contrast to KSFM, is serum- and BPE-free, and is complemented with a non-proprietary supplement and 96 ng/mL hydrocortisone.

Similar to KSFM (Figure **6.12**), the fluorescent positive cells population is approximately 10% and could be observed at 24h pvi (Figure **6.13**A and B). *IP-10* expression is detected at 2h pvi at a higher levels compared to KSFM (Figure **6.13**C vs Figure **6.12**C), which peaks at 8h (Figure **6.13**C) compared to 24h in the KSFM condition (Figure **6.12**C).

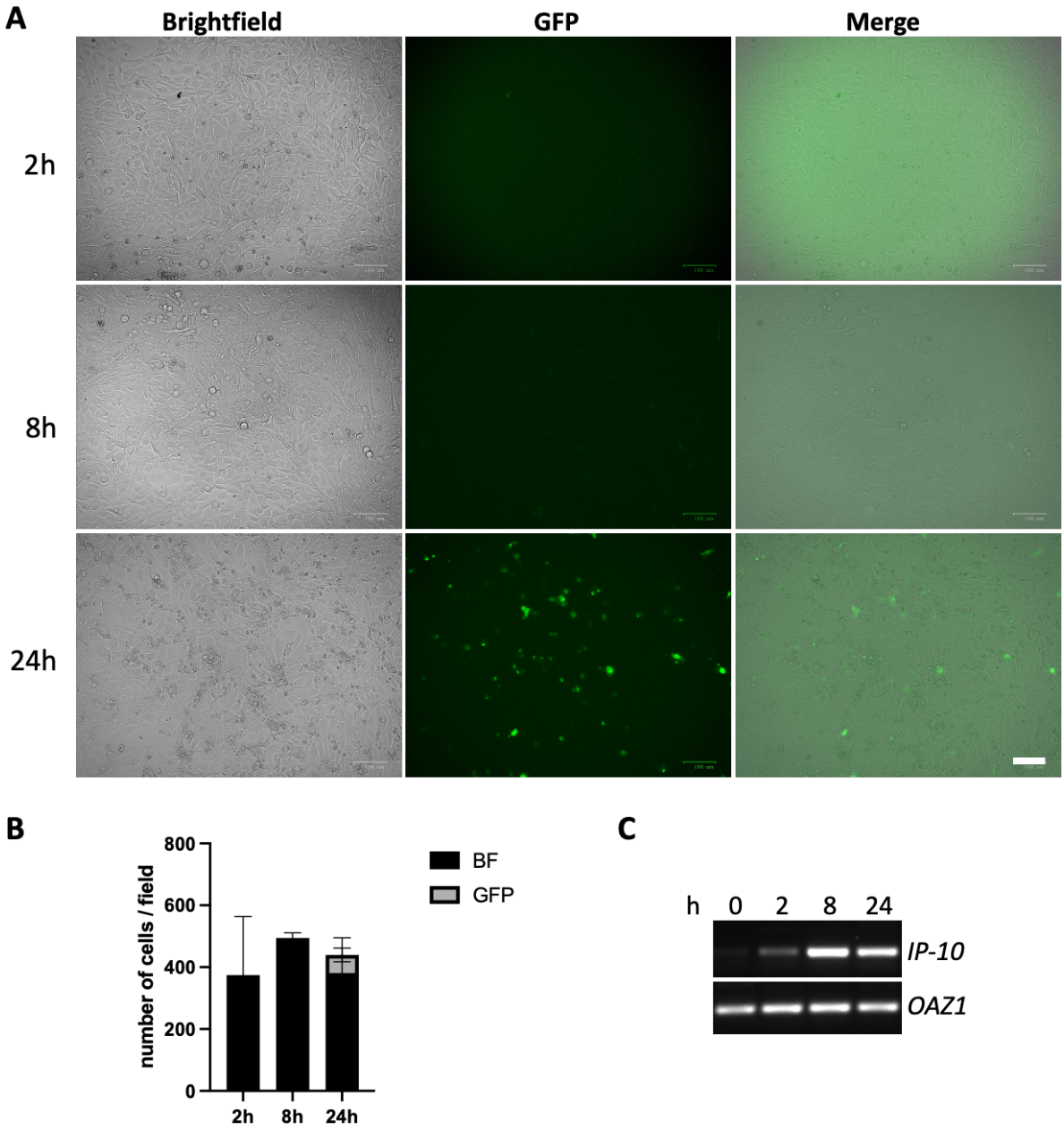


Figure 6.13: Time course of PR8 viral infection in ExPlus media-maintained cells. Undifferentiated parental HBEC3-KT cells were cultured in ExPlus media prior PR8 viral infection at a MOI ratio of 1. **A)** Cell morphology and virus induced GFP co-expression was monitored by brightfield microscopy or the green-fluorescent channel, respectively, 2, 8 and 24 hours post infection. **B)** Bar graph indicates detected cells per brightfield area in A and corresponding green-fluorescent positive cells. **C)** End point PCR analysis of isolated RNA, 2, 8 and 24h post viral infection was used to assess *IP-10* (190 bp) levels, an indicator of viral infection. *OAZ1*(164 bp) served as internal loading control. Bar indicates 100 μm.

In order to assess if PR8 infection leads to a differential cytokine expression following 24 hours post PR8 infection, a commercial cytokine array was employed comparing both media - KSFM and ExPlus – in a basal condition and post 24 h PR8 challenge.

6.3.2.2 Cytokine production upon PR8 exposure in HBEC3-KT cells

After a 24h challenge of PR8 virus at MOI of 1, albeit the infected number of cells are only constituting a population of 10 % (Figure **6.12**A and B and Figure **6.13**A and B), a highly significant increase of IP-10 by 180 fold ($p<0.0001$) in KSFM cultured cells versus 129 fold ($p<0.0001$) in ExPlus conditions was detected (Figure **6.14**C and E). This was verified using endpoint PCR analysis, which showed increased *IP-10* mRNA levels as soon as 2h post infection in ExPlus media conditions (Figure **6.14**F) or 8 h in KFSM conditions (Figure **6.14**D). This increase further correlates with detected protein levels of IP-10 on the array, which is 5480 (± 20) in KFSM versus 10583 (± 677) in ExPlus basal media, highlighting increased detection by endpoint PCR in ExPlus samples (Figure **6.14**D and F).

In addition, ExPlus media results in a higher basal level of CXCL1 (9.2 fold), G-CSF (2.1 fold) and IL-18 (23.1 fold) compared to KSFM (Figure **6.14**C and E). Notably, the protein levels of CXCL1 ($p=0.0157$), G-CSF ($p=0.0016$), IL-8 ($p=0.0193$) and IL-18 ($p=0.0292$) only in ExPlus maintained cells led to a significant change when challenged with the PR8 Influenza A strain compared to mock control treated cells (Figure **6.14**E). Inversely, macrophage migration inhibitory factor (MIF) was detected in both media conditions but is only reported significantly changed ($p=0.0202$) upon PR8 infection in KSFM but not ExPlus media (Figure **6.14**C and E). In contrast, serine protease inhibitor serpin family E member 1 (SERPIN E1) is present in high levels in both media tested but not significantly altered upon PR8 infection in both cases (Figure **6.14**C and E).

In order to establish if KSFM or ExPlus media is more suitable to maintain and expand cells in an undifferentiated state for subsequent experiments, a direct comparison showed that there is no significant difference in terms of PR8 infection efficiency (Figure **6.15**A and B). However, analysis by cytokine secretion (Figure **6.14**) or gene expression on mRNA level (Figure **6.15**C) showed that ExPlus media displays low level differentiation properties as observed significantly enhanced cytokine response upon PR8 challenge, or *BPIFA1* expression stimulation compared to KSFM but not *BPIFB1* or *TEKT* in this condition (Figure **6.15**C). This indicates that the ExPlus media contains one or more unidentified components,

which potentially stimulates HBEC3-KT cells and therefore, for all subsequent experiments KSFM was utilised.

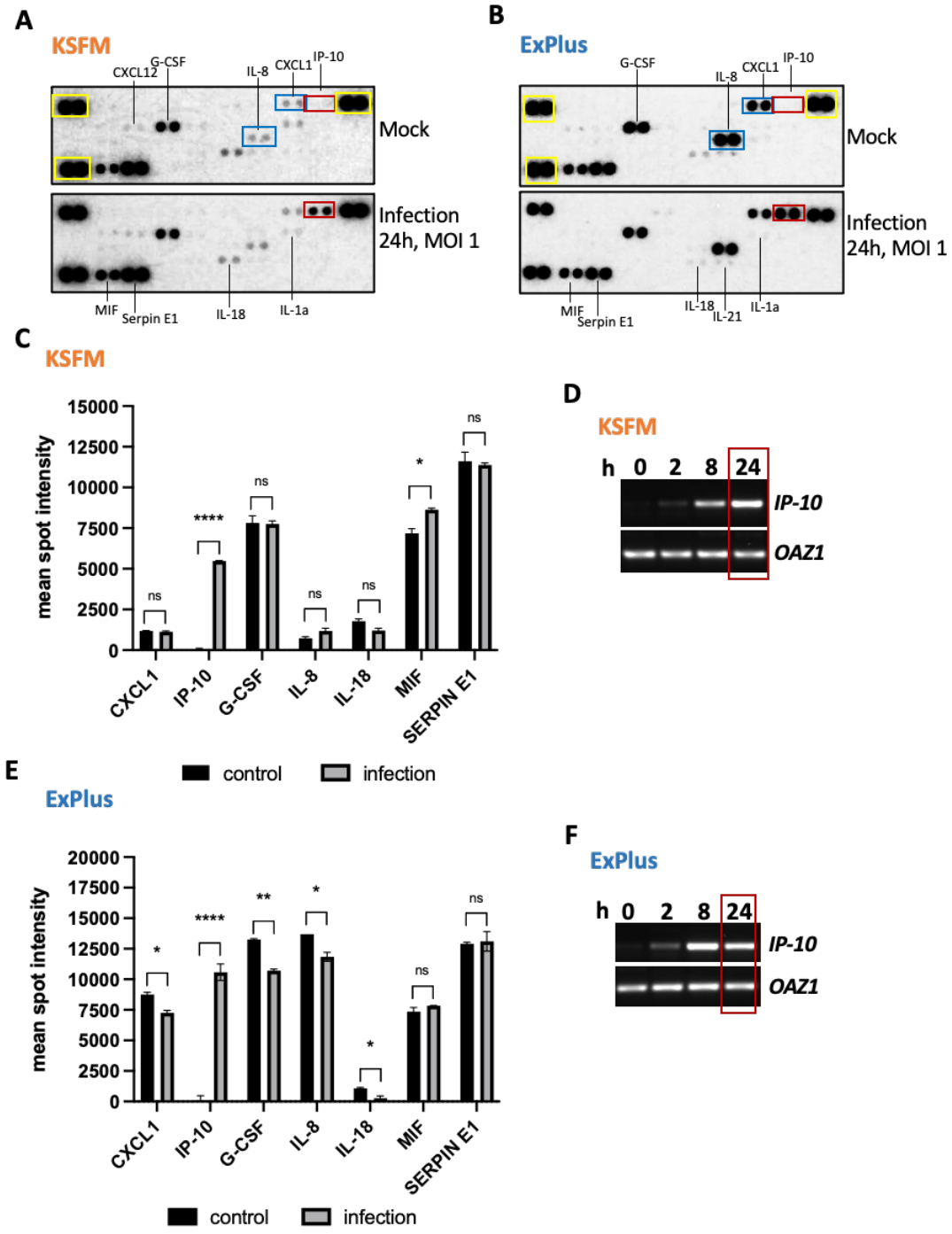


Figure 6.14: Human cytokine array to detect relative levels of selected cytokines and chemokines. Undifferentiated HBEC3-KT cells were cultured in a well format using KFSM or ExPlus media and subsequently infected with the PR8 strain at MOI 1 (infection) or left untreated (mock). After 24h viral challenge, KSFM **(A)** or ExPlus media **(B)** was collected and applied to a cytokine array membrane to detect differential secreted immune mediators detected and quantified by densitometry. **C, E**) Quantification and statistical analysis of detected cytokines. **D, F**) End point PCR analysis of isolated RNA, 2, 8 and 24h post viral

infection was used to assess *IP-10* (190 bp) levels. *OAZ1* (164 bp) served as internal loading control. Red square indicates highest differential expressed cytokine within each media condition, whereas blue squares highlight different basal cytokine levels between KFSM and ExPlus media. Yellow squares indicate array specific reference points.

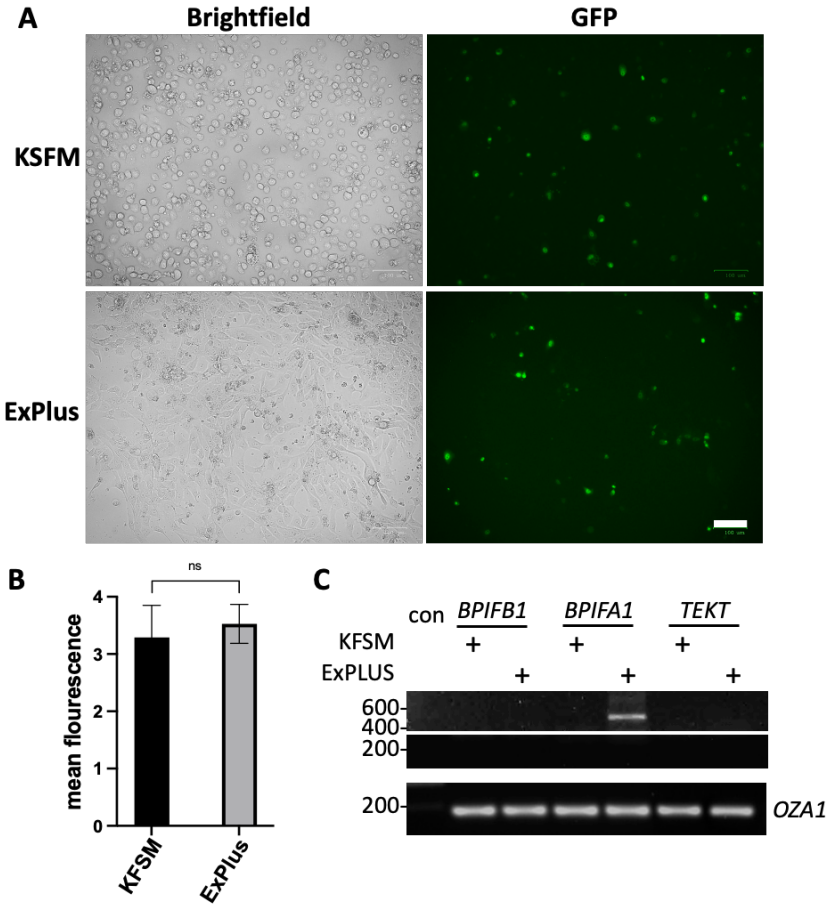


Figure 6.15: Comparison of KSFM versus ExPlus media upon HBEC3-KT PR8 infection and cellular differentiation. **A)** Phenotypical observation of HBEC3-KT cells maintained in indicated media, 24h post infection at MOI 1, by brightfield and green fluorescence. **B)** Quantification of mean fluorescence per recorded field. **C)** Endpoint PCR analysis of *BPIFB1* (207 bp), *BPIFA1* (466 bp) and *TEKT* (198 bp). *OZA1* (164 bp) served as internal control. ns – non-significant. Bar indicates 100 μm.

After establishing an experimental system which allows for PR8 challenge and read out of infection efficiency, CRISPR generated clonal cell population expressing wild type BPIFB1 (clone 7) or the knockout counterpart (clone 13) (Figure **6.7**) were used to generate well-differentiated cell populations at the ALI. These established cultures were then used to collect lavage specimen to be employed in subsequent PR8 infection experiments.

6.3.2.3 Effect of CRISPR clone derived lavage treated HBEC3-KT upon PR8 infection

In order to understand if BPIFB1 confers a protective role against IAV, undifferentiated HBEC3-KT cells were incubated with wild type (clone 7) or knock out (clone 13) ALI lavages and infected with PR8 virus at MOI 1 (Figure **6.16**).

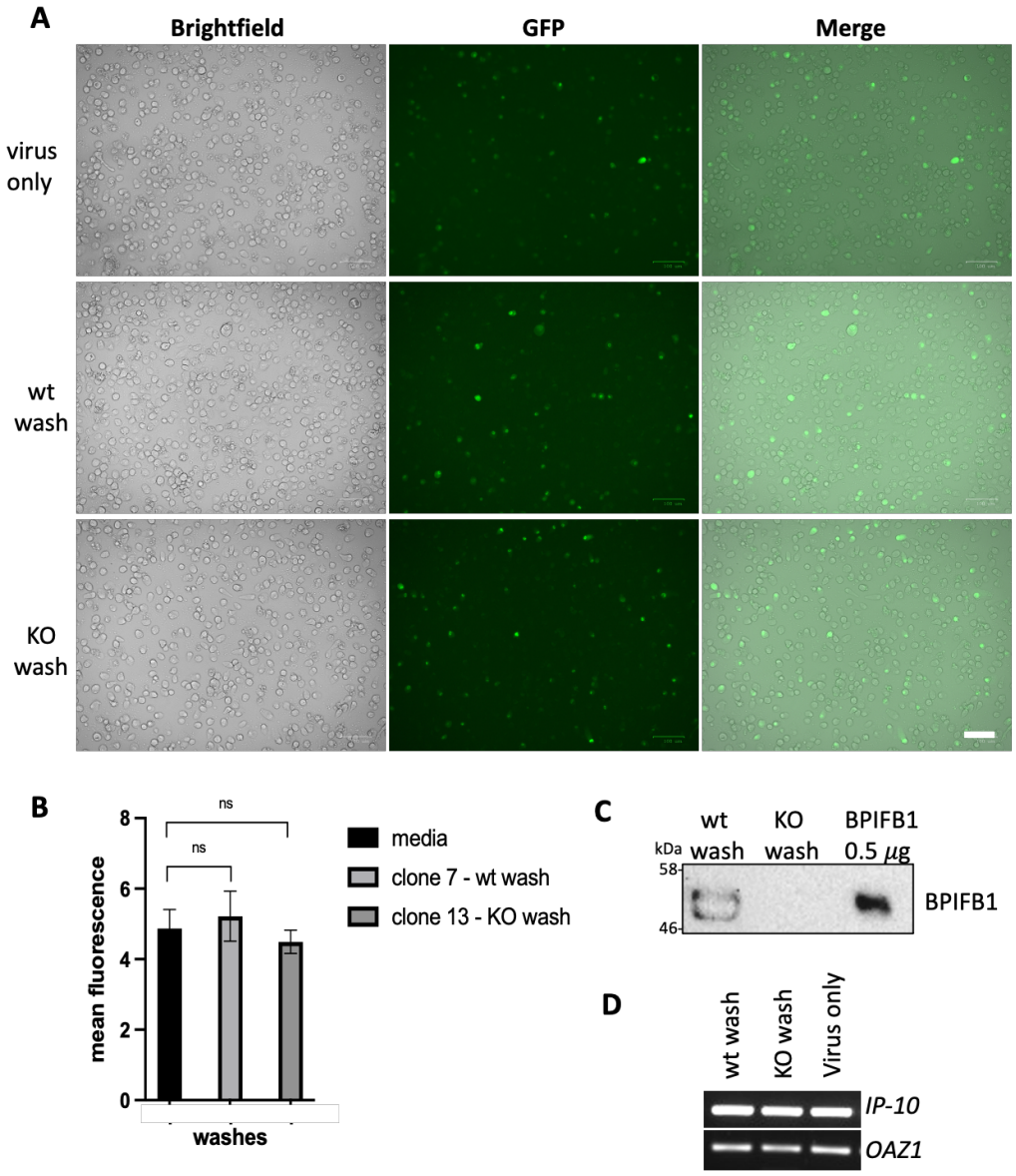


Figure 6.16: Effect of BPIFB1 wt or knock out conditioned lavage on PR8 viral infection efficiency. Undifferentiated HBEC3-KT cells were cultured in KSFM and CRISPR generated wt clone 7 lavage or BPIFB1 knockout clone 13 lavages applied at a concentration of 250 µg/ml total protein. Cells were challenged with PR8 virus at an MOI of 1. **A)** Cells were assessed microscopically by brightfield and green-fluorescent imaging 24h post infection. **B)** Plotted quantification of mean green fluorescence of representative fields of A. **C)** Western blot analysis of BPIFB1 levels in lavages or unconditioned media. **D)** Endpoint PCR analysis of *IP-10* (190 bp) expression across lavage treated and infected cells or virus only treated

(unconditioned media) cells. *OZA1* (164 bp) served as internal control. Data from three independent experiments, ns-non-significant. Bar indicates 100 μm.

After 24 h incubation, cells were assessed microscopically and infection efficiency of PR8 quantified by mean fluorescence intensity across all conditions (Figure **6.16**). Quantification of infection levels by fluorescence indicated that there is no significant difference between the wt and knock out lavage, which were comparable to only PR8 virus infected conditions (Figure **6.16**B). To assess that the wt lavage contains BPIFB1 protein, which should be absent in the knock out condition as established earlier (Figure **6.7**), western blot analysis confirmed presence of BPIFB1 in the wt lavage only (Figure **6.16**C). Consistent with the observation that wt versus knock out lavage did not lead to a significant infection differential, endpoint RNA analysis confirmed comparable *IP-10* expression levels, similar across all conditions (Figure **6.16**D). This would suggest that BPIFB1 does not have protective properties against IAV infections, however it has not been established if the lavage conditions contained sufficient BPIFB1 protein to have a significant biological effect. Due to time constrains, it was decided not to undertake further experiments to address lavage-based optimisations and purchase recombinant BPIFB1, which could be supplemented to media upon PR8 infection in a dose dependent manner.

6.3.2.4 Effect of recombinant BPIFB1 treated HBEC3-KT upon PR8 infection

Recombinant BPIFB1 was commercially obtained and supplemented to HBEC3-KT cells at various concentrations during PR8 infection. After 24 h, cells were observed microscopically and infection efficiency of PR8 quantified by mean fluorescence intensity across all protein concentrations (Figure **6.17**A). Quantification of infection levels by fluorescence indicated that there is a negative correlation between amount of present recombinant BPIFB1 and number of infected cells, showing a significant effect from 0.36 μg/ml across all four tested amounts to the highest level at 2.86 μg/ml (Figure **6.17**B). To test for presence of the recombinant protein in the media and stability across the incubation period over 24 h, western blot analysis was carried out, probing for BPIFB1. As expected, the protein shows a similar migration pattern as observed for the endogenous, glycosylated moiety (Figure **6.7**A), which notably decreases by about 50% over the course of 24 h (Figure **6.17**C). This indicates that the protein stability is affected to a certain degree over the incubation period, but this still results in a significant protective response towards PR8 infection. This effect correlates well

with viral induced expression of *IP-10*, which decreased significantly at higher BPIFB1 concentrations (Figure **6.17**D).

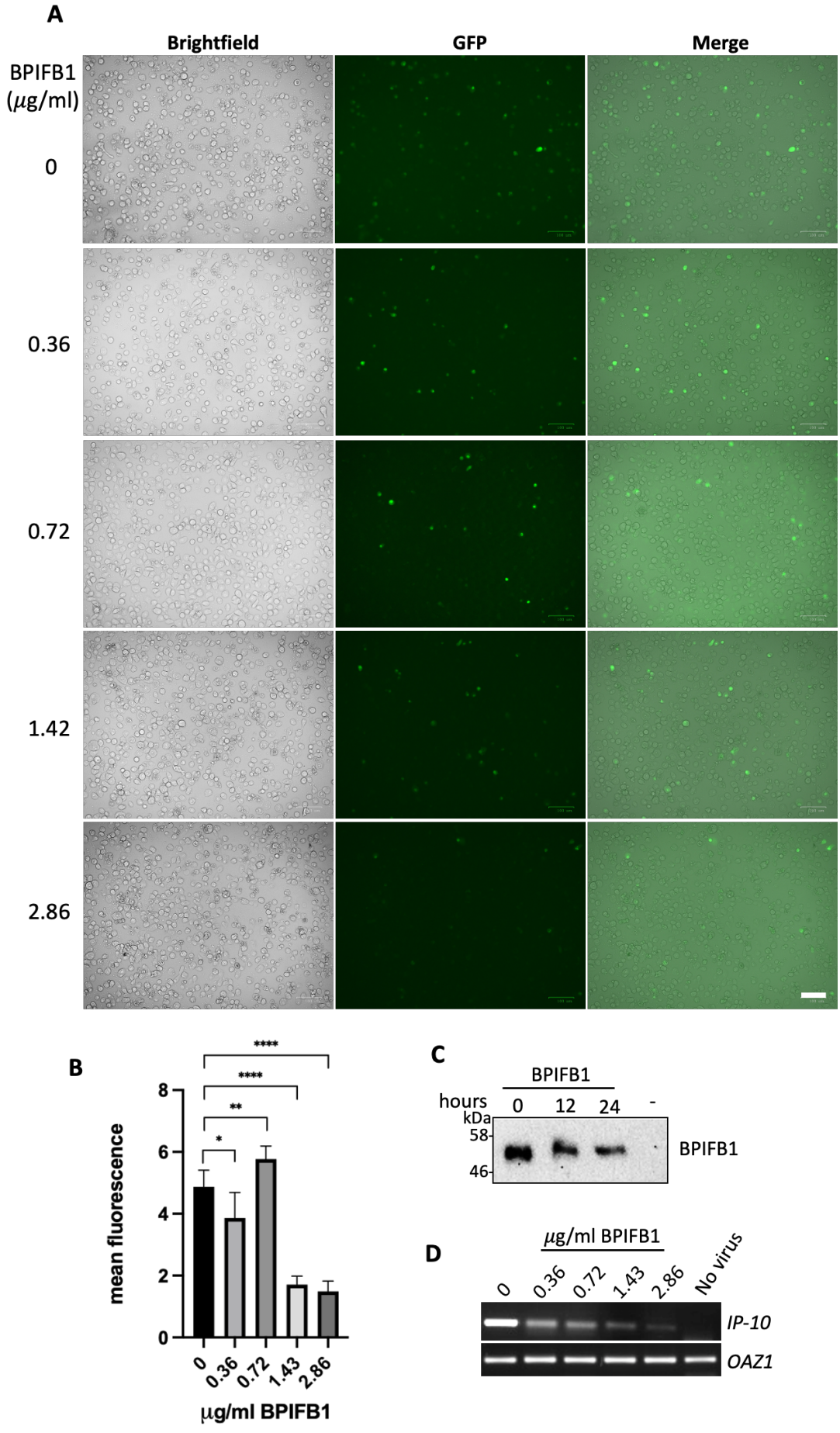

A
Brightfield
GFP
Merge
BPIFB1 (μg/ml)
0
0.36
0.72
1.42
2.86
B
mean fluorescence
μg/ml BPIFB1
C
BPIFB1
hours
0
12
24
-
kDa
58
46
BPIFB1
D
μg/ml BPIFB1
0
0.36
0.72
1.43
2.86
No virus
IP-10
OAZ1

Figure 6.17: Effect of recombinant BPIFB1 dose response on PR8 viral infection efficiency. Undifferentiated HBEC3-KT cells were cultured in KSFM and PR8 infected at MOI 1. **A)** Representative brightfield and green-fluorescent images were mages taken 24 h post infection. **B)** Analysis of fluorescence intensity plotted as mean intensity per field of cells across different concentrations of recombinant BPIFB1 as indicated. Data are presented as the mean ± SD of three independent experiments, *p=0.0309, **p=0.0095, ****p<0.0001 vs. no protein control (0). **C)** Western blot analysis of diluted recombinant BPIFB1 in KSFM (0 - 2.86 μg/ml) incubated at cell culture conditions for 24 h. Media only (-) served as negative control for BPIFB1 detection. **D)** Endpoint PCR analysis of *IP-10* (190 bp) expression across infected cells incubated with recombinant BPIFB1 protein concentrations as indicated or uninfected cells only (no virus). *OZA1* (164 bp) served as internal control. Data from three independent experiments, ns-non-significant. Bar indicates 100 μm.

At the highest tested BPIFB1 concentration of 2.86 μg/ml, hardly detectable IP-10 levels were recorded, similar to uninfected conditions (Figure **6.17**D). This result suggests a protective effect of BPIFB1 in IAV infections for the first time. To further investigate this observation CRISPR clonal populations of differentiated BPIFB1 wt or KO cells are needed to be PR8 infected and viral uptake visualised by confocal microscopy. However, due to time constrains these experiments could not be undertaken.

6.4 Discussion

Due to COVID-19 restrictions, the initial strategy of analysing a potential immune supportive role of BPIFB1 during IAV infections of mTEC cells could not be carried out due to unavailability of the required mice. Thus, an alternative approach was developed, which allowed the development of an assay for a subsequent assessment of a role of BPIFB1 during IAV challenge. The well-established CRISPR method was employed to generate stable *BPIFB1* knock out cell lines of human origin. This technology presents a powerful tool for high precision genome alteration, despite being simple and efficient (Cong *et al.;* Ran *et al.*, 2013c; Mali *et al.*, 2013; Lino *et al.*, 2018). During the last decade, this technology has been successfully applied to countless organisms, microorganisms and cell lines, including primary cells (Gundry *et al.*, 2016). The genome editing technology relies on the DNA repair machinery, activated by the double strand break introduced by Cas9. Upon incision it is possible to introduce new genetic information, however the DNA repair system is not designed to integrate DNA fragments in the genome, and thus targeted alleles often carry additional modifications, such as deletions, duplications, partial or multiple integrations of the targeting vector (Li *et al.*, 2015; Pavlovic *et al.*, 2016). With gene editing approaches comes the danger of off-target effects or modifications, and efforts have been made to reduce this effect by employing for example a Cas-9 nickase (Cas9n), a variant that induces single-stranded breaks (SSBs), rather than double strand breaks in the intended location through a combination of a sgRNA pair targeting both strands of the DNA (Ran *et al.*, 2013b; Uddin *et al.*, 2020). Other approaches resulted in a higher target specificity with no detectable off target activity employing the SpCas9-HF1 protein, which has been generated by modification of four residues involved in direct hydrogen bonding between Cas9 and the phosphate backbone of the target DNA (Kleinstiver *et al.*, 2016).

To study pathogen-host interactions of bronchial epithelia, primary cells are usually sourced in order to establish an *in vitro* model of airway epithelia that allows for differentiation and subsequent functional assays. Studies are typically performed on differentiated cells, which form a pseudo-stratified epithelium containing ciliated and secretory cells, upon air-liquid interface culture conditions for up to 28 days (Fulcher *et al.*, 2005; Bhowmick and Gappa-Fahlenkamp, 2016).

Although primary cells have been successfully modified using the CRISPR technology, including adaptations such as transfection of Cas9 protein pre-complexed with guide RNA in

order to achieve superior editing efficiencies and to reduce toxicity, experimental studies are very limited due to the restricted life span of primary bronchial epithelial cells (Chu *et al.*, 2015; Peters-Hall *et al.*, 2018; Everman *et al.*, 2019; Rapiteanu *et al.*, 2020). It is worth mentioning that methods have been developed which extend the lifespan of bronchial epithelial cells, whilst maintaining the normal cellular phenotype and transcriptome integrity through increased expression for example of the polycomb group protein BMI-1 (O'Loughlin *et al.*, 2021).

However, due to the short cellular lifespan, and to generate a larger pool of homogenous cell material for subsequent experiments, we chose to use the HBEC3-KT as model system for BPIFB1 ablation. While these cells are immortalized with CDK4 and hTERT, they retain characteristics of primary cells and do not display immortalised cell phenotypes such as disruption of the p53 pathway, extensive copy number changes or lack of contact-inhibition (Ramirez *et al.*, 2004; Sato *et al.*, 2006). In addition, immortalized human bronchial epithelial cells retain the ability to differentiate into normal bronchial epithelium (Vaughan *et al.*, 2006; Delgado *et al.*, 2011) and have been used extensively in our lab over the past 3 years.

The performed approach of two gRNAs releasing a specific genomic region including the exon 2 start codon allowed for successful gene knock out in clonal populations. Validation experiments confirmed that the KO cells do not differ phenotypically or by functional assays from the empty vector or wt counterpart. These validations demonstrated that BPIFB1 was not detectable at protein level in differentiated cells, and cells cycle and proliferate at levels comparable to control clonal cell lines such as the empty vector or parental line.

For our infection studies and establishing the HBEC3-KT based system to analyse BPIFB1 biological function *in vitro*, we choose to utilise the A/Puerto Rico/8/34 (PR8) (H1N1) strain over the previously tested A/HKx31 (X31) (H3N2) (section 4.3.3.1) on mTEC cells. The X31 virus contains the six internal genes of PR8 but expresses H3N2 surface proteins, whereas PR8 expresses H1N1 surface proteins (Kilbourne, 1969; Lamb and Krug, 1996). Askovich and colleagues carried out a study assessing PR8, X31 VN1203 IAV strains in a murine model with respect to viral growth rates and host transcriptional response (Askovich *et al.*, 2013). Although all strains demonstrated a similar replication rate over the first viral life-cycle, this strongly divided at later times, and separate networks of genes were discovered at early time points by specific strains (Askovich *et al.*, 2013).

It was observed that PR8 is able to infect HBEC3-KT cells, similar to levels as observed for the X31 strain on mTEC cells (chapter 4). Most studies are carried out on differentiated cells, which need to be ALI grown for 14-28 days. Being able to use undifferentiated cells for infection susceptibility studies would simplify and shorten experimental setups tremendously.

At a calculated MOI of 1, an infected cell population of 10% was observed. It was not possible to extract more information which specific cell type(s) were targeted as a co-staining for goblet or ciliated cell markers was not carried out. However, this could be performed in future experiments.

Influenza viruses target glycosylated glycan oligosaccharides that terminate in sialic acid (SA) residues (Rogers and Paulson, 1983; Matrosovich *et al.*, 1997), through α2,3, α2,6, or α2,8 linkages mediated by sialyltransferases, which are expressed in a cell- and species-specific manner (Angata and Varki, 2002; Gagneux *et al.*, 2003). IAV susceptibility, tropism transmissibility and clinical symptoms have been attributed to preferences in HA binding to specific membrane SA - galactose (Gal) molecules residing on epithelial cell surfaces (Connor *et al.*, 1994; van Riel *et al.*, 2007). The X31 (H3N2) strain has been reported to possess binding specificity for SAα2,6Gal, whereas PR8 (H1N1) is able to utilize two types of oligosaccharides receptors, SAα2,6Gal and SAα2,3Gal. Earlier studies demonstrated that influenza strains efficiently infect human airway epithelia through preferential binding to SAα2,3Gal receptors, despite reports showing that tracheal epithelia express sialyloligosaccharides with both SAα2,3Gal and SAα2,6Gal linkages (Baum and Paulson, 1990; Couceiro *et al.*, 1993; Slepushkin *et al.*, 2001). Goblet cells and respective mucus droplets stained with lectins specific for SAα2,3Gal, while SAα2,6Gal linkages were found to be specific to ciliated cells (Baum and Paulson, 1990; Couceiro *et al.*, 1993). In contrast, X31 has been identified to bind to the apical surface of airway epithelia, but was unable to enter the cells, based on minimal evidence of membrane fusion and inability of virus production (Slepushkin *et al.*, 2001) and therefore might not be a suitable virus for our studies on HBEC3-KT cells.

