CHALMERS
UNIVERSITY OF TECHNOLOGY

Investigation of ER Stress in Respiratory Diseases with *in Vitro* and *in Silico* Models

Induction of ER stress and the UPR pathway in human bronchial epithelial cells and analysis of UPR signature genes by utilizing public scRNA sequencing data from patients with IPF and COPD

Master's thesis in Biotechnology

JESSICA REIJER

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

Master's thesis 2022

Investigation of ER Stress in Respiratory Diseases with *in Vitro* and *in Silico* Models

Induction of ER stress and the UPR pathway in human bronchial epithelial cells and analysis of UPR signature genes by utilizing public scRNA sequencing data from patients with IPF and COPD

JESSICA REIJER

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Biology and Biological Engineering
Division of Systems and Synthetic Biology
Chalmers University of Technology
Gothenburg, Sweden 2022

Investigation of ER Stress in Respiratory Diseases with *in Vitro* and *in Silico* Models
JESSICA REIJER
Department of Biology and Biological engineering
Division of Systems and Synthetic Biology
Chalmers University of Technology

Supervisors:
Karin Björhall, Senior Research Scientist,
Bioscience COPD/IPF, Research and Early Development,
Respiratory and Immunology, BioPharmaceuticals R&D,
AstraZeneca, Gothenburg, Sweden,
karin.bjorhall@astrazeneca.com

Susan Monkley, Associate Principal Informatician,
Transitional Science and Experimental Medicine, Research and Early Development,
Respiratory and Immunology, BioPharmaceuticals R&D,
AstraZeneca, Gothenburg, Sweden,
susan.monkley1@astrazeneca.com

Pia Berntsson, Senior Research Scientist,
Bioscience COPD/IPF, Research and Early Development,
Respiratory and Immunology, BioPharmaceuticals R&D,
AstraZeneca, Gothenburg, Sweden,
pia.berntsson@astrazeneca.com

Stefan Vollmer, Associate Principal Scientist,
Bioscience COPD/IPF, Research and Early Development,
Respiratory and Immunology, BioPharmaceuticals R&D,
AstraZeneca, Gothenburg, Sweden,
stefan.vollmer@astrazeneca.com

Examiner:
Mikael Molin, Associate professor/Senior researcher,
Department of Biology and Biological Engineering,
Division of Systems and Synthetic Biology,
Chalmers University of Technology,
mikael.molin@chalmers.se

Master's Thesis 2022: BBTX03
Department of Biology and Biological Engineering
Division of Systems and Synthetic Biology
Chalmers University of Technology
SE-412 96 Gothenburg
Typeset in LaTeX, Printed by Chalmers Reproservice, Gothenburg, Sweden 2022

Investigation of ER Stress in Respiratory Diseases with *in Vitro* and *in Silico* Models
JESSICA REIJER
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

Chronic respiratory diseases are amongst the most common causes of death worldwide. Two of these are chronic obstructive pulmonary disease (COPD) and idiopathic pulmonary fibrosis (IPF). COPD is the third leading cause of death worldwide and characterized by irreversible respiratory airflow limitations, with cigarette smoke as one of the most common causes of disease. IPF is characterized by irreversible scarring of the distal lung and often advances into respiratory failure and death within five years of diagnosis.

Endoplasmic reticulum (ER) stress has been observed in these respiratory diseases. The aim of this project was to investigate ER stress and the unfolded protein response (UPR) with *in vitro* and *in silico* models. The *in vitro* models included induction of UPR in primary human bronchial epithelial cells (HBECs) by exposure to well characterized ER stress inducers, tunicamycin and thapsigargin, as well as stimuli associated with chronic respiratory disease, H_2O_2 and cigarette smoke extract. Gene and protein readouts of the UPR pathway, that are characterized markers used to monitor ER stress and UPR, were established and cell viability examined. Gene expressions of for example BiP, CHOP, spliced XBP1 and PPP1R15A and protein expressions of eIF2α, phospho-eIF2α, BiP and ATF6 were induced in HBECs after exposure to the stimuli used.

In the *in silico* model, an UPR gene signature was investigated in single cell RNA sequencing (scRNA-seq) datasets from patients with IPF and COPD, to link epithelial cell type specific transcriptomic markers of UPR to the diseases and investigate the chances in UPR response of different epithelial cell subtypes at gene expression level. Epithelial club cells and aberrant basaloid cells were the most abundant cell subtypes found in IPF and with the highest up-regulation of UPR. In COPD, UPR genes were mostly down-regulated across epithelial cell subtypes.

This project was performed at AstraZeneca in the Bioscience COPD IPF Department within the Early Respiratory and Immunology R&D unit located in Gothenburg, Sweden.

Keywords: COPD, IPF, ER stress, Unfolded protein response, *in vitro*, *in silico*, human bronchial epithelial cells, qPCR, Western blot, scRNA-seq

Acknowledgements

I would like to start by thanking my family and friends for being so supportive during this very intense but valuable and rewarding period of my life. Then, I would like to give a large thank you to my supervisors that has guided me throughout my master's thesis work. First, to Karin Björhall that has been my main supervisor and introduced me to AstraZeneca and the project, and supported me with knowledge and guidance when planning my project. Without this support, I would not have managed to accomplish the project with close as much knowledge gained or gotten the opportunity to get to know so many fantastic people during my master's thesis work. I would also like to thank Pia Berntsson, for giving me the best introduction I could have gotten to the laboratory work and methods used there, which has enabled the *in vitro* work performed. Furthermore, I want to thank Sue Monkley for being my supervisor in the bioinformatic part of the project, which would not have been possible without her and which has resulted in valuable knowledge of scRNA-seq analysis and the different platforms I have used for the analysis. I am also very thankful for the introduction to both traditional and automated Western blot by Karolina Öst, from which I have learned a lot of valuable knowledge and insights. Also, I want to thank Stefan Vollmer for all the valuable knowledge I have been provided with throughout this time. Finally, a big thank you to Linda Yrlid for letting me be a part of the team and giving me the opportunity to perform my master's thesis work at AstraZeneca.

Jessica Reijer, Gothenburg, June 2022

Contents

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

COPD	Chronic Obstructive Pulmonary Disease
CSE	Cigarette Smoke Extract
CT	Cycle Threshold
DE	Differential Expression
ECM	Extracellular Matrix
EMT	Epithelial–Mesenchymal Transition
ER	Endoplasmic Reticulum
ERAD	Endoplasmic Reticulum Associated Degradation
ERAF	Endoplasmic Reticulum Associated Folding
FACS	Flow-Activated Cell Sorting
HBECs	Human Bronchial Epithelial Cells
IPA	Ingenuity Pathway Analysis
IPF	Idiopathic Pulmonary Fibrosis
MMPs	matrix metalloproteinases
RIDD	Regulated IRE1-Dependent Decay
ROS	Reactive Oxygen Species
qPCR	Quantitative Polymerase Chain Reaction
SASP	senescence associated secretory phenotype
SFTPC	Surfactant Protein C
scRNA-seq	Single Cell RNA Sequencing
UPR	Unfolded Protein Response

List of Figures

List of Tables

1
Introduction

This chapter describes the background behind the project and why it was performed. It also includes the main objectives, limitations and ethical aspects concerning the project and an overview of the methods used.

1.1 Background

Some of the most common non-communicable diseases worldwide are chronic respiratory diseases, such as chronic obstructive pulmonary disease (COPD), asthma, idiopathic pulmonary disease (IPF), pulmonary sarcoidosis and cystic fibrosis, causing suffering and death of millions of people each year. The prevalence and global health burden of chronic diseases are large and research dedicated to them is crucial [1].

COPD is the third leading cause of death worldwide and caused 3.23 million deaths in 2019, which also contributes to it being a major economic and social burden [2], [3]. It is characterized by irreversible expiratory airflow limitations and caused by inhalation of noxious particles or gases, such as cigarette smoke [4]. IPF is another progressive lung disease with few treatment options available and with a high mortality rate [5]. It is characterized by irreversible scarring of the distal lung, leading to respiratory failure.

The endoplasmic reticulum (ER) plays a critical role in protein, lipid and glucose metabolism, and all secretory proteins must be folded and post-translationally processed inside the ER. When the protein homeostasis is perturbed, a situation named ER stress arises. One consequence of ER stress is activation of the unfolded protein response (UPR), which consist of transcriptional and signaling events that aims to restore the ER homeostasis via decreased protein translation, upregulation of ER chaperones and other molecules associated with increased degradation of misfolded proteins (ER-associated protein degradation, ERAD) and productive folding (ER-associated folding, ERAF). The UPR consist of three different arms which are regulated by different sensors located on the ER membrane. The first arm is regulated by the ER resident protein kinase RNA-like ER kinase (PERK) which phosphorylates eIF2α that inhibits translational responses. While phosphorylation of eIF2α represses global translation, it results in preferential translation of ATF4, a transcription factor regulating key genes for adaptive functions such as CHOP. The second arm is regulated by the endoribonucleases inositol requiring enzyme 1

(IRE1α), that upon activation splices X-box binding protein 1 (XBP1) mRNA, allowing it to become a functional transcription factor, regulating gene expression of ER chaperones and ERAD genes. Furthermore, IRE1α degrades RNA during high ER stress levels in a process called regulated IRE1α-dependent decay (RIDD). The last arm is regulated by the transcription factor ATF6, which enhances ER function by upregulating chaperones involved in protein folding. Besides regulations within each arm, the UPR arms regulate and interact with each other [5].

In the airways, ER stress and activation of the UPR can be triggered by a number of factors such as inhaled toxins (cigarette smoke, pollution) and pathogens (bacteria, viruses, fungi). Although the UPR aims to restore protein homeostasis for sustained cell function, in situations where this cannot be achieved it initiates death programs. The UPR have been found to be upregulated in many respiratory diseases, however its dysregulation and causal role in disease still need further investigations [4]. There are evidence of a connection between UPR and disease for both COPD and IPF [4]. Increased UPR responses are observed in IPF patients, and surfactant protein C gene mutations in pulmonary fibrosis indicates that ER stress can be directly causative of the disease. Furthermore, UPR-associated protein expression has been reported to increase in cells from patients with COPD [4].

Nevertheless, further research is needed to identify key factors that induce ER stress, both to understand critical downstream processes activated in these chronic respiratory diseases, to enable modulation of UPR pathways for therapeutic benefits and to find novel treatments [6]. In this project, COPD and IPF was investigated *in silico* by analysing epithelial cellular subtypes in the diseases and *in vitro* by developing protocols for ER stress induction and UPR activation in human primary bronchial epithelial cells (HBECs). Increased scientific knowledge around disease-causing molecular mechanisms could then be used to identify novel targets for treatment of the two respiratory diseases COPD and IPF.

1.2 Aim

The main objective of this project was to investigate and modulate the UPR pathway in lung epithelial cells. The project was divided into two parts. In the first part, bioinformatics was used to investigate UPR gene signatures in single cell RNA sequencing (scRNA-seq) datasets from studies including IPF/COPD patients. The bioinformatics part was performed in order to explore UPR in the airway epithelium and link transcriptomic markers of ER stress and the UPR pathway with disease. In the second part, protocols were established to induce ER stress and monitor UPR response and associated biological pathways *in vitro* in HBECs.

1.3 Specification of objective

The main objective can be divided into the following focus areas:

- Characterisation of cell type specific UPR signatures in the airway epithelium for IPF and COPD by investigating publicly available scRNA-seq datasets.
- Establishment of protocols for gene and protein readouts to induce ER stress and monitor the UPR and associated disease relevant biological pathways in primary lung cells by:
  - Investigation of cell viability and UPR induction in submerged HBECs challenged with either tunicamycin, thapsigargin, cigarette smoke or H_2O_2.

1.4 Limitations

The project will involve tests on healthy submerged HBECs. Other cell types are affected in IPF and COPD patients as well, but they will not be investigated in this thesis. When it comes to the bioinformatics component, there is a limited number of publicly available scRNA-seq datasets concerning these diseases, which will limit the establishment of cell type specific UPR signatures in the airway epithelium. Furthermore, different dissociation methods used for extracting lung cell populations affects cell types and number of cells extracted, which is a major limitation regarding both the proportion of different cell types and the actual number of epithelial cells in each study. Regarding experimental limitations, there is a limitation in only exposing cells to acute ER stress when investigating novel targets for chronic diseases. Thereby, it is of importance to also try to establish a 'chronic' ER stress model where cells are repeatedly challenged, which will be dealt with when analysing scRNA-seq datasets *in silico*.

1.5 Societal and ethical aspects

The project will only include *in vitro* experiments performed on HBECs. Hence, no harm will be inflicted on either animals or humans. The cells are purchased from Swiss pharmaceutical and biotech company Lonza and no additional approval by the Swedish Research Ethical Committee in Gothenburg will be needed.

The results from this thesis might be used for further testing *in vivo*. Therefore, ethical dilemmas concerning whether it is right or not to use animals for testing might arise in the future. However, it can be argued that the benefits outweigh the risks since the procedure of drug development follows the 3 Rs (replace, reduce, refine) in animal research. This is accomplished by performing experiments *in vitro* and *in silico* prior to *in vivo* testing, thereby replacing the use of animals with other methods and reducing the number of animals required. Since the research to be conducted will increase the scientific knowledge around respiratory diseases and might contribute to novel information about IPF and COPD, which are major causes of death worldwide, the positive impacts on society can be seen as of higher value than the number of animals that might be needed in future research.

1.6 Thesis structure

The following chapters of the thesis are divided into *theory, analytical procedure, materials and method, result, discussion* and *conclusion*. In the theory chapter, the respiratory system and epithelium are explained, as well as ER stress and UPR in relation to the diseases COPD and IPF and information about *in vitro* and *in silico* models. The analytical procedure chapter will include the analytical methods being used, namely qPCR, Western blot and an algorithm for scRNA analysis called Venice. The materials and method chapter will include explanations of the datasets used for the bioinformatics part and the analyses made, and the methods used in the experimental part. The result, discussion and conclusions will then be presented.

2
Theory

This chapter is divided into six sections. The respiratory system and its epithelial cell subtypes are explained in the first section, followed by an explanation of ER stress and the UPR pathway in the second section. Thereafter, the two respiratory diseases COPD and IPF are explained in section 3 and 4, followed by an explanation of *in silico* and *in vitro* models in section 5 and 6.

2.1 Respiratory system and epithelial cells

The respiratory tract is a complex organ system which primary responsibility is to efficiently carry out gas exchange between inhaled air and the bloodstream. It can be divided into two compartments. The upper respiratory tract that consists of the nasal cavity, pharynx and larynx and the lower respiratory tract that consist of the conducting airways and the respiratory zone. The conducting airways includes the bronchea, bronchi and bronchioles while the respiratory zone includes respiratory bronchioles and alveoli. The main purpose of the conducting airways is the inhalation and exhalation of air that passes it, while the main purpose of the respiratory zone is gas exchange between the lungs and the pulmonary capillaries [7]. For an illustrative overview of the conducting and respiratory zones, see Figure 2.1.