Innate immune cells, in particular myeloid cells, play a vital role in containing influenza infections (McGill *et al.*, 2009) and neutrophils can clear the virus *in vitro* and *in vivo* (Tecle *et al.*, 2007; Fujisawa, 2008). This has been established from observations that depletion of either alveolar macrophages or neutrophils prior sublethal infection with a pandemic virus resulted in uncontrolled viral replication and increased mortality (Tate *et al.*, 2009). However,

myeloid cells can be detrimental to the host, contributing to morbidity and mortality through inflammatory injury (Kobasa *et al.*, 2004; Kash *et al.*, 2006). Infection by highly pathogenic strains, such as the 1918 H1N1 and avian H5N1, caused immense recruitment of neutrophils and monocytes and macrophages (Guan *et al.*, 2004; Perrone *et al.*, 2008). A servere complication encountered during influenza infection is contracting viral pneumonia, which can result in acute respiratory distress syndrome (ARDS). In this case, infected epithelial cells produce cytokines that attract and activate neutrophils and macrophages, leading to damage to the epithelial-endothelial barrier, hampering gas exchange (Ito *et al.,* 2015).

Cytokines and chemokines need to be well balanced during influenza infection as they are critical for antiviral protection but could lead to a source of inflammatory pathology if not well regulated (De Jong *et al.*, 2006). For example, a high-pathological acute influenza virus infection is associated with a dysregulated $CD8^+$ T cell response, likely caused by the highly inflamed airway microenvironment during the initial days of infection (Rutigliano *et al.*, 2014).

The pro-inflammatory cytokine MIF mediates the innate immune response to bacterial pathogens due to is expression at sites of inflammation and facilitating the function of macrophages in host defence. In addition, MIF impairs antiviral host immunity and increases inflammation during influenza infection, as blocking of MIF during IAV challenge improved survival rates of mice (Smith *et al.*, 2019). Consistent with this observation, *Mif*-deficient mice display less inflammation, viral load, and mortality compared to control litter mates. Conversely, transgenic mice overexpressing MIF in alveolar epithelial cells showed a higher inflammation, viral load, and mortality (Smith *et al.*, 2019). Similar observations have been made from studies using H5N1 influenza virus where in a significant reduction in pulmonary inflammatory cytokines such as IL-1beta, IL-6 and TNF-alpha and chemokine interferon was seen using in a pneumonia murine model (Hou *et al.*, 2009). As MIF is a critical upstream mediator of immune and inflammatory responses it needs to be tightly controlled to prevent excessive infection-induced inflammation that can cause collateral tissue damage and in effect strike a balance between effective versus protective immunity. This is mediated by the anti-inflammatory cytokine IL-10, being crucial to counteract the strong MIF-induced pro-inflammatory response, leading to pathology control (Stijlemans *et al.*, 2022).

In contrast, G-CSF expression facilitates viral clearance and sustained mouse survival, and G-CSF levels has been shown to be increased during IAV lung infection, thus providing a

resistance to IAV (Huang *et al.*, 2011; Halstead *et al.*, 2018; Wang *et al.*, 2019a). Furthermore, it revealed a protective effect upon secondary bacterial infections of IAV challenged mice (Ishikawa *et al.*, 2016). The proinflammatory cytokine IL-18 is primarily involved in immune responses triggering synthesis of inflammatory mediators in polarized T-helper 1 and natural killer cells (Kato *et al.*, 2003; Tsutsumi *et al.*, 2014). It further synergizes with interleukin-12 (IL-12) to induce Interferon gamma (IFNγ) synthesis (Tominaga *et al.*, 2000; Kaplanski, 2018), however only IL- 18 not IL- 12 has been demonstrated to be protective in the early defences against influenza infections (Liu *et al.*, 2004; Denton *et al.*, 2007). Notably, the avian influenza strains H5N1 and H7N9 lead to an excess IL- 18 production and consequently induce a very damaging IFNγ- biased cytokine storm, attributed to ARDS pathogenesis (McGill *et al.*, 2009; Yang *et al.*, 2015; Guo *et al.*, 2015).

HBEC3-KT cells were typically cultured in serum free KSFM, which is supplemented with rEGF and bovine pituitary extract. Whereas PneumaCult™-Ex Plus medium is a defined, serum- and BPE-free cell culture medium, advertised to support more expansion of primary human airway epithelial cells at each passage, compared to other commercially available media. Additionally, this medium has been marketed to further supports at least two additional passages of cell expansion with a reported better differentiation potential, measured by the ability to form a pseudostratified mucociliary epithelium at the air-liquid interface (https://www.stemcell.com/products/pneumacult-ex-plus-medium.html).

In order to assess if the KSFM or ExPlus media is better suited for our purposes, establishment of the HBEC3-KT infection assay was carried out in parallel using both media independently. Although the infection efficiency of the PR8 strain on undifferentiated cells was comparable between both media sources, it was observed that gene expression of IP-10 was detected in ExPlus media-maintained cells at an earlier timepoint after virus challenge (2h). In addition, the response intensity seems to also peak earlier at 8h in ExPlus compared to 24h in KSFM. IP-10 is a pro-inflammatory cytokine that is involved in a wide variety of processes such as chemotaxis, differentiation, activation of immune cells, cell growth, apoptosis and modulation of angiogenesis (Angiolillo *et al.*, 1995; Romagnani *et al.*, 2001; Aksoy *et al.*, 2006; Sidahmed *et al.*, 2012). More specifically, it plays an important role during viral infections by stimulating leukocyte migration and activation to the infected sites through activation of phospholipase C-dependent pathway, an increase in intracellular calcium production and actin reorganization (Smit *et al.*, 2003; Gao *et al.*, 2009). In addition, IP-10 has been suggested to act as a good biomarker for infection, and as an indicator of

severity of acute respiratory infection in humans (Hayney *et al.*, 2017). This response is not just limited to airway epithelia, as for example only the active IAV strain A/Aichi/2/68 (H3N2) increased IP-10 levels in endothelial human umbilical vein endothelial cells (HUVEC) (Ishiguro *et al.*, 2003).

Notably, MIF expression levels were significantly elevated 24h post IAV PR8 encounter in KSFM-maintained cells, which was co-incidentally the only significant cytokine response besides IP-10. In stark contrast, CXCL1, G-CSF, IL-8 and IL-18 but not MIF expression levels were significantly reduced upon IAV PR8 exposure from cells maintained in ExPlus media. It seems that the ExPlus media contains undisclosed or proprietary supplement(s), which prime the undifferentiated HBEC3-KT cells to a committed differentiation state as confirmed by detection of low level *BPIFA1* expression by PCR, as well as cytokine array-based detection of higher basal levels of cytokines IL-8, CXCL1, G-CSF compared to KSFM.

The employed cytokine array is a fast and convenient method to gain a first understanding to which cytokine or chemokine is modulated during IAV encounter. However, these results would need to be validated with for example independent ELISA assays or other commonly used technologies, which vary in sensitivity, dynamic range, or robustness and are reviewed in (Platchek *et al.*, 2020).

The key aim of this chapter was to test the role of BPIFB1 in cell defence against IAV infection. Surprisingly, we did not observe a reduced IAV infection rate when HBEC3-KT BPIFB1 KO washes of differentiated cells were applied to undifferentiated populations in the presence of IAV PR8 viruses at an MOI of 1. A similar infection rate by mean fluorescence was achieved in all three tested conditions between unconditioned media or the wt conditioned wash, compared to BPIFB1 KO washes. This experiment was essentially undertaken blind as we did not employ a positive control in this setting to assess if BPIFB1 levels in the wt wash are sufficient to exert a biological effect. To attempt to overcome this limitation recombinant BPIFB1 was utilised in subsequent experiments. It was observed that there was a threshold of BPIFB1 concentration, which is required to produce a protective role against IAV infection by decreasing viral particle entry. This data is only preliminary and needs much further experimentation. Further studies need to be carried out to investigate the protective mechanism in detail. A hypothesis how BPIFB1 mediates the reduced IAV susceptibility could be through viral trapping as observed for other secreted reparatory proteins like the mucins. Respiratory epithelial cells constitutively express mucin glycoproteins at their surface, such as MUC5AC, MUC5B, and MUC1, which carry out a

vital role in restricting IAV infection (Ehre *et al.*, 2012; Roy *et al.*, 2014; McAuley *et al.*, 2017). For example, mucin glycoproteins are rich in SA, acting as viral receptor traps and thus preventing viral binding to target cells (Duez *et al.*, 2009; McAuley *et al.*, 2017; Chen *et al.*, 2018).

Due to severe COVID-19 related procurement restrictions of required cell culture items, which led to time constrains, it was not possible to carry out additional follow up experiments such as repetition of the cytokine assay on differentiated HBEC3-KT cells using an IAV PR8 challenge, at potentially different MOIs, or more mechanistic experiments analysing the protective role of BPIFB1. These experiments need to be addressed in the future. Overall, my data does provide some suggestion how BPIFB1 may influence IAV infection in human airway epithelium.

Chapter 7:

Final discussion

This thesis was designed to expand the knowledge of the BPIF member BPIFB1, with a focus on whether BPIFB1 influences the rate of infection and disease progression, following infection with IAV. It further explored the hypothesis that this protein functions as a modulator of respiratory tract homeostasis.

The initial study using multiple online platforms in conjunction with published studies in chapter three, demonstrated that BPIFB1 is found across mammal species, although with a low level of similarity consistent with previous publications (Bingle *et al.*, 2011a). It could be speculated that BPIFB1 is required in organisms with an airway based respiratory system and hence carries out an airway supporting function. However, *Bpifb1* knock out mice do not show any apparent respiratory defects, possibly as loss of *Bpifb1* could potentially be compensated by other family members, or more likely it could be speculated that BPIFB1 function is required in conditions of respiratory dysregulation caused by, for example inflammation.

My analysis showed that structurally, BPIFB1 contains a primate specific C-terminus. It seems likely that might contribute a particular biological function to the protein and requires further investigation. In addition, the presence of conserved regions such as the disulphide bridge suggests that these particular features are important for the structure of BPIFB1. This could be tested in the future using a site-directed mutagenesis or truncation approach targeting conserved regions in a functional BPIFB1 assay such as the established IAV HBEC3-KT infection assay.

I showed that the predicted structure of mouse and human BPIFB1 proteins are similar, although they are less than 60% identical. Moreover, the glycosylation sites in number and location are not very conserved across species and do not associate with specific regions of the protein, suggesting that these modifications might not be essential for biological structure and demonstrates the evolutionary change in protein function and structure (Kim *et al.*, 2015). With respect to RNA and protein expression, BPIFB1 demonstrates expression limited to specific sites, mostly associated with the airway and mucosal tissues, supporting the suggestion of its role in mucosal homeostasis. Within these tissues *BPIFB1* is highly expressed. Analysis of published scRNA data suggests that the gene is expressed in a restricted to number of cell types and is in certain instances found associated with *BPIFA1*. As it has previously been considered that BPIFB1 and BPIFA1 were not co-localised (Vargas *et al.*, 2008; Musa *et al.*, 2012), this constitutes a very significant finding. However, with regards to available RNA-Seq data, the sample size and material origin should be carefully

assessed to avoid bias or overinterpretation of results, as in certain cases where it is difficult to obtain human material, only a very limited sample size has been analysed.

Continuing from this analysis, further experimental studies are needed to investigate in more detail in which cell type or disease conditions BPIFB1 and BPIFA1 co-localise, as it was found over expressed in certain secretory cells in disease states. This could be achieved by a dual *in situ* hybridisation technique on RNA basis or by dual immunofluorescence for protein-based detection.

Work in my host lab has used the mTEC respiratory model system for a number of studies (Akram *et al.,* 2018). For my studies I showed that I was able to extract, expand and ALI differentiate *in vitro* these respiratory epithelial cells. I also established that I could infect cells cultured under these conditions with IAV in undifferentiated and differentiated states. This technique was intended to be the main experimental system used during my project. This model system was used to investigate differential expression patterns of genes by DNA microarray during the expansion and differentiation phase, a potential modulation of mTEC differentiation by IL-13, and the acute cellular response to IAV X31 strain infection in undifferentiated versus differentiated cells.

Comparison between the expansion and differentiation phase of mTECs showed a subset of DEGs, which are inversely correlated between both conditions. During the ALI differentiation, gene signatures of cilium and cell projection as well as differentiation genes are upregulated, alongside specific membrane associated proteins such as cadherin-related family members. In contrast, genes encoding for extracellular space and secreted proteins, keratinization or keratinocyte differentiation or epidermis development were found to be downregulated during mTEC ALI differentiation. This is consistent with a study carried out in mTEC cells during differentiation assessing the transcriptional regulator TAp73 compared wt cells, where a ciliogenesis gene signature was observed (Nemajerova *et al.*, 2016). Of note, *BPIFB1* has been found to be one of the highest expressed DEGs in a previous study of HBECs post 28 days of ALI culture (Ross *et al.*, 2007). However, in our current experiment using mTECs during 14 days of ALI, *Bpifb1* was not so significantly modulated. This may well represent a species specific difference of be due to differences in culture techniques.

Furthermore, unexpectedly, the array data from the IL-13 treatment experiment resulted in a very small number of DEGs, of which two out of three were below the log2 cut off of 1 (*Timp4* and *Bpifb1*). Merely *Olfm4* was the only significantly upregulated DEG. Modulated

expression of *Bpifb1* on the array or on protein level by western blotting served as positive control to ensure that IL-13 treatment was functional in our study. However, this should be further tested in terms of downstream signalling, for example assessing STAT3/6 activation or staining for enriched goblet cell populations (Matsukura *et al.*, 2001; Hershey, 2003; Eenjes *et al.*, 2018). The aim of IL-13 treatment was to increase BPIFB1 levels, and then test if this has an effect in response to IAV infection, as IL-13 has been reported to alter the differentiation process for ciliated cells and consequently strongly increases the secretory cell population (Laoukili *et al.*, 2001; Eenjes *et al.*, 2018). Furher studies are needed to investigate this further.

Alternatively, mTEC cells could be expanded in presence of the ROCK inhibitor Y-27632 and the γ-secretase inhibitor N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester (DAPT), which negatively impacts NOTCH signalling (Soni *et al.*, 2019) and has been reported to improve growth rate and morphology (Eenjes *et al.*, 2018). As IL-13 induces goblet cell differentiation, DAPT is able to change the capacity of basal cells to differentiate into ciliated cells (Eenjes *et al.*, 2018), therefore it might be of intertest to compare both conditions to shed more light into signalling pathways, which governs basal cell differentiation towards specific cell types.

Both cell states, differentiated and undifferentiated showed infection ability by the X31 IAV strain. However, the response was shown to be more restricted in undifferentiated cells and identified DEGs overlap completely with the differentiated gene signature, which contains a large number of ISGs. Having demonstrated that undifferentiated cells were susceptible to IAV infection, we concluded that this system could form the basis for a BPIFB1 functional assay. For example, it could be used to show if BPIFB1 presence modulates IAV infection ability or viral replication?

However, a drawback of my microarray study is the fact that only two replicates were available due to technical difficulties, and the conclusions drawn from the acquired data should be considered carefully. However, it should be mentioned that the replicates of conditions behaved similar for the ALI differentiation data set but showed a slightly higher biological variation for X31 infection conditions. This could be due to variability of the infection process, whereas mTEC expansion and ALI differentiation might be more synchronised or homogenous.

Microarray data analysis is very complex and requires comprehensive data analysis algorithms or pipelines to control for various factors, and to preform appropriate data normalization. Aside from the analysis and data processing side, a lot of factors can influence an DNA microarray study such as contamination or damage on the microarray's surface (Moffitt *et al.*, 2011; Jaksik *et al.*, 2015), incorrect assignments of probes to genes (Dai *et al.*, 2005), a wrong background level evaluation or presence of non-specific probe hybridization signals (Draghici *et al.*, 2006; Kroll *et al.*, 2008). In addition, the sample purity and stability (Opitz *et al.*, 2010) along with the amplification process (Croner *et al.*, 2009) can impact the estimates of gene expression. Apart from potential technical issues, another drawback of microarray studies is the high cost per experiment, combined with the very limited control over chip design with regards to probe designs for sequences.

An alternative to the microarray experiment techniques is RNA-Seq. This approach has been used extensively in the last decade and is paramount for transcriptome profiling (Wang *et al.*, 2009; Chu and Corey, 2012). The main difference between both methodologies is that RNA-Seq allows for complete sequencing of the whole transcriptome, while microarrays are only able to capture predefined transcripts or genes through hybridization. Therefore, RNA-Seq can detect in addition non-coding DEGs, novel genes, alternative or chimeric transcript variants, as well as gene mutations and single nucleotide polymorphisms (SNPs), allele-specific expression and alternative gene spliced transcripts (Maher *et al.*, 2009; Hong *et al.*, 2020). For example, studies comparing both technologies found that RNA-Seq identified more protein coding DEGs and was able to show a wider quantitative range of expression levels and therefore provided an immense amount of additional information compared to a DNA microarray (Sultan *et al.*, 2008; Ozsolak and Milos, 2011; Zhang *et al.*, 2015; Rao *et al.*, 2019).

To develop a functional assay, a key goal of my study was to generate murine constructs of wt BPIFB1 as well as glycosylation mutations and BPIFB1 truncation constructs, with the aim to generate and purify secreted recombinant proteins for functional infection assays or rescue experiments. For initial validation experiments I used the HEK293 mammalian expression system with a plasmid based transient expression approach. In the future, if larger protein amounts would be required, stable cell cultures could be established and cultured in a suspension mode for higher yields (Hunter *et al.*, 2019; Chapple and Dyson, 2021). The mammalian expression system offers the advantage that required post translational modifications such as N-glycosylation or disulphide bond formation occur, along appropriate

protein folding and secretion of recombinant proteins. Disadvantages of this system are high costs of protein production due to a slow cell growth, expensive growth medium and supplements such as FBS, required transfection reagents if a transient approach is taken, and more costly culture conditions. A well-established alternative is the insect cell expression system, which utilises cells of organism such as the fall armyworm *Spodoptera fugiperda (Sf9)* in either suspension or adherent culture conditions. Unfortunately, the setup for this system in terms of cloning requirements is time consuming and glycosylation modifications are different compared to the mammalian system (Irons *et al.*, 2018; Palomares *et al.*, 2018). A similar issue has been reported for yeast (e.g. *Saccharomyces cerevisiae*) derived recombinant proteins, which can be found hyperglycosylated (Wildt and Gerngross, 2005; Baghban *et al.*, 2019). This would be a problem for BPIFB1 studies involving analysis of the functional role of glycosylation.

For protein activity and structural studies, the procaryotic expression system using for example *E. coli,* is often employed due to its low associated costs, easy growth conditions, scalability, and high yields. However, bacteria are unable to perform post translational modifications or are able to efficiently secrete recombinant proteins into the extracellular medium. Furthermore, formation of correct disulphide bonds, which is vital for biological activity and function, is potentially not achieved and can lead to conformational protein misfolding and aggregation (Rosano and Ceccarelli, 2014). However, efforts are made to overcome these drawbacks (Ke and Berkmen, 2014; Kleiner-Grote *et al.*, 2018).

The cell free expression system is an alternative *in vitro* protein expression system, where proteins are expressed in a cell free environment using a derived cell extract, combined with a DNA template, amino acids and other cofactors (Chong, 2014). This method is simple and fast and allows for protein labelling for structural studies. However, this method is costly, and can only accommodate limited post translational modifications, and therefore is not considered for our studies.

Due to very significant lab restrictions during the COVID-19 pandemic, my original strategy to employ *Bpifb1*$^{-/-}$ cells was no longer feasible as our collaborator was unable to breed the mice required for the work. As an alternative, I had to develop a different system to allow me to undertake some functional studies. In the lab we were already using a HBEC-3KT cells as these have been used as model system for the respiratory airway (Delgado *et al.,* 2013). I generated a HBEC3-KT *BPIFB1* KO cell line successfully, which was shown by western

blotting to have lost BPIFB1 protein expression despite expressing a truncated RNA. Functional assays confirmed that this cell line is morphologically similar to the parental cell line, and cell cycle and proliferation parameters were comparable to parental cells. This line was only generated at the very end of my study, and I was only able to undertake limited studies with is. However, I established IAV infection assays to demonstrate that the PR8 IAV strain at MOI of 1 was able to infect approximately 10% of the cell population. Future studies can be undertaken to improve infection efficiencies by applying other viral strains or higher MOI amounts in order to have a larger population of cells where a biological function of BPIB1 could be studied.

Using BPIFB1 KO washes from HBEC3-KT cell ALI cultures did not have a significant effect on PR8 IAV infections of HBEC3-KT cells compared to that of WT washes. A limitation of this experiment is the fact that it was unknown how much BPIFB1 was produced in the WT wash. This could also be further investigated by concentrating the wash volume using concentrator columns as the current BPIFB1 concentration might have been below biological activity. In contrast, the commercial recombinant BPIFB1 protein appeared to show a biological activity above 1.43 μg/ml, and this assay format could be used to identify the role of specific regions within BPIFB1. This would require the production of human BPIFB1 expression constructs. Moving back to the mTEC experiments, the murine based generated glycosylation deficient mutations or the N/C-terminal BPIFB1 truncations could be tested in this context if they are still able to block or reduce IAV uptake, indicating essential protein regions required for biological function. Alternatively, a murine based BPIFB1 KO cell line could be generated by CRISPR, or isolated mTECs could be generated as initially intended from a murine *Bpifb1* $^{-/-}$ background and tested with the above-mentioned constructs or respective purified proteins.

A significant question which still needs to be addressed is how BPIFB1 might mechanistically exert a protective role against IAV infection. It could be speculated that BPIFB1 positive cells secrete the protein onto the mucus surface, and this could prevent infection of all cells in the epithelium, much like MUC proteins (Zanin *et al.*, 2016). Interestingly, there might be an emerging link between MUC5AC and MUC5B proteins and BPIFB1. *Bbifb1*-knockout mice appear to display abnormal mucin secretion, suggesting that the protein is associated with MUC5AC and MUC5B function (Donoghue *et al.,* 2017) and Bingle *et al.,* (2013) reported that *BPIFB1* and *MUC5B* are upregulated in IPF. Human MUC5AC secretion is observed in superficial epithelial goblet cells, whereas MUC5B is

predominantly secreted from SMG, which coincides with BPIFB1 expression in the airway epithelium. Gel-forming mucins are one of the largest and most complex proteins known, as large proportions of amino acid residues are organized into different domains with diverse post-translational modifications, including O- and N-glycosylation (Pérez-Vilar and Mabolo, 2007). Murine knockout models demonstrated that MUC5B is required for normal airway function, while MUC5AC is beneficial but not essential (Roy *et al.*, 2014). Notably, it has been shown for MUC5AC, that overexpression is protective against IAV infection (Ehre *et al.*, 2012). However, it is not clear if glycosylation forms a protective mechanism and would need to be experimentally investigated using the generated glycosylation deficient mutants.

Overall, this work aimed to study a role of BPIFB1 in immune defence, which has been implicated structurally but has not been shown formally. Bioinformatic expression analysis did highlight that BPIFB1 might be involved in respiratory homeostasis with respect to respiratory diseases, and it potentially exerts a protective role against IAV infection. However, these findings need to be examined in more detail. Firstly, my data need to be replicated in both the HBEC3-KT cells and as I originally intended, expanded into the mTEC system using both WT and *Bpifb1*$^{-/-}$ cells. Secondly, studies need to be expanded to other pathogens, of both viral or bacterial nature. This will address the homeostatic function of BPIFB1 in airway host defence. Furthermore, it is important to note that my study was limited to cell culture models only, which form a good base for initial research, but they lack the more three-dimensional cellular heterogeneity of an organ system, including immune response feedback (Rijsbergen *et al.*, 2021; Mifsud *et al.*, 2021).

Future experiments could further this research by extending it to an animal model for IAV infection research as carried out successfully for BPIFA1 (Akram *et al.,* 2018). *Bpifb1*$^{-/-}$ mice have been used in a number of studies and could be challenged *in vivo* in a similar manner.

References

ABOLINS, S., KING, E. C., LAZAROU, L., WELDON, L., HUGHES, L., DRESCHER, P., RAYNES, J. G., HAFALLA, J. C., VINEY, M. E. & RILEY, E. M. 2017. The comparative immunology of wild and laboratory mice, Mus musculus domesticus. *Nature communications,* 8**,** 1-13.

ADLI, M. 2018. The CRISPR tool kit for genome editing and beyond. *Nature Communications,* 9**,** 1911.

AEBI, M. 2013. N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1833**,** 2430-2437.

AHDIEH, M., VANDENBOS, T. & YOUAKIM, A. 2001. Lung epithelial barrier function and wound healing are decreased by IL-4 and IL-13 and enhanced by IFN-γ. *American Journal of Physiology-Cell Physiology,* 281**,** C2029-C2038.

AKRAM, K. M., MOYO, N. A., LEEMING, G. H., BINGLE, L., JASIM, S., HUSSAIN, S., SCHORLEMMER, A., KIPAR, A., DIGARD, P., TRIPP, R. A., SHOHET, R. V., BINGLE, C. D. & STEWART, J. P. 2018. An innate defense peptide BPIFA1/SPLUNC1 restricts influenza A virus infection. *Mucosal Immunol,* 11**,** 71-81.

AKSOY, M. O., YANG, Y., JI, R., REDDY, P., SHAHABUDDIN, S., LITVIN, J., ROGERS, T. J. & KELSEN, S. G. 2006. CXCR3 surface expression in human airway epithelial cells: cell cycle dependence and effect on cell proliferation. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 290**,** L909-L918.

ALEVY, Y. G., PATEL, A. C., ROMERO, A. G., PATEL, D. A., TUCKER, J., ROSWIT, W. T., MILLER, C. A., HEIER, R. F., BYERS, D. E. & BRETT, T. J. 2012. IL-13–induced airway mucus production is attenuated by MAPK13 inhibition. *The Journal of clinical investigation,* 122**,** 4555-4568.

ALEXOPOULOU, L., HOLT, A. C., MEDZHITOV, R. & FLAVELL, R. A. 2001. Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3. *Nature,* 413**,** 732.

ALLEN, I. C., SCULL, M. A., MOORE, C. B., HOLL, E. K., MCELVANIA-TEKIPPE, E., TAXMAN, D. J., GUTHRIE, E. H., PICKLES, R. J. & TING, J. P.-Y. 2009. The NLRP3 inflammasome mediates in vivo innate immunity to influenza A virus through recognition of viral RNA. *Immunity,* 30**,** 556-565.

ALLEN, J. R., GE, L., HUANG, Y., BRAUER, R., PARIMON, T., CASSEL, S. L., SUTTERWALA, F. S. & CHEN, P. 2018. TIMP-1 promotes the immune response in influenza-induced acute lung injury. *Lung,* 196**,** 737-743.

ALVES, D. B. M., BINGLE, L., BINGLE, C. D., LOURENCO, S. V., SILVA, A. A., PEREIRA, D. L. & VARGAS, P. A. 2017. BPI-fold (BPIF)

containing/plunc protein expression in human fetal major and minor salivary glands. *Brazilian oral research,* 31.
AN, H. J., FROEHLICH, J. W. & LEBRILLA, C. B. 2009. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Current opinion in chemical biology,* 13**,** 421-426.
ANGATA, T. & VARKI, A. 2002. Chemical diversity in the sialic acids and related α-keto acids: an evolutionary perspective. *Chemical reviews,* 102**,** 439-470.
ANGIOLILLO, A. L., SGADARI, C., TAUB, D. D., LIAO, F., FARBER, J. M., MAHESHWARI, S., KLEINMAN, H. K., REAMAN, G. H. & TOSATO, G. 1995. Human interferon-inducible protein 10 is a potent inhibitor of angiogenesis in vivo. *The Journal of experimental medicine,* 182**,** 155-162.
ASKOVICH, P. S., SANDERS, C. J., ROSENBERGER, C. M., DIERCKS, A. H., DASH, P., NAVARRO, G., VOGEL, P., DOHERTY, P. C., THOMAS, P. G. & ADEREM, A. 2013. Differential host response, rather than early viral replication efficiency, correlates with pathogenicity caused by influenza viruses. *PloS one,* 8**,** e74863.
ATANASOVA, K. R. & REZNIKOV, L. R. 2018. Neuropeptides in asthma, chronic obstructive pulmonary disease and cystic fibrosis. *Respiratory Research,* 19**,** 1-18.
ATHERTON, H. C., JONES, G. & DANAHAY, H. 2003. IL-13-induced changes in the goblet cell density of human bronchial epithelial cell cultures: MAP kinase and phosphatidylinositol 3-kinase regulation. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 285**,** L730-L739.
AYDIN, M., NAUMOVA, E. A., BELLM, A., BEHRENDT, A.-K., GIACHERO, F., BAHLMANN, N., ZHANG, W., WIRTH, S., PAULSEN, F. & ARNOLD, W. H. 2021. From submerged cultures to 3D cell culture models: evolution of nasal epithelial cells in asthma research and virus infection. *Viruses,* 13**,** 387.
AYERS, M. & JEFFERY, P. 1988. Proliferation and differentiation in mammalian airway epithelium. *European Respiratory Journal,* 1**,** 58-80.
AZZEH, M., FLICK, R. & HOBOM, G. 2001. Functional analysis of the influenza A virus cRNA promoter and construction of an ambisense transcription system. *Virology,* 289**,** 400-410.
BAGDONAS, E., RAUDONIUTE, J., BRUZAUSKAITE, I. & ALDONYTE, R. 2015. Novel aspects of pathogenesis and regeneration mechanisms in COPD. *International journal of chronic obstructive pulmonary disease,* 10**,** 995.
BAGHBAN, R., FARAJNIA, S., RAJABIBAZL, M., GHASEMI, Y., MAFI, A., HOSEINPOOR, R., RAHBARNIA, L. & ARIA, M. 2019. Yeast

expression systems: overview and recent advances. *Molecular biotechnology,* 61**,** 365-384.
BAINES, K. J., SIMPSON, J. L., WOOD, L. G., SCOTT, R. J., FIBBENS, N. L., POWELL, H., COWAN, D. C., TAYLOR, D. R., COWAN, J. O. & GIBSON, P. G. 2014. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. *Journal of allergy and clinical immunology,* 133**,** 997-1007.
BAKER, O. J., CAMDEN, J. M., REDMAN, R. S., JONES, J. E., SEYE, C. I., ERB, L. & WEISMAN, G. A. 2008. Proinflammatory cytokines tumor necrosis factor-α and interferon-γ alter tight junction structure and function in the rat parotid gland Par-C10 cell line. *American Journal of Physiology-Cell Physiology,* 295**,** C1191-C1201.
BALS, R. & HIEMSTRA, P. 2004. Innate immunity in the lung: how epithelial cells fight against respiratory pathogens. *European Respiratory Journal,* 23**,** 327-333.
BALS, R., WANG, X., ZASLOFF, M. & WILSON, J. M. 1998. The peptide antibiotic LL-37/hCAP-18 is expressed in epithelia of the human lung where it has broad antimicrobial activity at the airway surface. *Proc Natl Acad Sci U S A,* 95**,** 9541-6.
BARNES, F. A., BINGLE, L. & BINGLE, C. D. 2008. Pulmonary genomics, proteomics, and PLUNCs. *American journal of respiratory cell and molecular biology,* 38**,** 377-379.
BARRATT, S. L., CREAMER, A., HAYTON, C. & CHAUDHURI, N. 2018. Idiopathic pulmonary fibrosis (IPF): an overview. *Journal of clinical medicine,* 7**,** 201.
BARRON, S. L., SAEZ, J. & OWENS, R. M. 2021. In Vitro Models for Studying Respiratory Host–Pathogen Interactions. *Advanced Biology,* 5**,** 2000624.
BAUM, A., SACHIDANANDAM, R. & GARCÍA-SASTRE, A. 2010. Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proceedings of the National Academy of Sciences,* 107**,** 16303-16308.
BAUM, L. & PAULSON, J. 1990. Sialyloligosaccharides of the respiratory epithelium in the selection of human influenza virus receptor specificity. *Acta histochemica. Supplementband,* 40**,** 35-38.
BAYES, H., RITCHIE, N., IRVINE, S. & EVANS, T. 2016. A murine model of early Pseudomonas aeruginosa lung disease with transition to chronic infection. *Scientific reports,* 6**,** 1-10.
BAYLY-JONES, C., PANG, S. S., SPICER, B. A., WHISSTOCK, J. C. & DUNSTONE, M. A. 2020. Ancient but not forgotten: New insights into MPEG1, a macrophage perforin-like immune effector. *Frontiers in Immunology,* 11**,** 581906.

BEAMER, L. 2003. Structure of human BPI (bactericidal/permeability-increasing protein) and implications for related proteins. *Biochemical Society Transactions,* 31**,** 791-794.
BEAMER, L. J., CARROLL, S. F. & EISENBERG, D. 1997. Crystal structure of human BPI and two bound phospholipids at 2.4 angstrom resolution. *Science,* 276**,** 1861-1864.
BEAMER, L. J., CARROLL, S. F. & EISENBERG, D. 1998. The BPI/LBP family of proteins: a structural analysis of conserved regions. *Protein Science,* 7**,** 906-914.
BECKER, M. N., DIAMOND, G., VERGHESE, M. W. & RANDELL, S. H. 2000. CD14-dependent lipopolysaccharide-induced beta-defensin-2 expression in human tracheobronchial epithelium. *J Biol Chem,* 275**,** 29731-6.
BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The protein data bank. *Nucleic acids research,* 28**,** 235-242.
BEURA, L. K., HAMILTON, S. E., BI, K., SCHENKEL, J. M., ODUMADE, O. A., CASEY, K. A., THOMPSON, E. A., FRASER, K. A., ROSATO, P. C. & FILALI-MOUHIM, A. 2016. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature,* 532**,** 512-516.
BEZENÇON, C., FÜRHOLZ, A., RAYMOND, F., MANSOURIAN, R., MÉTAIRON, S., LE COUTRE, J. & DAMAK, S. 2008. Murine intestinal cells expressing Trpm5 are mostly brush cells and express markers of neuronal and inflammatory cells. *Journal of Comparative Neurology,* 509**,** 514-525.
BHOWMICK, R. & GAPPA-FAHLENKAMP, H. 2016. Cells and culture systems used to model the small airway epithelium. *Lung,* 194**,** 419-428.
BHOWMIK, A., CHAHAL, K., AUSTIN, G. & CHAKRAVORTY, I. 2009. Improving mucociliary clearance in chronic obstructive pulmonary disease. *Respiratory medicine,* 103**,** 496-502.
BINGLE, C. & CRAVEN, C. 2003. Comparative analysis of the PLUNC (palate, lung and nasal epithelium clone) protein families. Portland Press Limited.
BINGLE, C. D., ARAUJO, B., WALLACE, W. A., HIRANI, N. & BINGLE, L. 2013. What is top of the charts? BPIFB1/LPLUNC1 localises to the bronchiolised epithelium in the honeycomb cysts in UIP. *Thorax,* 68**,** 1167-8.
BINGLE, C. D. & BINGLE, L. 2000. Characterisation of the human plunc gene, a gene product with an upper airways and nasopharyngeal restricted expression pattern. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression,* 1493**,** 363-367.