The respiratory system also plays an important part in maintaining respiratory homeostasis, which if dysregulated might lead to disease. It is also important for host defence, since it supports a multifaceted frontline defence system together with immune cells and maintains homeostasis by protecting against particles, toxins, allergens and pathogens in the airways [8]. For example, the epithelium has to manage toxins within the inhaled environment and it has been shown that epithelial dysfunction drives several chronic respiratory diseases [7].

2.1.1 Epithelial cell subtypes

The respiratory epithelial cells lining the airways are phenotypically and functionally different depending on location in the respiratory tract. The nasal and tracheal airways mainly consist of ciliated, goblet, basal, club, neuroendocrine and serous cells, the bronchiolar epithelium mainly consist of bronchioalveolar stem cells and club cells, and the alveoli consist of alveolar type 1 (AT1) and type 2 (AT2) cells. For an illustrative overview of the cell subtypes, see Figure 2.1.

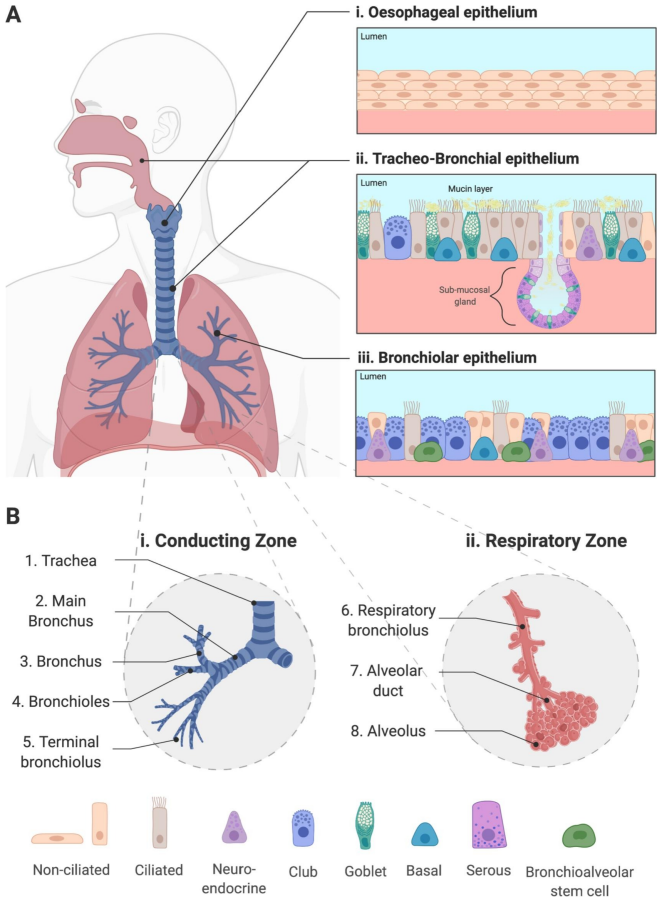

Figure 2.1: (A) Human epithelial cell subtypes. (B) The conducting (i) and respiratory (ii) zone of the lower respiratory tract. Reprinted from [9], with permission from Elsevier.

Basal cells have a cuboidal shape and are key modulators of respiratory homeostasis and epithelial regeneration when the epithelium has been injured. They can be seen as principal stem cells of the airway due to their ability of differentiating into for example goblet, club and ciliated cells. Club cells on the other hand are dome-shaped but also act as stem cells. They can give rise to ciliated and goblet cells and in a study by Habermann et al., it was observed that they differentiate into transitional AT2 cells as well [8], [10]. Club cells can contribute to epithelial repair, when acting as stem cells, and act as secretory cells due to expressing secretory proteins such as SCGB1A1, which is the most abundant protein in the airway lining fluid [8]. Hence, they are important for maintenance of the respiratory homeostasis and dysregulation of these cells contributes to respiratory diseases, such as COPD and IPF [8].

Goblet cells are the main mucus producing cells which contributes to effective MCC together with ciliated cells. The mucus is a collection of different products, amongst others mucins like MUC5AC and MUC5B that protects the lung by lining them with a mucus layer. However, an increased mucus production is a common feature in several pulmonary diseases, such as COPD and asthma. Ciliated cells can be found throughout the airways and are descendants of goblet and club cells. They are also important for homeostasis through mucociliary clearance (MCC), which is

a process where microorganisms and other particles are trapped in the mucus and expelled through rhythmic beating of hair-like cilia. Impairment of ciliated cells, resulting in shortening of cilia and reduction of cell numbers, leading to impaired MCC, is a consequence of smoking and a main feature of COPD [8].

AT2 cells have a cuboidal structure and their main function is to synthesize and secrete pulmonary surfactants. However, they also act as alveolar stem cells that can differentiate into AT1 cells as a repair mechanism to restore alveolar homeostasis. AT1 cells in turn are much larger squamous cells, covering 95 % of the alveolar surface area and are an important component for the epithelial part of the thin air-blood barrier and for gas exchange since it allows oxygen into the bloodstream by passive diffusion [11].

2.2 Endoplasmic reticulum stress and unfolded protein response

The ER consist of a membranous tubular network and is important for processes such as protein and lipid biosynthesis and calcium homeostasis. Together with the Golgi apparatus, endosomes, lysosomes and secretory vesicles, the ER is part of the protein secretory pathway. Secretory proteins are synthesized, folded and assembled in the ER before transportation to the Golgi apparatus where they are directed to their final compartment. The folding and processing in the ER take place in a highly oxidizing environment with the help of specific ER resident proteins that support the folding, disulfide bond formatting and oligomerization [12].

Conditions that perturb protein folding and homeostasis, such as protein load exaggeration and incorrect amino acid sequences, will lead to accumulation of unfolded or misfolded proteins and ER stress [4]. Furthermore, ER stress and UPR contribute to production of reactive oxygen species (ROS) through dysregulated disulfide bond formation, which lead to oxidative stress. The highly oxidizing environment in the ER is important for protein folding, but ER stress and oxidative stress can accentuate each other and result in cell dysfunction and apoptotic signals when the protein folding load is increased [13].

The stress conditions are recognised by activation of the three stress sensors located in the ER membrane, that are part of the UPR. The activation occurs when the chaperones called BiP are dissociated from the sensors and instead bind to unfolded proteins in the ER [14]. There are three main purposes with activated ER stress response: to decrease the rate of protein synthesis while enhancing the ER folding capacity, to stimulate disposal of terminally misfolded proteins and lastly, if unable to restore homeostasis, to trigger cell death [14]. The UPR pathway and the downstream cellular events as result of the three stress sensors can be seen in Figure 2.2.

The first UPR arm is initiated by the ER stress sensor called protein kinase RNA-

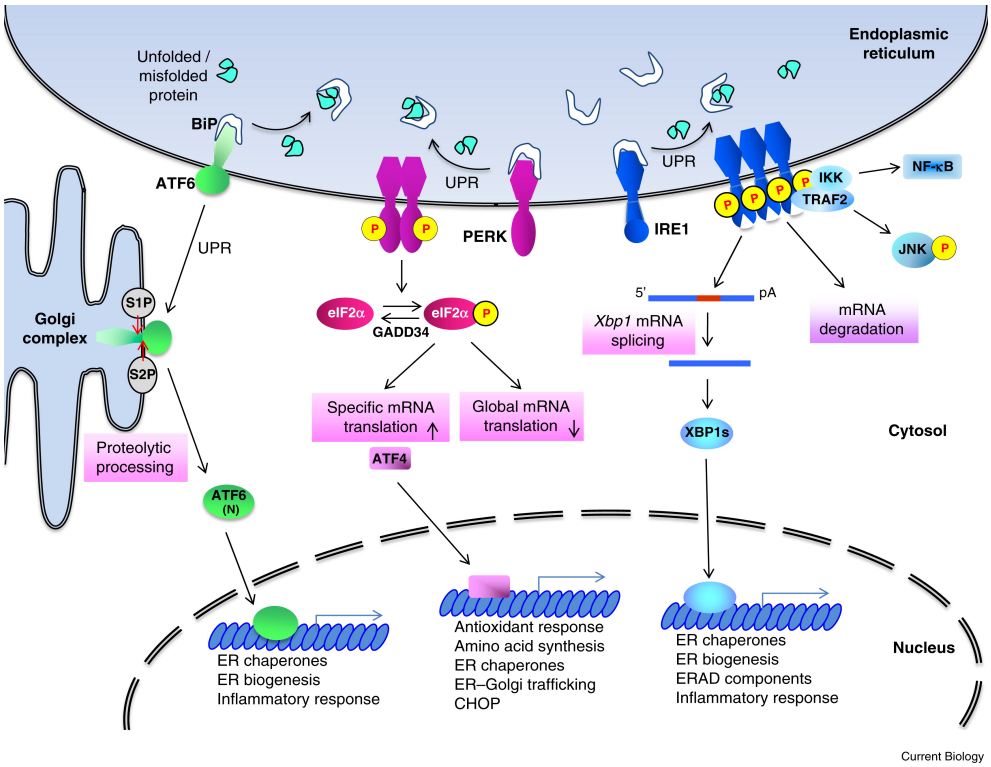

Figure 2.2: The pathways of the unfolded protein response. Reprinted from [15], with permission from Elsevier.

like ER kinase (PERK/EF2AK3), which is responsible for several different downstream processes. First, PERK phosphorylates the translation initiation factor eIF2α, which then inhibits translation of most proteins but paradoxally enhances the translation of others, such as the transcription factor ATF4. ATF4 in turn controls genes that regulate protective and apoptotic pathways, such as its target CCAAT/enhancer-binding protein-homologous protein (CHOP/DDIT3), which can contribute to oxidative stress and ROS production. ATF4 together with CHOP induce the protein phosphatase 1 regulatory subunit 15A (PPP1R15A/GADD34) which dephosphorylates eIF2α selectively and contributes to a negative feed-back loop [4]. Since the block on protein translation is reversed by this negative feedback loop, normal ribosomal activity can be restored by increased levels of PPP1R15A. Furthermore, a study by Monkley et al. showed reduction of PPP1R15A in IPF patients, in lung fibroblasts, and that restoration of PPP1R15A potentially could be used as a therapeutic strategy in IPF, for example to interrupt the deposition of ECM that occurs [16].

Besides from PERK, that primarily is activated by ER stress and accumulation of unfolded proteins, there is another eIF2α kinase involved in UPR called GCN2 that senses the level of available amino acids and regulates translation initiation. GCN2 is activated by amino acid deprivation through binding of uncharged transfer RNAs (tRNAs). These tRNAs accumulates when there is a lack of essential amino acids or inhibition of non-essential amino acids. Besides from this, GCN2 can also be activated by glucose deprivation, UV-irradiation and viral infections [17].

The second UPR arm is initiated by the endoribonucleases inositol requiring enzyme 1 (IRE1α/ERN1), which is the most conserved transducer of UPR. It splices the X-box binding protein (XBP1) mRNA and translates the active transcription factor sXBP1 that stimulates expression of chaperones and ER resident proteins necessary for folding and the ER-associated protein degradation (ERAD) [5]. When the level of ER stress is high, the activity of the IRE1α kinase domains leads to activation of the endoribonuclease domain which in turn facilitates splicing of XBP1 and degradation of several mRNAs and micro-RNAs in a process called Regulated IRE1α-Dependent Decay (RIDD). RIDD has shown to both induce cell death and preserve ER homeostasis. When comparing RIDD activity to XBP1 splicing during ER stress, the regulatory function of the UPR transducer IRE1 differs and the two processes may have opposite effects on cell fate [18].

Finally, the third UPR arm is initiated by transcription factor ATF6. When activated, the ATF6 protein exits the ER and migrates to the Golgi apparatus membrane where it is integrated and cleaved to an active form by membrane bound transcription factor site-1 and 2 proteases. The active form then translocates to the nucleus where it acts as a transcription factor and increases the transcription of UPR chaperone proteins such as BiP, for improved protein folding capacity, and of ERAD components, that degrade misfolded proteins [4].

Since ATF6, PERK and IRE1α are primarily activated by ER stress and accumulation of unfolded proteins through dissociation of BiP, while GCN2 primarily responds to starvation of amino acids, the focus in this study is on these three sensors.

2.3 COPD

COPD is a respiratory disorder where expiratory airflow limitations are developed due to inhalation of noxious stimuli, such as cigarette smoke. These airflow limitations are irreversible and the disease is the third leading cause of death worldwide and caused 3.23 million deaths in 2019 [3]. The disease is heterogeneous and can include several different clinical phenotypes; 1) emphysema, where the airspaces distal to terminal bronchioles have been destructively enlarged and the alveoles destructed, 2) chronic bronchitis, where the airways are remodeled with mucous hypersecretion, 3) small airway disease and 4) fibrosis. For example, emphysema often coexist with small airway disease [2].

Some of the molecular pathways that has been linked to smoking, airflow limitations and emphysema are chronic immune response and inflammation, tissue damage, oxidative stress, cellular senescence and lung epithelial cell apoptosis [2]. One of the most important groups of proteins involved in COPD are cytokines, which are released by inflammatory cells such as alveolar macrophages, contributing to disease pathology. There is an increase in number of macrophages in the lungs of COPD patients and several inflammatory proteins are expressed that are related to

the senescence associated secretory phenotype (SASP), including TGFβ and several matrix metalloproteinases (MMPs). SASP is also found in structural cells, such as epithelial cells and fibroblasts. Furthermore, an accelerated aging is observed in patients with COPD [19].

Macrophages stimulate the immune response by releasing inflammatory mediators, such as TNF-α, IL1-β, CXCL1, CXCL8 and LTB4, thereby recruiting monocytes, lymphocytes and neutrophils to the inflammatory site and inducing MMP secretion by epithelial cells. Increased macrophage number in the lung is associated with increased ROS-induced oxidative stress and dysfunctional macrophage response contributes to alveolar wall destruction and induced ECM degradation [19]. Neutrophilic inflammation is another feature of COPD and is induced by cigarette smoke, oxidative stress, bacteria and viruses. These inducers activate inflammation via nuclear factor-kβ and p38 mitogen protein kinase (MAPK) signaling in airway epithelial cells, resulting in neutrophilic mediators and induction of mucus hypersecretion, MMPs, oxidative stress and release of proteases that cause tissue damage [2]. See Figure 2.3 for an overview of the pathophysiology of COPD.

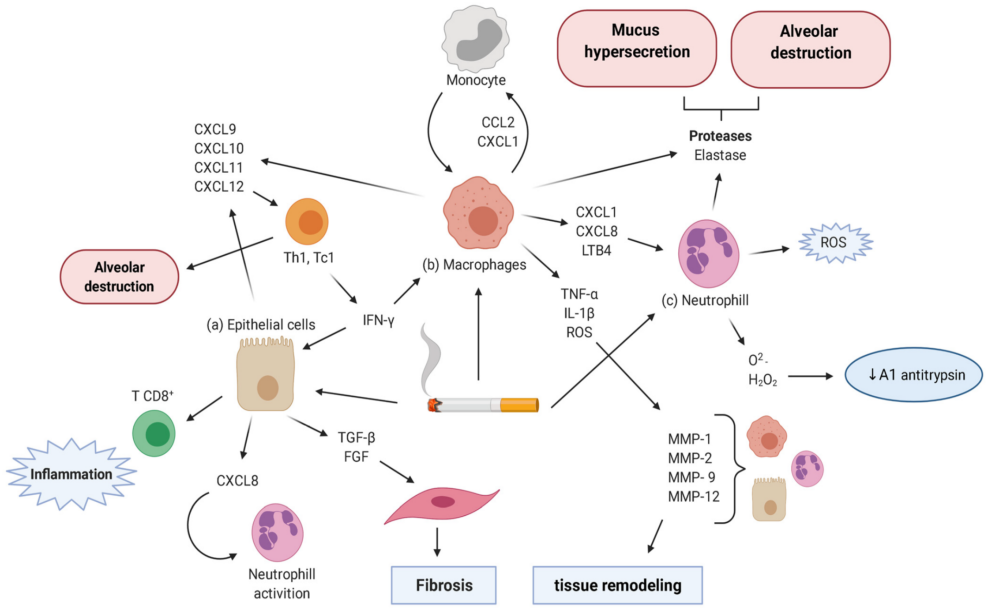

Figure 2.3: Patophysiology of COPD. From [19].

There are currently no treatments available to reverse COPD pathology and investigated drug target mechanisms often do not reach the clinical phase. Two commonly used treatments to reduce the symptoms of COPD are bronchodilators and corticosteroids. Bronchodilators relieve dyspnea, which is one of the main symptoms of COPD and defined by shortness of breath. The result of using bronchodilators is reduction of airway resistance, muscle relaxation in the lung and widening of the airway. However, they are not very efficient in patients with COPD and can only be used for a limited time. Corticosteroids are an anti-inflammatory drug shown to be very efficient against asthma but only have a limited anti-inflammatory effect in

COPD [19].

2.3.1 COPD and unfolded protein response

Activation of the UPR pathway has been observed in COPD but the pathogenic contribution of UPR to COPD is not fully understood [20]. For example, an increase in expression of CHOP and eIF2α has been reported from immunoblotting in lungs of COPD patients and decreased expression of miR199a-5p has been correlated to increased expression of sXBP1, BiP and ATF6 in monocytes [21]. The presence of these UPR markers could correlate with adverse airway remodeling and possibly be used as biomarkers and therapeutic targets. Opposite to the above mentioned correlations, there are also studies about AT2 cells that has been derived from COPD lungs and only shown low expression of these genes [20]. Forced exploratory volume decline (FEV1) has also been investigated in patients with COPD and in one study, FEV1 was shown to be correlated with an airway gene expression signature which led to identification of an UPR signature of genes downstream of sXBP1 [22]. Furthermore, acute UPR activation has been reported in airway epithelial cell lines and HBECs when exposed to cigarette smoke [4].

2.4 IPF

Pulmonary fibrosis is a chronic lung disease with progressive accumulation of extracellular matrix (ECM) in the peripheral lung, leading to matrix stiffness, stress, strain forces and loss of functional alveolar gas exchange units. It reflects the later stages of chronic interstitial lung diseases, the most severe form being IPF. IPF is a chronically scarring lung disease known to advance into respiratory failure and death within 5 years of diagnosis, with lung transplantation as one of few available treatments for prolonged survival [10]. However, the survival rate at 5 years of diagnosis has been estimated to 45 % in a study by Kaunisto et al. and depend on factors such as age and lung function at diagnosis [23]. The cause and mechanisms behind IPF are not fully known, but of clinical interest due to the high mortality rate and to the fact that it often is misdiagnosed or that patients are diagnosed with the disease in a late state of disease development [24].

The pathogenesis of IPF is partly explained by recurrent epithelial cell injuries and accelerated epithelial aging that causes abnormal repair of the alveolar cells, which is the current model for disease initiation. Another typical phenotype is senescence of alveolar epithelial cells and fibroblasts, where shortened telomeres, ER stress, oxidative injury and mitochondrial dysfunction are features that contribute to decreased cell proliferation. Furthermore, IPF is characterised by ECM deposition and abnormal mucociliary clearance. Another finding in patients with IPF are single-nucleotide polymorphism in the promoter region of MUC5B, which has shown to increase the risk of IPF since it leads to over expression of mucin in small-airway epithelial cells [24]. Epithelial–mesenchymal transition (EMT) is also a result of

ECM deposition. It contributes to loss of epithelial cell characteristics and markers, and gain of mesenchymal ones such as invasion, migration and production of ECM. EMT can be divided into three different types and the one occurring in IPF is called type II. Type II is characterised by normal wound healing and are important for reduntant tissue repair, which can be observed in IPF where the EMT-signals are over-induced. The over activation in turn causes ECM accumulation, with organ pathology and tissue remodeling as result [25].

Furthermore, innate immune cells, such as monocyte derived alveolar macrophages, has shown to be critical for the disease development and an increased bacterial burden has shown to increase disease progression in patients with IPF. Another connection between IPF and microbes are the genetic variation by a gene called TOLLIP and innate immune response in gene variants of the TOLLIP gene that have been associated with increased IPF susceptibility. The gene encodes a protein in the toll-like receptor pathway, which inhibits responses to microbes and contributes to the complex relation between IPF, host defense pathways and microbiome diversity [24].

Advances has been made when it comes to pharmacotherapeutic approaches to IPF, even if there still are no drug available that cures the disease, but rather prevent disease progression. Two of the medications that have shown to be effective, safe and to reduce severe respiratory events, such as acute exacerbations, are nintedanib and pirfenidone. There are also data suggesting that they contribute to reduction of the mortality rate of IPF. Nintedanib is a tyrokine kinase inhibitor that targets growth factor pathways while pirfenidone has antifibriotic and anti-inflammatory effects such as downregulation of TGF-β, inhibition of collagen synthesis and reduction in fibroblast proliferation [24]. An overview of the pathogenesis of IPF can be seen in Figure 2.4.

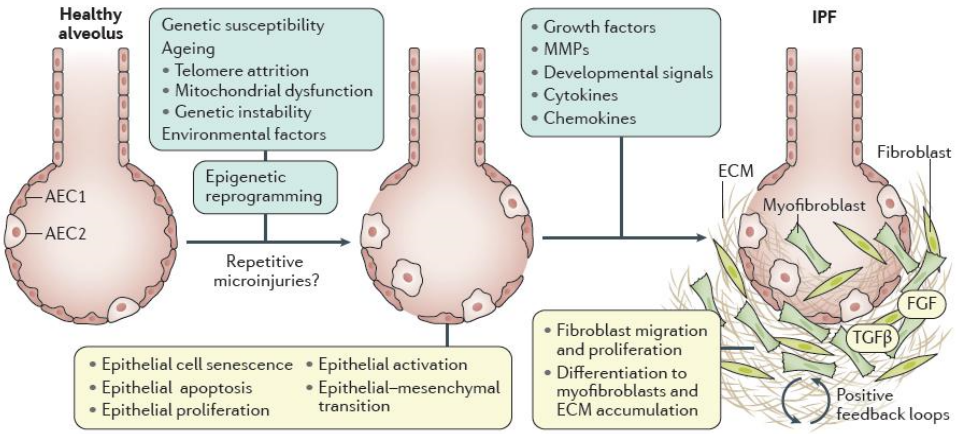

Figure 2.4: Pathobiologic features of IPF. Reprinted from [26], © 2017 with permission from Springer nature.

2.4.1 IPF and unfolded protein response

ER stress has been shown to be upregulated in IPF and all three arms of the UPR has previosuly been connected to the disease. There are several genetic links between pulmonary fibrosis and UPR. One link is through AT2 cells that secrete surfactants and where mutations in surfactant protein C (SFTPC) has been related to pulmonary fibrosis and impaired SFTPC folding, which lead to activation of the UPR and to protein aggregation [27]. Another population-wide genetic risk factor is the MUC5b promoter variant rs35705950 that leads to increased expression of MUC5B in the distal airways [24]. XBP1 is co-expressed with MUC5B and has been shown to induce expression of MUC5B in the distal bronchial epithelial by activating the promoter variant, contributing to a positive feedback mechanism that links the induction of UPR expression to IPF [28].

Evidence of UPR in pulmonary fibrosis has been observed in mouse models, where mutations in SFTPC has shown to drive pulmonary fibrosis directly by disrupted lung morphogenesis. Furthermore, markers of UPR has been observed in patients with sporadic or genetically determined IPF. One example is in alveolar epithelial cells in fibroblastic foci, where activation och upregulation of ATF6, expression of ATF4, CHOP and the IRE1 pathway has been observed and linked UPR to apoptosis [29]. Finally, ER-stress induced EMT has been related to IPF pathogenesis through upregulation of UPR markers such as BiP and XBP1, together with reduction of epithelial cell markers and a change in cell morphology to fibroblast-like structures [30], [31]. Nevertheless, the exact biological mechanisms and connections between IPF and UPR are not fully understood [4].

2.5 scRNA-seq and *in silico* models

An emerging tool used to map different pathways to disease states that has evolved recently is scRNA-seq. While bulk RNA is used to analyse pooled cell populations or tissue sections, scRNA-seq enables analysis of cells at individual cell level. The process starts with isolation of cells which is the main step that differs between bulk RNA-seq and scRNA-seq [32].

There are several different single-cell isolation techniques available and which one to use depends on cells of interest and number of cells available. If the cell number is low, micromanipulation can be used, where cells are collected by microscope-guided capillary pipettes. The main advantage with micromanipulation is the precise cell isolation, but it comes with the disadvantages of having a low throughput and being time-consuming. If the cell number instead is higher, preferably $> 10\ 000$, a more commonly used and preferred isolation method called flow-activated cell sorting (FACS) can be used, where flourescent monoclonal antibodies tag the cells, since they recognize specific surface markers. The main advantages with FACS is the possibility to receive highly purified single cells, its high-throughput and the possibility to target cells with low level of cell type markers. However, FACS requires a larger number of cells and specific monoclonal antibodies for isolation of cell types of in-

terest [33].

After the cell isolation, the cells are lysed before applying reverse transcription into first-strand cDNA, second-strand synthesis and cDNA amplification. To cope with the large amounts of transcripts generated, either 3' or 5' ends of transcripts are usually sequenced. To remove PCR bias and improve the accuracy of the reverse transcription step, unique molecular identifiers (UMIs) or barcodes can be used where reads are assigned to its original cell and enables a higher reproducibility due to the molecular counting achieved. However, the problem with sequencing either 3' or 5' end of transcripts remains, which makes the method unsuitable for analysing isoforms or allele-specific expression. See figure 2.5 for an overview of commonly used single cell isolation methods and an example of cDNA amplification using barcode primer beads [33].

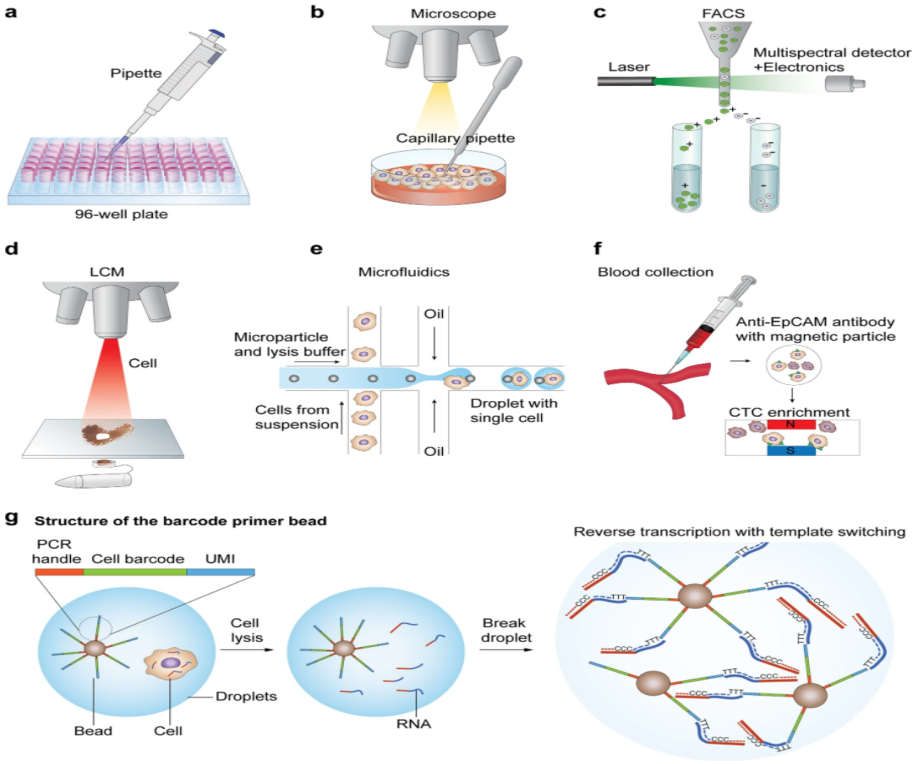

Figure 2.5: Single cell RNA isolation methods (a-f) and a droplet-based library generation technique with uniquely barcoded beads used for reverse transcription (g). From [33].

There are several advantages with using scRNA-seq compared to bulk RNA. One of the major advantages is the possibility of unbiased identification of cellular subpopulations from heterogeneous populations. Using scRNA-seq also allows for a more comprehensive knowledge of the diversity of molecular processes and states of different cell types and is an important tool for understanding the complexity and heterogeneity of different diseases [34].

One of the main challenges with scRNA-seq is to quickly and accurately capture and isolate single cells. Furthermore, the cells are exposed to a non-native environment when being isolated, which increases the risk of inducing cell stress, transcriptomic changes and a decreased viability. Another large challenge with scRNA-seq is to enable amplification of the often very small amounts of mRNA that exist in each cell and to achieve higher throughput, which is easier to accomplish with bulk RNA [34].