BINGLE, C. D., BINGLE, L. & CRAVEN, C. J. 2011a. Distant cousins: genomic and sequence diversity within the BPI fold-containing (BPIF)/PLUNC protein family. *Biochem Soc Trans,* 39**,** 961-5.

BINGLE, C. D. & CRAVEN, C. J. 2002. PLUNC: a novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. *Human Molecular Genetics,* 11**,** 937-943.

BINGLE, C. D., LECLAIR, E. E., HAVARD, S., BINGLE, L., GILLINGHAM, P. & CRAVEN, C. J. 2004. Phylogenetic and evolutionary analysis of the PLUNC gene family. *Protein Science,* 13**,** 422-430.

BINGLE, C. D., SEAL, R. L. & CRAVEN, C. J. 2011b. Systematic nomenclature for the PLUNC/PSP/BSP30/SMGB proteins as a subfamily of the BPI fold-containing superfamily. *Biochem Soc Trans,* 39**,** 977-83.

BINGLE, C. D., SEAL, R. L. & CRAVEN, C. J. 2011c. Systematic nomenclature for the PLUNC/PSP/BSP30/SMGB proteins as a subfamily of the BPI fold-containing superfamily. Portland Press Ltd.

BINGLE, C. D., WILSON, K., LUNN, H., BARNES, F. A., HIGH, A. S., WALLACE, W. A., RASSL, D., CAMPOS, M. A., RIBEIRO, M. & BINGLE, L. 2010. Human LPLUNC1 is a secreted product of goblet cells and minor glands of the respiratory and upper aerodigestive tracts. *Histochemistry and cell biology,* 133**,** 505-515.

BINGLE, L., BARNES, F. A., CROSS, S. S., RASSL, D., WALLACE, W. A., CAMPOS, M. A. & BINGLE, C. D. 2007. Differential epithelial expression of the putative innate immune molecule SPLUNC1 in cystic fibrosis. *Respiratory research,* 8**,** 1-10.

BINGLE, L. & BINGLE, C. D. 2011. Distribution of human PLUNC/BPI fold-containing (BPIF) proteins. *Biochemical Society Transactions,* 39**,** 1023-1027.

BINGLE, L., CROSS, S. S., HIGH, A. S., WALLACE, W. A., DEVINE, D. A., HAVARD, S., CAMPOS, M. A. & BINGLE, C. D. 2005. SPLUNC1 (PLUNC) is expressed in glandular tissues of the respiratory tract and in lung tumours with a glandular phenotype. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland,* 205**,** 491-497.

BINGLE, L., WILSON, K., MUSA, M., ARAUJO, B., RASSL, D., WALLACE, W. A., LECLAIR, E. E., MAUAD, T., ZHOU, Z., MALL, M. A. & BINGLE, C. D. 2012. BPIFB1 (LPLUNC1) is upregulated in cystic fibrosis lung disease. *Histochem Cell Biol,* 138**,** 749-58.

BOERS, J. E., AMBERGEN, A. W. & THUNNISSEN, F. B. 1998. Number and proliferation of basal and parabasal cells in normal human airway epithelium. *American journal of respiratory and critical care medicine,* 157**,** 2000-2006.

BOERS, J. E., AMBERGEN, A. W. & THUNNISSEN, F. B. 1999. Number and proliferation of clara cells in normal human airway epithelium. *American journal of respiratory and critical care medicine,* 159**,** 1585-1591.

BOERSEMA, P. J., GEIGER, T., WIŚNIEWSKI, J. R. & MANN, M. 2013. Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Molecular & cellular proteomics,* 12**,** 158-171.

BONSER, L. R. & ERLE, D. J. 2017. Airway Mucus and Asthma: The Role of MUC5AC and MUC5B. *J Clin Med,* 6.

BORNSTEIN, C., NEVO, S., GILADI, A., KADOURI, N., POUZOLLES, M., GERBE, F., DAVID, E., MACHADO, A., CHUPRIN, A. & TÓTH, B. 2018. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature,* 559**,** 622-626.

BRANCHFIELD, K., NANTIE, L., VERHEYDEN, J. M., SUI, P., WIENHOLD, M. D. & SUN, X. 2016. Pulmonary neuroendocrine cells function as airway sensors to control lung immune response. *Science,* 351**,** 707-710.

BRASS, A. L., HUANG, I.-C., BENITA, Y., JOHN, S. P., KRISHNAN, M. N., FEELEY, E. M., RYAN, B. J., WEYER, J. L., VAN DER WEYDEN, L. & FIKRIG, E. 2009. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell,* 139**,** 1243-1254.

BREEZE, R. G. & WHEELDON, E. B. 1977. The cells of the pulmonary airways. *American Review of Respiratory Disease,* 116**,** 705-777.

BREMNES, R. M., AL-SHIBLI, K., DONNEM, T., SIRERA, R., AL-SAAD, S., ANDERSEN, S., STENVOLD, H., CAMPS, C. & BUSUND, L.-T. 2011. The role of tumor-infiltrating immune cells and chronic inflammation at the tumor site on cancer development, progression, and prognosis: emphasis on non-small cell lung cancer. *Journal of Thoracic Oncology,* 6**,** 824-833.

BRITTO, C. J. & COHN, L. 2015. Bactericidal/permeability-increasing protein fold–containing family member A1 in airway host protection and respiratory disease. *American journal of respiratory cell and molecular biology,* 52**,** 525-534.

BROWN, J., GREAVES, M. & MOLGAARD, H. 1991. The gene encoding the stem cell antigen, CD34, is conserved in mouse and expressed in haemopoietic progenitor cell lines, brain, and embryonic fibroblasts. *International immunology,* 3**,** 175-184.

BRUCE, C., BEAMER, L. J. & TALL, A. R. 1998. The implications of the structure of the bactericidal/permeability-increasing protein on the lipid-transfer function of the cholesteryl ester transfer protein. *Current opinion in structural biology,* 8**,** 426-434.

BUSTAMANTE-MARIN, X. M. & OSTROWSKI, L. E. 2017. Cilia and mucociliary clearance. *Cold Spring Harbor perspectives in biology,* 9**,** a028241.

BUTLER, C. R., HYNDS, R. E., GOWERS, K. H., LEE, D. D. H., BROWN, J. M., CROWLEY, C., TEIXEIRA, V. H., SMITH, C. M., URBANI, L. & HAMILTON, N. J. 2016. Rapid expansion of human epithelial stem cells suitable for airway tissue engineering. *American journal of respiratory and critical care medicine,* 194**,** 156-168.

BUTTON, B., CAI, L.-H., EHRE, C., KESIMER, M., HILL, D. B., SHEEHAN, J. K., BOUCHER, R. C. & RUBINSTEIN, M. 2012. A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. *Science,* 337**,** 937-941.

CAMPOS, M. A., ABREU, A. R., NLEND, M. C., COBAS, M. A., CONNER, G. E. & WHITNEY, P. L. 2004. Purification and characterization of PLUNC from human tracheobronchial secretions. *American journal of respiratory cell and molecular biology,* 30**,** 184-192.

CANDIANO, G., BRUSCHI, M., PEDEMONTE, N., MUSANTE, L., RAVAZZOLO, R., LIBERATORI, S., BINI, L., GALIETTA, L. J. & ZEGARRA-MORAN, O. 2007. Proteomic analysis of the airway surface liquid: modulation by proinflammatory cytokines. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 292**,** L185-L198.

CARRARO, G., LANGERMAN, J., SABRI, S., LORENZANA, Z., PURKAYASTHA, A., ZHANG, G., KONDA, B., AROS, C. J., CALVERT, B. A. & SZYMANIAK, A. 2021. Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. *Nature medicine,* 27**,** 806-814.

CASADO, B., PANNELL, L. K., IADAROLA, P. & BARANIUK, J. N. 2005. Identification of human nasal mucous proteins using proteomics. *Proteomics,* 5**,** 2949-2959.

CASALINO-MATSUDA, S. M., MONZON, M. E., DAY, A. J. & FORTEZA, R. M. 2009. Hyaluronan Fragments/CD44 Mediate Oxidative Stress–Induced MUC5B Up-Regulation in Airway Epithelium. *American journal of respiratory cell and molecular biology,* 40**,** 277-285.

CHAN, R. W., YUEN, K. M., YU, W. C., HO, C. C., NICHOLLS, J. M., PEIRIS, J. M. & CHAN, M. C. 2010. Influenza H5N1 and H1N1 virus replication and innate immune responses in bronchial epithelial cells are influenced by the state of differentiation. *PloS one,* 5**,** e8713.

CHAPPLE, S. D. & DYSON, M. R. 2021. High-Throughput Expression Screening in Mammalian Suspension Cells. *Structural Genomics.* Springer.

CHEN, G., KORFHAGEN, T. R., XU, Y., KITZMILLER, J., WERT, S. E., MAEDA, Y., GREGORIEFF, A., CLEVERS, H. & WHITSETT, J. A. 2009. SPDEF is required for mouse pulmonary goblet cell differentiation

and regulates a network of genes associated with mucus production. *The Journal of clinical investigation,* 119**,** 2914-2924.
CHEN, X., CHO, D.-B. & YANG, P.-C. 2010. Double staining immunohistochemistry. *North American journal of medical sciences,* 2**,** 241.
CHEN, X., LIU, S., GORAYA, M. U., MAAROUF, M., HUANG, S. & CHEN, J.-L. 2018. Host immune response to influenza A virus infection. *Frontiers in immunology,* 9**,** 320.
CHEN, Y., LUN, A. T. & SMYTH, G. K. 2016. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research,* 5.
CHEN, Y.-W., HUANG, S. X., DE CARVALHO, A. L. R. T., HO, S.-H., ISLAM, M. N., VOLPI, S., NOTARANGELO, L. D., CIANCANELLI, M., CASANOVA, J.-L. & BHATTACHARYA, J. 2017. A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nature cell biology,* 19**,** 542-549.
CHIANG, S.-C., VELDHUIZEN, E. J., BARNES, F. A., CRAVEN, C. J., HAAGSMAN, H. P. & BINGLE, C. D. 2011. Identification and characterisation of the BPI/LBP/PLUNC-like gene repertoire in chickens reveals the absence of a LBP gene. *Developmental & Comparative Immunology,* 35**,** 285-295.
CHONG, S. 2014. Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Current protocols in molecular biology,* 108**,** 16.30. 1-16.30. 11.
CHU, H. W., RIOS, C., HUANG, C., WESOLOWSKA-ANDERSEN, A., BURCHARD, E. G., O'CONNOR, B. P., FINGERLIN, T. E., NICHOLS, D., REYNOLDS, S. D. & SEIBOLD, M. A. 2015. CRISPR–Cas9-mediated gene knockout in primary human airway epithelial cells reveals a proinflammatory role for MUC18. *Gene therapy,* 22**,** 822-829.
CHU, H. W., THAIKOOTTATHIL, J., RINO, J. G., ZHANG, G., WU, Q., MOSS, T., REFAELI, Y., BOWLER, R., WENZEL, S. E. & CHEN, Z. 2007. Function and regulation of SPLUNC1 protein in Mycoplasma infection and allergic inflammation. *The Journal of Immunology,* 179**,** 3995-4002.
CHU, Y. & COREY, D. R. 2012. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic acid therapeutics,* 22**,** 271-274.
CIESIELSKA, A., MATYJEK, M. & KWIATKOWSKA, K. 2021. TLR4 and CD14 trafficking and its influence on LPS-induced pro-inflammatory signaling. *Cellular and molecular life sciences,* 78**,** 1233-1261.
CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W. & MARRAFFINI, L. a.(2013) Multiplex

Genome Engineering Using CRISPR/Cas Systems. *Science,* 339**,** 819-823.
CONNOR, R. J., KAWAOKA, Y., WEBSTER, R. G. & PAULSON, J. C. 1994. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology,* 205**,** 17-23.
CORMIER, S. A., YOU, D. & HONNEGOWDA, S. 2010. The use of a neonatal mouse model to study respiratory syncytial virus infections. *Expert review of anti-infective therapy,* 8**,** 1371-1380.
COSTA-SILVA, J., DOMINGUES, D. & LOPES, F. M. 2017. RNA-Seq differential expression analysis: An extended review and a software tool. *PloS one,* 12**,** e0190152.
COUCEIRO, J. N. S., PAULSON, J. C. & BAUM, L. G. 1993. Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus research,* 29**,** 155-165.
CRONER, R. S., LAUSEN, B., SCHELLERER, V., ZEITTRAEGER, I., WEIN, A., SCHILDBERG, C., PAPADOPOULOS, T., DIMMLER, A., HAHN, E. G. & HOHENBERGER, W. 2009. Comparability of microarray data between amplified and non amplified RNA in colorectal carcinoma. *Journal of biomedicine and biotechnology,* 2009.
CROW, M., DENG, T., ADDLEY, M. & BROWNLEE, G. G. 2004. Mutational analysis of the influenza virus cRNA promoter and identification of nucleotides critical for replication. *Journal of virology,* 78**,** 6263-6270.
CRYSTAL, R. G. 2014. Airway basal cells. The "smoking gun" of chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine,* 190**,** 1355-1362.
DAI, M., WANG, P., BOYD, A. D., KOSTOV, G., ATHEY, B., JONES, E. G., BUNNEY, W. E., MYERS, R. M., SPEED, T. P. & AKIL, H. 2005. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic acids research,* 33**,** e175-e175.
DAVIDSON, D. J., KILANOWSKI, F. M., RANDELL, S. H., SHEPPARD, D. N. & DORIN, J. R. 2000. A primary culture model of differentiated murine tracheal epithelium. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 279**,** L766-L778.
DAVIS, J. D. & WYPYCH, T. P. 2021. Cellular and functional heterogeneity of the airway epithelium. *Mucosal immunology,* 14**,** 978-990.
DE BOECK, K. 2020. Cystic fibrosis in the year 2020: A disease with a new face. *Acta paediatrica,* 109**,** 893-899.
DE JONG, M. D., SIMMONS, C. P., THANH, T. T., HIEN, V. M., SMITH, G. J., CHAU, T. N. B., HOANG, D. M., VAN VINH CHAU, N., KHANH, T. H. & DONG, V. C. 2006. Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nature medicine,* 12**,** 1203-1207.

DE SMET, E. G., SEYS, L. J., VERHAMME, F. M., VANAUDENAERDE, B. M., BRUSSELLE, G. G., BINGLE, C. D. & BRACKE, K. R. 2018. Association of innate defense proteins BPIFA1 and BPIFB1 with disease severity in COPD. *Int J Chron Obstruct Pulmon Dis,* 13**,** 11-27.

DECKMANN, K. & KUMMER, W. 2016. Chemosensory epithelial cells in the urethra: sentinels of the urinary tract. *Histochemistry and cell biology,* 146**,** 673-683.

DELGADO, O., KAISANI, A. A., SPINOLA, M., XIE, X.-J., BATTEN, K. G., MINNA, J. D., WRIGHT, W. E. & SHAY, J. W. 2011. Multipotent capacity of immortalized human bronchial epithelial cells. *PloS one,* 6**,** e22023.

DELLER, M. C., KONG, L. & RUPP, B. 2016. Protein stability: a crystallographer's perspective. *Acta Crystallographica Section F: Structural Biology Communications,* 72**,** 72-95.

DENNIS, G., SHERMAN, B. T., HOSACK, D. A., YANG, J., GAO, W., LANE, H. C. & LEMPICKI, R. A. 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome biology,* 4**,** 1-11.

DENNIS, J. W., GRANOVSKY, M. & WARREN, C. E. 1999. Protein glycosylation in development and disease. *Bioessays,* 21**,** 412-421.

DENTENER, M. A., VREUGDENHIL, A. C., HOET, P. H., VERNOOY, J. H., NIEMAN, F. H., HEUMANN, D., JANSSEN, Y. M., BUURMAN, W. A. & WOUTERS, E. F. 2000. Production of the acute-phase protein lipopolysaccharide-binding protein by respiratory type II epithelial cells: implications for local defense to bacterial endotoxins. *American journal of respiratory cell and molecular biology,* 23**,** 146-153.

DENTON, A. E., DOHERTY, P. C., TURNER, S. J. & LA GRUTA, N. L. 2007. IL-18, but not IL-12, is required for optimal cytokine production by influenza virus-specific CD8+ T cells. *European journal of immunology,* 37**,** 368-375.

DEPREZ, M., ZARAGOSI, L.-E., TRUCHI, M., BECAVIN, C., RUIZ GARCÍA, S., ARGUEL, M.-J., PLAISANT, M., MAGNONE, V., LEBRIGAND, K. & ABELANET, S. 2020. A single-cell atlas of the human healthy airways. *American journal of respiratory and critical care medicine,* 202**,** 1636-1645.

DERRIEN, M., VAN PASSEL, M. W., VAN DE BOVENKAMP, J. H., SCHIPPER, R., DE VOS, W. & DEKKER, J. 2010. Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut microbes,* 1**,** 254-268.

DIEBOLD, S. S., KAISHO, T., HEMMI, H., AKIRA, S. & E SOUSA, C. R. 2004. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science,* 303**,** 1529-1531.

DOBNIKAR, L., TAYLOR, A. L., CHAPPELL, J., OLDACH, P., HARMAN, J. L., OERTON, E., DZIERZAK, E., BENNETT, M. R., SPIVAKOV, M. & JØRGENSEN, H. F. 2018. Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nature communications,* 9**,** 1-17.

DOBRICA, M.-O., LAZAR, C. & BRANZA-NICHITA, N. 2020. N-glycosylation and N-glycan processing in HBV biology and pathogenesis. *Cells,* 9**,** 1404.

DOHRMAN, A., MIYATA, S., GALLUP, M., LI, J.-D., CHAPELIN, C., COSTE, A., ESCUDIER, E., NADEL, J. & BASBAUM, C. 1998. Mucin gene (MUC 2 and MUC 5AC) upregulation by Gram-positive and Gram-negative bacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1406**,** 251-259.

DONOGHUE, L. J., LIVRAGHI-BUTRICO, A., MCFADDEN, K. M., THOMAS, J. M., CHEN, G., GRUBB, B. R., O'NEAL, W. K., BOUCHER, R. C. & KELADA, S. N. P. 2017. Identification of trans Protein QTL for Secreted Airway Mucins in Mice and a Causal Role for Bpifb1. *Genetics,* 207**,** 801-812.

DRAGHICI, S., KHATRI, P., EKLUND, A. C. & SZALLASI, Z. 2006. Reliability and reproducibility issues in DNA microarray measurements. *TRENDS in Genetics,* 22**,** 101-109.

DUEZ, J.-M., SIXT, N. & PÉCHINOT, A. 2009. Influenza virus infection: don't forget the role of the mucociliary system! *The Journal of antimicrobial chemotherapy,* 63**,** 421-422.

DURANTE, M. A., KURTENBACH, S., SARGI, Z. B., HARBOUR, J. W., CHOI, R., KURTENBACH, S., GOSS, G. M., MATSUNAMI, H. & GOLDSTEIN, B. J. 2020. Single-cell analysis of olfactory neurogenesis and differentiation in adult humans. *Nature neuroscience,* 23**,** 323-326.

DURAZZO, M., ADRIANI, A., FAGOONEE, S., SARACCO, G. M. & PELLICANO, R. 2021. Helicobacter pylori and respiratory diseases: 2021 update. *Microorganisms,* 9**,** 2033.

DVORAK, A., TILLEY, A. E., SHAYKHIEV, R., WANG, R. & CRYSTAL, R. G. 2011. Do airway epithelium air–liquid cultures represent the in vivo airway epithelium transcriptome? *American journal of respiratory cell and molecular biology,* 44**,** 465-473.

EENJES, E., MERTENS, T. C., BUSCOP-VAN KEMPEN, M. J., VAN WIJCK, Y., TAUBE, C., ROTTIER, R. J. & HIEMSTRA, P. S. 2018. A novel method for expansion and differentiation of mouse tracheal epithelial cells in culture. *Scientific reports,* 8**,** 1-12.

EHRE, C., WORTHINGTON, E. N., LIESMAN, R. M., GRUBB, B. R., BARBIER, D., O'NEAL, W. K., SALLENAVE, J.-M., PICKLES, R. J. & BOUCHER, R. C. 2012. Overexpressing mouse model demonstrates

the protective role of Muc5ac in the lungs. *Proceedings of the National Academy of Sciences,* 109**,** 16528-16533.
ELIEH ALI KOMI, D., SHARMA, L. & DELA CRUZ, C. S. 2018. Chitin and its effects on inflammatory and immune responses. *Clinical reviews in allergy & immunology,* 54**,** 213-223.
ELLISON, R. T., 3RD & GIEHL, T. J. 1991. Killing of gram-negative bacteria by lactoferrin and lysozyme. *J Clin Invest,* 88**,** 1080-91.
ELSBACH, P. & WEISS, J. 1993. The bactericidal/permeability-increasing protein (BPI), a potent element in host-defense against gram-negative bacteria and lipopolysaccharide. *Immunobiology,* 187**,** 417-429.
ELSBACH, P. & WEISS, J. 1998. Role of the bactericidal/permeability-increasing protein in host defence. *Current opinion in immunology,* 10**,** 45-49.
ERHABOR, G., ADENIYI, B., ARAWOMO, A., AKINWALERE, O., ADETONA, G., FAGBOHUN, F., AIGBIRIOR, J. & ERHABOR, J. 2021. Acute Exacerbation of COPD: Clinical Perspectives and Literature Review. *West African Journal of Medicine,* 38**,** 1129-1142.
ERMUND, A., MEISS, L. N., RODRIGUEZ-PINEIRO, A. M., BAHR, A., NILSSON, H. E., TRILLO-MUYO, S., RIDLEY, C., THORNTON, D. J., WINE, J. J., HEBERT, H., KLYMIUK, N. & HANSSON, G. C. 2017. The normal trachea is cleaned by MUC5B mucin bundles from the submucosal glands coated with the MUC5AC mucin. *Biochem Biophys Res Commun,* 492**,** 331-337.
EVANS, C. M., RACLAWSKA, D. S., TTOFALI, F., LIPTZIN, D. R., FLETCHER, A. A., HARPER, D. N., MCGING, M. A., MCELWEE, M. M., WILLIAMS, O. W. & SANCHEZ, E. 2015. The polymeric mucin Muc5ac is required for allergic airway hyperreactivity. *Nature communications,* 6**,** 1-11.
EVANS, M. J., CABRAL-ANDERSON, L. J. & FREEMAN, G. 1978. Role of the Clara cell in renewal of the bronchiolar epithelium. *Lab Invest,* 38**,** 648-53.
EVANS, M. J., VAN WINKLE, L. S., FANUCCHI, M. V. & PLOPPER, C. G. 2001. Cellular and molecular characteristics of basal cells in airway epithelium. *Experimental lung research,* 27**,** 401-415.
EVERMAN, J. L., SAJUTHI, S., SAEF, B., RIOS, C., STONER, A. M., NUMATA, M., HU, D., ENG, C., OH, S. & RODRIGUEZ-SANTANA, J. 2019. Functional genomics of CDHR3 confirms its role in HRV-C infection and childhood asthma exacerbations. *Journal of Allergy and Clinical Immunology,* 144**,** 962-971.
FANG, S., ZHANG, S., DAI, H., HU, X., LI, C. & XING, Y. 2019. The role of pulmonary mesenchymal cells in airway epithelium regeneration during injury repair. *Stem Cell Research & Therapy,* 10**,** 1-12.

FERNANDES, D. J., BONACCI, J. V. & STEWART, A. G. 2006. Extracellular matrix, integrins, and mesenchymal cell function in the airways. *Current drug targets,* 7**,** 567-577.

FINA, L., MOLGAARD, H. V., ROBERTSON, D., BRADLEY, N. J., MONAGHAN, P., DELIA, D., SUTHERLAND, D. R., BAKER, M. A. & GREAVES, M. F. 1990. Expression of the CD34 gene in vascular endothelial cells.

FLETCHER, R. B., DAS, D., GADYE, L., STREET, K. N., BAUDHUIN, A., WAGNER, A., COLE, M. B., FLORES, Q., CHOI, Y. G. & YOSEF, N. 2017. Deconstructing olfactory stem cell trajectories at single-cell resolution. *Cell stem cell,* 20**,** 817-830. e8.

FORREST, I., MURPHY, D., WARD, C., JONES, D., JOHNSON, G., ARCHER, L., GOULD, F., CAWSTON, T., LORDAN, J. & CORRIS, P. 2005. Primary airway epithelial cell culture from lung transplant recipients. *European Respiratory Journal,* 26**,** 1080-1085.

FUJII, K., NAKAMURA, H. & NISHIMURA, T. 2017. Recent mass spectrometry-based proteomics for biomarker discovery in lung cancer, COPD, and asthma. *Expert Review of Proteomics,* 14**,** 373-386.

FUJISAWA, H. 2008. Neutrophils play an essential role in cooperation with antibody in both protection against and recovery from pulmonary infection with influenza virus in mice. *Journal of virology,* 82**,** 2772-2783.

FUJISAWA, T., IDE, K., HOLTZMAN, M. J., SUDA, T., SUZUKI, K., KUROISHI, S., CHIDA, K. & NAKAMURA, H. 2008. Involvement of the p38 MAPK pathway in IL-13-induced mucous cell metaplasia in mouse tracheal epithelial cells. *Respirology,* 13**,** 191-202.

FUKUYAMA, S., KATSURA, H., ZHAO, D., OZAWA, M., ANDO, T., SHOEMAKER, J. E., ISHIKAWA, I., YAMADA, S., NEUMANN, G. & WATANABE, S. 2015. Multi-spectral fluorescent reporter influenza viruses (Color-flu) as powerful tools for in vivo studies. *Nature communications,* 6**,** 1-8.

FULCHER, M. L., GABRIEL, S., BURNS, K. A., YANKASKAS, J. R. & RANDELL, S. H. 2005. Well-differentiated human airway epithelial cell cultures. *Human cell culture protocols.* Springer.

GAGNEUX, P., CHERIYAN, M., HURTADO-ZIOLA, N., VAN DER LINDEN, E. C. B., ANDERSON, D., MCCLURE, H., VARKI, A. & VARKI, N. M. 2003. Human-specific regulation of α2–6-linked sialic acids. *Journal of Biological Chemistry,* 278**,** 48245-48250.

GANZ, T. 2002. Antimicrobial polypeptides in host defense of the respiratory tract. *J Clin Invest,* 109**,** 693-7.

GANZ, T. 2003. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol,* 3**,** 710-20.

GAO, J., OHLMEIER, S., NIEMINEN, P., TOLJAMO, T., TIITINEN, S., KANERVA, T., BINGLE, L., ARAUJO, B., RONTY, M., HOYHTYA, M., BINGLE, C. D., MAZUR, W. & PULKKINEN, V. 2015. Elevated sputum BPIFB1 levels in smokers with chronic obstructive pulmonary disease: a longitudinal study. *Am J Physiol Lung Cell Mol Physiol,* 309**,** L17-26.

GAO, J.-M., XIANG, R.-L., JIANG, L., LI, W.-H., FENG, Q.-P., GUO, Z.-J., SUN, Q., ZENG, Z.-P. & FANG, F.-D. 2009. Sulfated tyrosines 27 and 29 in the N-terminus of human CXCR3 participate in binding native IP-10. *Acta Pharmacologica Sinica,* 30**,** 193-201.

GARCIA-CABALLERO, A., RASMUSSEN, J. E., GAILLARD, E., WATSON, M. J., OLSEN, J. C., DONALDSON, S. H., STUTTS, M. J. & TARRAN, R. 2009. SPLUNC1 regulates airway surface liquid volume by protecting ENaC from proteolytic cleavage. *Proceedings of the National Academy of Sciences,* 106**,** 11412-11417.

GARD, A., LUU, R., MILLER, C., MALONEY, R., CAIN, B., MARR, E., BURNS, D., GAIBLER, R., MULHERN, T. & WONG, C. 2021. High-throughput human primary cell-based airway model for evaluating influenza, coronavirus, or other respiratory viruses in vitro. *Scientific reports,* 11**,** 1-18.

GAZZANO-SANTORO, H., PARENT, J., GRINNA, L., HORWITZ, A., PARSONS, T., THEOFAN, G., ELSBACH, P., WEISS, J. & CONLON, P. 1992. High-affinity binding of the bactericidal/permeability-increasing protein and a recombinant amino-terminal fragment to the lipid A region of lipopolysaccharide. *Infection and immunity,* 60**,** 4754-4761.

GEIGER, T., VELIC, A., MACEK, B., LUNDBERG, E., KAMPF, C., NAGARAJ, N., UHLEN, M., COX, J. & MANN, M. 2013. Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. *Molecular & Cellular Proteomics,* 12**,** 1709-1722.

GENTER, M. B., VAN VELDHOVEN, P. P., JEGGA, A. G., SAKTHIVEL, B., KONG, S., STANLEY, K., WITTE, D. P., EBERT, C. L. & ARONOW, B. J. 2003. Microarray-based discovery of highly expressed olfactory mucosal genes: potential roles in the various functions of the olfactory system. *Physiological genomics,* 16**,** 67-81.

GERACE, E. & MOAZED, D. 2015. Affinity pull-down of proteins using anti-FLAG M2 agarose beads. *Methods in enzymology.* Elsevier.

GERBE, F. & JAY, P. 2016. Intestinal tuft cells: epithelial sentinels linking luminal cues to the immune system. *Mucosal immunology,* 9**,** 1353-1359.

GEYER, H. & GEYER, R. 2006. Strategies for analysis of glycoprotein glycosylation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1764**,** 1853-1869.

GHAFOURI, B., STÅHLBOM, B., TAGESSON, C. & LINDAHL, M. 2002. Newly identified proteins in human nasal lavage fluid from non-smokers

and smokers using two-dimensional gel electrophoresis and peptide mass fingerprinting. *PROTEOMICS: International Edition,* 2**,** 112-120.
GHOSH, M., HELM, K. M., SMITH, R. W., GIORDANENGO, M. S., LI, B., SHEN, H. & REYNOLDS, S. D. 2011. A single cell functions as a tissue-specific stem cell and the in vitro niche-forming cell. *American journal of respiratory cell and molecular biology,* 45**,** 459-469.
GHOSH, N., DUTTA, M., SINGH, B., BANERJEE, R., BHATTACHARYYA, P. & CHAUDHURY, K. 2016. Transcriptomics, proteomics and metabolomics driven biomarker discovery in COPD: an update. *Expert review of molecular diagnostics,* 16**,** 897-913.
GOLD, D. L., MIECZNIKOWSKI, J. C. & LIU, S. 2009. Error control variability in pathway-based microarray analysis. *Bioinformatics,* 25**,** 2216-2221.
GOLDFARBMUREN, K. C., JACKSON, N. D., SAJUTHI, S. P., DYJACK, N., LI, K. S., RIOS, C. L., PLENDER, E. G., MONTGOMERY, M. T., EVERMAN, J. L. & BRATCHER, P. E. 2020. Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. *Nature communications,* 11**,** 1-21.
GOLDMAN, M. J., ANDERSON, G. M., STOLZENBERG, E. D., KARI, U. P., ZASLOFF, M. & WILSON, J. M. 1997. Human beta-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell,* 88**,** 553-60.
GONZÁLEZ-ARRIAGADA, W. A., SANTOS-SILVA, A. R., ITO, F. A., VARGAS, P. A., SPEIGHT, P. M., BINGLE, L. & LOPES, M. A. 2012. Expression pattern of PLUNC proteins as an auxiliary tool for the diagnosis of high-grade mucoepidermoid carcinoma of the salivary gland. *Journal of oral pathology & medicine,* 41**,** 589-597.
GONZALO SANZ, R. & SÁNCHEZ-PLA, A. 2019. Statistical analysis of microarray data. *Microarray Bioinformatics.* Springer.
GOTO, Y., NIWA, Y., SUZUKI, T., UEMATSU, S., DOHMAE, N. & SIMIZU, S. 2014. N-glycosylation is required for secretion and enzymatic activity of human hyaluronidase1. *FEBS Open Bio,* 4**,** 554-559.
GOWERS, K. H., HYNDS, R. E., THAKRAR, R. M., CARROLL, B., BIRCHALL, M. A. & JANES, S. M. 2018. Optimized isolation and expansion of human airway epithelial basal cells from endobronchial biopsy samples. *Journal of tissue engineering and regenerative medicine,* 12**,** e313-e317.
GRAY, P. W., CORCORRAN, A. E., EDDY JR, R. L., BYERS, M. G. & SHOWS, T. B. 1993. The genes for the lipopolysaccharide binding protein (LBP) and the bactericidal permeability increasing protein (BPI) are encoded in the same region of human chromosome 20. *Genomics,* 15**,** 188-190.

GREANEY, A. M., ADAMS, T. S., RAREDON, M. S. B., GUBBINS, E., SCHUPP, J. C., ENGLER, A. J., GHAEDI, M., YUAN, Y., KAMINSKI, N. & NIKLASON, L. E. 2020. Platform effects on regeneration by pulmonary basal cells as evaluated by single-cell RNA sequencing. *Cell reports,* 30**,** 4250-4265. e6.

GRIGGS, T. F., BOCHKOV, Y. A., BASNET, S., PASIC, T. R., BROCKMAN-SCHNEIDER, R. A., PALMENBERG, A. C. & GERN, J. E. 2017. Rhinovirus C targets ciliated airway epithelial cells. *Respiratory research,* 18**,** 1-11.

GRUBE, B. J., COCHANE, C. G., YE, R. D., GREEN, C. E., MCPHAIL, M. E., ULEVITCH, R. J. & TOBIAS, P. S. 1994. Lipopolysaccharide binding protein expression in primary human hepatocytes and HepG2 hepatoma cells. *Journal of Biological Chemistry,* 269**,** 8477-8482.

GUAN, Y., POON, L., CHEUNG, C., ELLIS, T., LIM, W., LIPATOV, A., CHAN, K., STURM-RAMIREZ, K., CHEUNG, C. & LEUNG, Y. 2004. H5N1 influenza: a protean pandemic threat. *Proceedings of the National Academy of Sciences,* 101**,** 8156-8161.

GUNDRY, M. C., BRUNETTI, L., LIN, A., MAYLE, A. E., KITANO, A., WAGNER, D., HSU, J. I., HOEGENAUER, K. A., ROONEY, C. M. & GOODELL, M. A. 2016. Highly efficient genome editing of murine and human hematopoietic progenitor cells by CRISPR/Cas9. *Cell reports,* 17**,** 1453-1461.

GUO, C.-J., LEUNG, P. S., ZHANG, W., MA, X. & GERSHWIN, M. E. 2018. The immunobiology and clinical features of type 1 autoimmune polyglandular syndrome (APS-1). *Autoimmunity reviews,* 17**,** 78-85.

GUO, J., HUANG, F., LIU, J., CHEN, Y., WANG, W., CAO, B., ZOU, Z., LIU, S., PAN, J. & BAO, C. 2015. The serum profile of hypercytokinemia factors identified in H7N9-infected patients can predict fatal outcomes. *Scientific reports,* 5**,** 1-10.

GUPTA, R. & BRUNAK, S. 2001. Prediction of glycosylation across the human proteome and the correlation to protein function.

HABERMANN, A. C., GUTIERREZ, A. J., BUI, L. T., YAHN, S. L., WINTERS, N. I., CALVI, C. L., PETER, L., CHUNG, M.-I., TAYLOR, C. J. & JETTER, C. 2020. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Science advances,* 6**,** eaba1972.

HAIMOVICH, G. & GERST, J. E. 2018. Single-molecule fluorescence in situ hybridization (smFISH) for RNA detection in adherent animal cells. *Bio-protocol,* 8**,** e3070-e3070.

HALLER, O. & KOCHS, G. 2020. Mx genes: host determinants controlling influenza virus infection and trans-species transmission. *Human genetics,* 139**,** 695-705.

HALSTEAD, E. S., UMSTEAD, T. M., DAVIES, M. L., KAWASAWA, Y. I., SILVEYRA, P., HOWYRLAK, J., YANG, L., GUO, W., HU, S. & HEWAGE, E. K. 2018. GM-CSF overexpression after influenza a virus infection prevents mortality and moderates M1-like airway monocyte/macrophage polarization. *Respiratory research,* 19**,** 1-14.

HAN, X., ZHOU, Z., FEI, L., SUN, H., WANG, R., CHEN, Y., CHEN, H., WANG, J., TANG, H. & GE, W. 2020. Construction of a human cell landscape at single-cell level. *Nature,* 581**,** 303-309.

HANSSON, G. C. 2019. Mucus and mucins in diseases of the intestinal and respiratory tracts. *Journal of internal medicine,* 285**,** 479-490.

HARTIKKA, J., SAWDEY, M., CORNEFERT-JENSEN, F., MARGALITH, M., BARNHART, K., NOLASCO, M., VAHLSING, H. L., MEEK, J., MARQUET, M. & HOBART, P. 1996. An improved plasmid DNA expression vector for direct injection into skeletal muscle. *Human gene therapy,* 7**,** 1205-1217.

HARVEY, D. J. 2005. Proteomic analysis of glycosylation: structural determination of N-and O-linked glycans by mass spectrometry. *Expert review of proteomics,* 2**,** 87-101.

HAYE, K., BURMAKINA, S., MORAN, T., GARCÍA-SASTRE, A. & FERNANDEZ-SESMA, A. 2009. The NS1 protein of a human influenza virus inhibits type I interferon production and the induction of antiviral responses in primary human dendritic and respiratory epithelial cells. *Journal of virology,* 83**,** 6849-6862.

HAYNEY, M. S., HENRIQUEZ, K. M., BARNET, J. H., EWERS, T., CHAMPION, H. M., FLANNERY, S. & BARRETT, B. 2017. Serum IFN-γ-induced protein 10 (IP-10) as a biomarker for severity of acute respiratory infection in healthy adults. *Journal of Clinical Virology,* 90**,** 32-37.

HEDLUND, E. & DENG, Q. 2018. Single-cell RNA sequencing: technical advancements and biological applications. *Molecular aspects of medicine,* 59**,** 36-46.

HEER, A. K., SHAMSHIEV, A., DONDA, A., UEMATSU, S., AKIRA, S., KOPF, M. & MARSLAND, B. J. 2007. TLR signaling fine-tunes anti-influenza B cell responses without regulating effector T cell responses. *The journal of immunology,* 178**,** 2182-2191.

HEIGWER, F., KERR, G. & BOUTROS, M. 2014. E-CRISP: fast CRISPR target site identification. *Nature methods,* 11**,** 122-123.

HERR, C., BEISSWENGER, C., HESS, C., KANDLER, K., SUTTORP, N., WELTE, T., SCHROEDER, J. M. & VOGELMEIER, C. 2009. Suppression of pulmonary innate host defence in smokers. *Thorax,* 64**,** 144-9.

HERSHEY, G. K. K. 2003. IL-13 receptors and signaling pathways: an evolving web. *Journal of Allergy and Clinical Immunology,* 111**,** 677-690.
HEWITT, R. J. & LLOYD, C. M. 2021. Regulation of immune responses by the airway epithelial cell landscape. *Nature Reviews Immunology,* 21**,** 347-362.
HÖFER, D., PÜSCHEL, B. & DRENCKHAHN, D. 1996. Taste receptor-like cells in the rat gut identified by expression of alpha-gustducin. *Proceedings of the National Academy of Sciences,* 93**,** 6631-6634.
HOGAN, B. & TATA, P. R. 2019. Cellular organization and biology of the respiratory system. *Nature Cell Biology*.
HONG, K. U., REYNOLDS, S. D., WATKINS, S., FUCHS, E. & STRIPP, B. R. 2004. Basal cells are a multipotent progenitor capable of renewing the bronchial epithelium. *The American journal of pathology,* 164**,** 577-588.
HONG, M., TAO, S., ZHANG, L., DIAO, L.-T., HUANG, X., HUANG, S., XIE, S.-J., XIAO, Z.-D. & ZHANG, H. 2020. RNA sequencing: new technologies and applications in cancer research. *Journal of hematology & oncology,* 13**,** 1-16.
HONG, S.-J. 2012. The role of Mycoplasma pneumoniae infection in asthma. *Allergy, asthma & immunology research,* 4**,** 59-61.
HONIGFORT, D. J., ALTMAN, M. O., GAGNEUX, P. & GODULA, K. 2021. Glycocalyx crowding with mucin mimetics strengthens binding of soluble and virus-associated lectins to host cell glycan receptors. *Proceedings of the National Academy of Sciences,* 118**,** e2107896118.
HOU, J., YASHIRO, K., OKAZAKI, Y., SAIJOH, Y., HAYASHIZAKI, Y. & HAMADA, H. 2004. Identification of a novel left–right asymmetrically expressed gene in the mouse belonging to the BPI/PLUNC superfamily. *Developmental dynamics: an official publication of the American Association of Anatomists,* 229**,** 373-379.
HOU, X., GAO, Y., YANG, S., WANG, C., MA, Z. & XIA, X. 2009. Role of macrophage migration inhibitory factor in influenza H5N1 virus pneumonia. *Acta virologica,* 53**,** 225-231.
HOWITT, M. R., CAO, Y. G., GOLOGORSKY, M. B., LI, J. A., HABER, A. L., BITON, M., LANG, J., MICHAUD, M., REGEV, A. & GARRETT, W. S. 2020. The taste receptor TAS1R3 regulates small intestinal tuft cell homeostasis. *Immunohorizons,* 4**,** 23-32.
HOWITT, M. R., LAVOIE, S., MICHAUD, M., BLUM, A. M., TRAN, S. V., WEINSTOCK, J. V., GALLINI, C. A., REDDING, K., MARGOLSKEE, R. F. & OSBORNE, L. C. 2016. Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut. *Science,* 351**,** 1329-1333.
HSU, P. D., SCOTT, D. A., WEINSTEIN, J. A., RAN, F., KONERMANN, S., AGARWALA, V., LI, Y., FINE, E. J., WU, X. & SHALEM, O. 2013.

DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology,* 31**,** 827-832.
HU, Y., SHAH, P., CLARK, D. J., AO, M. & ZHANG, H. 2018. Reanalysis of global proteomic and phosphoproteomic data identified a large number of glycopeptides. *Analytical chemistry,* 90**,** 8065-8071.
HUANG, D. W., SHERMAN, B. T., TAN, Q., COLLINS, J. R., ALVORD, W. G., ROAYAEI, J., STEPHENS, R., BASELER, M. W., LANE, H. C. & LEMPICKI, R. A. 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology,* 8**,** 1-16.
HUANG, F.-F., BARNES, P. F., FENG, Y., DONIS, R., CHRONEOS, Z. C., IDELL, S., ALLEN, T., PEREZ, D. R., WHITSETT, J. A. & DUNUSSI-JOANNOPOULOS, K. 2011. GM-CSF in the lung protects against lethal influenza infection. *American journal of respiratory and critical care medicine,* 184**,** 259-268.
HUANG, Y.-H., KLINGBEIL, O., HE, X.-Y., WU, X. S., ARUN, G., LU, B., SOMERVILLE, T. D., MILAZZO, J. P., WILKINSON, J. E. & DEMERDASH, O. E. 2018. POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. *Genes & development,* 32**,** 915-928.
HUBACEK, J. A., BÜCHLER, C., ASLANIDIS, C. & SCHMITZ, G. 1997. The genomic organization of the genes for human lipopolysaccharide binding protein (LBP) and bactericidal permeability increasing protein (BPI) is highly conserved. *Biochemical and biophysical research communications,* 236**,** 427-430.
HUH, D., MATTHEWS, B. D., MAMMOTO, A., MONTOYA-ZAVALA, M., HSIN, H. Y. & INGBER, D. E. 2010. Reconstituting organ-level lung functions on a chip. *Science,* 328**,** 1662-1668.
HUNTER, M., YUAN, P., VAVILALA, D. & FOX, M. 2019. Optimization of protein expression in mammalian cells. *Current protocols in protein science,* 95**,** e77.
HURST, L. D. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *TRENDS in Genetics,* 18**,** 486-487.
IBRAHIM, H. R., AOKI, T. & PELLEGRINI, A. 2002. Strategies for new antimicrobial proteins and peptides: lysozyme and aprotinin as model molecules. *Curr Pharm Des,* 8**,** 671-93.
IMPERIALI, B. & O'CONNOR, S. E. 1999. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Current opinion in chemical biology,* 3**,** 643-649.
IOANNIDIS, I., MCNALLY, B., WILLETTE, M., PEEPLES, M. E., CHAUSSABEL, D., DURBIN, J. E., RAMILO, O., MEJIAS, A. & FLAÑO, E. 2012. Plasticity and virus specificity of the airway epithelial cell immune response during respiratory virus infection. *Journal of virology,* 86**,** 5422-5436.

IOVINE, N., EASTVOLD, J., ELSBACH, P., WEISS, J. P. & GIOANNINI, T. L. 2002. The carboxyl-terminal domain of closely related endotoxin-binding proteins determines the target of protein-lipopolysaccharide complexes. *Journal of Biological Chemistry,* 277**,** 7970-7978.
IOVINE, N. M., ELSBACH, P. & WEISS, J. 1997. An opsonic function of the neutrophil bactericidal/permeability-increasing protein depends on both its N-and C-terminal domains. *Proceedings of the National Academy of Sciences,* 94**,** 10973-10978.
IRONS, S. L., CHAMBERS, A. C., LISSINA, O., KING, L. A. & POSSEE, R. D. 2018. Protein production using the baculovirus expression system. *Current Protocols in Protein Science,* 91**,** 5.5. 1-5.5. 22.
ISHIGURO, N., TAKADA, A., YOSHIOKA, M., MA, X., KIKUTA, H., KIDA, H. & KOBAYASHI, K. 2003. Induction of interferon-inducible protein-10 and monokine induced by interferon-γ from human endothelial cells infected with influenza A virus. *Archives of virology,* 149**,** 17-34.
ISHIKAWA, H., FUKUI, T., INO, S., SASAKI, H., AWANO, N., KOHDA, C. & TANAKA, K. 2016. Influenza virus infection causes neutrophil dysfunction through reduced G-CSF production and an increased risk of secondary bacteria infection in the lung. *Virology,* 499**,** 23-29.
ITO, S., YAMAMOTO, Y. & KIMURA, K. 2020. Analysis of ciliogenesis process in the bovine oviduct based on immunohistochemical classification. *Molecular biology reports,* 47**,** 1003-1012.
IVASHKIV, L. B. & DONLIN, L. T. 2014. Regulation of type I interferon responses. *Nature Reviews Immunology,* 14**,** 36-49.
JAKSIK, R., IWANASZKO, M., RZESZOWSKA-WOLNY, J. & KIMMEL, M. 2015. Microarray experiments and factors which affect their reliability. *Biology direct,* 10**,** 1-14.
JANEWAY, C. A. Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harbor symposia on quantitative biology, 1989. Cold Spring Harbor Laboratory Press, 1-13.
JAYAPRAKASH, N. G. & SUROLIA, A. 2017. Role of glycosylation in nucleating protein folding and stability. *Biochemical Journal,* 474**,** 2333-2347.
JAYAWARDANE, G., RUSSELL, G. C., THOMSON, J., DEANE, D., COX, H., GATHERER, D., ACKERMANN, M., HAIG, D. M. & STEWART, J. P. 2008. A captured viral interleukin 10 gene with cellular exon structure. *Journal of General Virology,* 89**,** 2447-2455.
JEFFERY, P. & LI, D. 1997. Airway mucosa: secretory cells, mucus and mucin genes. *European Respiratory Journal,* 10**,** 1655-1662.
JIANG, D., WENZEL, S. E., WU, Q., BOWLER, R. P., SCHNELL, C. & CHU, H. W. 2013. Human neutrophil elastase degrades SPLUNC1 and impairs airway epithelial defense against bacteria. *PloS one,* 8**,** e64689.

JIANG, M., ÖSTERLUND, P., SARIN, L. P., PORANEN, M. M., BAMFORD, D. H., GUO, D. & JULKUNEN, I. 2011. Innate immune responses in human monocyte-derived dendritic cells are highly dependent on the size and the 5′ phosphorylation of RNA molecules. *The Journal of immunology,* 187**,** 1713-1721.

JIANG, X., DENG, X., WANG, J., MO, Y., SHI, L., WEI, F., ZHANG, S., GONG, Z., HE, Y. & XIONG, F. 2022. BPIFB1 inhibits vasculogenic mimicry via downregulation of GLUT1-mediated H3K27 acetylation in nasopharyngeal carcinoma. *Oncogene,* 41**,** 233-245.

JOHNSTON, S. L., GOLDBLATT, D. L., EVANS, S. E., TUVIM, M. J. & DICKEY, B. F. 2021. Airway Epithelial Innate Immunity. *Frontiers in physiology***,** 2090.

JOO, N. S., EVANS, I. A. T., CHO, H.-J., PARK, I.-H., ENGELHARDT, J. F. & WINE, J. J. 2015. Proteomic analysis of pure human airway gland mucus reveals a large component of protective proteins. *PloS one,* 10**,** e0116756.

JOSHI, N., WALTER, J. M. & MISHARIN, A. V. 2018. Alveolar macrophages. *Cellular immunology,* 330**,** 86-90.

KALENIK, B. M., GÓRA-SOCHACKA, A. & SIRKO, A. 2018. B-defensins–Underestimated peptides in influenza combat. *Virus Research,* 247**,** 10-14.

KALLIO, H., HILVO, M., RODRIGUEZ, A., LAPPALAINEN, E.-H., LAPPALAINEN, A.-M. & PARKKILA, S. 2010. Global transcriptional response to carbonic anhydrase IX deficiency in the mouse stomach. *BMC genomics,* 11**,** 1-14.

KANOH, S., TANABE, T. & RUBIN, B. 2011. IL-13-induced MUC5AC production and goblet cell differentiation is steroid resistant in human airway cells. *Clinical & Experimental Allergy,* 41**,** 1747-1756.

KAO, W. W.-Y., FUNDERBURGH, J. L., XIA, Y., LIU, C.-Y. & CONRAD, G. W. 2006. Focus on molecules: lumican. *Experimental eye research,* 82**,** 3.

KAPLANSKI, G. 2018. Interleukin-18: Biological properties and role in disease pathogenesis. *Immunological reviews,* 281**,** 138-153.

KASH, J. C., TUMPEY, T. M., PROLL, S. C., CARTER, V., PERWITASARI, O., THOMAS, M. J., BASLER, C. F., PALESE, P., TAUBENBERGER, J. K. & GARCÍA-SASTRE, A. 2006. Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature,* 443**,** 578-581.

KATO, H., SATO, S., YONEYAMA, M., YAMAMOTO, M., UEMATSU, S., MATSUI, K., TSUJIMURA, T., TAKEDA, K., FUJITA, T. & TAKEUCHI, O. 2005. Cell type-specific involvement of RIG-I in antiviral response. *Immunity,* 23**,** 19-28.

KATO, K., CHANG, E. H., CHEN, Y., LU, W., KIM, M. M., NIIHORI, M., HECKER, L. & KIM, K. C. 2020. MUC1 contributes to goblet cell metaplasia and MUC5AC expression in response to cigarette smoke in vivo. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 319**,** L82-L90.

KATO, Z., JEE, J., SHIKANO, H., MISHIMA, M., OHKI, I., OHNISHI, H., LI, A., HASHIMOTO, K., MATSUKUMA, E. & OMOYA, K. 2003. The structure and binding mode of interleukin-18. *Nature Structural & Molecular Biology,* 10**,** 966-971.

KATSOULIDIS, E., CARAYOL, N., WOODARD, J., KONIECZNA, I., MAJCHRZAK-KITA, B., JORDAN, A., SASSANO, A., EKLUND, E. A., FISH, E. N. & PLATANIAS, L. C. 2009. Role of Schlafen 2 (SLFN2) in the generation of interferon α-induced growth inhibitory responses. *Journal of Biological Chemistry,* 284**,** 25051-25064.

KATSOULIDIS, E., MAVROMMATIS, E., WOODARD, J., SHIELDS, M. A., SASSANO, A., CARAYOL, N., SAWICKI, K. T., MUNSHI, H. G. & PLATANIAS, L. C. 2010. Role of interferon α (IFNα)-inducible Schlafen-5 in regulation of anchorage-independent growth and invasion of malignant melanoma cells. *Journal of Biological Chemistry,* 285**,** 40333-40341.

KAVSAN, V. M., IERSHOV, A. V. & BALYNSKA, O. V. 2011. Immortalized cells and one oncogene in malignant transformation: old insights on new explanation. *BMC cell biology,* 12**,** 1-2.

KAWAI, T. & AKIRA, S. 2008. Toll-like receptor and RIG-1-like receptor signaling. *Annals of the New York Academy of Sciences,* 1143**,** 1-20.

KE, N. & BERKMEN, M. 2014. Production of disulfide-bonded proteins in Escherichia coli. *Current protocols in molecular biology,* 108**,** 16.1 B. 1-16.1 B. 21.

KELL, A. M. & GALE JR, M. 2015. RIG-I in RNA virus recognition. *Virology,* 479**,** 110-121.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols,* 10**,** 845-858.

KILBOURNE, E. D. 1969. Future influenza vaccines and the use of genetic recombinants. *Bulletin of the World Health Organization,* 41**,** 643.

KILLIP, M. J., FODOR, E. & RANDALL, R. E. 2015. Influenza virus activation of the interferon system. *Virus research,* 209**,** 11-22.

KIM, D. S., CHOI, D. & HAHN, Y. 2015. Loss of ancestral N-glycosylation sites in conserved proteins during human evolution. *International Journal of Molecular Medicine,* 36**,** 1685-1692.

KIM, J. Y., SAJJAN, U. S., KRASAN, G. P. & LIPUMA, J. J. 2005. Disruption of tight junctions during traversal of the respiratory epithelium by Burkholderia cenocepacia. *Infection and immunity,* 73**,** 7107-7112.

KIM, S.-Y., HWANG, J.-S. & HAN, I.-O. 2013. Tunicamycin inhibits Toll-like receptor-activated inflammation in RAW264. 7 cells by suppression of NF-κB and c-Jun activity via a mechanism that is independent of ER-stress and N-glycosylation. *European journal of pharmacology,* 721**,** 294-300.

KIRSCHNING, C. J., AU-YOUNG, J., LAMPING, N., REUTER, D., PFEIL, D., SEILHAMER, J. J. & SCHUMANN, R. R. 1997. Similar organization of the lipopolysaccharide-binding protein (LBP) and phospholipid transfer protein (PLTP) genes suggests a common gene family of lipid-binding proteins. *Genomics,* 46**,** 416-425.

KLEINER-GROTE, G. R., RISSE, J. M. & FRIEHS, K. 2018. Secretion of recombinant proteins from E. coli. *Engineering in Life Sciences,* 18**,** 532-550.

KLEINSTIVER, B. P., PATTANAYAK, V., PREW, M. S., TSAI, S. Q., NGUYEN, N. T., ZHENG, Z. & JOUNG, J. K. 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature,* 529**,** 490-495.

KNIGHT, D. A. & HOLGATE, S. T. 2003. The airway epithelium: structural and functional properties in health and disease. *Respirology,* 8**,** 432-446.

KNOWLES, M. R. & BOUCHER, R. C. 2002. Mucus clearance as a primary innate defense mechanism for mammalian airways. *The Journal of clinical investigation,* 109**,** 571-577.

KOBASA, D., TAKADA, A., SHINYA, K., HATTA, M., HALFMANN, P., THERIAULT, S., SUZUKI, H., NISHIMURA, H., MITAMURA, K. & SUGAYA, N. 2004. Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. *Nature,* 431**,** 703-707.

KOCH, S. & NUSRAT, A. 2009. Dynamic regulation of epithelial cell fate and barrier function by intercellular junctions. *Annals of the New York Academy of Sciences,* 1165**,** 220-227.

KONDO, M., TAMAOKI, J., TAKEYAMA, K., ISONO, K., KAWATANI, K., IZUMO, T. & NAGAI, A. 2006. Elimination of IL-13 reverses established goblet cell metaplasia into ciliated epithelia in airway epithelial cell culture. *Allergology International,* 55**,** 329-336.

KOYAMA, S., ISHII, K. J., COBAN, C. & AKIRA, S. 2008. Innate immune response to viral infection. *Cytokine,* 43**,** 336-341.

KRAMMER, F., SMITH, G., FOUCHIER, R., PEIRIS, M., KEDZIERSKA, K., DOHERTY, P., PALESE, P., SHAW, M., TREANOR, J. & WEBSTER, R. 2018a. Influenza. vol. 4. *Nat. Rev. Dis. Primers***,** 3.

KRAMMER, F., SMITH, G. J. D., FOUCHIER, R. A. M., PEIRIS, M., KEDZIERSKA, K., DOHERTY, P. C., PALESE, P., SHAW, M. L., TREANOR, J., WEBSTER, R. G. & GARCÍA-SASTRE, A. 2018b. Influenza. *Nature Reviews Disease Primers,* 4**,** 3.

KRASITY, B. C., TROLL, J. V., WEISS, J. P. & MCFALL-NGAI, M. J. 2011. LBP/BPI proteins and their relatives: conservation over evolution and roles in mutualism. *Biochem Soc Trans,* 39**,** 1039-44.

KRASTEVA, G., HARTMANN, P., PAPADAKIS, T., BODENBENNER, M., WESSELS, L., WEIHE, E., SCHÜTZ, B., LANGHEINRICH, A., CHUBANOV, V. & GUDERMANN, T. 2012. Cholinergic chemosensory cells in the auditory tube. *Histochemistry and cell biology,* 137**,** 483-497.

KRISHNAN, A., LI, X., KAO, W.-Y., VIKER, K., BUTTERS, K., MASUOKA, H., KNUDSEN, B., GORES, G. & CHARLTON, M. 2012. Lumican, an extracellular matrix proteoglycan, is a novel requisite for hepatic fibrosis. *Laboratory investigation,* 92**,** 1712-1725.

KROEKER, A. L., EZZATI, P., HALAYKO, A. J. & COOMBS, K. M. 2012. Response of primary human airway epithelial cells to influenza infection: a quantitative proteomic study. *Journal of proteome research,* 11**,** 4132-4146.

KROLL, K. M., BARKEMA, G. T. & CARLON, E. 2008. Modeling background intensity in DNA microarrays. *Physical Review E,* 77**,** 061915.

KUMAR, A., HAQUE, J., LACOSTE, J., HISCOTT, J. & WILLIAMS, B. 1994. Double-stranded RNA-dependent protein kinase activates transcription factor NF-kappa B by phosphorylating I kappa B. *Proceedings of the National Academy of Sciences,* 91**,** 6288-6292.

KUPERMAN, D. A., HUANG, X., KOTH, L. L., CHANG, G. H., DOLGANOV, G. M., ZHU, Z., ELIAS, J. A., SHEPPARD, D. & ERLE, D. J. 2002. Direct effects of interleukin-13 on epithelial cells cause airway hyperreactivity and mucus overproduction in asthma. *Nature medicine,* 8**,** 885-889.

LAMB, R. & KRUG, R. 1996. Orthomyxoviridae: the viruses and their replication. Fields Virology. *Fields, BN, Knipe, RM & Chanock, MS Lippincott-Raven, Philadelphia*.

LAMBKIN-WILLIAMS, R., NOULIN, N., MANN, A., CATCHPOLE, A. & GILBERT, A. S. 2018. The human viral challenge model: accelerating the evaluation of respiratory antivirals, vaccines and novel diagnostics. *Respiratory research,* 19**,** 1-15.

LAOUKILI, J., PERRET, E., WILLEMS, T., MINTY, A., PARTHOENS, E., HOUCINE, O., COSTE, A., JORISSEN, M., MARANO, F. & CAPUT, D. 2001. IL-13 alters mucociliary differentiation and ciliary beating of human respiratory epithelial cells. *The Journal of clinical investigation,* 108**,** 1817-1824.

LAW, C. W., CHEN, Y., SHI, W. & SMYTH, G. K. 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology,* 15**,** 1-17.

LECLAIR, E., NGUYEN, L., BINGLE, L., MACGOWAN, A., SINGLETON, V., WARD, S. & BINGLE, C. 2001. Genomic organization of the mouse plunc gene and expression in the developing airways and thymus. *Biochemical and biophysical research communications,* 284**,** 792-797.

LECLAIR, E., NOMELLINI, V., BAHENA, M., SINGLETON, V., BINGLE, L., CRAVEN, C. & BINGLE, C. 2004. Cloning and expression of a mouse member of the PLUNC protein family exclusively expressed in tongue epithelium. *Genomics,* 83**,** 658-666.

LEEMING, G. H., KIPAR, A., HUGHES, D. J., BINGLE, L., BENNETT, E., MOYO, N. A., TRIPP, R. A., BIGLEY, A. L., BINGLE, C. D. & SAMPLE, J. T. 2015. Gammaherpesvirus infection modulates the temporal and spatial expression of SCGB1A1 (CCSP) and BPIFA1 (SPLUNC1) in the respiratory tract. *Laboratory investigation,* 95**,** 610-624.

LEHMANN, R., LEE, C. M., SHUGART, E. C., BENEDETTI, M., CHARO, R. A., GARTNER, Z., HOGAN, B., KNOBLICH, J., NELSON, C. M. & WILSON, K. M. 2019. Human organoids: a new dimension in cell biology. *Molecular biology of the cell,* 30**,** 1129-1137.

LI, J., SHOU, J., GUO, Y., TANG, Y., WU, Y., JIA, Z., ZHAI, Y., CHEN, Z., XU, Q. & WU, Q. 2015. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *Journal of molecular cell biology,* 7**,** 284-298.

LI, J., XU, P., WANG, L., FENG, M., CHEN, D., YU, X. & LU, Y. 2020. Molecular biology of BPIFB1 and its advances in disease. *Annals of Translational Medicine,* 8.

LI, J. & YE, Z. 2020. The potential role and regulatory mechanisms of MUC5AC in chronic obstructive pulmonary disease. *Molecules,* 25**,** 4437.

LI, M., KAO, E., GAO, X., SANDIG, H., LIMMER, K., PAVON-ETERNOD, M., JONES, T. E., LANDRY, S., PAN, T. & WEITZMAN, M. D. 2012. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature,* 491**,** 125-128.

LI, N., QI, Y., ZHANG, F., YU, X., WU, Y., CHEN, Y., JIANG, C. & KONG, W. 2011. Overexpression of α-2, 6 sialyltransferase stimulates propagation of human influenza viruses in Vero cells. *Acta virologica,* 55**,** 147-153.

LI, N., QIU, L., ZENG, C., FANG, Z., CHEN, S., SONG, X., SONG, H. & ZHANG, G. 2021a. Bioinformatic analysis of differentially expressed genes and pathways in idiopathic pulmonary fibrosis. *Annals of Translational Medicine,* 9.

LI, S. & XIANG, M. 2011. Foxn4 influences alveologenesis during lung development. *Developmental Dynamics,* 240**,** 1512-1517.

LI, W., ZHANG, Q., WANG, X., WANG, H., ZUO, W., XIE, H., TANG, J., WANG, M., ZENG, Z. & CAI, W. 2021b. Comparative Proteomic Analysis to Investigate the Pathogenesis of Oral Adenoid Cystic Carcinoma. *ACS omega,* 6**,** 18623-18634.

LI, Y., JIANG, Z., XUE, D., DENG, G., LI, M., LIU, X. & WANG, Y. 2016. Mycoplasma ovipneumoniae induces sheep airway epithelial cell apoptosis through an ERK signalling-mediated mitochondria pathway. *BMC microbiology,* 16**,** 1-13.

LIAO, Q., ZENG, Z., GUO, X., LI, X., WEI, F., ZHANG, W., LI, X., CHEN, P., LIANG, F., XIANG, B., MA, J., WU, M., TANG, H., DENG, M., ZENG, X., TANG, K., XIONG, W. & LI, G. 2014. LPLUNC1 suppresses IL-6-induced nasopharyngeal carcinoma cell proliferation via inhibiting the Stat3 activation. *Oncogene,* 33**,** 2098-109.

LINO, C. A., HARPER, J. C., CARNEY, J. P. & TIMLIN, J. A. 2018. Delivering CRISPR: a review of the challenges and approaches. *Drug delivery,* 25**,** 1234-1257.

LIU, B., LAGUE, J. R., NUNES, D. P., TOSELLI, P., OPPENHEIM, F. G., SOARES, R. V., TROXLER, R. F. & OFFNER, G. D. 2002. Expression of membrane-associated mucins MUC1 and MUC4 in major human salivary glands. *Journal of Histochemistry & Cytochemistry,* 50**,** 811-820.

LIU, B., MORI, I., HOSSAIN, M. J., DONG, L., TAKEDA, K. & KIMURA, Y. 2004. Interleukin-18 improves the early defence system against influenza virus infection by augmenting natural killer cell-mediated cytotoxicity. *Journal of general virology,* 85**,** 423-428.

LIU, G., BETTS, C., CUNOOSAMY, D. M., ÅBERG, P. M., HORNBERG, J. J., SIVARS, K. B. & COHEN, T. S. 2019. Use of precision cut lung slices as a translational model for the study of lung biology. *Respiratory Research,* 20**,** 162.

LIU, M., GUO, S., HIBBERT, J. M., JAIN, V., SINGH, N., WILSON, N. O. & STILES, J. K. 2011. CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. *Cytokine & growth factor reviews,* 22**,** 121-130.

LIU, M., HUANG, L. Z., SMITS, A. A., BÜLL, C., NARIMATSU, Y., VAN KUPPEVELD, F. J., CLAUSEN, H., DE HAAN, C. A. & DE VRIES, E. 2022. Human-type sialic acid receptors contribute to avian influenza A virus binding and entry by hetero-multivalent interactions. *Nature Communications,* 13**,** 1-12.

LIU, Q., ZHOU, Y.-H. & YANG, Z.-Q. 2016. The cytokine storm of severe influenza and development of immunomodulatory therapy. *Cellular & molecular immunology,* 13**,** 3-10.

LIU, S.-Y., SANCHEZ, D. J., ALIYARI, R., LU, S. & CHENG, G. 2012. Systematic identification of type I and type II interferon-induced antiviral

factors. *Proceedings of the National Academy of Sciences,* 109**,** 4239-4244.
LIU, W. & RODGERS, G. P. Olfactomedin 4 is a biomarker for the severity of infectious diseases. Open Forum Infectious Diseases, 2022. Oxford University Press US, ofac061.
LIU, Y., BARTLETT, J. A., DI, M. E., BOMBERGER, J. M., CHAN, Y. R., GAKHAR, L., MALLAMPALLI, R. K., MCCRAY JR, P. B. & DI, Y. P. 2013. SPLUNC1/BPIFA1 contributes to pulmonary host defense against Klebsiella pneumoniae respiratory infection. *The American journal of pathology,* 182**,** 1519-1531.
LOPEZ-SOUZA, N., AVILA, P. C. & WIDDICOMBE, J. H. 2003. Polarized cultures of human airway epithelium from nasal scrapings and bronchial brushings. *In Vitro Cellular & Developmental Biology-Animal,* 39**,** 266-269.
LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology,* 15**,** 1-21.
LUCIANO, L., CASTELLUCCI, M. & REALE, E. 1981. The brush cells of the common bile duct of the rat. *Cell and tissue research,* 218**,** 403-420.
LUCIANO, L. & REALE, E. 1997. Presence of brush cells in the mouse gallbladder. *Microscopy research and technique,* 38**,** 598-608.
LUND, J. M., ALEXOPOULOU, L., SATO, A., KAROW, M., ADAMS, N. C., GALE, N. W., IWASAKI, A. & FLAVELL, R. A. 2004. Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 5598-5603.
LUNDBERG, A. S., RANDELL, S. H., STEWART, S. A., ELENBAAS, B., HARTWELL, K. A., BROOKS, M. W., FLEMING, M. D., OLSEN, J. C., MILLER, S. W. & WEINBERG, R. A. 2002. Immortalization and transformation of primary human airway epithelial cells by gene transfer. *Oncogene,* 21**,** 4577-4586.
LUO, W., TIAN, P., WANG, Y., XU, H., CHEN, L., TANG, C., SHU, Y., ZHANG, S., WANG, Z. & ZHANG, J. 2018. Characteristics of genomic alterations of lung adenocarcinoma in young never-smokers. *International journal of cancer,* 143**,** 1696-1705.
LUO, Y., ZHANG, M., CHEN, Y., CHEN, Y. & ZHU, D. 2021. Application of human induced pluripotent stem cell-derived cellular and organoid models for COVID-19 research. *Frontiers in Cell and Developmental Biology,* 9.
MAHER, C. A., KUMAR-SINHA, C., CAO, X., KALYANA-SUNDARAM, S., HAN, B., JING, X., SAM, L., BARRETTE, T., PALANISAMY, N. & CHINNAIYAN, A. M. 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature,* 458**,** 97-101.