In this study, scRNA-seq is used to map UPR to IPF/COPD and to investigate the role of different epithelial cell subtypes in the two diseases. As previously mentioned, progress has been made in identifying IPF regulating factors in experimental models. However, in order to understand the diversity of different cell types, their states, molecular programs and the central mechanisms driving pathological epithelial remodelling in IPF lungs more comprehensively, alternative *in silico* models based on scRNA-seq are required. The cellular mechanisms of the lung in respiratory diseases are complex and heterogeneous, with a large variation between and within different cell subtypes. By scRNA analyses of these cells, the inherent spatial heterogeneity of disease pathology and cell type specific changes in diseased tissue can be investigated. This will enable identification of key mechanistic mediators and contribute to overcoming the limitations with bulk RNA studies [10].

In previous studies, novel cell subtypes has been discovered when analysing IPF with scRNA-seq. One example is an epithelial cell subtype called aberrant basaloid that has been found in two different studies, performed by Habermann et al. [10] and Adams et al. [35], and has been further investigated in this study. The cells express extracellular matrix components across a subset of histopathologic patterns of pulmonary fibrosis and cell markers of basal epithelial cells, epithelial-mesenchymal transition, cellular senescence, ECM production and IPF. They were shown to be disease-enriched and almost only discovered in IPF lungs in both datasets [10], [35].

Pseudo-time trajectory can also be utilized to investigate the connection between different epithelial cell types. These trajectories are performed by identifying a path through a high dimensional expression space that shows the different cellular states of the cells in a continuous process. It can be used to analyse differentiation patterns of cell types to multiple differentiated cellular states [36]. In Habermann et al., one pseudo-time trajectory analysis was performed of the aberrant basaloid cells, which led to the discovery that these cells probably evolve from transitional AT2 cells. Another pseudo-time trajectory performed resulted in the discovery that a SCGB3A2+ cell subtype, ie secretory club cells, could differentiate into AT2 cells [10]. These discoveries could be made thanks to utilization of scRNA-seq data. Furthermore, scRNA-seq has been used to investigate the role of different UPR components in specific cell subtypes in IPF, such as the gene PPP1R15A in lung fibroblasts, where it was observed that the gene have a distinct role in both senescent and proliferate states of these cells [16].

Regarding COPD, there are only a few scRNA-seq studies that have been executed. One of these is a study by Li et al. [2] where scRNA-seq was used to identify lung

cell subtype specific gene expression signatures by profiling lung tissue samples from patients with and without COPD. One of the observations made was that a decline of macrophages, AT2 cells, endothelial cells and fibroblasts could be seen in patients with COPD [2]. However, the data from the study by Li et al. was not accessible and thereby not analysed in this study. Instead, a study performed by Adams et al. [35] was used to analyse COPD, which is a study where IPF was compared against COPD in order to create an IPF cell atlas and investigate cell specific changes in gene expression [35].

2.6 *In vitro* models of unfolded protein response

In vitro models are commonly developed and used to investigate specific disease mechanisms and the effect of modulation of specific pathways important in disease.

In this project, primary cells were used. Primary cells are very similar to the tissue of origin since they are taken from the tissue directly and processed to enable optimal culture conditions without modifying the cells. Hence, they imitate the *in vivo* state well and exhibit normal physiology, which is desirable when for example studying biochemistry, such as aging and signaling, and effects of drugs and toxic compounds. From start, the cultures are heterogeneous and represent several cell types from the tissue, but can only be maintained in that state for a limited period of time. They can also be transformed into cell lines that has the ability to proliferate indefinitely. The downside with this transformation is that they then no longer represent the *in vivo* state to the same extent due to genetic modifications made that changes the physiological properties [37].

Besides reflecting the *in vivo* state, primary cells allow experiments on human tissue that could not have been done *in vivo*, ethical dilemmas that may arise with animal experiments can be avoided and it is cost effective due to reducing the need of *in vivo* animal models. However, primary cells require longer growth time than cell lines, have limited growth potential and life span, their phenotype change once in culture and their characteristics may change between passages if not cultured optimally. Furthermore, primary cells are hard to obtain, are often not pure and cells from different donors might respond differently to for example pro-inflammatory cytokines [37].

Some commonly used primary cells are epithelial cells, endothelial cells, fibroblasts and different stem cells. In this project, primary human lung epithelial cells were used, that are available either as bronchial epithelial cells, which were used in this study, or small airway epithelial cells. The HBECs were used to enable investigation of ER stress responses and the UPR pathway in epithelial cells when exposed to different stimuli. As previously mentioned, ER stress activates the UPR to resolve pathological alterations in protein folding. The UPR in turn causes cellular transcriptional, translational and degradation pathway changes to take place through the activation of PERK, IRE1 and ATF6. By studying these responses with *in vitro* models, the pathology and development of diseases where ER stress is activated can

be uncovered [38].

Two commonly used ER stress inducers that outlines the pathological alterations caused by ER stress *in vitro* are tunicamycin and thapsigargin. Tunicamycin inhibits UDP-N-acetylglucosamine-dolichol phosphate N-acetylglucosamine-1- phosphate transferase (GPT), which blocks the first step of glycoprotein biosynthesis in the ER. The blocking leads to accumulation of unfolded glycoproteins inside the ER and activation of UPR. Thapsigargin causes ER stress by inhibiting sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) and thereby decreases the levels of calcium in the ER. ER-dependent chaperones, which assist in the proper folding of proteins, lose their activity when the levels of calcium are lowered and thereby thapsigargin causes accumulation of unfolded proteins [39]. ER stress and UPR activation has previously been observed by upregulation of CHOP and BiP when challenging cells with tunicamycin or thapsigargin [38].

Besides from tunicamycin and thapsigargin, there are several compounds that can be used to induce ER stress and the UPR by oxidative stress. Oxidative stress occurs when there is an imbalance between the antioxidant capacity of the cells and the free radical production. This imbalance occurs due to accumulation of ROS that overwhelms antioxidant defense mechanisms. While hydrogen peroxide (H_2O_2) treatment has various effects in cells, ER stress and a subsequent UPR have been observed too. In fact, several studies demonstrate PERK-mediated upregulation of CHOP in H_2O_2 stress-induced senescence models [40], [41].

Cigarette smoke extract (CSE) exposure of cells *in vitro* has shown to induce the UPR in several studies as well. Firstly, PERK-dependent phosphorylation of eIF2α and induction of ATF4 and PPP1R15A was observed when HBECs from non-smoking individuals were exposed to CSE [42]. Furthermore, CHOP has been shown to be upregulated in HBECs when exposed to CSE, together with ROS generation and induction of apoptosis. When these HBECs thereafter were exposed to antioxidants, the CSE-induced apoptosis was inhibited and CHOP expression decreased, indicating that oxidative stress and UPR occur when cells are exposed to CSE [41].

3

Analytical procedures

3.1 Quantification of gene expression

Quantitative polymerase chain reaction (qPCR) is a PCR method that monitors the progress of the PCR run throughout the running time and thereby collects data continuously. TaqMan assays are commonly used for qPCR, which are based on targeting of specific primers and probes optimized for the specific gene. For gene expression measurements with TaqMan, 5' nuclease chemistry is applied and a fluorogenic probe is used that enables detection of the chosen PCR product as it accumulates. The assay process for TaqMan gene expression can be observed in Figure 3.1 [43].

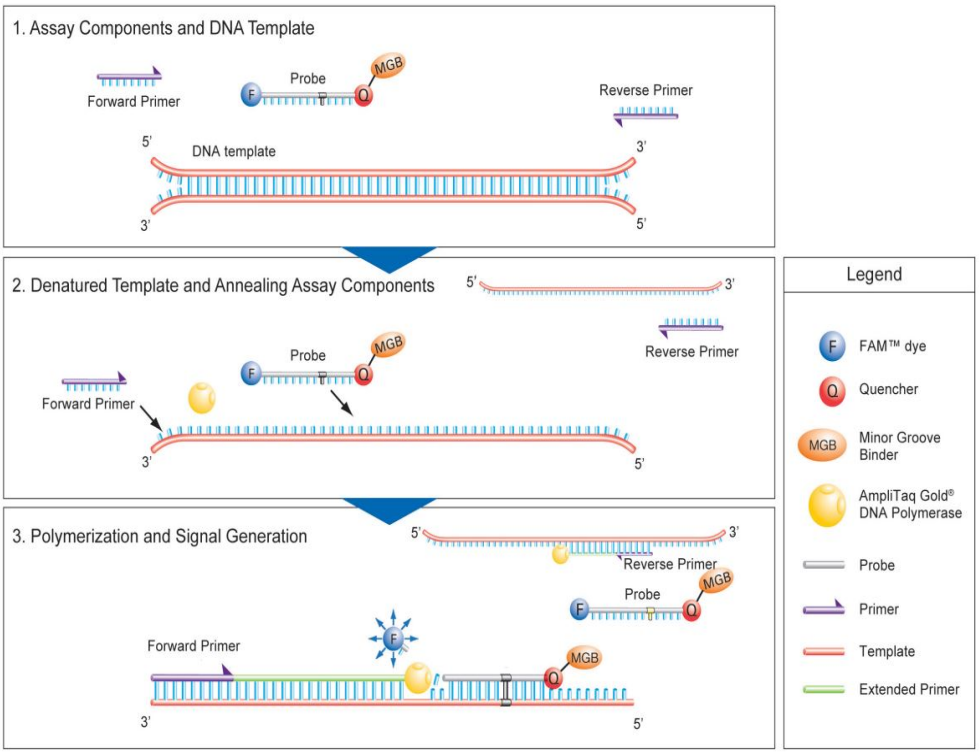

Figure 3.1: TaqMan Gene Expression Assay reaction steps. From [43].

In step 1, double stranded cDNA is denatured by a temperature increase and the fluorescent dye signal on the 5' end of the probe is quenched by a non-fluorescent quencher (NFQ) at the 3' end. In step 2, the temperature of the reaction is lowered to enable the primers and probe to anneal to their specific target sequences. Lastly,

when the Taqman DNA polymerase reaches the bound probe, it can thanks to its exonuclease activity chop the probe apart and continue the elongation. The separation of the dye from the quencher results in a fluorescent signal proportional to the amount of templates. When the fluorescence emitted is greater than the background signal, the cycle threshold (Ct) is reached. Hence, the emitted light reaches this threshold faster if the gene of interest is highly expressed and results in a lower Ct-value than a lower expressed gene [43].

3.2 Quantification of protein expression

Western blot is a routine technique used for protein analysis. It uses specific interactions between antibody and antigen to detect the protein of interest from a complex mixture and can produce both qualitative and semi-quantitative data [44].

Western blotting can be divided into different steps. First, gel electrophoresis is used to separate proteins in a sample based on their physical properties. This is commonly done with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) which separates proteins on a polyacrylamide gel based on mass due to using the negative charge that is imparted on the proteins when bound to the ionic SDS detergent. Thereafter, the separated proteins are transferred onto a second membrane, often made of nitrocellulose or polyvinylidene diflouride (PVDF), using a sandwich construction where the gel is placed in direct contact to the membrane with porous pads and filter papers around it that facilitates the transfer. The sandwich of layers is exposed to an electric field allowing the proteins to transfer from the gel and onto the membrane surface where the proteins attach. The transfer efficiency is dependent on several factors, such as the gel composition, the complete contact between gel and membrane, the electrode position, transfer time, protein size and composition, field strength and detergents in the buffer used. Successful protein transfer can be assessed by reversible Ponceau S staining, which visualises the total protein amount [44].

Before probing the proteins of interest with antibodies, it is important to block nonspecific binding on the membrane. The probing can then be performed and is either based on direct or indirect detection. The indirect detection consist of a primary antibody specific for the protein of interest, but without being detectable, and a secondary antibody, conjugated with a detectable tag, which binds to the primary antibody through specificity against host species of the primary antibody. The tag can either consist of an enzyme together with a substrate that emits a detectable signal, or a fluorophore detectable by an instrument that can capture fluorescence. In the direct method, the primary antibody is conjugated with a detectable tag directly, eliminating the need for a secondary antibody. Even if the direct method only requires one antibody, the indirect method is often preferred due to the signal amplification achieved by using a secondary antibody and the possibility of multiplexing [44]. The two detection methods can be seen in Figure 3.2. In this study, the indirect method was used.

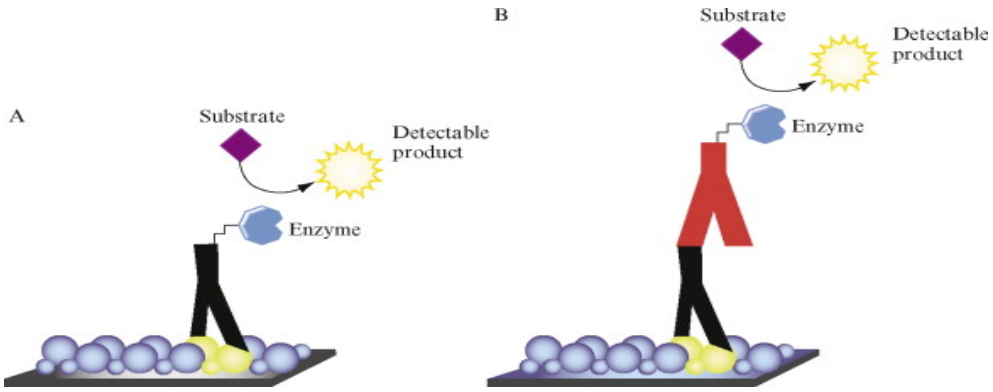

Figure 3.2: Direct detection of proteins with primary antibody conjugated with detectable tag (A) and indirect detection of proteins with primary and secondary antibodies (B), where the secondary antibody is conjugated with a detectable tag. From [44].

3.3 Analysis *in silico*

The most commonly used methods for finding marker genes in RNA data are based on bulk RNA sequencing where differentially expressed genes are defined as genes with different means of expression value. In these methods, a small group of homogeneous cells are compared to the rest of the data (outside group) to find DE genes. However, to enable capturing the large cell heterogeneity of scRNA-seq data, which consist of many different cell types/states, a single parameter (mean value) cannot represent the complete population of cells and the outside group cannot be said to have a specific distribution [45].

Venice has been developed to meet these challenges with scRNA-seq data and is the method used in this study for DE analysis of scRNA-seq data. It is based on a non-parametric test where marker genes in a group of cells are the ones that can be used to distinguish these cells from the rest. The significance of a marker gene is scored using accuracy of classification as metric, which is estimated by how well a certain gene can be used to predict cells inside or outside the chosen cluster of cells. Thereafter, this accuracy metric is used for marker gene ranking and to classify DE genes as up- or down-regulated [45].

In a study by Vuong H. et al., Venice was compared to 14 other methods by validating precision and performance with two simulated datasets as test data. Venice had the highest true positive rate (AUC score > 0.98) and amongst the methods with similar rate, Venice performed both faster and used significantly less memory [45].

4
Materials and Methods

This chapter is divided into two main parts. Part one consists of the methods used for the *in silico* analysis of scRNA-seq datasets and are explained in section 4.1. Thereafter, the *in vitro* part is explained in section 4.2-4.5.

4.1 Analysis of scRNA-seq data

The scRNA-seq analysis of IPF and COPD was performed in BBROWSER, a platform for single cell transcriptome analysis and spatial transcriptomics [46]. DE analysis of epithelial cells from subjects with IPF or COPD was performed and the datasets and methods used for analysis in BBROWSER are explained in this section.

4.1.1 Datasets

Three datasets were used for scRNA-seq analysis of IPF. The first dataset was collected from a study by Habermann et al. [10] which had the objective to investigate cell types/states in IPF and the mechanisms driving pathologic epithelial remodeling and ECM expansion in the human lung. Single cell suspensions from peripheral lung tissue of healthy and diseased subjects were made from 12 IPF lungs and 10 non-fibrotic lungs. Thereafter, sequencing was performed with the 10x Genomics Chromium platform. Canonical lineage defining data was used to annotate clusters and a total of 31 cell types was found, whereof 12 epithelial subtypes in a subcluster of 37 325 cells. One of these cell subtypes was a previously unrecognized one, called KRT5-/KRT17+ in the study. This subset expressed collagen and ECM components and was shown to be enriched in pulmonary fibrosis [10].

The second dataset came from a study by Reyfman et al. [47] which had the objective to determine if scRNA-seq could be used to reveal disease-related heterogeneity within human lung cells from patients with pulmonary fibrosis. Lung biopsies from healthy donors and lung explants from patients with pulmonary fibrosis, whereof 4 patients with IPF, were used and scRNA-seq performed. The assignment of cell subtypes to each cluster of cells was done based on expression of established markers from the databases ImmGen and LungMap. A total of 13 cell types were found, whereof five epithelial subtypes. A single-cell atlas of pulmonary fibrosis was generated and heterogeneity within epithelial cells from patients with pulmonary fibrosis was observed [47].

The third dataset was collected from a study by Adams et al. [35] which had the objective to develop a single-cell atlas for IPF of aberrant cellular populations from the epithelium, endothelium and stroma of the IPF lung. A total of 32 IPF lungs, 18 COPD lungs and 28 control donor lungs were profiled and the computational analysis resulted in identification of 38 cell types by using markers without batch or cell cycle effects. Besides previously known cell subtypes, a new population of cells was identified in IPF lungs but not in controls or COPD lungs, called aberrant basaloid cells. The aberrant basaloid cell subgroup is similar to the KRT5-/KRT17+ subgroup annotated by Habermann et al. and both these groups will therefore be referred to as aberrant basaloid cells. The aberrant basaloid cells express both epithelial markers, basal markers such as KRT17, epithelial-mesenchymal markers such as COL1A1, senescence related markers such as CDKN1A and IPF biomarkers such as MMP7. To validate this new cell subtype, the authors compared it with the data published by Habermann et al., which resulted in a similar marker profile and a correlation coefficient (Spearman's rho) close to 1 for each cell subtype [35].

The three studies analysed will hereafter be referred to by the surnames of the authors, namely Habermann, Reyfman and Adams. To summarize their characteristics, a comparison was made that can be seen in Table 4.1 where several differences can be observed that might effect the analysis. For example, the number of cells where there are more IPF cells than control cells in Habermann and Adams, the methods and markers used to annotate cell types and the GO enrichment. The GO enrichment affects the representation of cells for different cell types and depending on method used, the true number of cells will be more or less reflected in the data.

4.1.2 Differential expression analysis of IPF

DE analysis of IPF versus control cells was performed with the Venice algorithm in BBROWSER on each of the annotated epithelial cell subtypes from the studies by Habermann, Reyfman and Adams. Prior to the DE analysis, an UPR gene signature was generated through the use of Ingenuity Pathway Analysis, which is a platform for visualization and data analysis of omics data [48], and by adding some additional genes related to the IRE pathway from an article by Becker et al. [22]. The signature was then used to analyse the UPR pathway for each cell subtype by observing the fold-changes of genes with FDR < 0.1 from the DE analysis. Only cell subtypes with close to or more than 100 cells in both IPF and the control group were analysed.

4.1.3 Analysis of aberrant basaloid cells

To investigate the connection between the aberrant basaloid cells found in Habermann, IPF and the UPR pathway, DE analysis was performed comparing the cell subgroup against the other epithelial cell subgroups. The most upregulated genes (fold-change > 1.4) were selected and the percentage difference in number of cells expressing these markers was calculated. The 12 most upregulated genes were chosen, which had percentage differences larger than 66 % between aberrant basaloid and

Table 4.1: Comparison of datasets from Habermann [10], Reyfman [47] and Adams [35].

	Habermann	Reyfman	Adams
Nr of subjects	IPF: 12 Control: 10	IPF: 4 Control: 8	IPF: 32 Control: 28 COPD: 18
Nr of total cells	114 396	79 242	312 928
Nr of epithelial cells	IPF: 24 891 Control: 8430	IPF: 3257 Control: 22 926	IPF: 13 223 Control: 4807 COPD: 3154
Cell type focus	Epithelial	Epithelial and macrophages	Endothelial, epithelial, fibroblasts, myofibroblasts, macrophages
Source of samples	**IPF**: Lung explants obtained at time of transplantation **Control**: Non-fibrotic lungs declined for organ donation	**IPF:** Lung explants obtained at time of transplantation **Control**: Non-fibrotic lungs from donor	**IPF/COPD**: Lung explants obtained at time of transplantation **Control**: Non-fibrotic lungs from donor
Library preparation (barcoding)	10x chromium platform 3' v2 or 5'	10x chromium platform 3' v2	10x chromium platform 3' v2
Cut-off QC	Removed: • Cells with < 1000 transcripts • Cells with > 25 % of reads from mitochondrial genes • Reads with read quality < 30	No details provided	Removed: • Cells with < 1000 transcripts • Cells with > 20 % of reads from mitochondrial genes • If trimmed below 30 bp • If > 12 % transcripts from intron spanning or unspliced reads
Enzymatic digestion	Collagenase I/dispase II tissue or Miltenyi Multi-Tissue Dissociation Kit	Collagenase, Dnase I	Elastase, deoxyribonuclease I, liberase
Cell type annotation	Immune, epithelial, endothelial and mesenchymal	Myeloid leukocyte, epithelial, progenitor, T-cells, B-cells, endothelial	Epithelial, stromal/mesenchymal, myeloid, and lymphoid.
Cell type markers	Canonical lineage-defining markers	Established markers from the LungMAP and ImmGen databases	Distinct markers with no discernible batch or cell cycle effects on data architecture
Enrichment of cells	**Serial columns**. Enrichment of immune (CD45+) and other (CD45-) cells mixed 1:2	• **FACSAria III instrument.** Enrichment of macrophages and AT2 cells • Dead cell removal with viability dye	• **No enrichment** • Dead cell removal with MS columns

other epithelial cells. The gene signature was first applied on the already annotated aberrant basaloid subgroups in Habermann and Adams, to confirm its existence, investigate their similarities and validate the usage of these markers to sub-cluster a similar group in the third dataset by Reyfman. The gene signature was then used to investigate if a similar subgroup could be found in Reyfman by sub-clustering cells in that dataset that had a large ratio between gene expression in the created subgroup and the rest of the epithelial cells. DE between the created subgroup and the rest of the epithelial cells was performed and the fold-changes of the UPR signature genes observed.

Furthermore, GO enrichment was performed on the most upregulated genes in the created aberrant basaloid cell subgroup in Reyfman, to investigate the biological processes and cellular components that the most upregulated genes was related to. The GO enrichment was performed within BBROWSER with the Gene Set Enrichment Analysis (GSEA) method, where the rest of the epithelial cells was used as background, and outside of BBROWSER with DAVID, a database for annotation, visualization and integrated discovery of gene lists [49].

4.1.4 Differential expression analysis of COPD

To investigate the regulation of UPR genes in COPD, the dataset from Adams was used. Similarly as for IPF, only cell subgroups with close to or more than 100 cells

in both COPD and the control group were analysed.

4.2 Experimental layout

The *in vitro* part of the project consisted of three individual experiments. Gene readouts for the UPR pathway were established in all three experiments, protein readouts in the first experiment and cell viability was measured in the third experiment. An overview of the experimental layouts can be seen in Table 4.2. In addition, already available samples were used for establishment of protein readouts. In these samples, cells had been stimulated with 0.1 µg/ml tunicamycin, with or without an inhibitor of the PERK pathway. For detailed protocols of the experimental part of the project, see section B.1 in Appendix B.

Table 4.2: Experimental layouts of experiments performed.

	Cell donor	Stimuli	Concentration	Cell density (cells/well)	Cell media	Stimulation (h)	Gene readouts	Protein readouts	Cell viability
Experiment 1	N1936	Tunicamycin	0.01-0.5 µg/ml	10000 (96 well plates) 30 000 (48 well plates)	ExPlus	3, 6, 24	BiP, CHOP, XBP1, sXBP1, ATF4, ATF6, PPP1R15A	ATF6, CHOP, eIF2α, phospho-eIF2α, BiP	N/A
Experiment 2	N3375	H2O2	1000-50 µM	30000 (96 well plates)	BEGM	6, 24	BiP, CHOP, XBP1, sXBP1, ATF4, ATF6	N/A	N/A
Experiment 3	N3375	Tunicamycin H2O2 Thapsigargin CSE	0.01-0.5 µg/ml 400-50 µM 5-500 nM 1-10 %	12500 (96 well plates)	BEGM	3, 6, 24	BiP, CHOP, XBP1, sXBP1, ATF4, ATF6, PPP1R15A, HSP90B1	N/A	Apoptosis Proliferation Total cells

4.2.1 Primary human bronchial epithelial cells

HBECs, isolated from the surface epithelium of human bronchi, were obtained from LONZA. The cells used in this project are from healthy donors (Lonza, CC-2540) and will be referred to as N3375 (batch number 0000613375) and N1936 (batch number 0000501936). The cells were received at passage 3 and have been tested for mycoplasma and viability post-freezing before use.

4.2.2 Culture conditions

In the first experiment, cells from donor N1936 were seeded onto three 48 well plates (Costar, 3548) in 300 µl media and with a cell density of 30 000 cells/well, and three 96 well plates (Costar, 3596) in 100 µl media per well and with a corresponding cell density of 10 000 cells/well. The medium used was PneumaCult™-Ex Plus Medium (STEMCELL™ Technologies, 05040), which is designed for airway epithelial cell research and primary HBECs, and the cells were cultivated over night at 37 °C prior to stimulation.

In the second and third experiment, the cells from donor N3375 were seeded into two different 96 well plates. In the second experiment, two 96 well Corning™ Costar™ Flat Bottom Cell Culture Plates (Costar, 3595) were used and cells were seeded in 100 µl media and with a cell density of 30 000 cells/well. In the third experiment, six of these 96 well plates and one PhenoPlate™ 96-well microplate (PerkinElmer, 6055302) with black cell walls were used and cells were seeded in 100 µl media with a

cell density of 12 500 cells/well. The cells were then cultivated over night at 37 °C in BEGM™ Bronchial Epithelial Cell Growth Medium (CC-3171), which is specifically designed for growth of primary HBECs. The media was used in experiment 2 and 3 based on previous observations that ExPlus media rendered the cells unresponsive to H_2O_2 and that BEGM media might contain less antioxidants and thereby decrease the risk of H_2O_2 neutralisation caused by culture media.

4.2.3 Stimulation of cells

Before stimulation of cells, the media was changed in the first experiment, but not in the second and third. The cells were stimulated with tunicamycin in experiment 1, H_2O_2 in experiment 2 and tunicamycin, thapsigargin, H_2O_2 and cigarette smoke extract (CSE) in experiment 3. Concentrations of thapsigargin were chosen based on previous studies by Schadewijk et al. [50] and Kemp et al. [51], where HBECs were cultivated in BEGM media. Regarding concentrations of tunicamycin, H_2O_2 and CSE, these were selected based on previous experiments performed at AstraZeneca. The concentrations and stimulation timepoints used can be seen in Table 4.2 in the beginning of section 4.2.

In the third experiment, cells stimulated with tunicamycin, thapsigargin and H_2O_2 were exposed to a mixture of IncuCyte NucLight Rapid Red Reagent (Essen Bioscience, 4717) and IncuCyte Caspase 3/7 Green Reagent (Essen Bioscience, 4440) for 2 hrs before addition of stimuli, to enable measurement of cell proliferation and apoptosis. After the exposure time, these cells were stimulated with the same concentrations as the other plates in the experiment. Finally, the plate was imaged for 48 h in an Incucyte® Live-Cell Analysis instrument for cell imaging.

4.2.4 Lysis for RNA analysis by qPCR

In all three experiments, the cells in the 96 wells, except for the plate analysed with the Incucyte instrument, were lysed with RLT lysis Plus buffer (Qiagen, 79216) and kept in -80 °C freezer prior to RNA purification and subsequent gene expression analysis with qPCR.

4.2.5 Lysis for protein analysis by Western blot

The cells in the 48 well plates from the first experiment were lysed with a lysis buffer consisting of M-PER lysis buffer (ThermoFisher, 78501), DNAse I solution (Sigma Aldrich, D4527), cOmplete™ Mini Protease Inhibitor Cocktail (Roche, 11836170001) and PhosSTOP™ (Roche,4906837001). Lysed cells were then kept in -80 °C freezer prior to protein expression analysis with Western blot.

4.3 Gene expression analysis

The methods used for the gene expression analysis is presented in this section and was performed in the same way for all experiments. Complete protocol can be seen in section B.2 in Appendix B.

4.3.1 RNA purification

Lysed cells were thawed on ice and then purified according to the manufacturer's protocol using the RNeasy 96 Kit (QIAGEN, 74192). Most of the Taqman probes used for qPCR span several exons, which excludes the need of a gDNA elimination step prior to the RNA purification since the probability of detecting genomic DNA is small. After the purification steps, the RNA concentrations was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNase-free water was used as blank prior to quantification of the RNA content by adding 1 µl of the sample on the NanoDrop. The cDNA synthesis was thereafter performed, without prior normalisation of the RNA concentrations in the samples.