MAKARENKOVA, H. P. & DARTT, D. A. 2015. Myoepithelial cells: their origin and function in lacrimal gland morphogenesis, homeostasis, and repair. *Current molecular biology reports,* 1**,** 115-123.
MALEY, F., TRIMBLE, R. B., TARENTINO, A. L. & PLUMMER JR, T. H. 1989. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Analytical biochemistry,* 180**,** 195-204.
MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-826.
MANNING, G., PLOWMAN, G. D., HUNTER, T. & SUDARSANAM, S. 2002a. Evolution of protein kinase signaling from yeast to man. *Trends in biochemical sciences,* 27**,** 514-520.
MANNING, G., WHYTE, D. B., MARTINEZ, R., HUNTER, T. & SUDARSANAM, S. 2002b. The protein kinase complement of the human genome. *Science,* 298**,** 1912-1934.
MANSELL, A. & TATE, M. D. 2018. In vivo infection model of severe influenza A virus. *Inflammation and Cancer.* Springer.
MARIB, E. N. 2001. *Human Anatomy and Phsiology*.
MARTINEZ-ANTON, A., SOKOLOWSKA, M., KERN, S., DAVIS, A. S., ALSAATY, S., TAUBENBERGER, J. K., SUN, J., CAI, R., DANNER, R. L. & EBERLEIN, M. 2013. Changes in microRNA and mRNA expression with differentiation of human bronchial epithelial cells. *American journal of respiratory cell and molecular biology,* 49**,** 384-395.
MARUYAMA, E. O., AURE, M. H., XIE, X., MYAL, Y., GAN, L. & OVITT, C. E. 2016. Cell-specific Cre strains for genetic manipulation in salivary glands. *PloS one,* 11**,** e0146711.
MATROSOVICH, M., GAMBARYAN, A., TENEBERG, S., PISKAREV, V., YAMNIKOVA, S., LVOV, D., ROBERTSON, J. & KARLSSON, K.-A. 1997. Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology,* 233**,** 224-234.
MATSUKURA, S., STELLATO, C., GEORAS, S. N., CASOLARO, V., PLITT, J. R., MIURA, K., KUROSAWA, S., SCHINDLER, U. & SCHLEIMER, R. P. 2001. Interleukin-13 upregulates eotaxin expression in airway epithelial cells by a STAT6-dependent mechanism. *American journal of respiratory cell and molecular biology,* 24**,** 755-761.
MATSUYA, Y., HAMADA, N., YACHI, Y., SATOU, Y., ISHIKAWA, M., DATE, H. & SATO, T. 2022. Inflammatory Signaling and DNA Damage Responses after Local Exposure to an Insoluble Radioactive Microparticle. *Cancers,* 14**,** 1045.

MCAULEY, J., CORCILIUS, L., TAN, H., PAYNE, R., MCGUCKIN, M. & BROWN, L. 2017. The cell surface mucin MUC1 limits the severity of influenza A virus infection. *Mucosal immunology,* 10**,** 1581-1593.
MCELVANEY, N. G., NAKAMURA, H., BIRRER, P., HEBERT, C. A., WONG, W. L., ALPHONSO, M., BAKER, J. B., CATALANO, M. A. & CRYSTAL, R. G. 1992. Modulation of airway inflammation in cystic fibrosis. In vivo suppression of interleukin-8 levels on the respiratory epithelial surface by aerosolization of recombinant secretory leukoprotease inhibitor. *J Clin Invest,* 90**,** 1296-301.
MCGILL, J., HEUSEL, J. W. & LEGGE, K. L. 2009. Innate immune control and regulation of influenza virus infections. *Journal of leukocyte biology,* 86**,** 803-812.
MCGILLIVARY, G. & BAKALETZ, L. O. 2010. The multifunctional host defense peptide SPLUNC1 is critical for homeostasis of the mammalian upper airway. *PloS one,* 5**,** e13224.
MEDZHITOV, R. 2001. Toll-like receptors and innate immunity. *Nature Reviews Immunology,* 1**,** 135.
MERCER, R. R., RUSSELL, M. L. & CRAPO, J. 1994a. Alveolar septal structure in different species. *Journal of Applied Physiology,* 77**,** 1060-1066.
MERCER, R. R., RUSSELL, M. L., ROGGLI, V. L. & CRAPO, J. D. 1994b. Cell number and distribution in human and rat airways. *American journal of respiratory cell and molecular biology,* 10**,** 613-624.
MERTENS, T. C., HIEMSTRA, P. S. & TAUBE, C. 2016. Azithromycin differentially affects the IL-13-induced expression profile in human bronchial epithelial cells. *Pulmonary pharmacology & therapeutics,* 39**,** 14-20.
MIAH, K. M., HYDE, S. C. & GILL, D. R. 2019. Emerging gene therapies for cystic fibrosis. *Expert Review of Respiratory Medicine,* 13**,** 709-725.
MIFSUD, E. J., KUBA, M. & BARR, I. G. 2021. Innate immune responses to influenza virus infections in the upper respiratory tract. *Viruses,* 13**,** 2090.
MILLER, A. J., YU, Q., CZERWINSKI, M., TSAI, Y.-H., CONWAY, R. F., WU, A., HOLLOWAY, E. M., WALKER, T., GLASS, I. A. & TREUTLEIN, B. 2020. In vitro and in vivo development of the human airway at single-cell resolution. *Developmental cell,* 53**,** 117-128. e6.
MILLER, J. A., MENON, V., GOLDY, J., KAYKAS, A., LEE, C.-K., SMITH, K. A., SHEN, E. H., PHILLIPS, J. W., LEIN, E. S. & HAWRYLYCZ, M. J. 2014. Improving reliability and absolute quantification of human brain microarray data by filtering and scaling probes using RNA-Seq. *BMC genomics,* 15**,** 1-14.
MOFFITT, R. A., YIN-GOEN, Q., STOKES, T. H., PARRY, R. M., TORRANCE, J. H., PHAN, J. H., YOUNG, A. N. & WANG, M. D. 2011. caCORRECT2: Improving the accuracy and reliability of

microarray data in the presence of artifacts. *BMC bioinformatics,* 12**,** 1-17.
MONTORO, D. T., HABER, A. L., BITON, M., VINARSKY, V., LIN, B., BIRKET, S. E., YUAN, F., CHEN, S., LEUNG, H. M. & VILLORIA, J. 2018. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature,* 560**,** 319-324.
MOORE, M., CHUA, W., CHARLES, K. & CLARKE, S. 2010. Inflammation and cancer: causes and consequences. *Clinical Pharmacology & Therapeutics,* 87**,** 504-508.
MORIMOTO, M., NISHINAKAMURA, R., SAGA, Y. & KOPAN, R. 2012. Different assemblies of Notch receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells. *Development,* 139**,** 4365-4373.
MOSKWA, P., LORENTZEN, D., EXCOFFON, K. J., ZABNER, J., MCCRAY, P. B., JR., NAUSEEF, W. M., DUPUY, C. & BANFI, B. 2007. A novel host defense system of airways is defective in cystic fibrosis. *Am J Respir Crit Care Med,* 175**,** 174-83.
MOU, H., VINARSKY, V., TATA, P. R., BRAZAUSKAS, K., CHOI, S. H., CROOKE, A. K., ZHANG, B., SOLOMON, G. M., TURNER, B. & BIHLER, H. 2016. Dual SMAD signaling inhibition enables long-term expansion of diverse epithelial basal cells. *Cell stem cell,* 19**,** 217-231.
MOU, H., YANG, Y., RIEHS, M. A., BARRIOS, J., SHIVARAJU, M., HABER, A. L., MONTORO, D. T., GILMORE, K., HAAS, E. A. & PAUNOVIC, B. 2021. Airway basal stem cells generate distinct subpopulations of PNECs. *Cell reports,* 35**,** 109011.
MUHL, L., GENOVÉ, G., LEPTIDIS, S., LIU, J., HE, L., MOCCI, G., SUN, Y., GUSTAFSSON, S., BUYANDELGER, B. & CHIVUKULA, I. V. 2020. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nature communications,* 11**,** 1-18.
MULAY, A., AKRAM, K. M., WILLIAMS, D., ARMES, H., RUSSELL, C., HOOD, D., ARMSTRONG, S., STEWART, J. P., BROWN, S. D. & BINGLE, L. 2016. An in vitro model of murine middle ear epithelium. *Disease Models & Mechanisms,* 9**,** 1405-1417.
MULERO, J. J., BOYLE, B. J., BRADLEY, S., BRIGHT, J. M., NELKEN, S. T., HO, T. T., MIZE, N. K., CHILDS, J. D., BALLINGER, D. G. & FORD, J. E. 2002. Three new human members of the lipid transfer/lipopolysaccharide binding protein family (LT/LBP). *Immunogenetics,* 54**,** 293-300.
MÜLLER, L., BRIGHTON, L. E., CARSON, J. L., FISCHER II, W. A. & JASPERS, I. 2013. Culturing of human nasal epithelial cells at the air liquid interface. *JoVE (Journal of Visualized Experiments)***,** e50646.

MURZIN, A. G., BRENNER, S. E., HUBBARD, T. & CHOTHIA, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology,* 247**,** 536-540.
MUSA, M., WILSON, K., SUN, L., MULAY, A., BINGLE, L., MARRIOTT, H. M., LECLAIR, E. E. & BINGLE, C. D. 2012. Differential localisation of BPIFA1 (SPLUNC1) and BPIFB1 (LPLUNC1) in the nasal and oral cavities of mice. *Cell and tissue research,* 350**,** 455-464.
MUTA, T. & TAKESHIGE, K. 2001. Essential roles of CD14 and lipopolysaccharide-binding protein for activation of toll-like receptor (TLR) 2 as well as TLR4: reconstitution of TLR2-and TLR4-activation by distinguishable ligands in LPS preparations. *European journal of biochemistry,* 268**,** 4580-4589.
NAGAOKA, I., TAMURA, H. & REICH, J. 2020. Therapeutic potential of cathelicidin peptide LL-37, an antimicrobial agent, in a murine sepsis model. *International Journal of Molecular Sciences,* 21**,** 5973.
NAKAUCHI, M., NAGATA, N., TAKAYAMA, I., SAITO, S., KUBO, H., KAIDA, A., OBA, K., ODAGIRI, T. & KAGEYAMA, T. 2019. Propagation of rhinovirus C in differentiated immortalized human airway HBEC3-KT epithelial cells. *Viruses,* 11**,** 216.
NAYAK, D. P., BALOGUN, R. A., YAMADA, H., ZHOU, Z. H. & BARMAN, S. 2009. Influenza virus morphogenesis and budding. *Virus research,* 143**,** 147-161.
NEGEWO, N. A., GIBSON, P. G. & MCDONALD, V. M. 2015. COPD and its comorbidities: impact, measurement and mechanisms. *Respirology,* 20**,** 1160-1171.
NEMAJEROVA, A., KRAMER, D., SILLER, S. S., HERR, C., SHOMRONI, O., PENA, T., SUAZO, C. G., GLASER, K., WILDUNG, M. & STEFFEN, H. 2016. TAp73 is a central transcriptional regulator of airway multiciliogenesis. *Genes & development,* 30**,** 1300-1312.
NICHOLAS, B., SKIPP, P., MOULD, R., RENNARD, S., DAVIES, D. E., O'CONNOR, C. D. & DJUKANOVIĆ, R. 2006. Shotgun proteomic analysis of human-induced sputum. *Proteomics,* 6**,** 4390-4401.
NICHOLS, J. E., NILES, J. A., VEGA, S. P., ARGUETA, L. B., EASTAWAY, A. & CORTIELLA, J. 2014. Modeling the lung: Design and development of tissue engineered macro-and micro-physiologic lung models for research use. *Experimental biology and medicine,* 239**,** 1135-1169.
O'LOUGHLIN, J., HALL, R. J., BHAKER, S., PORTELLI, M. A., HENRY, A., PANG, V., BATES, D. O., SHARP, T. V. & SAYERS, I. 2021. Extended lifespan of bronchial epithelial cells maintains normal cellular phenotype and transcriptome integrity. *ERJ open research,* 7.
OKUDA, K., CHEN, G., SUBRAMANI, D. B., WOLF, M., GILMORE, R. C., KATO, T., RADICIONI, G., KESIMER, M., CHUA, M. & DANG, H. 2019. Localization of secretory mucins MUC5AC and MUC5B in

normal/healthy human airways. *American journal of respiratory and critical care medicine,* 199**,** 715-727.
OOI, C. E., WEISS, J., ELSBACH, P., FRANGIONE, B. & MANNION, B. 1987. A 25-kDa NH2-terminal fragment carries all the antibacterial activities of the human neutrophil 60-kDa bactericidal/permeability-increasing protein. *Journal of Biological Chemistry,* 262**,** 14891-14894.
OPITZ, L., SALINAS-RIESTER, G., GRADE, M., JUNG, K., JO, P., EMONS, G., GHADIMI, B. M., BEISSBART, T. & GAEDCKE, J. 2010. Impact of RNA degradation on gene expression profiling. *BMC medical genomics,* 3**,** 1-14.
OYAMA, Y., SHUFF, S. R., BURNS, N., VOHWINKEL, C. U. & ECKLE, T. 2022. Intense light-elicited alveolar type 2-specific circadian PER2 protects from bacterial lung injury via BPIFB1. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 322**,** L647-L661.
OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics,* 12**,** 87-98.
PACK, R., AL-UGAILY, L. H., MORRIS, G. & WIDDICOMBE, J. 1980. The distribution and structure of cells in the tracheal epithelium of the mouse. *Cell and tissue research,* 208**,** 65-84.
PALESE, P., TOBITA, K., UEDA, M. & COMPANS, R. W. 1974. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology,* 61**,** 397-410.
PALOMARES, L. A., SRIVASTAVA, I. K., RAMÍREZ, O. T. & COX, M. M. 2018. Glycobiotechnology of the insect cell-baculovirus expression system technology. *Advances in Glycobiotechnology.* Springer.
PANG, I. K., ICHINOHE, T. & IWASAKI, A. 2013. IL-1R signaling in dendritic cells replaces pattern-recognition receptors in promoting CD8+ T cell responses to influenza A virus. *Nature immunology,* 14**,** 246.
PANG, I. K. & IWASAKI, A. 2012. Control of antiviral immunity by pattern recognition and the microbiome. *Immunological reviews,* 245**,** 209-226.
PARAMESWARAN, G. I., SETHI, S. & MURPHY, T. F. 2011. Effects of bacterial infection on airway antimicrobial peptides and proteins in COPD. *Chest,* 140**,** 611-617.
PATTERSON, M. C. Metabolic mimics: the disorders of N-linked glycosylation. Seminars in pediatric neurology, 2005. Elsevier, 144-151.
PAVLOVIC, G., ERBS, V., ANDRÉ, P., SYLVIE, J., EISENMAN, B., DREYER, D., LORENTZ, R., WATTENHOFER-DONZE, M., BIRLING, M.-C. & HERAULT, Y. 2016. Generation of targeted overexpressing models by CRISPR/Cas9 and need of careful validation of your knock-in line obtained by nuclease genome editing. *Transgenic Res,* 25**,** 254-5.
PÉREZ-VILAR, J. & MABOLO, R. 2007. Gel-forming mucins. Notions from in vitro studies. *Histology and histopathology*.

PERRONE, L. A., PLOWDEN, J. K., GARCÍA-SASTRE, A., KATZ, J. M. & TUMPEY, T. M. 2008. H5N1 and 1918 pandemic influenza virus infection results in early and excessive infiltration of macrophages and neutrophils in the lungs of mice. *PLoS pathogens,* 4**,** e1000115.

PESTKA, S., KRAUSE, C. D. & WALTER, M. R. 2004. Interferons, interferon-like cytokines, and their receptors. *Immunological reviews,* 202**,** 8-32.

PETERS-HALL, J. R., COQUELIN, M. L., TORRES, M. J., LARANGER, R., ALABI, B. R., SHO, S., CALVA-MORENO, J. F., THOMAS, P. J. & SHAY, J. W. 2018. Long-term culture and cloning of primary human bronchial basal cells that maintain multipotent differentiation capacity and CFTR channel function. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 315**,** L313-L327.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry,* 25**,** 1605-1612.

PEZZULO, A. A., STARNER, T. D., SCHEETZ, T. E., TRAVER, G. L., TILLEY, A. E., HARVEY, B.-G., CRYSTAL, R. G., MCCRAY JR, P. B. & ZABNER, J. 2011. The air-liquid interface and use of primary cell cultures are important to recapitulate the transcriptional profile of in vivo airway epithelia. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 300**,** L25-L31.

PHIPSON, B., LEE, S., MAJEWSKI, I. J., ALEXANDER, W. S. & SMYTH, G. K. 2016. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The annals of applied statistics,* 10**,** 946.

PICHLMAIR, A., SCHULZ, O., TAN, C. P., NÄSLUND, T. I., LILJESTRÖM, P., WEBER, F. & E SOUSA, C. R. 2006. RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science,* 314**,** 997-1001.

PILLOUX, L., LEROY, D., BRUNEL, C., ROGER, T. & GREUB, G. 2016. Mouse model of respiratory tract infection induced by Waddlia chondrophila. *Plos one,* 11**,** e0150909.

PINKENBURG, O., MEYER, T., BANNERT, N., NORLEY, S., BOLTE, K., CZUDAI-MATWICH, V., HEROLD, S., GESSNER, A. & SCHNARE, M. 2016. The human antimicrobial protein bactericidal/permeability-increasing protein (BPI) inhibits the infectivity of influenza a virus. *PloS one,* 11**,** e0156929.

PLASSCHAERT, L. W., ŽILIONIS, R., CHOO-WING, R., SAVOVA, V., KNEHR, J., ROMA, G., KLEIN, A. M. & JAFFE, A. B. 2018. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature,* 560**,** 377-381.

PLATCHEK, M., LU, Q., TRAN, H. & XIE, W. 2020. Comparative analysis of multiple immunoassays for cytokine profiling in drug discovery. *SLAS DISCOVERY: Advancing the Science of Drug Discovery,* 25**,** 1197-1213.
PLOTKOWSKI, M., BAJOLET-LAUDINAT, O. & PUCHELLE, E. 1993. Cellular and molecular mechanisms of bacterial adhesion to respiratory mucosa. *European Respiratory Journal,* 6**,** 903-916.
POHL, C., HERMANNS, M. I., UBOLDI, C., BOCK, M., FUCHS, S., DEI-ANANG, J., MAYER, E., KEHE, K., KUMMER, W. & KIRKPATRICK, C. J. 2009. Barrier functions and paracellular integrity in human cell culture models of the proximal respiratory unit. *European Journal of Pharmaceutics and Biopharmaceutics,* 72**,** 339-349.
POROTTO, M., FERREN, M., CHEN, Y.-W., SIU, Y., MAKHSOUS, N., RIMA, B., BRIESE, T., GRENINGER, A., SNOECK, H.-W. & MOSCONA, A. 2019. Authentic modeling of human respiratory virus infection in human pluripotent stem cell-derived lung organoids. *MBio,* 10**,** e00723-19.
POTHLICHET, J., MEUNIER, I., DAVIS, B. K., TING, J. P., SKAMENE, E., VON MESSLING, V. & VIDAL, S. M. 2013. Type I IFN triggers RIG-I/TLR3/NLRP3-dependent inflammasome activation in influenza A virus infected cells. *PLoS pathogens,* 9**,** e1003256.
POZAROWSKI, P. & DARZYNKIEWICZ, Z. 2004. Analysis of cell cycle by flow cytometry. *Checkpoint controls and cancer.* Springer.
PUCHELLE, E., DE BENTZMANN, S. & ZAHM, J. 1995. Physical and functional properties of airway secretions in cystic fibrosis–therapeutic approaches. *Respiration,* 62**,** 2-12.
PUCHELLE, E., ZAHM, J.-M., TOURNIER, J.-M. & CORAUX, C. 2006. Airway epithelial repair, regeneration, and remodeling after injury in chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society,* 3**,** 726-733.
QUIRT, J., HILDEBRAND, K. J., MAZZA, J., NOYA, F. & KIM, H. 2018. Asthma. *Allergy, Asthma & Clinical Immunology,* 14**,** 50.
RADICIONI, G., CAO, R., CARPENTER, J., FORD, A. A., WANG, T., LI, Y. & KESIMER, M. 2016. The innate immune properties of airway mucosal surfaces are regulated by dynamic interactions between mucins and interacting proteins: the mucin interactome. *Mucosal immunology,* 9**,** 1442-1454.
RADIGAN, K. A., MISHARIN, A. V., CHI, M. & BUDINGER, G. S. 2015. Modeling human influenza infection in the laboratory. *Infection and drug resistance,* 8**,** 311.
RAFEEQ, M. M. & MURAD, H. A. S. 2017. Cystic fibrosis: current therapeutic targets and future approaches. *Journal of translational medicine,* 15**,** 1-9.

RAJAVELU, P., CHEN, G., XU, Y., KITZMILLER, J. A., KORFHAGEN, T. R. & WHITSETT, J. A. 2015. Airway epithelial SPDEF integrates goblet cell differentiation and pulmonary Th2 inflammation. *The Journal of clinical investigation,* 125**,** 2021-2031.

RAMACHANDRAN, P., BOONTHEUNG, P., PANG, E., YAN, W., WONG, D. T. & LOO, J. A. 2008. Comparison of N-linked glycoproteins in human whole saliva, parotid, submandibular, and sublingual glandular secretions identified using hydrazide chemistry and mass spectrometry. *Clinical proteomics,* 4**,** 80.

RAMIREZ, R. D., SHERIDAN, S., GIRARD, L., SATO, M., KIM, Y., POLLACK, J., PEYTON, M., ZOU, Y., KURIE, J. M. & DIMAIO, J. M. 2004. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer research,* 64**,** 9027-9034.

RAMOS, I., SMITH, G., RUF-ZAMOJSKI, F., MARTÍNEZ-ROMERO, C., FRIBOURG, M., CARBAJAL, E. A., HARTMANN, B. M., NAIR, V. D., MARJANOVIC, N. & MONTEAGUDO, P. L. 2019. Innate immune response to influenza virus at single-cell resolution in human epithelial cells revealed paracrine induction of interferon lambda 1. *Journal of virology,* 93**,** e00559-19.

RAN, F., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013a. Genome engineering using the CRISPR-Cas9 system. *Nature protocols,* 8**,** 2281-2308.

RAN, F. A., HSU, P. D., LIN, C.-Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S. & ZHANG, Y. 2013b. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-1389.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013c. Genome engineering using the CRISPR-Cas9 system. *Nature protocols,* 8**,** 2281-2308.

RAO, M. S., VAN VLEET, T. R., CIURLIONIS, R., BUCK, W. R., MITTELSTADT, S. W., BLOMME, E. A. & LIGUORI, M. J. 2019. Comparison of RNA-Seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies. *Frontiers in genetics,* 9**,** 636.

RAPITEANU, R., KARAGYOZOVA, T., ZIMMERMANN, N., SINGH, K., WAYNE, G., MARTUFI, M., BELYAEV, N. N., HESSEL, E. M., MICHALOVICH, D. & MACARRON, R. 2020. Highly efficient genome editing in primary human bronchial epithelial cells differentiated at air–liquid interface. *European Respiratory Journal,* 55.

RAREDON, M. S. B., ADAMS, T. S., SUHAIL, Y., SCHUPP, J. C., POLI, S., NEUMARK, N., LEIBY, K. L., GREANEY, A. M., YUAN, Y. & HORIEN, C. 2019. Single-cell connectomic analysis of adult mammalian lungs. *Science advances,* 5**,** eaaw3851.

RAWLINS, E. L., OKUBO, T., XUE, Y., BRASS, D. M., AUTEN, R. L., HASEGAWA, H., WANG, F. & HOGAN, B. L. 2009. The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell stem cell,* 4**,** 525-534.

REDDLE, R. 1988. Transformation of human bronchial epithelial cells by infection with SV40 or adenovirus-12 SV40 hybrid virus, or transfection via strontium phosphate coprecipitation with a plasmid containing SV40 early region genes. *Cancer Res,* 48**,** 1904-1909.

RENTHAL, R., SCHNEIDER, B. G., MILLER, M. M. & LUDUEÑA, R. F. 1993. βIV is the major β-tubulin isotype in bovine cilia. *Cell motility and the cytoskeleton,* 25**,** 19-29.

REVINSKI, D. R., ZARAGOSI, L.-E., BOUTIN, C., RUIZ-GARCIA, S., DEPREZ, M., THOMÉ, V., ROSNET, O., GAY, A.-S., MERCEY, O. & PAQUET, A. 2018. CDC20B is required for deuterosome-mediated centriole production in multiciliated cells. *Nature Communications,* 9**,** 4668.

REYNOLDS, S. D., GIANGRECO, A., POWER, J. H. & STRIPP, B. R. 2000. Neuroepithelial bodies of pulmonary airways serve as a reservoir of progenitor cells capable of epithelial regeneration. *Am J Pathol,* 156**,** 269-78.

RIJSBERGEN, L. C., SCHMITZ, K. S., BEGEMAN, L., DREW-BEAR, J., GOMMERS, L., LAMERS, M. M., GRENINGER, A. L., HAAGMANS, B. L., POROTTO, M. & DE SWART, R. L. 2022. Modeling Infection and Tropism of Human Parainfluenza Virus Type 3 in Ferrets. *Mbio,* 13**,** e03831-21.

RIJSBERGEN, L. C., VAN DIJK, L. L., ENGEL, M. F., DE VRIES, R. D. & DE SWART, R. L. 2021. In Vitro Modelling of Respiratory Virus Infections in Human Airway Epithelial Cells–A Systematic Review. *Frontiers in Immunology***,** 3301.

RISHI, A., JOYCE-BRADY, M., FISHER, J., DOBBS, L., FLOROS, J., VANDERSPEK, J., BRODY, J. & WILLIAMS, M. 1995. Cloning, characterization, and developmental expression of a rat lung alveolar type I cell gene in embryonic endodermal and neural derivatives. *Developmental biology,* 167**,** 294-306.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. & SMYTH, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research,* 43**,** e47-e47.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics,* 26**,** 139-140.

ROBINSON, M. D. & OSHLACK, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology,* 11**,** 1-9.

ROCK, J. R., ONAITIS, M. W., RAWLINS, E. L., LU, Y., CLARK, C. P., XUE, Y., RANDELL, S. H. & HOGAN, B. L. 2009. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proceedings of the National Academy of Sciences,* 106**,** 12771-12775.

ROCK, J. R., RANDELL, S. H. & HOGAN, B. L. 2010. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Disease models & mechanisms,* 3**,** 545-556.

ROGERS, G. N. & PAULSON, J. C. 1983. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology,* 127**,** 361-373.

ROGERS, T. D., BUTTON, B., KELADA, S. N., OSTROWSKI, L. E., LIVRAGHI-BUTRICO, A., GUTAY, M. I., ESTHER JR, C. R. & GRUBB, B. R. 2022. Regional Differences in Mucociliary Clearance in the Upper and Lower Airways. *Frontiers in physiology***,** 187.

ROKICKI, W., ROKICKI, M., WOJTACHA, J. & DŻELJIJLI, A. 2016. The role and importance of club cells (Clara cells) in the pathogenesis of some respiratory diseases. *Kardiochirurgia i torakochirurgia polska= Polish journal of cardio-thoracic surgery,* 13**,** 26.

ROLLINS, B. M., GARCIA-CABALLERO, A., STUTTS, M. J. & TARRAN, R. 2010. SPLUNC1 expression reduces surface levels of the epithelial sodium channel (ENaC) in Xenopus laevis oocytes. *Channels,* 4**,** 255-259.

ROMAGNANI, P., ANNUNZIATO, F., LAZZERI, E., COSMI, L., BELTRAME, C., LASAGNI, L., GALLI, G., FRANCALANCI, M., MANETTI, R. & MARRA, F. 2001. Interferon-inducible protein 10, monokine induced by interferon gamma, and interferon-inducible T-cell alpha chemoattractant are produced by thymic epithelial cells and attract T-cell receptor (TCR) αβ+ CD8+ single-positive T cells, TCRγδ+ T cells, and natural killer–type cells in human thymus. *Blood, The Journal of the American Society of Hematology,* 97**,** 601-607.

ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology,* 5**,** 172.

ROSE, M. C. & VOYNOW, J. A. 2006. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiological reviews,* 86**,** 245-278.

ROSS, A. J., DAILEY, L. A., BRIGHTON, L. E. & DEVLIN, R. B. 2007. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. *American journal of respiratory cell and molecular biology,* 37**,** 169-185.

ROXO-ROSA, M., DA COSTA, G., LUIDER, T. M., SCHOLTE, B. J., COELHO, A. V., AMARAL, M. D. & PENQUE, D. 2006. Proteomic analysis of nasal cells from cystic fibrosis patients and non-cystic fibrosis control individuals: Search for novel biomarkers of cystic fibrosis lung disease. *Proteomics,* 6**,** 2314-2325.

ROY, M. G., LIVRAGHI-BUTRICO, A., FLETCHER, A. A., MCELWEE, M. M., EVANS, S. E., BOERNER, R. M., ALEXANDER, S. N., BELLINGHAUSEN, L. K., SONG, A. S. & PETROVA, Y. M. 2014. Muc5b is required for airway defence. *Nature,* 505**,** 412-416.

RUIZ GARCÍA, S., DEPREZ, M., LEBRIGAND, K., CAVARD, A., PAQUET, A., ARGUEL, M.-J., MAGNONE, V., TRUCHI, M., CABALLERO, I. & LEROY, S. 2019. Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. *Development,* 146**,** dev177428.

RUTIGLIANO, J. A., SHARMA, S., MORRIS, M. Y., OGUIN III, T. H., MCCLAREN, J. L., DOHERTY, P. C. & THOMAS, P. G. 2014. Highly pathological influenza A virus infection is associated with augmented expression of PD-1 by functionally compromised virus-specific CD8+ T cells. *Journal of virology,* 88**,** 1636-1651.

RYAN, D. M., VINCENT, T. L., SALIT, J., WALTERS, M. S., AGOSTO-PEREZ, F., SHAYKHIEV, R., STRULOVICI-BAREL, Y., DOWNEY, R. J., BURO-AURIEMMA, L. J. & STAUDT, M. R. 2014. Smoking dysregulates the human airway basal cell transcriptome at COPD risk locus 19q13. 2. *PloS one,* 9**,** e88051.

SAATIAN, B., REZAEE, F., DESANDO, S., EMO, J., CHAPMAN, T., KNOWLDEN, S. & GEORAS, S. N. 2013. Interleukin-4 and interleukin-13 cause barrier dysfunction in human airway epithelial cells. *Tissue barriers,* 1**,** e24333.

SAFERALI, A., OBEIDAT, M. E., BÉRUBÉ, J.-C., LAMONTAGNE, M., BOSSÉ, Y., LAVIOLETTE, M., HAO, K., NICKLE, D. C., TIMENS, W. & SIN, D. D. 2015. Polymorphisms associated with expression of BPIFA1/BPIFB1 and lung disease severity in cystic fibrosis. *American journal of respiratory cell and molecular biology,* 53**,** 607-614.

SAFERALI, A., TANG, A. C., STRUG, L. J., QUON, B. S., ZLOSNIK, J., SANDFORD, A. J. & TURVEY, S. E. 2020. Immunomodulatory function of the cystic fibrosis modifier gene BPIFA1. *Plos one,* 15**,** e0227067.

SAJJAN, U., WANG, Q., ZHAO, Y., GRUENERT, D. C. & HERSHENSON, M. B. 2008. Rhinovirus disrupts the barrier function of polarized airway epithelial cells. *American journal of respiratory and critical care medicine,* 178**,** 1271-1281.

SAJJAN, U. S., COREY, M., KARMALI, M. A. & FORSTNER, J. F. 1992. Binding of Pseudomonas cepacia to normal human intestinal mucin and

respiratory mucin from patients with cystic fibrosis. *The Journal of clinical investigation,* 89**,** 648-656.
SALAUN, B., DE SAINT-VIS, B., PACHECO, N., PACHECO, Y., RIESLER, A., ISAAC, S., LEROUX, C., CLAIR-MONINOT, V., PIN, J.-J. & GRIFFITH, J. 2004. CD208/dendritic cell-lysosomal associated membrane protein is a marker of normal and transformed type II pneumocytes. *The American journal of pathology,* 164**,** 861-871.
SATELLI, A. & LI, S. 2011. Vimentin in cancer and its potential as a molecular target for cancer therapy. *Cellular and molecular life sciences,* 68**,** 3033-3046.
SATO, M., VAUGHAN, M. B., GIRARD, L., PEYTON, M., LEE, W., SHAMES, D. S., RAMIREZ, R. D., SUNAGA, N., GAZDAR, A. F. & SHAY, J. W. 2006. Multiple oncogenic changes (K-RASV12, p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer research,* 66**,** 2116-2128.
SAULER, M., MCDONOUGH, J. E., ADAMS, T. S., KOTHAPALLI, N., BARNTHALER, T., WERDER, R. B., SCHUPP, J. C., NOUWS, J., ROBERTSON, M. J. & COARFA, C. 2022. Characterization of the COPD alveolar niche using single-cell RNA sequencing. *Nature communications,* 13**,** 1-17.
SCHEETZ, T. E., ZABNER, J., WELSH, M. J., COCO, J., EYESTONE, M., DE FATIMA BONALDO, M., KUCABA, T., CASAVANT, T. L., SOARES, M. B. & MCCRAY JR, P. B. 2004. Large-scale gene discovery in human airway epithelia reveals novel transcripts. *Physiological genomics,* 17**,** 69-77.
SCHILDERS, K. A., EENJES, E., VAN RIET, S., POOT, A. A., STAMATIALIS, D., TRUCKENMÜLLER, R., HIEMSTRA, P. S. & ROTTIER, R. J. 2016. Regeneration of the lung: Lung stem cells and the development of lung mimicking devices. *Respiratory research,* 17**,** 1-16.
SCHNEIDER, C., O'LEARY, C. E. & LOCKSLEY, R. M. 2019. Regulation of immune responses by tuft cells. *Nature Reviews Immunology,* 19**,** 584-593.
SCHOGGINS, J. W. 2014. Interferon-stimulated genes: roles in viral pathogenesis. *Current opinion in virology,* 6**,** 40-46.
SCHOGGINS, J. W. 2019. Interferon-stimulated genes: what do they all do? *Annual review of virology,* 6**,** 567-584.
SCHULTZ, H. & WEISS, J. P. 2007. The bactericidal/permeability-increasing protein (BPI) in infection and inflammatory disease. *Clinica Chimica Acta,* 384**,** 12-23.
SCHUMANN, R. R., LEONG, S. R., FLAGGS, G. W., GRAY, P. W., WRIGHT, S. D., MATHISON, J. C., TOBIAS, P. S. & ULEVITCH, R.