4.3.2 cDNA synthesis

After RNA purification, samples were prepared for cDNA synthesis with a High Capacity cDNA reverse Transcription kit (Applied Biosystems, 4368813). For each sample, 20 µL master mix was prepared on ice with reagents from the kit. Master mix was prepared for the number of samples in each experiment and added together with 20 µL sample to non-skirted 96 well PCR plates (VWR, 82006-636), to a total volume of 40 µL. Two negative controls were used, one with master mix and Rnase-free water instead of RNA and one with RNA but without the MultiScribe Reverse Transcriptase in the master mix in order to detect background signal and potential signals from genomic DNA. The samples in the PCR plate were mixed and spun down to the bottom of the wells. A Veriti 96 well Thermal Cykler was then used to synthesise cDNA.

4.3.3 Quantitative polymerase chain reaction

Quantitative polymerase chain reaction (qPCR) was performed according to the TaqMan Gene Expression Assay (ThermoFisher, Waltham, MA, USA). The cDNA was thawed on ice and nuclease-free water was added for a 2X or 3X dilution of cDNA, depending on the amount of RNA previously quantified. Taqman Fast Advanced master mix (Applied Biosystems, 4444554) was prepared for each TaqMan probe in triplicates and with 30 % extra volume. TaqMan plates were prepared using an automated BioMek NX pipetting station (Beckman Coulter) to add 3 uL cDNA and 7 uL prepared master mix into a 384-well plate. The qPCR was performed with the Quantstudio Real-Time PCR system for comparative CT method. The genes analysed with the TaqMan Gene Expression assay can be seen in Table 4.3. The last part of the assay IDs indicate whether the probes span an exon junction (m) or if they might be found within a single exon (g).

Table 4.3: Genes analysed with the TaqMan Gene Expression assay.

Gene	Target	Species	Assay ID
HSPA5 (BiP, GRP78)	UPR pathway	Human	Hs00607129_gH
DDIT3 (CHOP)	UPR pathway	Human	Hs00358796_g1
ATF4	UPR pathway	Human	Hs00909569_g1
sXBP1	UPR pathway	Human	Hs03929085_g1
XBP1	UPR pathway	Human	Hs002856596_m1
ATF6	UPR pathway	Human	Hs00232586_m1
PPP1R15A (Gadd34)	UPR pathway	Human	Hs00169585_m1
HSP90B1 (Grp94)	UPR pathway	Human	Hs00427665_g1
GAPDH	House-keeping gene	Human	Hs99999905_m1
RPLP0	House-keeping gene	Human	Hs99999902_m1

4.3.4 Data analysis

For each tested gene and sample, a Ct (cycle threshold) value is returned, which is defined by the number of cycles required to reach a threshold where the fluorescent signal overreaches the background signal. The Ct value is reversely proportional to the amount of target nucleic acid in the sample, meaning that a higher amount of target is indicated by a lower Ct value. Hence, high Ct values are an indication of low gene expression.

The mean Ct value calculated from the triplicates of each sample, was used to calculate the fold-change between the reference and the target. The geometric means of the reference genes run on the samples were calculated to then normalize the target mean Ct values against. The resulting ΔCt-value is calculated with equation 4.1.

$$\Delta Ct_{target} = Ct_{target} - Ct_{reference} \tag{4.1}$$

ΔCt values of the non-stimulated controls are thereafter used to calculate the ΔΔCt value with equation 4.2.

$$\Delta\Delta Ct_{target} = \Delta Ct_{target} - \Delta Ct_{control} \tag{4.2}$$

Finally, the fold changes are calculated according to equation 4.3.

$$\text{Foldchange} = 2^{-\Delta\Delta Ct_{target}} \tag{4.3}$$

4.4 Protein expression analysis

The Western blot methods used for the protein expression analysis in experiment 1 are presented in this section. The analysis was performed on the samples from

experiment 1 in this study, but also on previously generated samples from an in-house experiment where cells were stimulated with 0.1 µg/ml tunicamycin for 24 h with and without a PERK inhibitor. For protocols of protein purification and Western blot, see section B.3-B.5 in Appendix B.

4.4.1 Protein quantification

Protein lysates from the 48 well plates were thawed on ice and protein quantification was performed with Pierce™ BCA Protein Assay Kit (Pierce, 23225) according to the manufacturer's protocol. Albumin standards for protein determination were prepared in M-PER and samples further diluted in M-PER. A working reagent mixture consisting of BCA reagent A and B was prepared and added to Nunc™ clear polystyrene (PS) 96-well MicroWell™ plates (Thermo Scientific Nunc, 269620) together with diluted standards or samples. The OD was measured in a spectrophotometer at 562 nm and a standard curve created based on the albumin standards' OD-values. The protein concentration in the samples were calculated from the standard curve. The remaining protein samples were stored at -80 °C until Western blot analysis was performed. See Appendix B for protocol, table with concentrations of BSA standards and resulting standard-curves used for protein determination.

4.4.2 Automated Western blot

An automated, capillary based Western blot method called JESS (ProteinSimple) was used for relative protein detection. JESS has a high protein sensitivity, enabling multiplexing for both chemiluminescence and fluorescence detection, and requires low sample volume. JESS also allows a much higher throughput than traditional Western blot since it is possible to run 24 samples at the same time and can complete protein separation, reagent additions, incubations, detection steps and analysis during a 3 h run. There are several possible detection methods, for example chemiluminescence detection and fluorescence detection with Stellar NIR/IR modules, which can be used to measure several proteins in the same capillary [52].

JESS was first optimised on samples from another in-house experiment and then used on time-course samples from experiment 1. Information about the proteins measured and the antibodies used can be seen in Table 4.4 and 4.5. Two JESS-plates were run, one with 3 and 6 h samples, with 0.98 µg total protein per well, and one with 24 h samples, with 1.83 µg total protein per well. For the 3 and 6 h plate, samples were run in singlets to fit into one plate, while the 24 h samples were run in duplicates.

The automated Western blot was performed with Jess 12-230 kDa separation module kit (ProteinSimple, SM-W004), according to the manufacturer's protocol. Samples were diluted in 0.1X sample buffer and water, and combined with a master mix consisting of DTT solution and 10X sample buffer. The antibodies were diluted in milk-free antibody diluent and chemiluminescent substrate was prepared by combining Luminol-S and Peroxidase. Ladder, samples, antibodies, antibody diluent,

Table 4.4: Proteins used for automated Western blot, their molecular weights and information about primary antibodies used.

Protein	MW (kDa)	Primary antibodies	Source	Dilutions used	Cat No	Supplier
eIF2	38	eIF2α (D7D3) XP® Rabbit mAb	Rabbit IgG	1:100 & 1:1000	5324S	Cell Signaling Technologies
peIF2	38	P-eIF2α (Ser51) (D9G8) XP® Rabbit mAb	Rabbit IgG	1:5 & 1:10	3398S	Cell Signaling Technologies
BiP	78	BiP (C50B12) Rabbit mAb	Rabbit IgG	1:100 & 1:1000	3177S	Cell Signaling Technologies
ATF6	90-100	ATF-6 (D4Z8V) Rabbit mAb	Rabbit IgG	1:25 & 1:100	65880S	Cell Signaling Technologies
CHOP	27	CHOP (L63F7) mouse mAb	Mouse IgG2a	1:25 & 1:100	2895S	Cell Signaling Technologies
b-actin	45	b-Actin mouse mAb	mouse IgG1	1:20	A5441	Sigma-Aldrich
b-actin	45	b-Actin rabbit pAb	Rabbit IgG	1:50	4967	Cell Signaling Technologies

Table 4.5: Secondary antibodies used in automated Western blot with JESS.

Secondary antibodies	Species	Cat no	Supplier	Dilution	Primary ab target
anti-mouse 800CW	donkey	926-32212	LiCOR	1:100	CHOP (L63F7) mouse mAb
anti-rabbit 800CW	donkey	926-32211	LiCOR	1:100	eIF2α (D7D3) XP® Rabbit mAb
anti-rabbit HRP	goat	P0448	Dako	1:20	b-Actin rabbit pAb
anti-mouse HRP	goat	P0447	Dako	1:20	b-Actin mouse mAb

streptadivin HRP, chemiluminescent substrate and wash buffer were added to the JESS plate. The plate and capillaries were inserted into the JESS machine and the standard protocol modified by changing separation time to 30 min and primary antibody time to 60 min. The JESS run was started and ran for approximately 4 hrs.

4.4.3 Traditional Western blot

To further investigate if it is possible to establish readouts for ATF6 and CHOP, since detection failed in automated Western blot, the same in-house samples were used. Western blot was thereafter performed on samples from experiment 1 in this study, following the same procedure. Information about the proteins measured and the antibodies used can be seen in Table 4.6 and 4.7.

Table 4.6: Proteins used for traditional Western blot, their molecular weights and information about primary antibodies used.

Protein	MW (kDa)	Primary ab	Source	Dilution	Cat No	Company
ATF6a	90-100	ATF6 (D4Z8V) Rabbit mAb	Rabbit IgG	1:1000	65880S	Cell Signaling Technologies
CHOP	27	CHOP (L63F7) mouse mAb	mouse IgG2a	1:1000	2895S	Cell Signaling Technologies
b-Actin	45	b-actin mouse	mouse IgG1	1:5000	A5441	Sigma-Aldrich
b-Actin	45	b-actin rabbit	Rabbit IgG	1:1000	4967	Cell Signaling Technologies

Purified protein samples were thawed on ice and diluted with water and a mixture of NuPAGE LDS Sample Buffer (4x) (InVitrogen, NP0007) and NuPAGE Sample Reducing Agent (10x) (InVitrogen, NP0004). Samples and ladder, Chameleon Duo (LiCOR, 928-60000), were loaded into 1.5 mm thick 10 well NuPAGE 4-12 % Bis-Tris Protein gels (InVitrogen, NP0335BOX).

SDS-PAGE was run with NuPAGE® MES SDS Running Buffer (Invitrogen, NP0335 - BOX), to resolve CHOP, and NuPAGE™ MOPS SDS Running Buffer (20X) (Invitrogen, NP0001), to resolve ATF6. PVDF membrane from Invitrolon™ PVDF/-Filter Paper Sandwiches (InVitrogen, LC2005) was activated in 100 % methanol

Table 4.7: Secondary antibodies used in traditional Western blot.

Secondary ab	Species	Cat no	Supplier	Dilution	Primary ab target
anti-rabbit 800 CW	goat	926-32211	LiCOR	1:10000	ATF6 (D4Z8V) Rabbit
anti-mouse 800 CW	goat	926-32210	LiCOR	1:10000	CHOP (L63F7) mouse
anti-rabbit 680 RD	goat	926-68071	LiCOR	1:10000	b-actin rabbit
anti-mouse 680 RD	goat	926-68070	LiCOR	1:10000	b-actin mouse

(Honeywell, 34966) and a transfer sandwich prepared for the protein transfer by putting together sponges, filter paper, gel and PVDF membrane. The Sandwich was added to a transfer box and filled with NuPAGE™ Transfer Buffer (20X) (In-Vitrogen, NP00061). The membranes were blocked in Intercept® (TBS) Blocking Buffer (LiCOR, 927-50000) after transfer.

The membranes were incubated with primary antibodies overnight at 4 °C for CHOP/β-actin and ATF6/β-actin. The membranes were washed in Tris buffered saline with Tween® 20 (Sigma ,91414-100TAB) prior to incubation with secondary antibodies for 2 hrs at room temperature and then washed again. Finally, the proteins were quantified by measuring infrared signal from the secondary antibodies by a LI-COR Odyssey platform. β-actin was used as normalization in between wells and conditions. See Appendix B for detailed protocols, including timepoints and settings used for SDS-PAGE and protein transfer.

4.5 Statistical analysis

The data from the qPCR was analysed with GraphPad Prism® v9.0 (GraphPad Software, Inc.). Nonlinear regressions was performed to examine EC_{80}-values for the dose response curves of stimulation with tunicamycin in experiment 1. For experiment 2 and 3, the different concentrations of stimuli was plotted in bar plots to compare stimulation times and stimuli against each other.

Cell viability measurements performed in experiment 3 was analysed by first transferring raw data images to MATLAB to transform the data. The transformed data was then transferred into Columbus software to analyse the number of live, apoptotic and total cells. The percentages of live and apoptotic cells were thereafter plotted in GraphPad Prism to compare apoptosis and proliferation between different time points and stimuli.

5
Results

In this chapter, the results from the scRNA-seq analysis and the experimental results are presented.

5.1 scRNA-seq analysis on epithelial cells

The results from the analysis of the three scRNA-seq data sets, including lungs originating from control, IPF and COPD patients, are presented in this section. The aim was to characterize UPR gene expression profiles in both COPD and IPF lung epithelial cells, but due to lack of thorough scRNA-seq datasets from COPD lungs, the main focus will be on the analysis of IPF versus control cells. The results from the DE of IPF are shown for data from Habermann, Reyfman and Adams, while Adams' scRNA-seq dataset was the only that contained data from COPD patients.

5.1.1 Datasets

To get an overview of cell distribution between the studies, BBROWSER [46] was used to extract the distribution of cells within IPF, COPD and control subjects, and between different smoking conditions, age and gender. Cell distributions were calculated from epithelial cells in the datasets from Habermann, Reyfman and Adams, which had different coverages of IPF and control cells. The result can be seen in Figure 5.1.

The gender distribution differed in between studies, but seemed to be quite consistent between control and diseased lung. All three studies had a substantial age difference between control and diseased subjects and the smoking history differs both between datasets and between the conditions within each dataset. A difference between the smoking history of controls is that the majority are never-smokers in Adams.