J. 1990. Structure and function of lipopolysaccharide binding protein. *Science,* 249**,** 1429-1431.
SCHUMANN, R. R., RIETSCHEL, E. T. & LOPPNOW, H. 1994. The role of CD14 and lipopolysaccharide-binding protein (LBP) in the activation of different cell types by endotoxin. *Medical microbiology and immunology,* 183**,** 279-297.
SCULL, M. A., GILLIM-ROSS, L., SANTOS, C., ROBERTS, K. L., BORDONALI, E., SUBBARAO, K., BARCLAY, W. S. & PICKLES, R. J. 2009. Avian influenza virus glycoproteins restrict virus replication and spread through human airway epithelium at temperatures of the proximal airways. *PLoS pathogens,* 5**,** e1000424.
SEIBOLD, M. A. 2018. Interleukin-13 stimulation reveals the cellular and functional plasticity of the airway epithelium. *Annals of the American Thoracic Society,* 15**,** S98-S102.
SENTANI, K., OUE, N., SAKAMOTO, N., ARIHIRO, K., AOYAGI, K., SASAKI, H. & YASUI, W. 2008. Gene expression profiling with microarray and SAGE identifies PLUNC as a marker for hepatoid adenocarcinoma of the stomach. *Modern Pathology,* 21**,** 464-475.
SETHI, S. 2000. Bacterial infection and the pathogenesis of COPD. *Chest,* 117**,** 286S-291S.
SHAPIRO, G., GURNEY, T. & KRUG, R. 1987. Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *Journal of virology,* 61**,** 764-773.
SHARMA, P., YI, R., NAYAK, A. P., WANG, N., TANG, F., KNIGHT, M. J., PAN, S., OLIVER, B. & DESHPANDE, D. A. 2017. Bitter taste receptor agonists mitigate features of allergic asthma in mice. *Scientific Reports,* 7**,** 1-14.
SHEEHAN, J. K., KESIMER, M. & PICKLES, R. Innate immunity and mucus structure and function. Novartis Found Symp, 2007. 155-166.
SHERMAN, A. C., MEHTA, A., DICKERT, N. W., ANDERSON, E. J. & ROUPHAEL, N. 2019. The future of flu: a review of the human challenge model and systems biology for advancement of influenza vaccinology. *Frontiers in cellular and infection microbiology,* 9**,** 107.
SHIN, K., FOGG, V. C. & MARGOLIS, B. 2006. Tight junctions and cell polarity. *Annu. Rev. Cell Dev. Biol.,* 22**,** 207-235.
SHIN, O. S., UDDIN, T., CITORIK, R., WANG, J. P., DELLA PELLE, P., KRADIN, R. L., BINGLE, C. D., BINGLE, L., CAMILLI, A. & BHUIYAN, T. R. 2011. LPLUNC1 modulates innate immune responses to Vibrio cholerae. *Journal of Infectious Diseases,* 204**,** 1349-1357.
SHIRAISHI, K., SHICHINO, S., UEHA, S., NAKAJIMA, T., HASHIMOTO, S., YAMAZAKI, S. & MATSUSHIMA, K. 2019. Mesenchymal-

epithelial interactome analysis reveals essential factors required for fibroblast-free alveolosphere formation. *Iscience,* 11**,** 318-333.
SHUM, A. K., ALIMOHAMMADI, M., TAN, C. L., CHENG, M. H., METZGER, T. C., LAW, C. S., LWIN, W., PERHEENTUPA, J., BOUR-JORDAN, H., CAREL, J. C., HUSEBYE, E. S., DE LUCA, F., JANSON, C., SARGUR, R., DUBOIS, N., KAJOSAARI, M., WOLTERS, P. J., CHAPMAN, H. A., KAMPE, O. & ANDERSON, M. S. 2013. BPIFB1 is a lung-specific autoantigen associated with interstitial lung disease. *Sci Transl Med,* 5**,** 206ra139.
SIDAHMED, A. M., LEÓN, A. J., BOSINGER, S. E., BANNER, D., DANESH, A., CAMERON, M. J. & KELVIN, D. J. 2012. CXCL10 contributes to p38-mediated apoptosis in primary T lymphocytes in vitro. *Cytokine,* 59**,** 433-441.
SIDNEY, L. E., BRANCH, M. J., DUNPHY, S. E., DUA, H. S. & HOPKINSON, A. 2014. Concise review: evidence for CD34 as a common marker for diverse progenitors. *Stem cells,* 32**,** 1380-1389.
SIMONIN, J., BILLE, E., CRAMBERT, G., NOEL, S., DREANO, E., EDWARDS, A., HATTON, A., PRANKE, I., VILLERET, B. & COTTART, C.-H. 2019. Airway surface liquid acidification initiates host defense abnormalities in Cystic Fibrosis. *Scientific reports,* 9**,** 1-11.
SINGH, P. K., JIA, H. P., WILES, K., HESSELBERTH, J., LIU, L., CONWAY, B. A., GREENBERG, E. P., VALORE, E. V., WELSH, M. J., GANZ, T., TACK, B. F. & MCCRAY, P. B., JR. 1998. Production of beta-defensins by human airway epithelia. *Proc Natl Acad Sci U S A,* 95**,** 14961-6.
SKEHEL, J. J. & WILEY, D. C. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual review of biochemistry,* 69**,** 531-569.
SLEPUSHKIN, V. A., STABER, P. D., WANG, G., MCCRAY JR, P. B. & DAVIDSON, B. L. 2001. Infection of human airway epithelia with H1N1, H2N2, and H3N2 influenza A virus strains. *Molecular Therapy,* 3**,** 395-402.
SMIT, M. J., VERDIJK, P., VAN DER RAAIJ-HELMER, E. M., NAVIS, M., HENSBERGEN, P. J., LEURS, R. & TENSEN, C. P. 2003. CXCR3-mediated chemotaxis of human T cells is regulated by a Gi-and phospholipase C–dependent pathway and not via activation of MEK/p44/p42 MAPK nor Akt/PI-3 kinase. *Blood,* 102**,** 1959-1965.
SMITH, C. A., TYRELL, D. J., KULKARNI, U. A., WOOD, S., LENG, L., ZEMANS, R. L., BUCALA, R. & GOLDSTEIN, D. R. 2019. Macrophage migration inhibitory factor enhances influenza-associated mortality in mice. *JCI insight,* 4.

SMYTH, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology,* 3.

SONI, H., MATTHEWS, A. T., PALLIKKUTH, S., GANGARAJU, R. & ADEBIYI, A. 2019. γ-secretase inhibitor DAPT mitigates cisplatin-induced acute kidney injury by suppressing Notch1 signaling. *Journal of Cellular and Molecular Medicine,* 23**,** 260-270.

SPINA, D. 1998. Epithelium smooth muscle regulation and interactions. *American journal of respiratory and critical care medicine,* 158**,** S141-S145.

SRIWILAIJAROEN, N. & SUZUKI, Y. 2012. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proceedings of the Japan Academy, Series B,* 88**,** 226-249.

STAEHELI, P., HALLER, O., BOLL, W., LINDENMANN, J. & WEISSMANN, C. 1986. Mx protein: constitutive expression in 3T3 cells transformed with cloned Mx cDNA confers selective resistance to influenza virus. *Cell,* 44**,** 147-158.

STEILING, K., KADAR, A. Y., BERGERAT, A., FLANIGON, J., SRIDHAR, S., SHAH, V., AHMAD, Q. R., BRODY, J. S., LENBURG, M. E. & STEFFEN, M. 2009. Comparison of proteomic and transcriptomic profiles in the bronchial airway epithelium of current and never smokers. *PLoS One,* 4**,** e5043.

STENVIK, J., SOLSTAD, T., STRAND, C., LEIROS, I. & JØRGENSEN, T. Ø. 2004. Cloning and analyses of a BPI/LBP cDNA of the Atlantic cod (Gadus morhua L.). *Developmental & Comparative Immunology,* 28**,** 307-323.

STIJLEMANS, B., SCHOOVAERTS, M., DE BAETSELIER, P., MAGEZ, S. & DE TREZ, C. 2022. The Role of MIF and IL-10 as Molecular Yin-Yang in the Modulation of the Host Immune Microenvironment During Infections: African Trypanosome Infections as a Paradigm. *Frontiers in Immunology,* 13.

STRINE, M. S. & WILEN, C. B. 2022. Tuft cells are key mediators of interkingdom interactions at mucosal barrier surfaces. *PLoS Pathogens,* 18**,** e1010318.

SUI, P., WIESNER, D. L., XU, J., ZHANG, Y., LEE, J., VAN DYKEN, S., LASHUA, A., YU, C., KLEIN, B. S. & LOCKSLEY, R. M. 2018. Pulmonary neuroendocrine cells amplify allergic asthma responses. *Science,* 360**,** eaan8546.

SULTAN, M., SCHULZ, M. H., RICHARD, H., MAGEN, A., KLINGENHOFF, A., SCHERF, M., SEIFERT, M., BORODINA, T., SOLDATOV, A. & PARKHOMCHUK, D. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science,* 321**,** 956-960.

SUN, Y. L., HURLEY, K., VILLACORTA-MARTIN, C., HUANG, J., HINDS, A., GOPALAN, K., CABALLERO, I. S., RUSSO, S. J., KITZMILLER, J. A. & WHITSETT, J. A. 2021. Heterogeneity in Human Induced Pluripotent Stem Cell–derived Alveolar Epithelial Type II Cells Revealed with ABCA3/SFTPC Reporters. *American Journal of Respiratory Cell and Molecular Biology,* 65**,** 442-460.

SUTTAPITUGSAKUL, S., TONG, M., SUN, F. & WU, R. 2021. Enhancing comprehensive analysis of secreted glycoproteins from cultured cells without serum starvation. *Analytical chemistry,* 93**,** 2694-2705.

SVENSSON, V., NATARAJAN, K. N., LY, L.-H., MIRAGAIA, R. J., LABALETTE, C., MACAULAY, I. C., CVEJIC, A. & TEICHMANN, S. A. 2017. Power analysis of single-cell RNA-sequencing experiments. *Nature methods,* 14**,** 381-387.

SWAIN, R. J., KEMP, S. J., GOLDSTRAW, P., TETLEY, T. D. & STEVENS, M. M. 2010. Assessment of cell line models of primary human cells by Raman spectral phenotyping. *Biophysical journal,* 98**,** 1703-1711.

TALL, A. R. 1993. Plasma cholesteryl ester transfer protein. *Journal of lipid research,* 34**,** 1255-1274.

TAO, L. & REESE, T. A. 2017. Making mouse models that reflect human immune responses. *Trends in immunology,* 38**,** 181-193.

TATA, A., KOBAYASHI, Y., CHOW, R. D., TRAN, J., DESAI, A., MASSRI, A. J., MCCORD, T. J., GUNN, M. D. & TATA, P. R. 2018. Myoepithelial cells of submucosal glands can function as reserve stem cells to regenerate airways after injury. *Cell stem cell,* 22**,** 668-683. e6.

TATA, P. R., MOU, H., PARDO-SAGANTA, A., ZHAO, R., PRABHU, M., LAW, B. M., VINARSKY, V., CHO, J. L., BRETON, S. & SAHAY, A. 2013. Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature,* 503**,** 218-223.

TATA, P. R. & RAJAGOPAL, J. 2017. Plasticity in the lung: making and breaking cell identity. *Development,* 144**,** 755-766.

TATE, M. D., DENG, Y.-M., JONES, J. E., ANDERSON, G. P., BROOKS, A. G. & READING, P. C. 2009. Neutrophils ameliorate lung injury and the development of severe disease during influenza infection. *The Journal of Immunology,* 183**,** 7441-7450.

TAUBE, C., DUEZ, C., CUI, Z.-H., TAKEDA, K., RHA, Y.-H., PARK, J.-W., BALHORN, A., DONALDSON, D. D., DAKHAMA, A. & GELFAND, E. W. 2002. The role of IL-13 in established allergic airway disease. *The Journal of Immunology,* 169**,** 6482-6489.

TECLE, T., WHITE, M. R., GANTZ, D., CROUCH, E. C. & HARTSHORN, K. L. 2007. Human neutrophil defensins increase neutrophil uptake of influenza A virus and bacteria and modify virus-induced respiratory burst responses. *The Journal of Immunology,* 178**,** 8046-8052.

THAISS, C. A., ZMORA, N., LEVY, M. & ELINAV, E. 2016. The microbiome and innate immunity. *Nature,* 535**,** 65-74.
THOMAS, P. G., DASH, P., ALDRIDGE JR, J. R., ELLEBEDY, A. H., REYNOLDS, C., FUNK, A. J., MARTIN, W. J., LAMKANFI, M., WEBBY, R. J. & BOYD, K. L. 2009. The intracellular sensor NLRP3 mediates key innate and healing responses to influenza A virus via the regulation of caspase-1. *Immunity,* 30**,** 566-575.
TIAN, B., YUAN, Y., YANG, Y., LUO, Z., SUI, B., ZHOU, M., FU, Z. F. & ZHAO, L. 2020. Interferon-inducible GTPase 1 impedes the dimerization of rabies virus phosphoprotein and restricts viral replication. *Journal of virology,* 94**,** e01203-20.
TITZ, B., SEWER, A., SCHNEIDER, T., ELAMIN, A., MARTIN, F., DIJON, S., LUETTICH, K., GUEDJ, E., VUILLAUME, G. & IVANOV, N. V. 2015. Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects. *Journal of proteomics,* 128**,** 306-320.
TOMINAGA, K., YOSHIMOTO, T., TORIGOE, K., KURIMOTO, M., MATSUI, K., HADA, T., OKAMURA, H. & NAKANISHI, K. 2000. IL-12 synergizes with IL-18 or IL-1β for IFN-γ production from human T cells. *International immunology,* 12**,** 151-160.
TRAVAGLINI, K. J., NABHAN, A. N., PENLAND, L., SINHA, R., GILLICH, A., SIT, R. V., CHANG, S., CONLEY, S. D., MORI, Y. & SEITA, J. 2020. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature,* 587**,** 619-625.
TREUTLEIN, B., BROWNFIELD, D. G., WU, A. R., NEFF, N. F., MANTALAS, G. L., ESPINOZA, F. H., DESAI, T. J., KRASNOW, M. A. & QUAKE, S. R. 2014. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature,* 509**,** 371-375.
TRIPATHI, S., TECLE, T., VERMA, A., CROUCH, E., WHITE, M. & HARTSHORN, K. L. 2013. The human cathelicidin LL-37 inhibits influenza A viruses through a mechanism distinct from that of surfactant protein D or defensins. *The Journal of general virology,* 94**,** 40.
TROMBETTA, E. S. 2003. The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology,* 13**,** 77R-91R.
TSUTSUMI, N., KIMURA, T., ARITA, K., ARIYOSHI, M., OHNISHI, H., YAMAMOTO, T., ZUO, X., MAENAKA, K., PARK, E. Y. & KONDO, N. 2014. The structural basis for receptor recognition of human interleukin-18. *Nature communications,* 5**,** 1-13.
TUGIZOV, S. 2021. Virus-associated disruption of mucosal epithelial tight junctions and its role in viral transmission and spread. *Tissue Barriers,* 9**,** 1943274.
TZOU, P., OHRESSER, S., FERRANDON, D., CAPOVILLA, M., REICHHART, J.-M., LEMAITRE, B., HOFFMANN, J. A. & IMLER, J.-

L. 2000. Tissue-specific inducible expression of antimicrobial peptide genes in Drosophila surface epithelia. *Immunity,* 13**,** 737-748.
UDDIN, F., RUDIN, C. M. & SEN, T. 2020. CRISPR gene therapy: applications, limitations, and implications for the future. *Frontiers in oncology,* 10**,** 1387.
ULLMAN, T. A. & ITZKOWITZ, S. H. 2011. Intestinal inflammation and cancer. *Gastroenterology,* 140**,** 1807-1816. e1.
UTHAIAH, R. C., PRAEFCKE, G. J., HOWARD, J. C. & HERRMANN, C. 2003. IIGP1, an interferon-γ-inducible 47-kDa GTPase of the mouse, showing cooperative enzymatic activity and GTP-dependent multimerization. *Journal of Biological Chemistry,* 278**,** 29336-29343.
VAN DE BOVENKAMP, J. H., MAHDAVI, J., KORTELAND-VAN MALE, A. M., BÜLLER, H. A., EINERHAND, A. W., BORÉN, T. & DEKKER, J. 2003. The MUC5AC glycoprotein is the primary receptor for Helicobacter pylori in the human stomach. *Helicobacter,* 8**,** 521-532.
VAN DEN BRINK, G., TYTGAT, K., VAN DER HULST, R., VAN DER LOOS, C., EINERHAND, A., BÜLLER, H. & DEKKER, J. 2000. H pylori colocalises with MUC5AC in the human stomach. *Gut,* 46**,** 601-607.
VAN RIEL, D., MUNSTER, V. J., DE WIT, E., RIMMELZWAAN, G. F., FOUCHIER, R. A., OSTERHAUS, A. D. & KUIKEN, T. 2007. Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *The American journal of pathology,* 171**,** 1215-1223.
VARGAS, P., SPEIGHT, P., BINGLE, C., BARRETT, A. & BINGLE, L. 2008. Expression of PLUNC family members in benign and malignant salivary gland tumours. *Oral diseases,* 14**,** 613-619.
VAUGHAN, M. B., RAMIREZ, R. D., WRIGHT, W. E., MINNA, J. D. & SHAY, J. W. 2006. A three-dimensional model of differentiation of immortalized human bronchial epithelial cells. *Differentiation,* 74**,** 141-148.
VAZIN, T. & FREED, W. J. 2010. Human embryonic stem cells: derivation, culture, and differentiation: a review. *Restorative neurology and neuroscience,* 28**,** 589-603.
VELDHUIS KROEZE, E., BAUER, L., CALIENDO, V. & VAN RIEL, D. 2021. In vivo models to study the pathogenesis of extra-respiratory complications of influenza A virus infection. *Viruses,* 13**,** 848.
VIANA, F., O'KANE, C. M. & SCHROEDER, G. N. 2022. Precision-cut lung slices: A powerful ex vivo model to investigate respiratory infectious diseases. *Molecular Microbiology,* 117**,** 578-588.
VIEIRA BRAGA, F. A., KAR, G., BERG, M., CARPAIJ, O. A., POLANSKI, K., SIMON, L. M., BROUWER, S., GOMES, T., HESSE, L. & JIANG,

J. 2019. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nature medicine,* 25**,** 1153-1163.
VILLENAVE, R., SHIELDS, M. D. & POWER, U. F. 2013. Respiratory syncytial virus interaction with human airway epithelium. *Trends in microbiology,* 21**,** 238-244.
VON MOLTKE, J., JI, M., LIANG, H.-E. & LOCKSLEY, R. M. 2016. Tuft-cell-derived IL-25 regulates an intestinal ILC2–epithelial response circuit. *Nature,* 529**,** 221-225.
VREUGDENHIL, A. C., DENTENER, M. A., SNOEK, A. P., GREVE, J.-W. M. & BUURMAN, W. A. 1999. Lipopolysaccharide binding protein and serum amyloid A secretion by human intestinal epithelial cells during the acute phase response. *The Journal of Immunology,* 163**,** 2792-2798.
WADDELL, A., VALLANCE, J. E., HUMMEL, A., ALENGHAT, T. & ROSEN, M. J. 2019. IL-33 induces murine intestinal goblet cell differentiation indirectly via innate lymphoid cell IL-13 secretion. *The Journal of Immunology,* 202**,** 598-607.
WANG, H., ALOE, C., WILSON, N. & BOZINOVSKI, S. 2019a. G-CSFR antagonism reduces neutrophilic inflammation during pneumococcal and influenza respiratory infections without compromising clearance. *Scientific Reports,* 9**,** 1-12.
WANG, H., ZHOU, Y., OYANG, L., HAN, Y., XIA, L., LIN, J., TANG, Y., SU, M., TAN, S. & TIAN, Y. 2019b. LPLUNC1 stabilises PHB1 by counteracting TRIM21-mediated ubiquitination to inhibit NF-κB activity in nasopharyngeal carcinoma. *Oncogene,* 38**,** 5062-5075.
WANG, R., HUME, A. J., BEERMANN, M. L., SIMONE-ROACH, C., LINDSTROM-VAUTRIN, J., LE SUER, J., HUANG, J., OLEJNIK, J., VILLACORTA-MARTIN, C. & BULLITT, E. 2022. Human airway lineages derived from pluripotent stem cells reveal the epithelial responses to SARS-CoV-2 infection. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 322**,** L462-L478.
WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics,* 10**,** 57-63.
WATERSTON, R. H., LINDBLAD-TOH, K., BIRNEY, E., ROGERS, J., ABRIL, J. F., AGARWAL, P., AGARWALA, R., AINSCOUGH, R., ALEXANDERSSON, M., AN, P., ANTONARAKIS, S. E., ATTWOOD, J., BAERTSCH, R., BAILEY, J., BARLOW, K., BECK, S., BERRY, E., BIRREN, B., BLOOM, T., BORK, P., BOTCHERBY, M., BRAY, N., BRENT, M. R., BROWN, D. G., BROWN, S. D., BULT, C., BURTON, J., BUTLER, J., CAMPBELL, R. D., CARNINCI, P., CAWLEY, S., CHIAROMONTE, F., CHINWALLA, A. T., CHURCH, D. M., CLAMP, M., CLEE, C., COLLINS, F. S., COOK, L. L., COPLEY, R. R., COULSON, A., COURONNE, O., CUFF, J., CURWEN, V., CUTTS, T., DALY, M., DAVID, R., DAVIES, J., DELEHAUNTY, K. D., DERI, J.,

DERMITZAKIS, E. T., DEWEY, C., DICKENS, N. J., DIEKHANS, M., DODGE, S., DUBCHAK, I., DUNN, D. M., EDDY, S. R., ELNITSKI, L., EMES, R. D., ESWARA, P., EYRAS, E., FELSENFELD, A., FEWELL, G. A., FLICEK, P., FOLEY, K., FRANKEL, W. N., FULTON, L. A., FULTON, R. S., FUREY, T. S., GAGE, D., GIBBS, R. A., GLUSMAN, G., GNERRE, S., GOLDMAN, N., GOODSTADT, L., GRAFHAM, D., GRAVES, T. A., GREEN, E. D., GREGORY, S., GUIGÓ, R., GUYER, M., HARDISON, R. C., HAUSSLER, D., HAYASHIZAKI, Y., HILLIER, L. W., HINRICHS, A., HLAVINA, W., HOLZER, T., HSU, F., HUA, A., HUBBARD, T., HUNT, A., JACKSON, I., JAFFE, D. B., JOHNSON, L. S., JONES, M., JONES, T. A., JOY, A., KAMAL, M., KARLSSON, E. K., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature,* 420**,** 520-62.

WATSON, A., MADSEN, J. & CLARK, H. W. 2021. SP-A and SP-D: dual functioning immune molecules with antiviral and immunomodulatory properties. *Frontiers in Immunology,* 11**,** 622598.

WEATHERHEAD, J. E., GAZZINELLI-GUIMARAES, P., KNIGHT, J. M., FUJIWARA, R., HOTEZ, P. J., BOTTAZZI, M. E. & CORRY, D. B. 2020. Host immunity and inflammation to pulmonary helminth infections. *Frontiers in Immunology,* 11**,** 594520.

WEBER, J. R., FREYER, D., ALEXANDER, C., SCHRÖDER, N. W., REISS, A., KÜSTER, C., PFEIL, D., TUOMANEN, E. I. & SCHUMANN, R. R. 2003. Recognition of pneumococcal peptidoglycan: an expanded, pivotal role for LPS binding protein. *Immunity,* 19**,** 269-279.

WEI, F., TANG, L., HE, Y., WU, Y., SHI, L., XIONG, F., GONG, Z., GUO, C., LI, X., LIAO, Q., ZHANG, W., NI, Q., LUO, J., LI, X., LI, Y., PENG, C., CHEN, X., LI, G., XIONG, W. & ZENG, Z. 2018a. BPIFB1 (LPLUNC1) inhibits radioresistance in nasopharyngeal carcinoma by inhibiting VTN expression. *Cell Death Dis,* 9**,** 432.

WEI, F., WU, Y., TANG, L., HE, Y., SHI, L., XIONG, F., GONG, Z., GUO, C., LI, X., LIAO, Q., ZHANG, W., ZHOU, M., XIANG, B., LI, X., LI, Y., LI, G., XIONG, W. & ZENG, Z. 2018b. BPIFB1 (LPLUNC1) inhibits migration and invasion of nasopharyngeal carcinoma by interacting with VTN and VIM. *Br J Cancer,* 118**,** 233-247.

WESTON, W. M., LECLAIR, E. E., TRZYNA, W., MCHUGH, K. M., NUGENT, P., LAFFERTY, C. M., MA, L., TUAN, R. S. & GREENE, R. M. 1999. Differential display identification of plunc, a novel gene expressed in embryonic palate, nasal epithelium, and adult lung. *Journal of Biological Chemistry,* 274**,** 13698-13703.

WHEELER, T. T., HAIGH, B. J., MCCRACKEN, J. Y., WILKINS, R. J., MORRIS, C. A. & GRIGOR, M. R. 2002. The BSP30 salivary proteins from cattle, LUNX/PLUNC and von Ebner's minor salivary gland protein

are members of the PSP/LBP superfamily of proteins. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression,* 1579**,** 92-100.
WHITSETT, J. A. 2018. Airway epithelial differentiation and mucociliary clearance. *Annals of the American Thoracic Society,* 15**,** S143-S148.
WHITSETT, J. A., KALIN, T. V., XU, Y. & KALINICHENKO, V. V. 2019. Building and regenerating the lung cell by cell. *Physiological reviews,* 99**,** 513-554.
WIDDICOMBE, J. H. 2019. Early studies on the surface epithelium of mammalian airways. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 317**,** L486-L495.
WIDDICOMBE, J. H. & WINE, J. J. 2015. Airway gland structure and function. *Physiological reviews,* 95**,** 1241-1319.
WILDT, S. & GERNGROSS, T. U. 2005. The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology,* 3**,** 119-128.
WJ VAN DER SCHAFT, D., WAGSTAFF, J., MAYO, K. H. & GRIFFIOEN, A. W. 2002. The antiangiogenic properties of bactericidal/permeability-increasing protein (BPI). *Annals of medicine,* 34**,** 19-27.
WOOD, P. R., HILL, V. L., BURKS, M. L., PETERS, J. I., SINGH, H., KANNAN, T. R., VALE, S., CAGLE, M. P., PRINCIPE, M. F. & BASEMAN, J. B. 2013. Mycoplasma pneumoniae in children with acute and refractory asthma. *Annals of Allergy, Asthma & Immunology,* 110**,** 328-334. e1.
WORLD HEALTH ORGANIZATION, W. 2018. *Up to 650 000 people die of respiratory diseases linked to seasonal flu each year* [Online]. [Accessed July, 2018].
WU, C., JIN, X., TSUENG, G., AFRASIABI, C. & SU, A. I. 2016a. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic acids research,* 44**,** D313-D316.
WU, J., KOBAYASHI, M., SOUSA, E. A., LIU, W., CAI, J., GOLDMAN, S. J., DORNER, A. J., PROJAN, S. J., KAVURU, M. S. & QIU, Y. 2005. Differential Proteomic Analysis of Bronchoalveolar Lavage Fluid in Asthmatics following Segmental Antigen Challenge* S. *Molecular & Cellular Proteomics,* 4**,** 1251-1264.
WU, N.-H., YANG, W., BEINEKE, A., DIJKMAN, R., MATROSOVICH, M., BAUMGÄRTNER, W., THIEL, V., VALENTIN-WEIGAND, P., MENG, F. & HERRLER, G. 2016b. The differentiated airway epithelium infected by influenza viruses maintains the barrier function despite a dramatic loss of ciliated cells. *Scientific reports,* 6**,** 1-12.
WU, T., HUANG, J., MOORE, P. J., LITTLE, M. S., WALTON, W. G., FELLNER, R. C., ALEXIS, N. E., PETER DI, Y., REDINBO, M. R. & TILLEY, S. L. 2017. Identification of BPIFA1/SPLUNC1 as an epithelium-derived smooth muscle relaxing factor. *Nature communications,* 8**,** 1-10.

XING, Y., LI, C., LI, A., SRIDURONGRIT, S., TIOZZO, C., BELLUSCI, S., BOROK, Z., KAARTINEN, V. & MINOO, P. 2010. Signaling via Alk5 controls the ontogeny of lung Clara cells. *Development,* 137**,** 825-833.

XU, Y., TAO, Z., JIANG, Y., LIU, T. & XIANG, Y. 2019. Overexpression of BPIFB1 promotes apoptosis and inhibits proliferation via the MEK/ERK signal pathway in nasopharyngeal carcinoma. *International Journal of Clinical and Experimental Pathology,* 12**,** 356.

YAMAGATA, M., ROOK, S. L., SASSA, Y., MA, R. C., GERALDES, P., GODDARD, L., CLERMONT, A., GAO, B., SALTI, H. & GUNDEL, R. 2006. Bactericidal/permeability-increasing protein's signaling pathways and its retinal trophic and anti-angiogenic effects. *The FASEB journal,* 20**,** 2058-2067.

YAMASHITA, S., SAKAI, N., HIRANO, K.-I., ISHIGAMI, M., MARUYAMA, T., NAKAJIMA, N. & MATSUZAWA, Y. 2001. Roles of plasma lipid transfer proteins in reverse cholesterol transport. *Front Biosci,* 6**,** D366-D387.

YANG, Z. F., MOK, C. K. P., LIU, X. Q., LI, X. B., HE, J. F., GUAN, W. D., XU, Y. H., PAN, W. Q., CHEN, L. Y. & LIN, Y. P. 2015. Clinical, virological and immunological features from patients infected with re-emergent avian-origin human H7N9 influenza disease of varying severity in Guangdong province. *PloS one,* 10**,** e0117846.

YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics,* 13**,** 1-11.

YONKER, L. M., MOU, H., CHU, K. K., PAZOS, M. A., LEUNG, H., CUI, D., RYU, J., HIBBLER, R. M., EATON, A. D. & FORD, T. N. 2017. Development of a primary human co-culture model of inflamed airway mucosa. *Scientific reports,* 7**,** 1-12.

YOU, Y. & BRODY, S. L. 2012. Culture and differentiation of mouse tracheal epithelial cells. *Epithelial Cell Culture Protocols*. Springer.

YOU, Y., HUANG, T., RICHER, E. J., SCHMIDT, J.-E. H., ZABNER, J., BOROK, Z. & BRODY, S. L. 2004. Role of f-box factor foxj1 in differentiation of ciliated airway epithelial cells. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 286**,** L650-L657.

YOU, Y., RICHER, E. J., HUANG, T. & BRODY, S. L. 2002. Growth and differentiation of mouse tracheal epithelial cells: selection of a proliferative population. *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 283**,** L1315-L1321.

ZANIN, M., BAVISKAR, P., WEBSTER, R. & WEBBY, R. 2016. The interaction between respiratory pathogens and mucus. *Cell host & microbe,* 19**,** 159-168.

ZHAI, Y., FRANCO, L. M., ATMAR, R. L., QUARLES, J. M., ARDEN, N., BUCASAS, K. L., WELLS, J. M., NINO, D., WANG, X. & ZAPATA,

G. E. 2015. Host transcriptional response to influenza and other acute respiratory viral infections–a prospective cohort study. *PLoS pathogens,* 11**,** e1004869.

ZHANG, B., NIE, X., XIAO, B., XIANG, J., SHEN, S., GONG, J., ZHOU, M., ZHU, S., ZHOU, J. & QIAN, J. 2003. Identification of tissue-specific genes in nasopharyngeal epithelial tissue and differentially expressed genes in nasopharyngeal carcinoma by suppression subtractive hybridization and cDNA microarray. *Genes, Chromosomes and Cancer,* 38**,** 80-90.

ZHANG, W., YU, Y., HERTWIG, F., THIERRY-MIEG, J., ZHANG, W., THIERRY-MIEG, D., WANG, J., FURLANELLO, C., DEVANARAYAN, V. & CHENG, J. 2015. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome biology,* 16**,** 1-12.

ZHENG, D., SOH, B.-S., YIN, L., HU, G., CHEN, Q., CHOI, H., HAN, J., CHOW, V. T. & CHEN, J. 2017. Differentiation of club cells to alveolar epithelial cells in vitro. *Scientific reports,* 7**,** 1-9.

ZHENG, S., XU, W., BOSE, S., BANERJEE, A. K., HAQUE, S. J. & ERZURUM, S. C. 2004. Impaired nitric oxide synthase-2 signaling pathway in cystic fibrosis airway epithelium. *Am J Physiol Lung Cell Mol Physiol,* 287**,** L374-81.

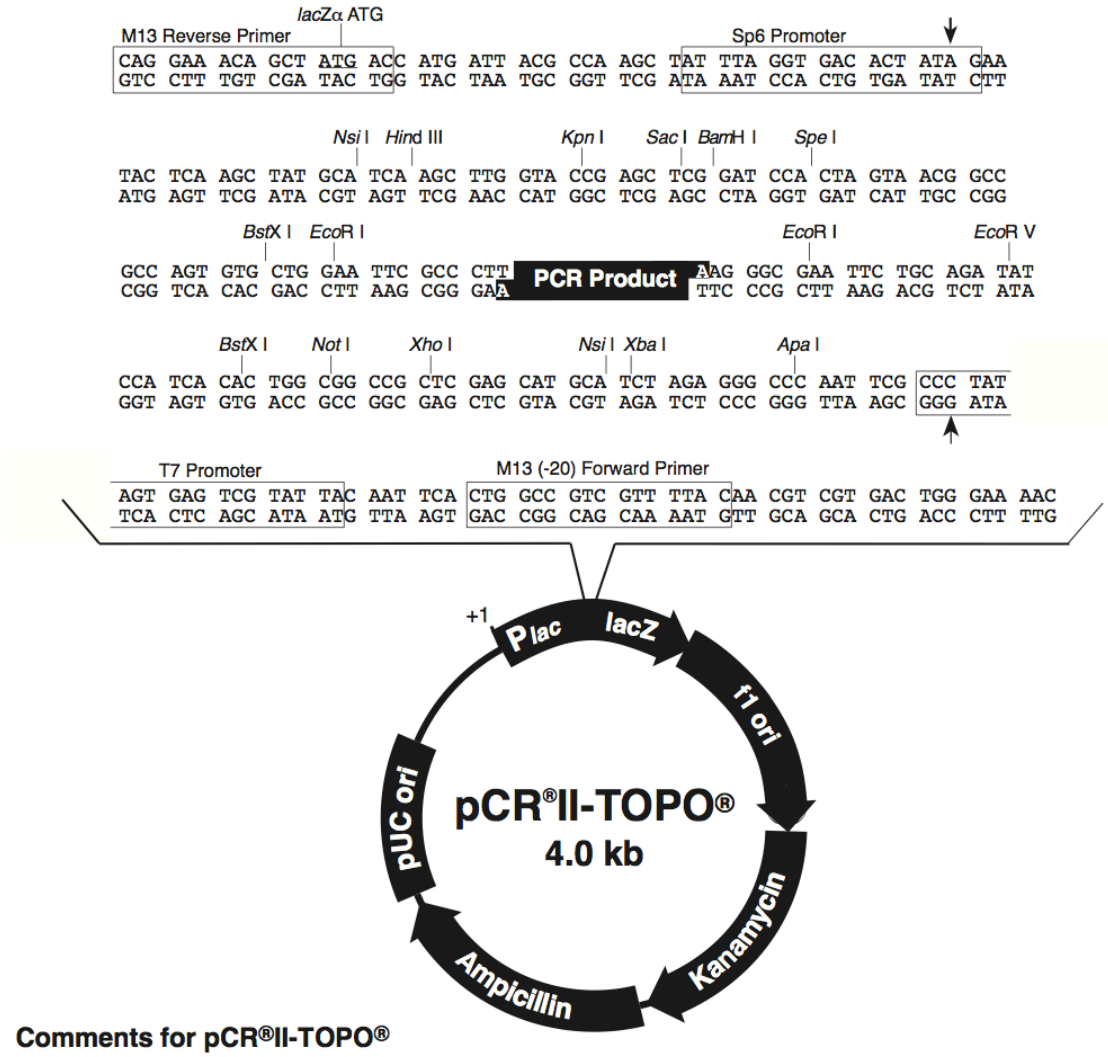
ZHOU, S., XU, B., QI, L., ZHU, D., LIU, B. & WEI, J. 2019. Next-generation sequencing reveals mutational accordance between cell-free DNA from plasma, malignant pleural effusion and ascites and directs targeted therapy in a gastric cancer patient. *Cancer biology & therapy,* 20**,** 15-20.