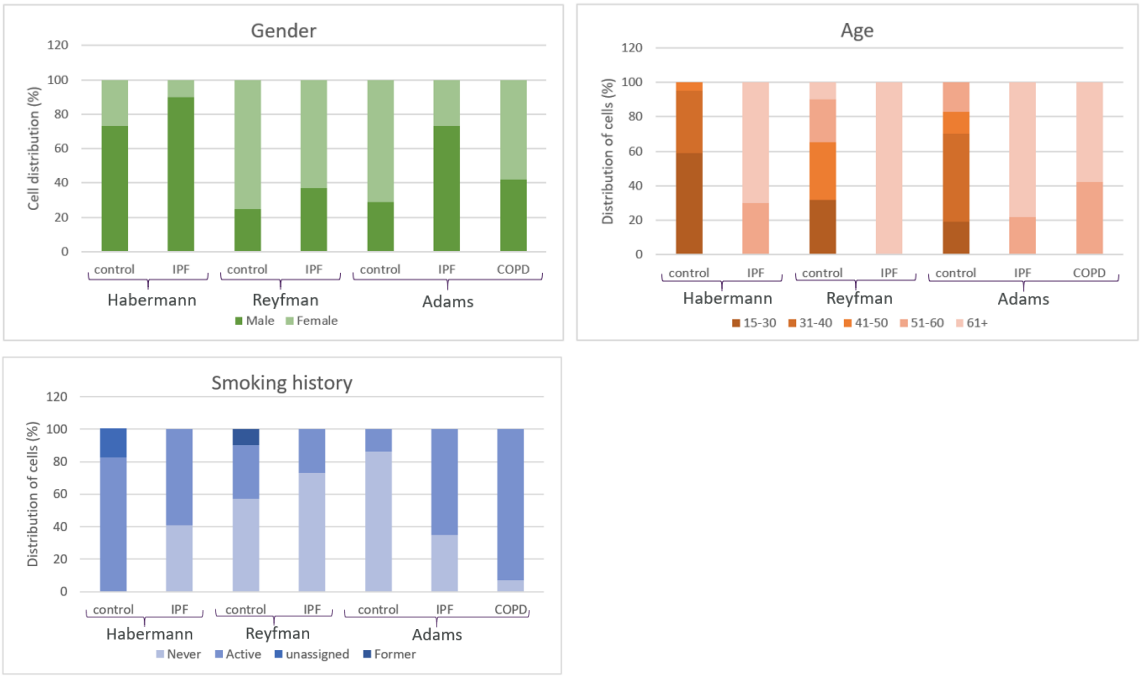

Figure 5.1: Comparison across the three datasets visualizing distributions of gender, age (male/female) and smoking history (never/active/unassigned/former) in control, IPF and COPD lungs.

5.1.2 Differential expression analysis of IPF

Results from DE analysis of IPF vs control cells for the different epithelial cell subtypes were analyzed if both groups, ie control and IPF, contained close to or more than 100 cells each. The fold-changes of genes in the UPR gene signature were investigated for predefined cell subtypes. The UPR gene signature collected from IPA and Becker et al. [22] can be seen in Table A.1 in Appendix A.

The analysed epithelial cell subtypes from all three datasets were AT1, AT2, club and ciliated cells. The genes with p-value < 0.1 were plotted and the result can be seen in Figure 5.2. The club cells from the study by Habermann, containing the largest number of analysed epithelial cells, had further been subdivided into 3 distinct populations: SCGB3A2+, SGB3A2+/SCGB1A1+ and MUC5B+, annotated Habermann 1, 2 and 3 respectively in Figure 5.2. This further division of club cells were made based on canonical lineage-defining markers and possible to accomplish due to the large amount of extracted IPF cells.

An upregulation of UPR gene expression can be seen for the majority of genes for AT1 and club cells, while a downregulation of UPR gene expression can be observed for AT2 cells, with exception for genes in the PERK pathway for the Habermann data. Concerning the ciliated cells, four of the genes are significantly downregulated in IPF (PPP1R15A, HSPA1B, HSPA1A and HSPA5/BiP) for the Habermann data and one for the Reyfman data (HSPH1), while the rest are similarly expressed in both IPF and control.

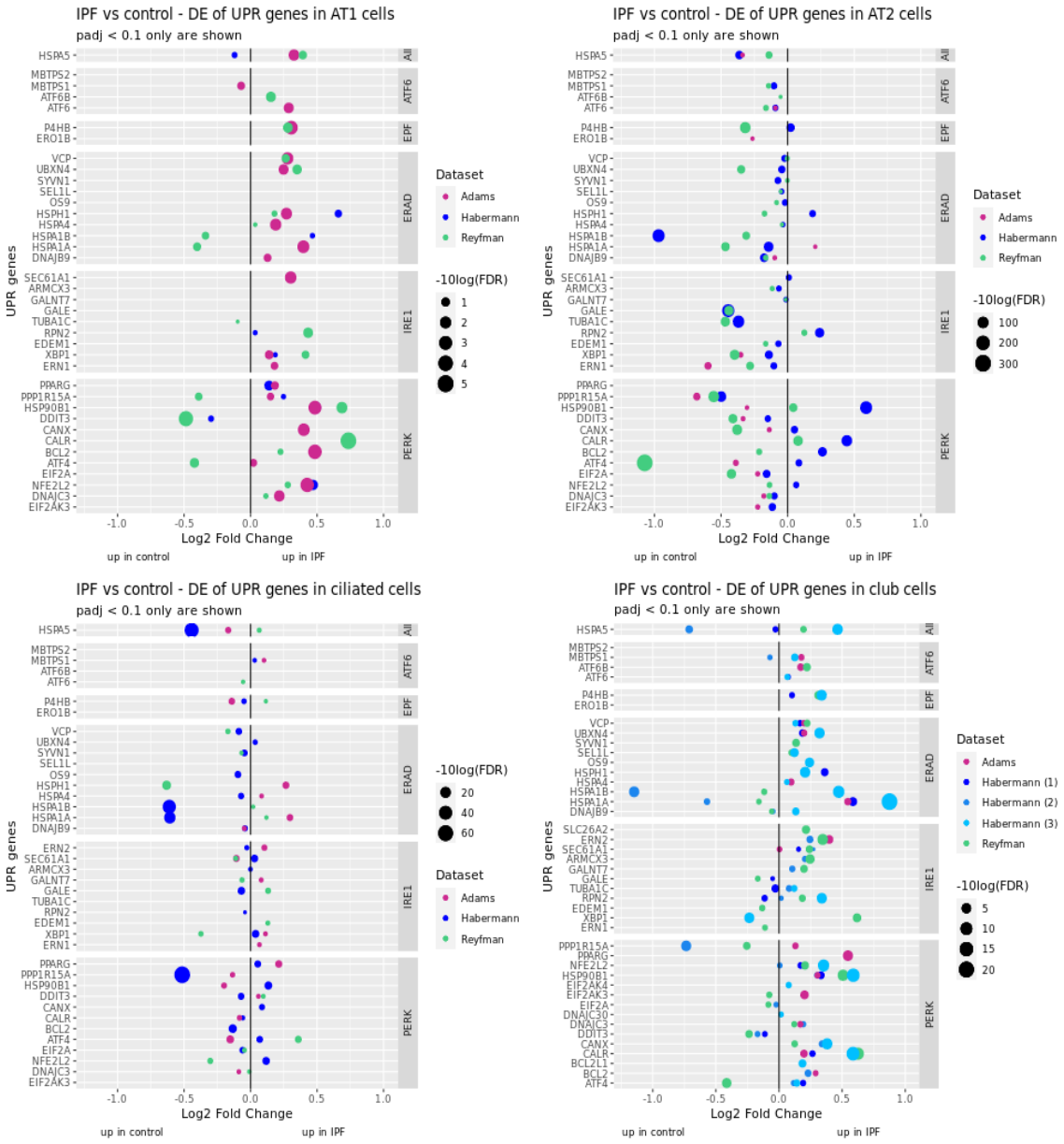

Figure 5.2: Differential expression of UPR signature genes across epithelial cell subtypes in IPF versus control cells. The UPR genes are divided into PERK, IRE1, ERAD, ER folding proteins (ERF), ATF6 and HSPA5 (All). The cell subtypes analysed across all three scRNA-seq datasets were AT1, AT2, ciliated and club. The club cells for Habermann was divided into SCGB3A2+/SCGB1A1+ (1), SCGB3A2+ (2) and MUC5B+ (3). The datasets are distinguished by colour and the significance of each gene expression are visualized by size, based on False Discovery Rate (FDR) where a higher FDR-value indicates a lower p-value and a more significantly expressed gene.

Most of the genes are similarly regulated across datasets within each cell subtype, with some exceptions, such as PERK and ERAD genes in the Reyfman data for AT1 cells and PERK genes for AT2 cells and some genes for ciliated cells and SCGB3A2+ club cells (Habermann (1)) in the Habermann data.

Further on, a comparison of the epithelial cell subtypes was made, where the cell distribution between IPF and control was calculated. An overview of cell distributions can be seen in Figure 5.3. For detailed cell distributions, see Figure A.2 in Appendix A.

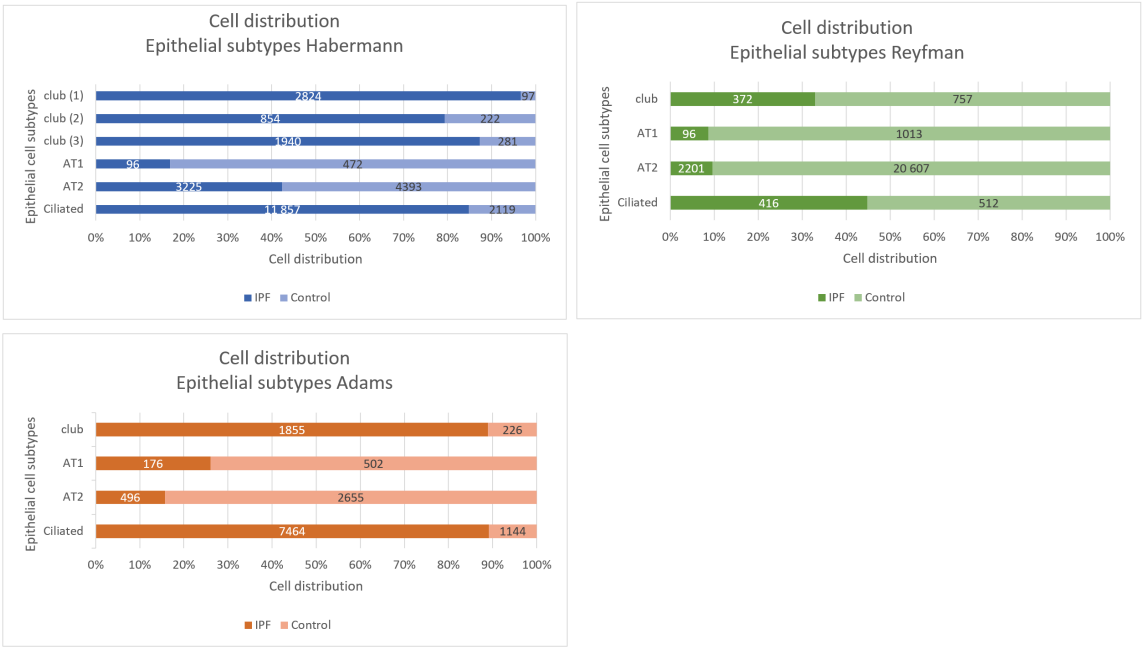

Figure 5.3: Distribution of epithelial cell subtypes in IPF versus control lungs in investigated scRNA-seq studies.

The majority of AT1 and AT2 cells was found in control, while the majority of club and ciliated cells were found in IPF. Epithelial cells in IPF are known from previous studies to consist of an increased proportion of club and ciliated epithelial cells and by a decline in alveolar epithelial cells, which was confirmed by the three scRNA-seq studies investigated in this study and consistent with the distribution seen in Figure 5.3.

5.1.3 Analysis aberrant basaloid cells

The aberrant basaloid cells discovered by Habermann and Adams were almost only found in patients with IPF (97 % and 93 % respectively). Based on the key features of this epithelial cell population, such as expression of collagen and ECM markers, this subgroup is likely to contribute to the disease pathogenesis. Thereby, this subgroup was further investigated in this study to examine its characteristics and the regulation of UPR in these cells. The gene signature created in Habermann of 12 upregulated genes for the aberrant basaloid subgroup can be observed in Table 5.1, together with p-value, FDR and percentage difference between the aberrant basaloid cells and other epithelial cells.

To first confirm that the gene signature can be used to annotate aberrant basaloid cells, the signature was applied on the epithelial cells in Habermann and Adams

to extract the cells with largest expression of these genes. In Habermann, the cell extraction resulted in a sub-cluster of 497 cells where 97 % were from IPF and 70 % of the cells were from the aberrant basaloid cell subgroup. For Adams, 448 cells were sub-clustered, 93 % of these were found in IPF and 80 % were annotated as aberrant basaloid cells or basal cells. For more details, see Figure A.1 and A.2 in Appendix A where the resulting cell distributions are visualized in T-SNE images and with graphs of the cell distribution between different subgroups.

Table 5.1: Gene signature of genes upregulated in aberrant basaloid cells compared to other epithelial cells from dataset by Habermann.

Gene	**p-value**	**FDR**	**perc diff (%)**
COL1A1	1.44E-214	210.16	82.59
FN1	4.68E-189	185.30	81.00
IL32	1.26E-191	187.82	79.56
KRT17	1.80E-185	181.76	77.40
PRSS2	3.27E-168	164.65	73.28
CALD1	4.52E-135	131.83	72.81
PTGS2	9.75E-137	133.48	70.24
PMEPA1	2.14E-126	123.19	69.34
DPYSL3	6.89E-122	118.74	68.92
MARCKSL1	4.13E-134	130.88	67.71
FHL2	1.05E-118	115.58	67.08
SLCO2A1	1.41E-107	104.58	67.02

The high percentages of IPF cells and aberrant basaloid cells led to the conclusion that the gene signature could be used to annotate an aberrant basaloid subgroup in Reyfman. In Reyfman, the epithelial cells with largest expression of the signature genes were chosen and the created sub-cluster consisted of 209 cells where 82 % of the cells were from IPF, and where the majority of cells were annotated as AT1 or AT2 (35 % and 48 % respectively). See Figure A.3 in Appendix A for T-SNE images and graph of cell distribution between different subgroups in Reyfman.

To compare the regulation of the gene signature from Habermann between cell subtypes in all three datasets and to investigate the reproducibility of the aberrant basaloid cells from Habermann in Adams and Reyfman, the proportions of IPF cells expressing these genes and the level of expression were plotted in heatmaps, which can be seen in Figure 5.4. The 12 signature genes are visibly more expressed in the aberrant basaloid cells, annotated KRT5-/KRT17+ for Habermann and Reyfman, than the rest of the epithelial cells for the subgroup in Habermann and Adams. For Reyfman, around half of the genes are higher expressed, but still quite low expression levels compared to Habermann and Adams since the maximal expression level is 1.5, opposed to 3.0 in Habermann and Adams. Additionally, it can be seen that KRT17, that is a basal cell marker, is expressed both in the aberrant basaloid cells and in the basal cells in Habermann and Adams.