ZHOU, Y., ZENG, Z., ZHANG, W., XIONG, W., WU, M., TAN, Y., YI, W., XIAO, L., LI, X. & HUANG, C. 2008. Lactotransferrin: A candidate tumor suppressor—Deficient expression in human nasopharyngeal carcinoma and inhibition of NPC cell proliferation by modulating the mitogen-activated protein kinase pathway. *International journal of cancer,* 123**,** 2065-2072.

ZUO, W.-L., ROSTAMI, M. R., LEBLANC, M., KANER, R. J., O'BEIRNE, S. L., MEZEY, J. G., LEOPOLD, P. L., QUAST, K., VISVANATHAN, S. & FINE, J. S. 2020. Dysregulation of club cell biology in idiopathic pulmonary fibrosis. *PLoS One,* 15**,** e0237529.

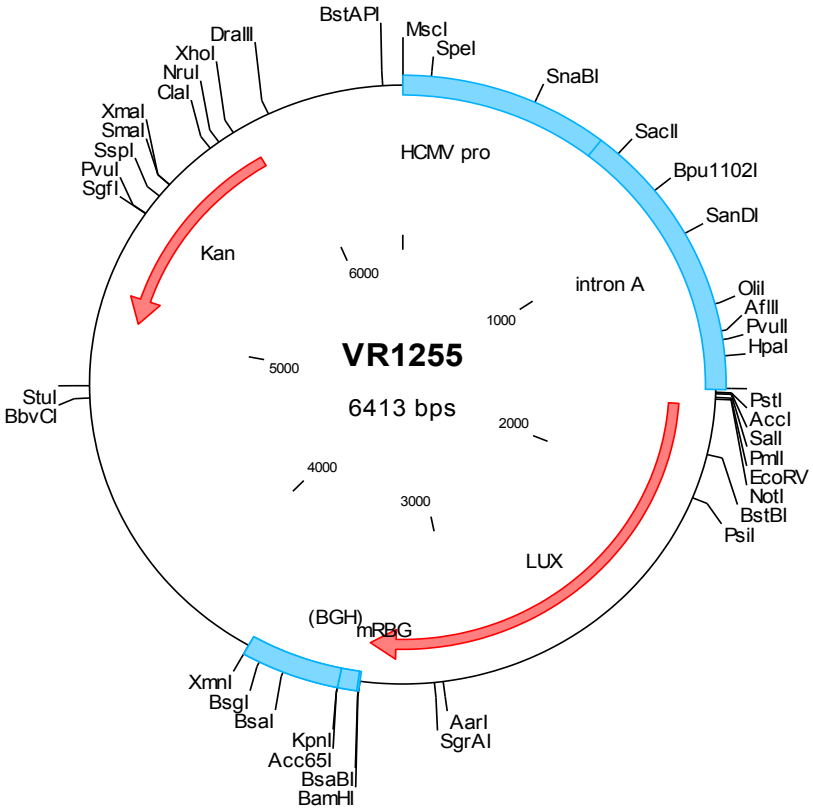
Appendix

Appendix I – Vector map of the pCRII-TOPO plasmid showing the sites of the SP6 and T7 primers used for sequencing.

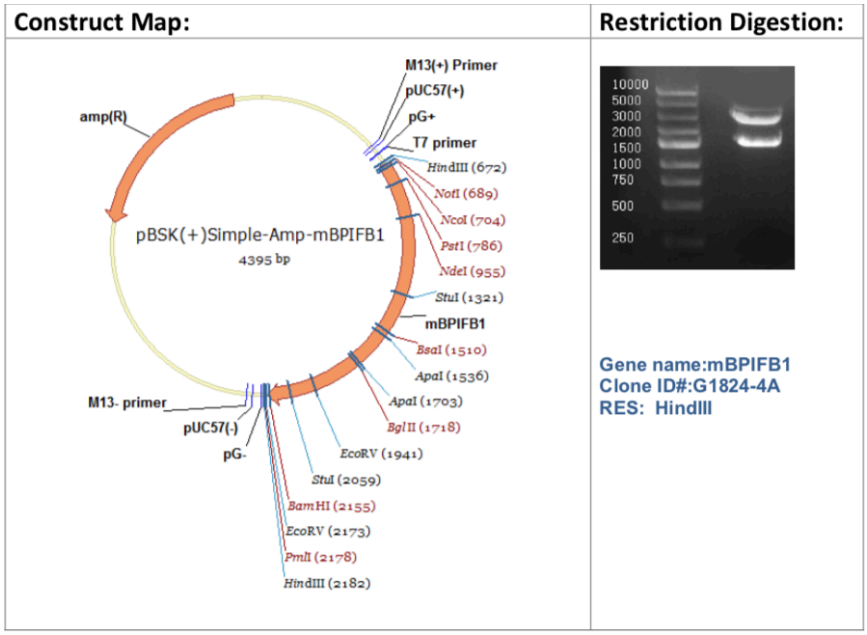


Comments for pCR®II-TOPO®
3973 nucleotides

Appendix II – Vector map of the VR1255 plasmid.



Appendix III – Vector map of pBSK(+) Simple mBPIFB1 and verification of pBSK(+) construct by sequencing and restriction digestion.



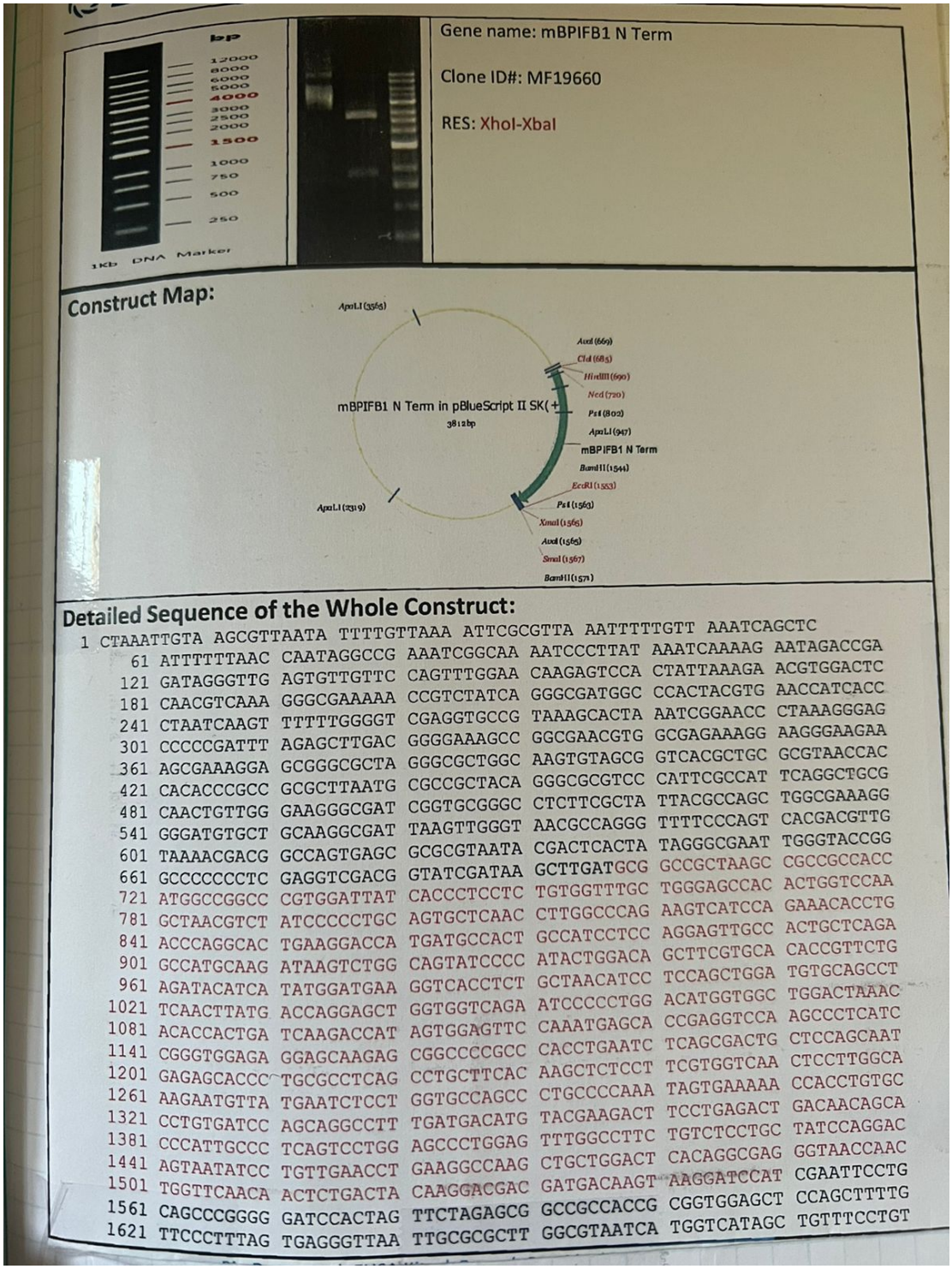
651	TGGGTACGGC	CGTCAAGGCC	AAGCTTCCCA	CTTATAGCGG	CCGCCGCCGC
701	CACCATGGCC	GGCCCGTGGA	TTATCACCCT	CCTCTGTGGT	TTGCTGGGAG
751	CCACACTGGT	CCAAGCTAAC	GTCTATCCCC	CTGCAGTGCT	CAACCTTGGC
801	CCAGAAGTCA	TCCAGAAACA	CCTGACCCAG	GCACTGAAGG	ACCATGATGC
851	CACTGCCATC	CTCCAGGAGT	TGCCACTGCT	CAGAGCCATG	CAAGATAAGT
901	CTGGCAGTAT	CCCCATACTG	GACAGCTTCG	TGCACACCGT	TCTGAGATAC
951	ATCATATGGA	TGAAGGTCAC	CTCTGCTAAC	ATCCTCCAGC	TGGATGTGCA
1001	GCCTTCAACT	TATGACCAGG	AGCTGGTGGT	CAGAATCCCC	CTGGACATGG
1051	TGGCTGGACT	AAACACACCA	CTGATCAAGA	CCATAGTGGA	GTTCCAAATG
1101	AGCACCGAGG	TCCAAGCCCT	CATCCGGGTG	GAGAGGAGCA	AGAGCGGCCC
1151	CGCCCACCTG	GTGCTCAGCG	ACTGCTCCAG	CAGCGAGAGC	ACCCTGCGCC
1201	TCAGCCTGCT	TCACAAGCTC	TCCTTCGTGG	TCAACTCCTT	GGCAAAGAAT
1251	GTTATGAATC	TCCTGGTGCC	AGCCCTGCCC	CAAATAGTGA	AAAACCACCT
1301	GTGCCCTGTG	ATCCAGCAGG	CCTTTGATGA	CATGTACGAA	GACTTCCTGA
1351	GACTGACAAC	AGCACCCATT	GCCCTCAGTC	CTGGAGCCCT	GGAGTTTGGC
1401	CTTCTGTCTC	CTGCTATCCA	GGACAGTAAT	ATCCTGTTGA	ACCTGAAGGC
1451	CAAGCTGCTG	GACTCACAGG	CGAGGGTAAC	CAACTGGTTC	AAGAACTCTG
1501	CGACTTCTCT	GATGGAGACC	ACCCCAGACA	GGGCCCCCTT	CAGCCTGACC
1551	GTGAGGCAGG	ACCTGGTGAA	TGCCATTGTG	ACCACCCTGG	TCCCCAAGGA
1601	GGAGCTTGTA	ATCCTGCTCA	GATTCGTGAT	TCCTGATGTG	GCCCGCCAGT
1651	TACAGATGGA	CATCAAGGAA	ATCAATGCAG	AGGCAGCCAA	CAAGCTGGGG
1701	CCCACCCAGA	TGTTGAAGAT	CTTCACCCAC	AGCACCCCCC	ACATTGTGCT
1751	GAACGAGGGC	AGTGCCAGGG	CAGCCCAAAG	TGTCGTGCTG	GAAGTGTTCC
1801	CTACCAACAC	TGATGTCCGG	CCCTTCTTCT	CTCTCGGCAT	TGAGGCCAGT
1851	TATGAAGCTC	AGTTCTTCAC	AGAAGACAAC	CGGCTTATGC	TCAACTTCAA
1901	TGAAGTCAGT	ATTGAGCGCA	TCAAGCTGAT	GATCTCAGAT	ATCAAACTAT
1951	TCGATCCTGA	AGTCCTGAAG	GACACCCTGA	CCAAGATCCT	TGAATACACA
2001	CTGCTGCCCA	ATGAGAATGG	CAAACTGAGA	ACTGGAGTCC	CCATGTCAAT
2051	GTCCAAGGCC	TTGGGATACG	AGAAAGCCAT	GTGGTCTGTG	AGCAAGGGTG
2101	CCCTCAAGCT	CACTCCAGCC	TCCTCCGACT	ACAAGGACGA	CGATGACAAG
2151	TAAGGATCCG	CCTAAGTGGG	ATATCACGTG	AAGCTTGCAA	GCTCCAGCTT

Appendix IV- DNA sequence and amino-acid sequence of the N-glycosylation mutant *mBPIFB1* gene. The DNA sequence was obtained from Biomatik and was translated into the amino-acid sequence using the ExPASy programme. Note the positions of the four mutated N-glycosylated sites are highlighted in yellow.

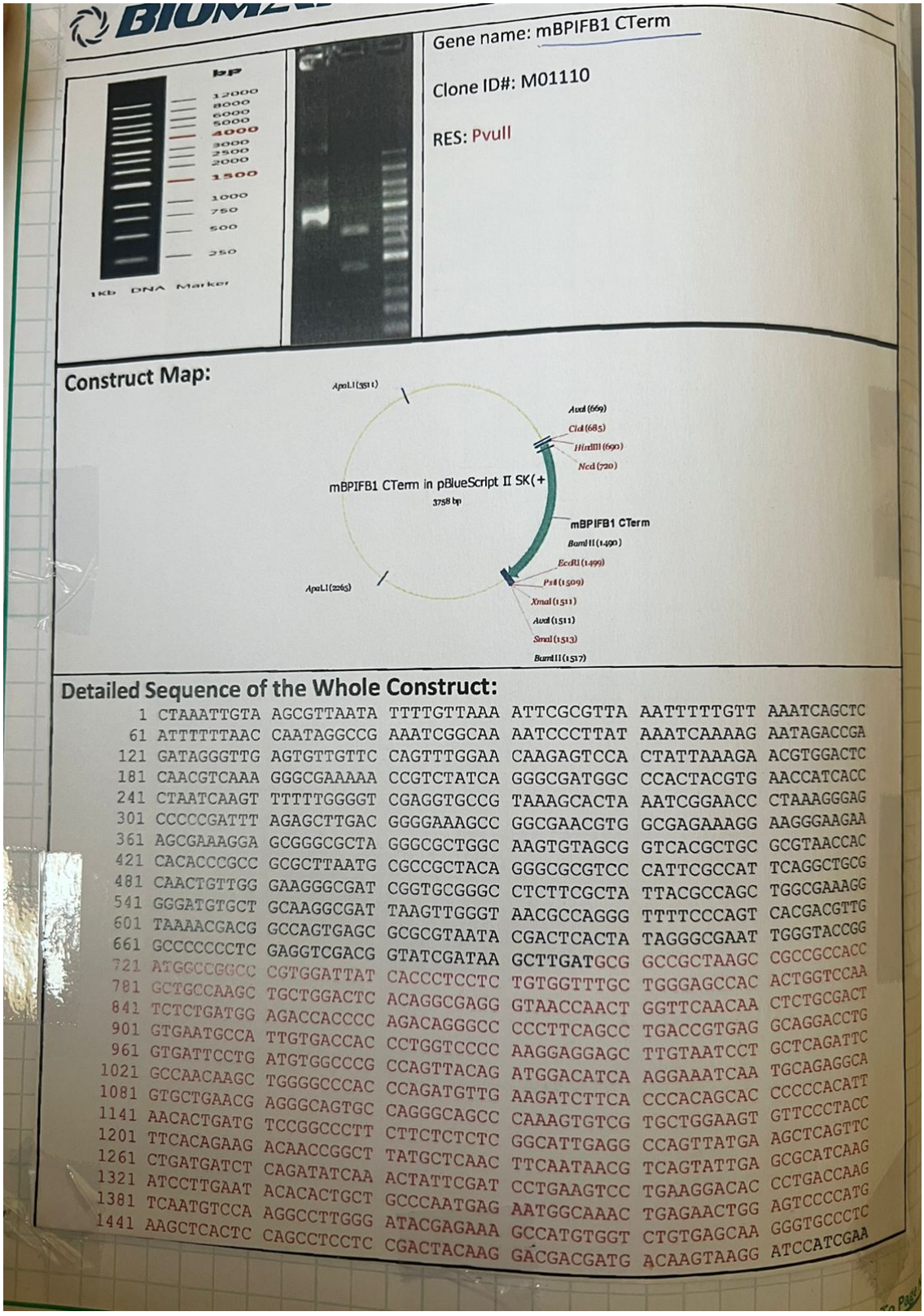
```
1 AGGAGAGGAGCAGACAGACCCAGGACTCTGGCACATTCAGGTCCTACTACCTGCCACCTA     60
.............................................................

61 CTTCTCAGCGACACCCAGGAAGATGGCCGGCCCGTGGATTATCACCCTCCTCTGTGGTTT    120
.............     .......-M--A--G--P--W--I--I--T--L--L--C--G--L    13

121 GCTGGGAGCCACACTGGTCCAAGCTAACGTCTATCCCCCTGCAGTGCTCAACCTTGGCCC    180

13 --L--G--A--T--L--V--Q--A--N--V--Y--P--P--A--V--L--N--L--G--P     33

181 AGAAGTCATCCAGAAACACCTGACCCAGGCACTGAAGGACCATGATGCCACTGCCATCCT    240

33 --E--V--I--Q--K--H--L--T--Q--A--L--K--D--H--D--A--T--A--I--L     53

241 CCAGGAGTTGCCACTGCTCAGAGCCATGCAAGATAAGTCTGGCAGTATCCCCATACTGGA    300

53 --Q--E--L--P--L--L--R--A--M--Q--D--K--S--G--S--I--P--I--L--D     73

301 CAGCTTCGTGCACACCGTTCTGAGATACATCATATGGATGAAGGTCACCTCTGCTAACAT    360

73 --S--F--V--H--T--V--L--R--Y--I--I--W--M--K--V--T--S--A--N--I     93

361 CCTCCAGCTGGATGTGCAGCCTTCAACTTATGACCAGGAGCTGGTGGTCAGAATCCCCCT    420

93 --L--Q--L--D--V--Q--P--S--T--Y--D--Q--E--L--V--V--R--I--P--L    113

421 GGACATGGTGGCTGGACTAAACACACCACTGATCAAGACCATAGTGGAGTTCCAAATGAG    480

113 --D--M--V--A--G--L--N--T--P--L--I--K--T--I--V--E--F--Q--M--S    133

481 CACCGAGGTCCAAGCCCTCATCCGGGTGGAGAGGAGCAAGAGCGGCCCCGCCCACCTGGT    540

133 --T--E--V--Q--A--L--I--R--V--E--R--S--K--S--G--P--A--H--L--V    153

541 GCTCAGCGACTGCTCCAGCAGCGAGAGCACCCTGCGCCTCAGCCTGCTTCACAAGCTCTC    600

153 --L--S--D--C--S--S--S--E--S--T--L--R--L--S--L--L--H--K--L--S    173

601 CTTCGTGGTCAACTCCTTGGCAAAGAATGTTATGAATCTCCTGGTGCCAGCCCTGCCCCA    660

173 --F--V--V--N--S--L--A--K--N--V--M--N--L--L--V--P--A--L--P--Q    193

661 AATAGTGAAAAACCACCTGTGCCCTGTGATCCAGCAGGCCTTTGATGACATGTACGAAGA    720

193 --I--V--K--N--H--L--C--P--V--I--Q--Q--A--F--D--D--M--Y--E--D    213

721 CTTCCTGAGACTGACAACAGCACCCATTGCCCTCAGTCCTGGAGCCCTGGAGTTTGGCCT    780

213 --F--L--R--L--T--T--A--P--I--A--L--S--P--G--A--L--E--F--G--L    233

781 TCTGTCTCCTGCTATCCAGGACAGTAATATCCTGTTGAACCTGAAGGCCAAGCTGCTGGA    840

233 --L--S--P--A--I--Q--D--S--N--I--L--L--N--L--K--A--K--L--L--D    253
```

```
841 CTCACAGGCGAGGGTAACCAACTGGTTCAAGAACTCTGCGACTTCTCTGATGGAGACCAC    900
253 --S--Q--A--R--V--T--N--W--F--K--N--S--A--T--S--L--M--E--T--T    273
901 CCCAGACAGGGCCCCCTTCAGCCTGACCGTGAGGCAGGACCTGGTGAATGCCATTGTGAC    960
273 --P--D--R--A--P--F--S--L--T--V--R--Q--D--L--V--N--A--I--V--T    293
961 CACCCTGGTCCCCAAGGAGGAGCTTGTAATCCTGCTCAGATTCGTGATTCCTGATGTGGC   1020
293 --T--L--V--P--K--E--E--L--V--I--L--L--R--F--V--I--P--D--V--A    313
1021 CCGCCAGTTACAGATGGACATCAAGGAAATCAATGCAGAGGCAGCCAACAAGCTGGGGCC   1080
313 --R--Q--L--Q--M--D--I--K--E--I--N--A--E--A--A--N--K--L--G--P    333
1081 CACCCAGATGTTGAAGATCTTCACCCACAGCACCCCCCACATTGTGCTGAACGAGGGCAG   1140
333 --T--Q--M--L--K--I--F--T--H--S--T--P--H--I--V--L--N--E--G--S    353
1141 TGCCAGGGCAGCCCAAAGTGTCGTGCTGGAAGTGTTCCCTACCAACACTGATGTCCGGCC   1200
353 --A--R--A--A--Q--S--V--V--L--E--V--F--P--T--N--T--D--V--R--P    373
1201 CTTCTTCTCTCTCGGCATTGAGGCCAGTTATGAAGCTCAGTTCTTCACAGAAGACAACCG   1260
373 --F--F--S--L--G--I--E--A--S--Y--E--A--Q--F--F--T--E--D--N--R    393
1261 GCTTATGCTCAACTTCAATGAAGTCAGTATTGAGCGCATCAAGCTGATGATCTCAGATAT   1320
393 --L--M--L--N--F--N--E--V--S--I--E--R--I--K--L--M--I--S--D--I    413
1321 CAAACTATTCGATCCTGAAGTCCTGAAGGACACCCTGACCAAGATCCTTGAATACACACT   1380
413 --K--L--F--D--P--E--V--L--K--D--T--L--T--K--I--L--E--Y--T--L    433
1381 GCTGCCCAATGAGAATGGCAAACTGAGAACTGGAGTCCCCATGTCAATGTCCAAGGCCTT   1440
433 --L--P--N--E--N--G--K--L--R--T--G--V--P--M--S--M--S--K--A--L    453
1441 GGGATACGAGAAAGCCATGTGGTCTGTGAGCAAGGGTGCCCTCAAGCTCACTCCAGCCTC   1500
453 --G--Y--E--K--A--M--W--S--V--S--K--G--A--L--K--L--T--P--A--S    473
1501 CTCCTAGAACCCTGGCTCCCCGCCCCTCCTGATGAAGACCTGGACAGTAGCAGCCAGCCC   1560
473 --S--*-.....................................................    474
1561 CGGCCAGGTCCCAGCCAGGAGTGTGGAAGCCTACTGTGTAGACTACCCCCCTGCAGTTAA   1620
```

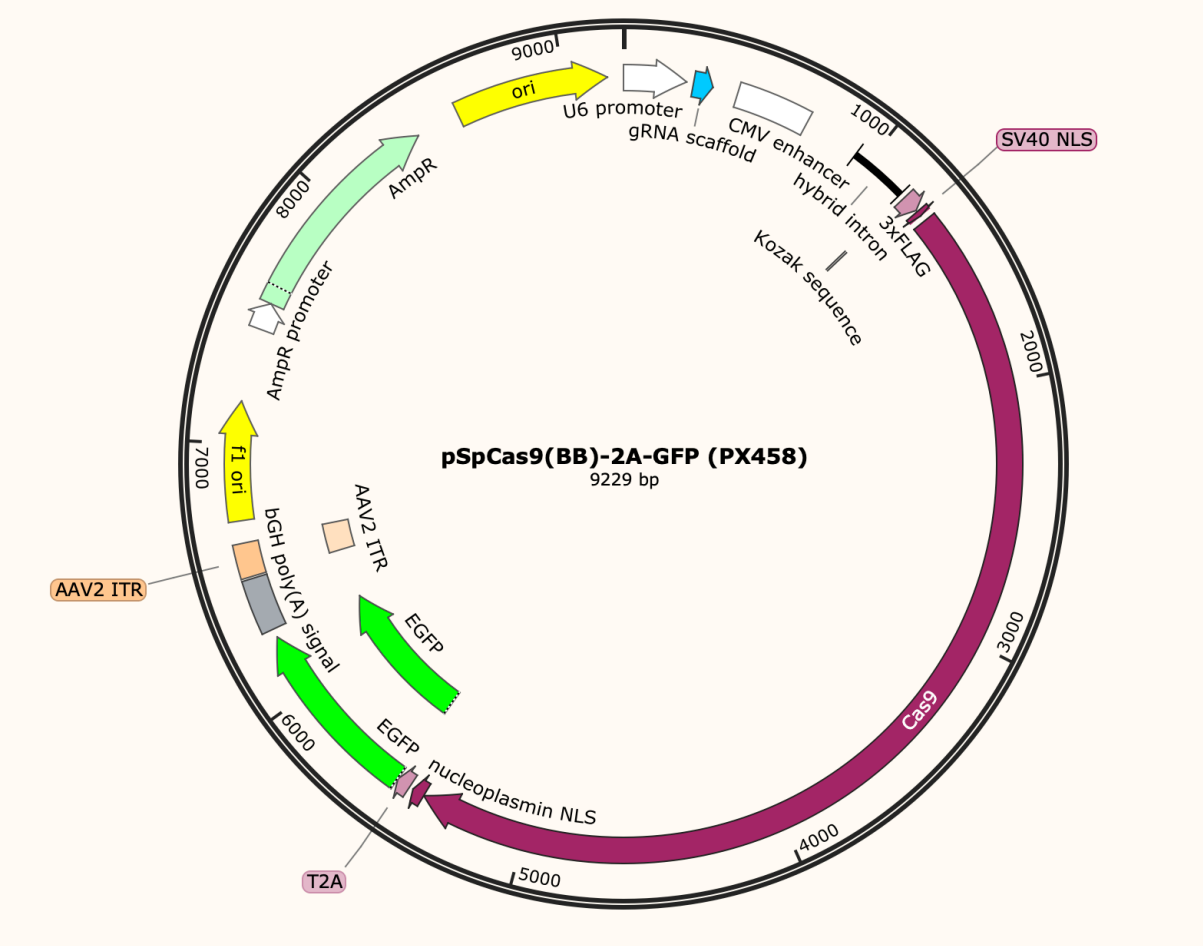
Appendix V - Vector map of pBlueScriptII SK construct containing the *mBPIFB1* N-terminal sequence (below in red)



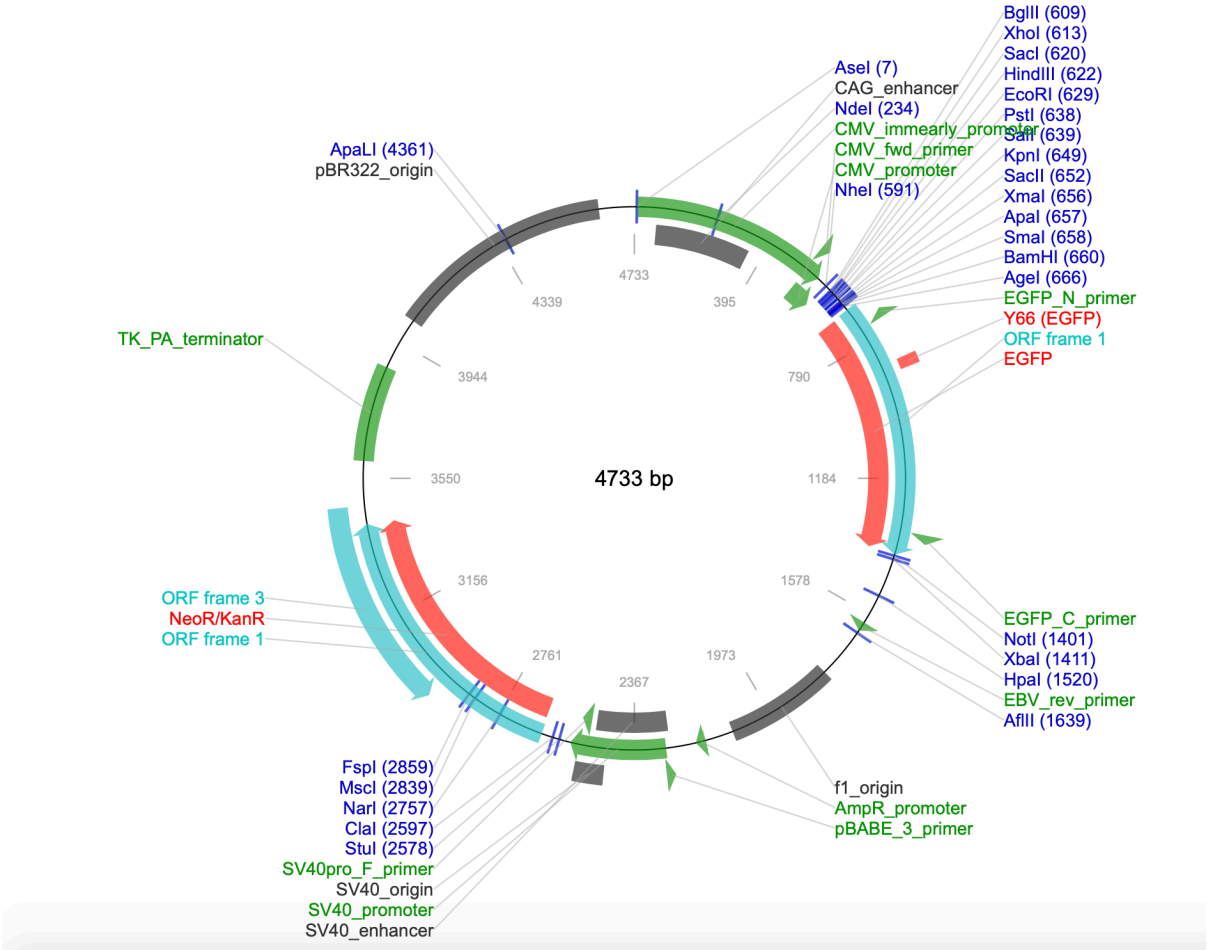
Appendix VI - Vector map of pBlueScriptII SK construct containing the *mBPIFB1* C-terminal sequence (below in red).



Appendix VII – CRISPR-Cas9 system delivery vector pX458



Appendix VIII – eGFP-N1 plasmid from Clonetech

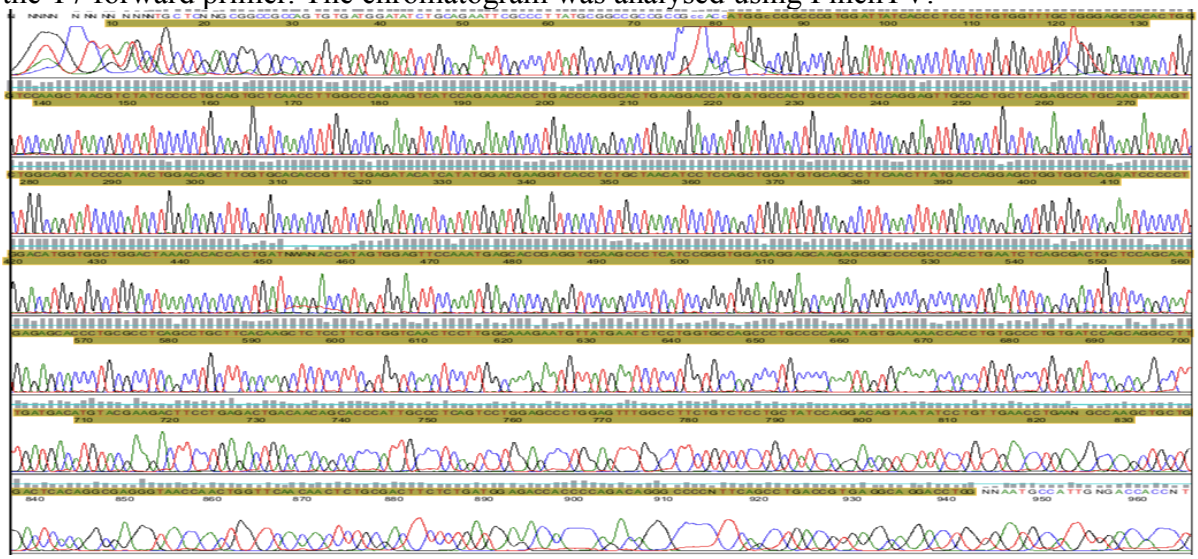


Appendix IX - Wild-type DNA sequence of mBPIFB1 and the translated protein sequence in single amino acid code. Start codon (ATG) is highlighted in green, N-glycosylation sites (AAT/AAC) in yellow and the stop codon in red (TAG).

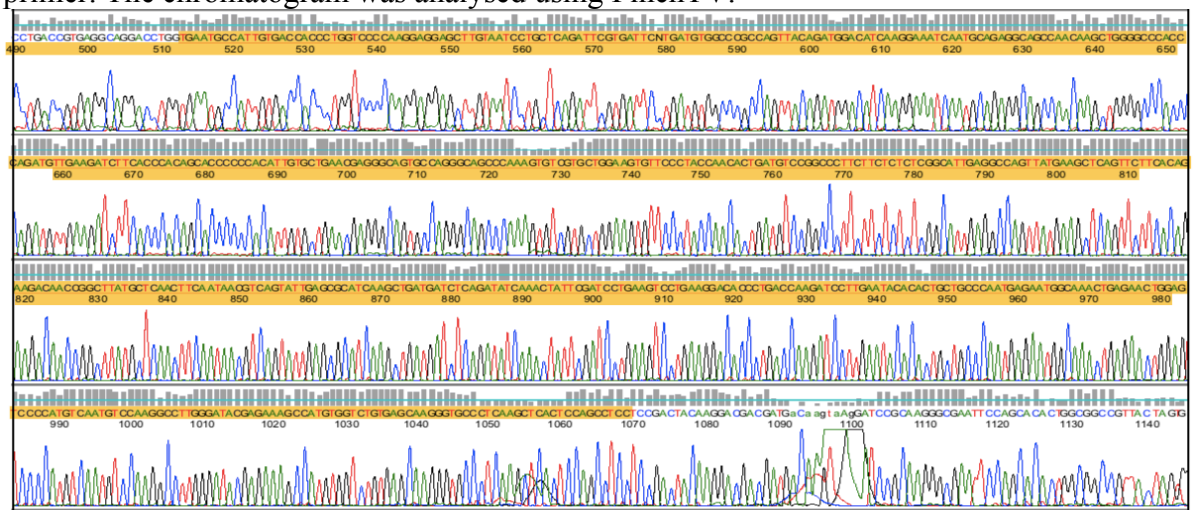
```
1 AGGAGAGGAGCAGACAGACCCAGGACTCTGGCACATTCAGGTCCTACTACCTGCCACCTA     60
.............................................................

61

840

233 --L--S--P--A--I--Q--D--S--N--I--L--L--N--L--K--A--
CTTCTCAGCGACACCCAGGAAGATGGCCGGCCCGTGGATTATCACCCTCCTCTGTGGTTT    120
.............     .......-M--A--G--P--W--I--I--T--L--L--C--G--L    13

121 GCTGGGAGCCACACTGGTCCAAGCTAACGTCTATCCCCCTGCAGTGCTCAACCTTGGCCC    180

13 --L--G--A--T--L--V--Q--A--N--V--Y--P--P--A--V--L--N--L--G--P     33

181 AGAAGTCATCCAGAAACACCTGACCCAGGCACTGAAGGACCATGATGCCACTGCCATCCT    240

33 --E--V--I--Q--K--H--L--T--Q--A--L--K--D--H--D--A--T--A--I--L     53

241 CCAGGAGTTGCCACTGCTCAGAGCCATGCAAGATAAGTCTGGCAGTATCCCCATACTGGA    300

53 --Q--E--L--P--L--L--R--A--M--Q--D--K--S--G--S--I--P--I--L--D     73

301 CAGCTTCGTGCACACCGTTCTGAGATACATCATATGGATGAAGGTCACCTCTGCTAACAT    360

73 --S--F--V--H--T--V--L--R--Y--I--I--W--M--K--V--T--S--A--N--I     93

361 CCTCCAGCTGGATGTGCAGCCTTCAACTTATGACCAGGAGCTGGTGGTCAGAATCCCCCT    420

93 --L--Q--L--D--V--Q--P--S--T--Y--D--Q--E--L--V--V--R--I--P--L    113

421 GGACATGGTGGCTGGACTAAACACACCACTGATCAAGACCATAGTGGAGTTCCAAATGAG    480

113 --D--M--V--A--G--L--N--T--P--L--I--K--T--I--V--E--F--Q--M--S    133

481 CACCGAGGTCCAAGCCCTCATCCGGGTGGAGAGGAGCAAGAGCGGCCCCGCCCACCTGAA    540

133 --T--E--V--Q--A--L--I--R--V--E--R--S--K--S--G--P--A--H--L--N    153

541 TCTCAGCGACTGCTCCAGCAATGAGAGCACCCTGCGCCTCAGCCTGCTTCACAAGCTCTC    600

153 --L--S--D--C--S--S--N--E--S--T--L--R--L--S--L--L--H--K--L--S    173

601 CTTCGTGGTCAACTCCTTGGCAAAGAATGTTATGAATCTCCTGGTGCCAGCCCTGCCCCA    660

173 --F--V--V--N--S--L--A--K--N--V--M--N--L--L--V--P--A--L--P--Q    193

661 AATAGTGAAAAACCACCTGTGCCCTGTGATCCAGCAGGCCTTTGATGACATGTACGAAGA    720

193 --I--V--K--N--H--L--C--P--V--I--Q--Q--A--F--D--D--M--Y--E--D    213

721 CTTCCTGAGACTGACAACAGCACCCATTGCCCTCAGTCCTGGAGCCCTGGAGTTTGGCCT    780

213 --F--L--R--L--T--T--A--P--I--A--L--S--P--G--A--L--E--F--G--L    233
```

```
781 TCTGTCTCCTGCTATCCAGGACAGTAATATCCTGTTGAACCTGAAGGCCAAGCTGCTGGA    K--L--L--D
253

841 CTCACAGGCGAGGGTAACCAACTGGTTCAACAACTCTGCGACTTCTCTGATGGAGACCAC    900

253 --S--Q--A--R--V--T--N--W--F--N--N--S--A--T--S--L--M--E--T--T    273

901 CCCAGACAGGGCCCCCTTCAGCCTGACCGTGAGGCAGGACCTGGTGAATGCCATTGTGAC    960

273 --P--D--R--A--P--F--S--L--T--V--R--Q--D--L--V--N--A--I--V--T    293

961 CACCCTGGTCCCCAAGGAGGAGCTTGTAATCCTGCTCAGATTCGTGATTCCTGATGTGGC   1020

293 --T--L--V--P--K--E--E--L--V--I--L--L--R--F--V--I--P--D--V--A    313

1021 CCGCCAGTTACAGATGGACATCAAGGAAATCAATGCAGAGGCAGCCAACAAGCTGGGGCC   1080

313 --R--Q--L--Q--M--D--I--K--E--I--N--A--E--A--A--N--K--L--G--P    333

1081 CACCCAGATGTTGAAGATCTTCACCCACAGCACCCCCCACATTGTGCTGAACGAGGGCAG   1140

333 --T--Q--M--L--K--I--F--T--H--S--T--P--H--I--V--L--N--E--G--S    353

1141 TGCCAGGGCAGCCCAAAGTGTCGTGCTGGAAGTGTTCCCTACCAACACTGATGTCCGGCC   1200

353 --A--R--A--A--Q--S--V--V--L--E--V--F--P--T--N--T--D--V--R--P    373

1201 CTTCTTCTCTCTCGGCATTGAGGCCAGTTATGAAGCTCAGTTCTTCACAGAAGACAACCG   1260

373 --F--F--S--L--G--I--E--A--S--Y--E--A--Q--F--F--T--E--D--N--R    393

1261 GCTTATGCTCAACTTCAATAACGTCAGTATTGAGCGCATCAAGCTGATGATCTCAGATAT   1320

393 --L--M--L--N--F--N--N--V--S--I--E--R--I--K--L--M--I--S--D--I    413

1321 CAAACTATTCGATCCTGAAGTCCTGAAGGACACCCTGACCAAGATCCTTGAATACACACT   1380

413 --K--L--F--D--P--E--V--L--K--D--T--L--T--K--I--L--E--Y--T--L    433

1381 GCTGCCCAATGAGAATGGCAAACTGAGAACTGGAGTCCCCATGTCAATGTCCAAGGCCTT   1440

433 --L--P--N--E--N--G--K--L--R--T--G--V--P--M--S--M--S--K--A--L    453

1441 GGGATACGAGAAAGCCATGTGGTCTGTGAGCAAGGGTGCCCTCAAGCTCACTCCAGCCTC   1500

453 --G--Y--E--K--A--M--W--S--V--S--K--G--A--L--K--L--T--P--A--S    473

1501 CTCCTAGAACCCTGGCTCCCCGCCCCTCCTGATGAAGACCTGGACAGTAGCAGCCAGCCC   1560

473 --S--*-.....................................................    474
1561 CGGCCAGGTCCCAGCCAGGAGTGTGGAAGCCTACTGTGTAGACTACCCCCCTGCAGTTAA    1620
```

Appendix X - (A) Sequence chromatogram of the pCRII-TOPO construct showing that the mBPIFB1 has been cloned into the vector. The sequence chromatogram was obtained using the T7 forward primer. The chromatogram was analysed using FinchTV.



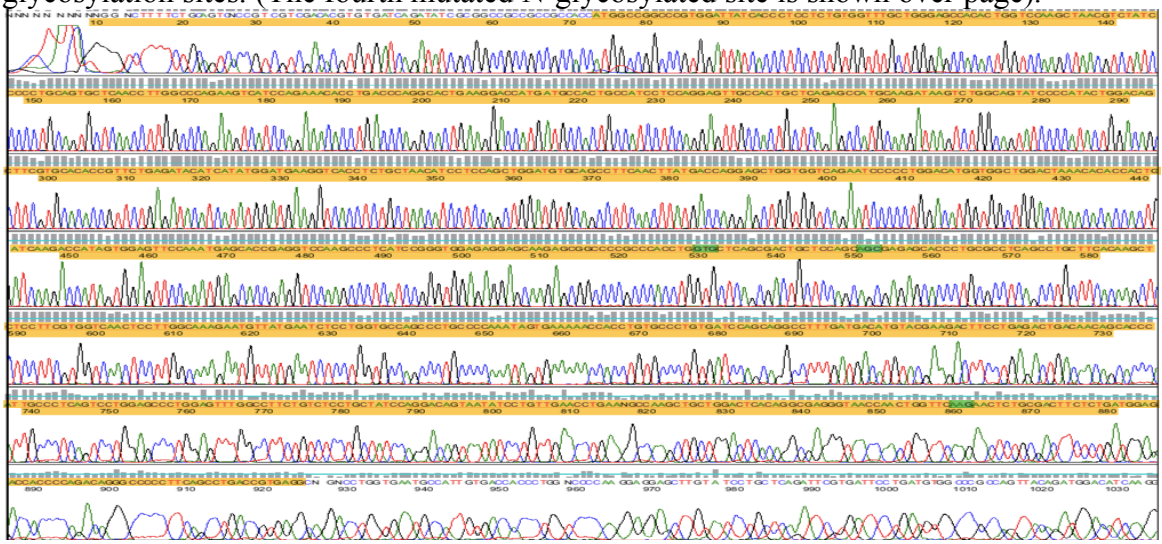
(B) Sequence chromatogram of the pCRII-TOPO construct showing that the mBPIFB1 has been cloned into the vector. The sequence chromatogram was obtained using the SP6 reverse primer. The chromatogram was analysed using FinchTV.



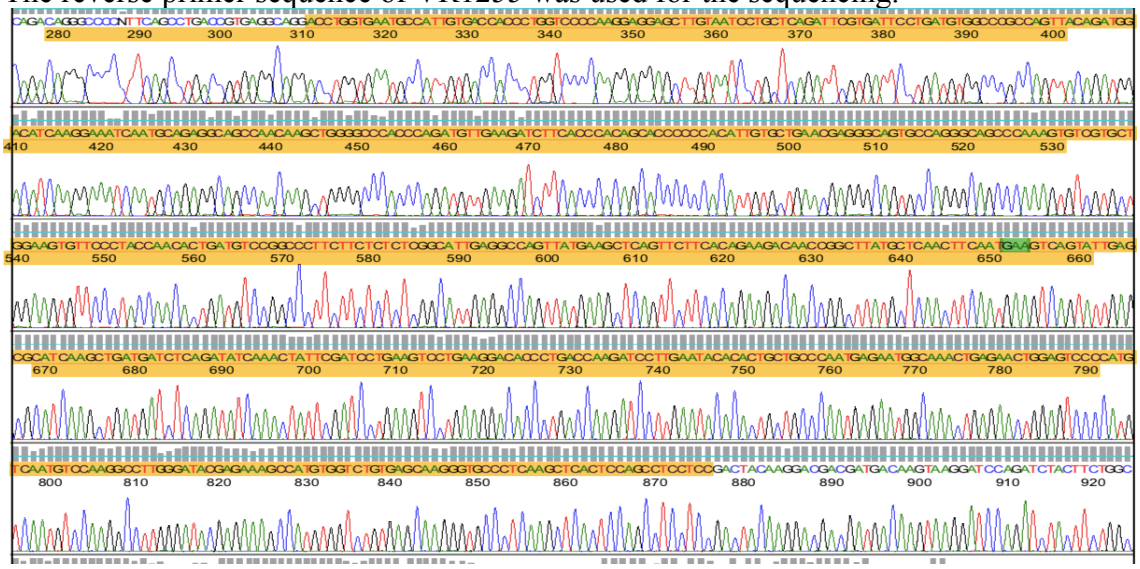
Below is a translation of the open reading frame. The DNA sequence of wild-type insert mBPIFB1 was translated into the amino-acid sequence using the ExPASy programme. Note the Flag peptide sequence is located C-terminally and highlighted in yellow.

MAGPWIITLLCGLLGATLVQANVYPPAVLNLGPEVIQKHLTQALKDHDATAILQE
LPLLRAMQDKSGSIPILDSFVHTVLRYIIWMKVTSANILQLDVQPSTYDQELVVR
IPLDMVAGLNTPLIKTIVEFQMSTEVQALIRVERSKSGPAHLNLSDCSSNESTLR
LSLLHKLSFVVNSLAKNVMNLLVPALPQIVKNHLCPVIQQAFDDMYEDFLRLTTA
PIALSPGALEFGLLSPAIQDSNILLNLKAKLLDSQARVTNWFNNSATSLMETTPD
RAPFSLTVRQDLVNAIVTTLVPKEELVILLRFVIPDVARQLQMDIKEINAEAANK
LGPTQMLKIFTHSTPHIVLNEGSARAAQSVVLEVFPTNTDVRPFFSLGIEASYEA
QFFTEDNRLMLNFNNVSIERIKLMISDIKLFDPEVLKDTLTKILEYTLLPNENGK
LRTGVPMSMSKALGYEKAMWSVSKGALKLTPASSDYKDDDDK

Appendix XI – (A) Sequence chromatogram showing mutant mBPIFB1 sequence, obtained using the forward primer of VR1255 vector. Note the positions of 3 of the 4 mutated N-glycosylation sites. (The fourth mutated N-glycosylated site is shown over page).



(B) Sequence chromatogram showing the position of the fourth mutated N-glycosylated site. The reverse primer sequence of VR1255 was used for the sequencing.



Below is a translation of the open reading frame. The DNA sequence of mutant insert mBPIFB1 was translated into the amino-acid sequence using the ExPASy programme. Note the yellow highlight is protein sequence of the flag peptide.

MAGPWIITLLCGLLGATLVQANVYPPAVLNLGPEVIQKHLTQALKDHDATAILQE
LPLLRAMQDKSGSIPILDSFVHTVLRYIIWMKVTSANILQLDVQPSTYDQELVVR
IPLDMVAGLNTPLIKTIVEFQMSTEVQALIRVERSKSGPAHLVLSDCSSSESTLR
LSLLHKLSFVVNSLAKNVMNLLVPALPQIVKNHLCPVIQQAFDDMYEDFLRLTTA
PIALSPGALEFGLLSPAIQDSNILLNLKAKLLDSQARVTNWFKNSATSLMETTPD
RAPFSLTVRQDLVNAIVTTLVPKEELVILLRFVIPDVARQLQMDIKEINAEAANK
LGPTQMLKIFTHSTPHIVLNEGSARAAQSVVLEVFPTNTDVRPFFSLGIEASYEA
QFFTEDNRLMLNFNEVSIERIKLMISDIKLFDPEVLKDTLTKILEYTLLPNENGK
LRTGVPMSMSKALGYEKAMWSVSKGALKLTPASSDYKDDDDK

Appendix XII – Online prediction of *mBPIFB*1 N-glycosylation based on sequence using the NetNGly tool (http://www.cbs.dtu.dk/services/NetNGlyc/)

```
MAGPWIITLLCGLLGATLVQANVYPPAVLNLGPEVIQKHLTQALKDHDATAILQELPLLRAMQDKSGSIPILDSFVHTVL      80
RYIIWMKVTSANILQLDVQPSTYDQELVVRIPLDMVAGLNTPLIKTIVEFQMSTEVQALIRVERSKSGPAHLNLSDCSSN     160
ESTLRLSLLHKLSFVVNSLAKNVMNLLVPALPQIVKNHLCPVIQQAFDDMYEDFLRLTTAPIALSPGALEFGLLSPAIQD     240
SNILLNLKAKLLDSQARVTNWFNNSATSLMETTPDRAPFSLTVRQDLVNAIVTTLVPKEELVILLRFVIPDVARQLQMDI     320
KEINAEAANKLGPTQMLKIFTHSTPHIVLNEGSARAAQSVVLEVFPTNTDVRPFFSLGIEASYEAQFFTEDNRLMLNFNN     400
VSIERIKLMISDIKLFDPEVLKDTLTKILEYTLLPNENGKLRTGVPMSMSKALGYEKAMWSVSKGALKLTPASS
.....                                                                                  80
.....N.....                                                                           160
.....                                                                                  240
.....                                                                                  320
.....N                                                                                 400
.....                                                                                  480

(Threshold=0.5)
-------------------------------------------------------------------------------
SeqName      Position  Potential   Jury    N-Glyc
                                  agreement result
-------------------------------------------------------------------------------
Sequence     153 NLSD   0.6753     (6/9)   +
Sequence     160 NEST   0.3299     (8/9)   -
Sequence     263 NNSA   0.4350     (6/9)   -
Sequence     400 NVSI   0.5560     (5/9)   +
-------------------------------------------------------------------------------
```

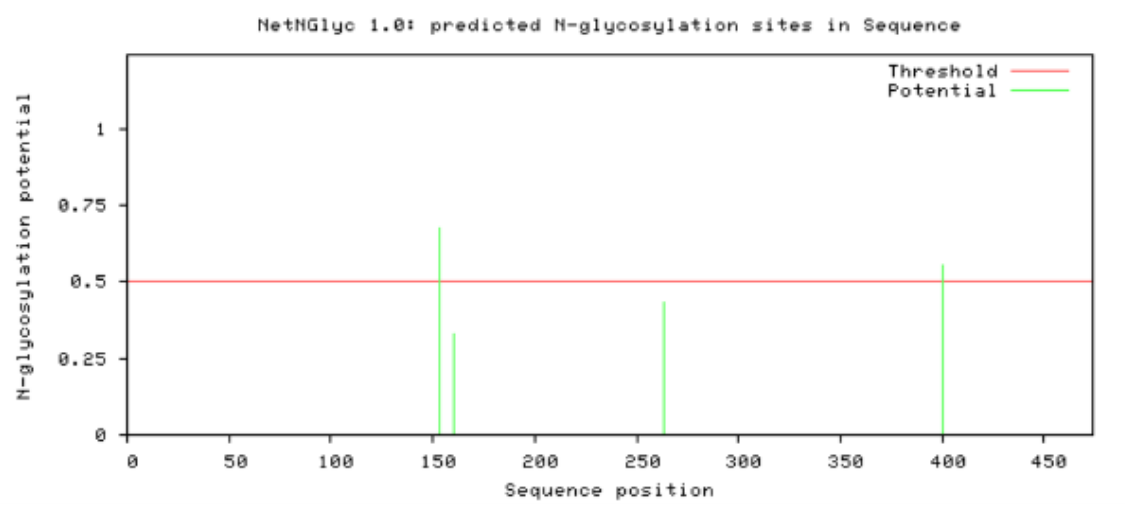


Table 9.1: DEGs across expansion/proliferative and differentiation conditions. *Genes were initially generally categorised using the DAVID Gene Functional Classification Tool (http://david.abcc.ncifcrf.gov) before manually refining them into categories listed below. Gene entrez IDs are listed in front of each gene. Genes are listed alphabetical rather than bt logFc values.*

Comparison wt-D0 *(mTEC proliferation/expansion)*

upregulated

Cell membrane and extracellular region

76942 *Ly6/Plaur domain containing 5 (Lypd5)*
241035 *polycystic kidney and hepatic disease 1 (Pkhd1)*
100340 *sphingomyelin phosphodiesterase, acid-like 3B (Smpdl3b)*
55925 *synaptotagmin VIII (Syt8)*
22361 *vanin 1 (Vnn1)*
12828 *collagen, type IV, alpha 3 (Col4a3)*
16819 *lipocalin 2 (Lcn2)*
20562 *slit guidance ligand 1 (Slit1)*

downregulated

Cell surface proteins and receptors

21940 *CD27 antigen (Cd27)*
12508 *CD53 antigen (Cd53)*
12562 *cadherin 5 (Cdh5)*
13805 *endoglin (Eng)*
14744 *G-protein coupled receptor 65(Gpr65)*
14961 *histocompatibility 2, class II antigen A, beta 1(H2-Ab1)*
14969 *histocompatibility 2, class II antigen E beta (H2-Eb1)*
16185 *interleukin 2 receptor, beta chain (Il2rb)*
16186 *interleukin 2 receptor, gamma chain (Il2rg)*
16197 *interleukin 7 receptor (Il7r)*
69774 *membrane-spanning 4-domains, subfamily A, member 6B (Ms4a6b)*
18124 *nuclear receptor subfamily 4, group A, member 3 (Nr4a3)*
78826 *purinergic receptor P2Y, G-protein coupled 10 (P2ry10)*

Hemoglobin complex

15122 *hemoglobin alpha, adult chain 1 (Hba-a1)*
110257 *hemoglobin alpha, adult chain 2 (Hba-a2)*
15129 *hemoglobin, beta adult major chain (Hbb-b1)*
15130 *hemoglobin, beta adult minor chain (Hbb-b2)*
100503605 *hemoglobin, beta adult s chain (Hbb-bs)*
101488143 *hemoglobin, beta adult t chain (Hbb-bt)*

Immunity

18843 *BPI fold containing family A, member 1 (Bpifa1)*
228801 *BPI fold containing family B, member 1 (Bpifb1)*
12478 *CD19 antigen (Cd19)*
16149 *CD74 antigen (Cd74)*
12518 *CD79A antigen (immunoglobulin-associated alpha) (Cd79a)*
15985 *CD79B antigen (Cd79b)*
12524 *CD86 antigen (Cd86)*
321019 *G protein-coupled receptor 183 (Gpr183)*
30925 *SLAM family member 6 (Slamf6)*

75345 *SLAM family member 7 (Slamf7)*
15002 *histocompatibility 2, O region beta locus (H2-Ob)*
14960 *histocompatibility 2, class II antigen A, alpha (H2-Aa)*
14961 *histocompatibility 2, class II antigen A, beta 1 (H2-Ab1)*
14969 *histocompatibility 2, class II antigen E beta (H2-Eb1)*
15000 *histocompatibility 2, class II, locus Mb2 (H2-DMb2)*
54167 *inducible T cell co-stimulator (Icos)*
16331 *inositol polyphosphate-5-phosphatase D (Inpp5d)*
17084 *lymphocyte antigen 86 (Ly86)*
240754 *lymphocyte transmembrane adaptor 1 (Lax1)*
83490 *phosphoinositide-3-kinase adaptor protein 1 (Pik3ap1)*
18751 *protein kinase C, beta (Prkcb)*
19264 *protein tyrosine phosphatase, receptor type, C (Ptprc)*
72049 *tumor necrosis factor receptor superfamily, member 13c (Tnfrsf13c)*

Chemokine signalling pathway
19354 *Rac family small GTPase 2 (Rac2)*
18829 *chemokine (C-C motif) ligand 21A (serine) (Ccl21a)*
100042493 *chemokine (C-C motif) ligand 21B (leucine) (Ccl21b)*
20299 *chemokine (C-C motif) ligand 22 (Ccl22)*
12458 *chemokine (C-C motif) receptor 6 (Ccr6)*
94176 *dedicator of cyto-kinesis 2 (Dock2)*
17969 *neutrophil cytosolic factor 1 (Ncf1)*
100504239 *predicted gene 10591 (Gm10591)*
100504346 *predicted gene 13304 (Gm13304)*

Lectins
12515 *CD69 antigen (Cd69)*
12517 *CD72 antigen (Cd72)*
80782 *killer cell lectin-like receptor subfamily B member 1B (Klrb1b)*
20343 *selectin, lymphocyte (Sell)*
20345 *selectin, platelet (p-selectin) ligand (Selplg)*
20234 *spermine binding protein (Sbp)*
20387 *surfactant associated protein A1 (Sftpa1)*

Comparison D0-D14 *(mTEC differentiation)*

upregulated
Cilium and cell projection
244653 *HYDIN, axonemal central pair apparatus protein (Hydin)*
74918 *IQ motif containing with AAA domain (Iqca)*
212190 *UBX domain protein 10 (Ubxn10)*
78801 *adenylate kinase 7(Ak7)*
74453 *cilia and flagella associated protein 53 (Cfap53)*
78774 *cilia and flagella associated protein 61 (Cfap61)*
207686 *cilia and flagella associated protein 69 (Cfap69)*
76670 *cilia and flagella associated protein 70 (Cfap70)*
51938 *coiled-coil domain containing 39 (Ccdc39)*
330830 *dynein regulatory complex subunit 7 (Drc7)*
13411 *dynein, axonemal, heavy chain 11 (Dnah11)*
110083 *dynein, axonemal, heavy chain 12 (Dnah12)*
110082 *dynein, axonemal, heavy chain 5 (Dnah5)*
330355 *dynein, axonemal, heavy chain 6 (Dnah6)*

627872 *dynein, axonemal, heavy chain 7A (Dnah7a)*
73873 *family with sequence similarity 161, member A (Fam161a)*
75050 *kinesin family member 27 (Kif27)*
74354 *leucine-rich repeats and guanylate kinase domain containing (Lrguk)*
22092 *radial spoke head 1 homolog (Chlamydomonas) (Rsph1)*
212892 *radial spoke head 4 homolog A (Chlamydomonas) (Rsph4a)*
19888 *retinitis pigmentosa 1 (human) (Rp1)*
66722 *sperm associated antigen 16 (Spag16)*
21645 *t-complex-associated testis expressed 1 (Tcte1)*
21689 *tektin 1 (Tekt1)*
73301 *tetratricopeptide repeat domain 29 (Ttc29)*

differentiation
12427 *cyclin A1(Ccna1)*
71827 *leucine rich repeat containing 34 (Lrrc34)*
104362 *meiosis expressed gene 1 (Meig1)*
73472 *spermatogenesis associated 18 (Spata18)*
74052 *tetratricopeptide repeat domain 21A (Ttc21a)*
21892 *tolloid-like (Tll1)*

membrane associated
68891 *CD177 antigen (Cd177)*
270162 *ELMO/CED-12 domain containing 1 (Elmod1)*
67483 *RIKEN cDNA 1700028P14 gene (1700028P14Rik)*
69398 *cadherin-related family member 4 (Cdhr4)*
329252 *leucine-rich repeat-containing G protein-coupled receptor 6 (Lgr6)*
238939 *predicted gene 281 (Gm281)*
229277 *stomatin (Epb7.2)-like 3 (Stoml3)*
381107 *transmembrane protein 232 (Tmem232)*
114602 *zinc finger, MYND domain containing 10 (Zmynd10)*

downregulated
Extracellular region/space and secreted
381493 *S100 calcium binding protein A7A (S100a7a)*
386463 *corneodesmosin (Cdsn)*
13179 *decorin (Dcn)*
13615 *endothelin 2 (Edn2)*
16592 *fatty acid binding protein 5, epidermal (Fabp5)*
19227 *parathyroid hormone-like peptide (Pthlh)*
69542 *RIKEN cDNA 2300002M23 gene (2300002M23Rik)*
282619 *suprabasin (Sbsn)*

Keratinization / keratinocyte differentiation / epidermis development
74591 *ATP-binding cassette, sub-family A (ABC1), member 12 (Abca12)*
72383 *Cornifelin (Cnfn)*
16175 *interleukin 1 alpha (Il1a)*
16666 *keratin 16 (Krt16)*
20753 *small proline-rich protein 1A (Sprr1a)*
20755 *small proline-rich protein 2A1 (Sprr2a1)*
20758 *small proline-rich protein 2D (Sprr2d)*
20759 *small proline-rich protein 2E (Sprr2e)*
20760 *small proline-rich protein 2F (Sprr2f)*
20761 *small proline-rich protein 2G (Sprr2g)*
20765 *small proline-rich protein 2K (Sprr2k)*
20766 *small proline-rich protein 3 (Sprr3)*
15218 *forkhead box N1 (Foxn1)*

64661 *keratinocyte differentiation associated protein (Krtdap)*

peptidase inhibitor activity

209294 *cystatin A1 (Csta1)*
76400 *phosphatidylethanolamine binding protein 2 (Pbp2)*
18788 *serine (or cysteine) peptidase inhibitor, clade B, member 2 (Serpinb2)*
18787 *serine (or cysteine) peptidase inhibitor, clade E, member 1 (Serpine1)*
72432 *serine peptidase inhibitor, Kazal type 5 (Spink5)*

Hydrolase activity

109254 *androgen dependent TFPI regulating protein (Adtrp)*
11846 *arginase, liver (Arg1)*
67855 *aspartic peptidase, retroviral-like 1 (Asprv1)*
71884 *chitinase 1 (chitotriosidase) (Chit1)*
23844 *chloride channel accessory 1 (Clca1)*
71932 *epoxide hydrolase 3 (Ephx3)*
626834 *kallikrein related-peptidase 13 (Klk13)*
78753 *lipase, family member M (Lipm)*
18792 *plasminogen activator, urokinase (Plau)*
213171 *protease, serine 27 (Prss27)*
319875 *transmembrane protease, serine 11B (Tmprss11b)*

Membrane components

74591 *ATP-binding cassette, sub-family A (ABC1), member 12 (Abca12)*
77596 *adhesion G protein-coupled receptor F1 (Adgrf1)*
109254 *androgen dependent TFPI regulating protein (Adtrp)*
67855 *aspartic peptidase, retroviral-like 1 (Asprv1)*
230099 *carbonic anhydrase 9 (Car9)*
23844 *chloride channel accessory 1 (Clca1)*
12737 *claudin 1 (Cldn1)*
225256 *desmoglein 1 beta (Dsg1b)*
66811 *dual oxidase maturation factor 2 (Duoxa2)*
75577 *dynactin associated protein (Dynap)*
71932 *epoxide hydrolase 3 (Ephx3)*
245128 *expressed sequence AU018091 (AU018091)*
72077 *glucosaminyl (N-acetyl) transferase 3, mucin type (Gcnt3)*
171285 *hepatitis A virus cellular receptor 2 (Havcr2)*
16409 *integrin alpha M (Itgam)*
70274 *lymphocyte antigen 6 complex, locus G6E (Ly6g6e)*
269328 *mucin 15 (Muc15)*
74090 *progestin and adipoQ receptor family member V (Paqr5)*
66183 *serine palmitoyltransferase, small subunit B (Sptssb)*
20970 *syndecan 3 (Sdc3)*
66561 *transmembrane epididymal family member 3 (Teddm3)*
56277 *transmembrane protein 45a (Tmem45a)*
235135 *transmembrane protein 45b (Tmem45b)*

Cytokine activity and negative regulation of cell proliferation

70045 *RIKEN cDNA 2610528A11 gene (2610528A11Rik)*
16175 *interleukin 1 alpha (Il1a)*
93672 *interleukin 24 (Il24)*
57349 *pro-platelet basic protein (Ppbp)*
57277 *secreted Ly6/Plaur domain containing 1 (Slurp1)*

wt-D0 expansion/proliferation upregulated or D0-D14 differentiation downregulated

Extracellular space or secreted
209232 *WAP four-disulfide core domain 5 (Wfdc5)*
244332 *defensin beta 14 (Defb14)*
380924 *olfactomedin 4 (Olfm4)*
116872 *serine (or cysteine) peptidase inhibitor, clade B, member 7 (Serpinb7)*
20706 *serine (or cysteine) peptidase inhibitor, clade B, member 9b (Serpinb9b)*

Hydrolases
12182 *bone marrow stromal cell antigen 1 (Bst1)*
72269 *cytidine deaminase (Cda)*
317653 *kallikrein related-peptidase 14 (Klk14)*
259277 *kallikrein related-peptidase 8 (Klk8)*
101533 *kallikrein related-peptidase 9 (Klk9)*

Glycoproteins
75690 *V-set and immunoglobulin domain containing 10 like (Vsig10l)*
76459 *carbonic anhydrase 12 (Car12)*
66601 *transmembrane and immunoglobulin domain containing 1 (Tmigd1)*
100647 *uroplakin 3B (Upk3b)*

wt-D0 expansion/proliferation downregulated or D0-D14 differentiation upregulated

Cytoskeleton
75465 *dynein light chain roadblock-type 2 (Dynlrb2)*
75429 *family with sequence similarity 183, member B (Fam183b)*
70113 *outer dense fiber of sperm tails 3B (Odf3b)*
320277 *sperm flagellar 2 (Spef2)*
67971 *tubulin polymerization-promoting protein family member 3 (Tppp3)*

Cytoplasm
380683 *SEC14-like lipid binding 3 (Sec14l3)*
17110 *lysozyme 1 (Lyz1)*
17105 *lysozyme 2 (Lyz2)*
626596 *regulator of G-protein signalling 22 (Rgs22)*

Cell membrane
68764 *cadherin-related family member 3 (Cdhr3)*
14262 *flavin containing monooxygenase 3 (Fmo3)*
226565 *flavin containing monooxygenase 6 (Fmo6)*
67664 *ring finger protein 125 (Rnf125)*

Appendix XIII – Violin plot representation of *BPIFA1* versus *BPIFB1* expression across a healthy human airway related to Figure 3.17. Single-cell RNA sequencing data comprising of a total of 77,969 cells, which were collected from 35 distinct locations, spanning from the nose to the 12th division of the airway tree. Image was generated using the Deprez et al., 2020 data set and were visualised through the interactive web tool (https://www.genomique.eu/cellbrowser/HCA/).