After the gene signature comparisons, DE analysis between IPF and control for the

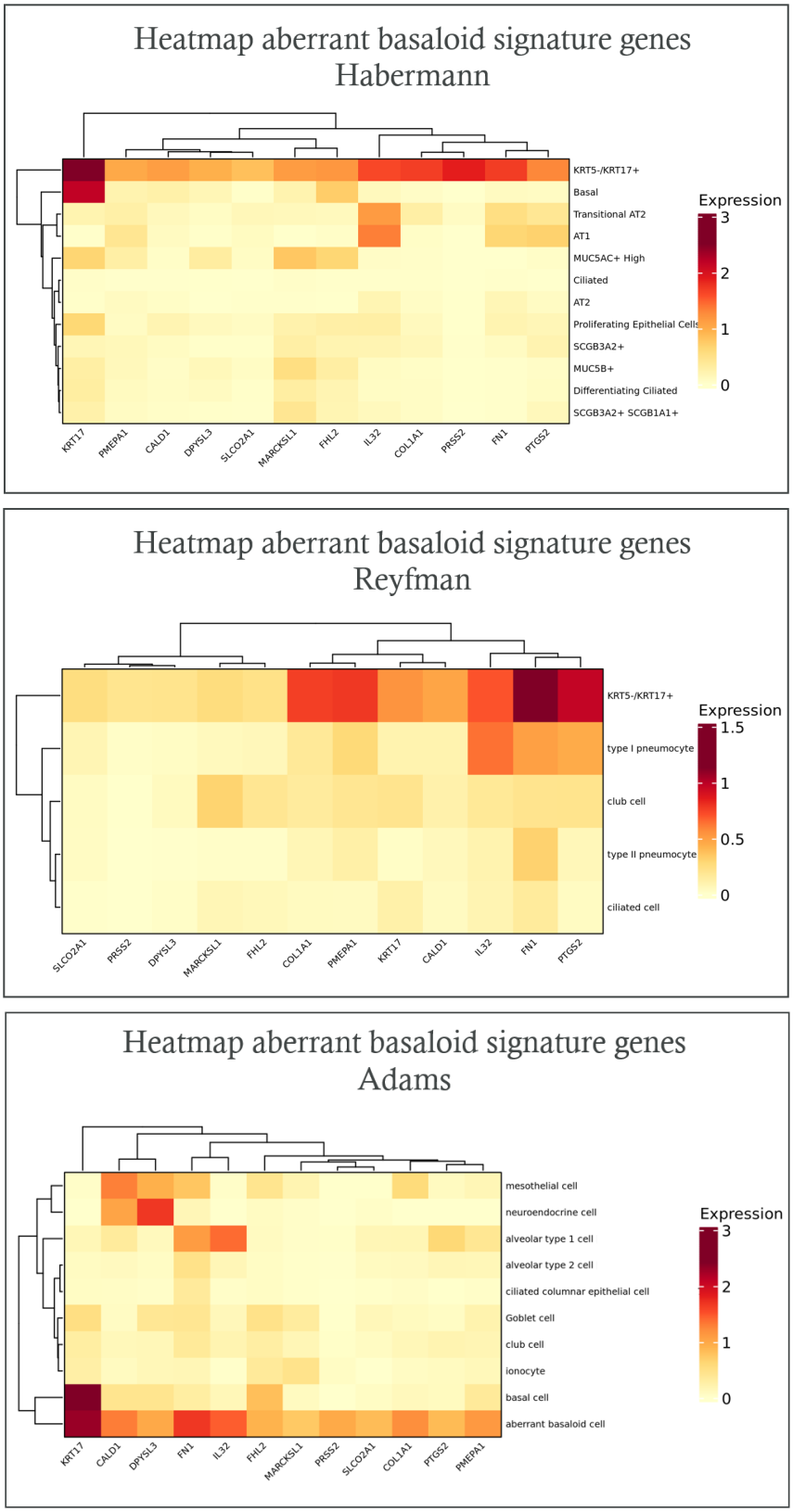

Figure 5.4: Heatmap of genes from the aberrant basaloid cell gene signature defined by data from Habermann. Gene expressions for the epithelial cell subtypes in IPF from Habermann, Reyfman and Adams are visualized. The aberrant basaloid cells are annotated as KRT5-/KRT17+ for Habermann and Reyfman. From [46].

aberrant basaloid cells was performed both on the created subgroup in Reyfman and the originally annotated aberrant basaloid cells in Habermann. However, the comparison was made on only 13 and 37 healthy cells respectively on the data from Habermann and Reyfman which is too few cells for the result to be significant or comparable.

To investigate if the UPR gene signature was especially upregulated in the aberrant basaloid cells compared to other epithelial cells, DE was performed with only IPF as condition, for all three datasets. For the subgroup in Habermann, there were 374 aberrant basaloid cells and 24891 other epithelial cells. For Adams, there were 448 aberrant basaloid cells and 12 775 other epithelial cells. Finally, for the created subgroup in Reyfman, there were 172 cells in the created aberrant basaloid subgroup and 3085 other epithelial cells. The DE analysis was performed both on the UPR gene signature from IPA and the study by Becker et al. [22], and on a terminal UPR (T-UPR) signature chosen from an article by Aueyung et al. [53], with genes that are associated with maladaptive UPR signaling.

Due to small or insignificant gene regulations for DE analysis performed on data from Reyfman, only the results from Habermann and Adams are shown in Figure 5.5, and only genes with p-value < 0.1. An upregulation of all UPR genes except ERN2/IRE1β, XBP1 and HSBH1 can be observed for the aberrant basaloid cells and for all T-UPR genes except for TXNIP in Habermann.

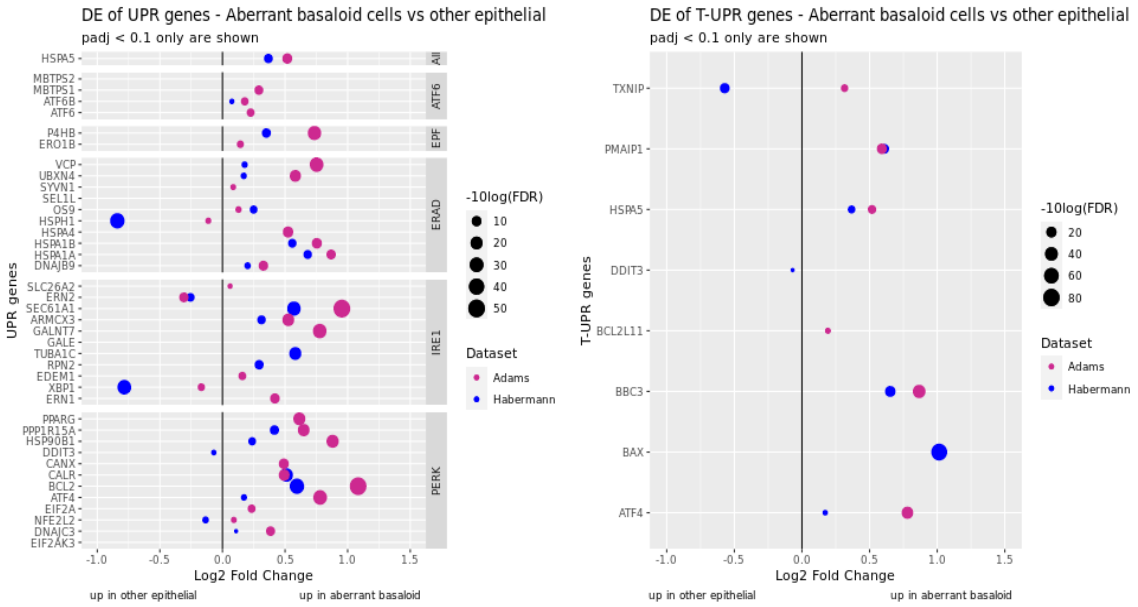

Figure 5.5: Differential expression of UPR genes in IPF for aberrant basaloid cells vs other epithelial cells from data by Habermann and Adams. DE analysis performed on data from Reyfman is not visualized due to lack of significantly regulated genes.

To further investigate the characteristics of the subgroup in the three datasets, the 28 most upregulated genes from the DE analyses between the subgroup and other epithelial cells was extracted from each dataset and compared against each other.

The comparison was made to investigate if similar gene regulations could be found across the datasets for the aberrant basaloid cells compared to all other epithelial cells. The majority of the most upregulated genes was found in all three datasets. Besides these, Habermann and Adams had the highest number of genes in common, while Reyfman had the highest number of unique genes, which could be due to fewer annotated cells in the subgroup in Reyfman.

As previously mentioned, the subgroup in Habermann and Adams expressed keratin genes, ECM components, collagens, markers of epithelial-mesenchymal transition, senescence marker genes, distal epithelial programs and the IPF biomarker MMP7. When comparing the cellular processes of the upregulated genes found in this study, they are consistent with these biomarkers. For example, upregulation of genes related to senescence, such as CDKN2A and GDF15 and upregulation of MMP7, which is a key biomarker for IPF, was found. Furthermore, upregulations of genes related to collagen and ECM were observed, such as the keratin genes and COL1A1 and FN1, where KRT7 and KRT17 are either the most regulated or amongst the top eight most regulated genes found. For more details, see Table A.3 in Appendix A where 28 of the most upregulated genes can be observed across the three datasets.

The characteristics of the created subgroup in Reyfman and its similarities to the annotated subgroups in Habermann and Adams were further investigated by performing GO enrichment for the most upregulated genes. Initially, DAVID was used for gene enrichment and the result for genes with fold-change > 0.7 can be seen in Table 5.2.

Table 5.2: GO enrichment of most upregulated genes for the created aberrant basaloid cell subgroup in Reyfman performed with DAVID and with the whole human atlas as background. Genes with fold-changes > 0.7 can be observed.

Gene	FC	P-value	Cellular component	Biological process
TAGLN	2.11	4.97E-08	Cytoskeleton	Protein binding, actin filament binding
KRT7	1.90	2.331E-06	Nucleus, cytoplasm, cytosol, intermediate filament, keratin filament	keratinization
CAV1	1.79	2.42E-08	ER, Golgi membrane	negative regulation of cytokine-mediated signalling pathway
TM4SF1	1.51	7.98E-11	Membrane	blastocyst formation
MYL9	1.39	0.00028	Cytoskeleton	Cell division and chromosome partitioning, Signal transduction mechanisms
S100A10	1.37	0.0072	Extracellular space	membrane raft assembly, membrane budding,
IL32	1.34	5.67E-07	extracellular space, cytosol, membrane	Cell adhesion, cytokine mediated signalling pathway, cell adhesion
CEACAM6	1.33	3.22E-10	Membrane, Cell membrane	Cell adhesion, signal transduction, positive regulation of gene expression, cytokine-mediated signalling pathway
EMP2	1.33	0.00002	Golgi membrane, nucleus, cytoplasm, Golgi apparatus, cytosol	Apoptosis, Cell adhesion,
TMSB10	1.16	0.0004	cytoplasm, cytoskeleton	actin filament organization, regulation of cell migration, sequestering of actin monomers, actin binding
CAV2	1.04	3.88E-06	Golgi membrane, acrosomal membrane, caveolar macromolecular signaling complex, nucleus, Golgi apparatus	regulation of cell-matrix adhesion, integrin binding
ANXA2	1.04	0.034	extracellular region, basement membrane, extracellular space, nucleus, cytoplasm	angiogenesis, membrane raft assembly, positive regulation of receptor recycling
CCND2	1.01	1.31E-08	Membrane, nucleus, cytoplasm	Cyclins and Cell Cycle Regulation, Cell cycle, Cell division
CTSE	1.01	7.93E-11	endosome	proteolysis, protein autoprocessing
KRT17	0.96	3.20E-08	cytosol, cytoskeleton, intermediate filament, keratin filament, intermediate filament cytoskeleton	positive regulation of cell growth, epithelial cell differentiation, keratinization
COL1A1	0.75	2.29E-09	extracellular matrix structural constituent	cell adhesion
MMP7	0.91	2.75E-09	extracellular region/space/matrix, cell surface	aging, extracellular matrix disassembly, extracellular matrix organization
DST	0.79	8.08E-08	basement membrane, nucleus, nuclear envelope, nucleoplasm, cytoplasm, ERmembrane	microtubule cytoskeleton organization, cytoskeleton organization, cell adhesion
FN1	0.98	5.14E-08	Extracellular matrix	cell adhesion, cell-matrix adhesion, calcium-independent cell-matrix adhesion, integrin-mediated signaling pathway

The majority of genes have characteristics similar to the annotated aberrant basaloid cell subgroup in Habermann and Adams. For example, there is an enrichment of cellular components such as cytoskeleton, extracellular space and extracellular matrix and of biological processes such as actin filament binding/organization, cell adhesion, cytokine mediated signaling, keratinization and integrin binding.

However, the analysis in DAVID was made with the whole human atlas as background, which might be too broad when looking at scRNA-seq data, hence a second GO enrichment analysis was performed in BBROWSER with GSEA where only the epithelial cells from the study was used as background. IDs with enrichment score $<$ 0.8 and FDR $<$ 0.13 was used. Examples of enriched cellular components are protein complex involved in cell adhesion, filamentous actin, keratin filament, myofilament and collagen trimer and of biological processes are cell matrix adhesion, cell-cell adhesion and integrin mediated signaling and processes connected to collagen.

5.1.4 Differential expression analysis of COPD

A comparison of cell distribution between COPD and control was calculated based on the data from the study by Adams, which was the only available scRNA-seq dataset with COPD patients, and the result can be seen in Figure 5.6. For a more detailed overview of the exact cell numbers, see Table A.2 in Appendix A.

There is an equal distribution between COPD and control cells for cells annotated as AT1, ciliated, basal and goblet, while the majority of annotated AT2 cells were found in control and the majority of annotated club cells were found in COPD.

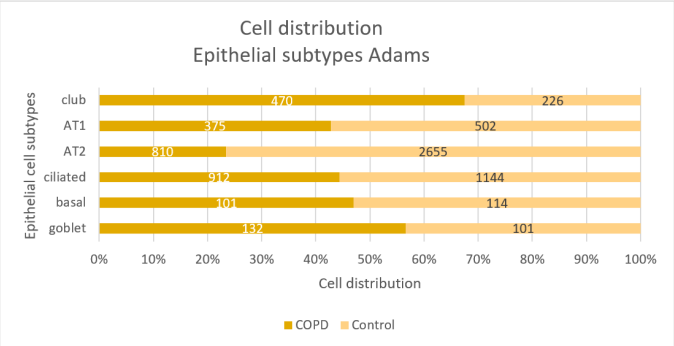

Figure 5.6: Cell distribution between COPD and control for the analysed epithelial cell subtypes annotated in the data from Adams.

For the DE analysis of COPD, the same UPR gene signature was used as for IPF and the data used was extracted from Adams. Only cell subtypes with around 100 cells or more in both COPD and control were analysed, which resulted in five epithelial cell subtypes; AT1, AT2, basal, ciliated, club and goblet. The genes with p-value $<$ 0.1 was plotted and the result can be seen in Figure 5.7.

In contrast to IPF, a downregulation can be observed for almost all UPR genes in each of the analysed cell subtypes, with exception of AT1 cells where most genes are upregulated. Due to low numbers of available cells for the different cell subtypes

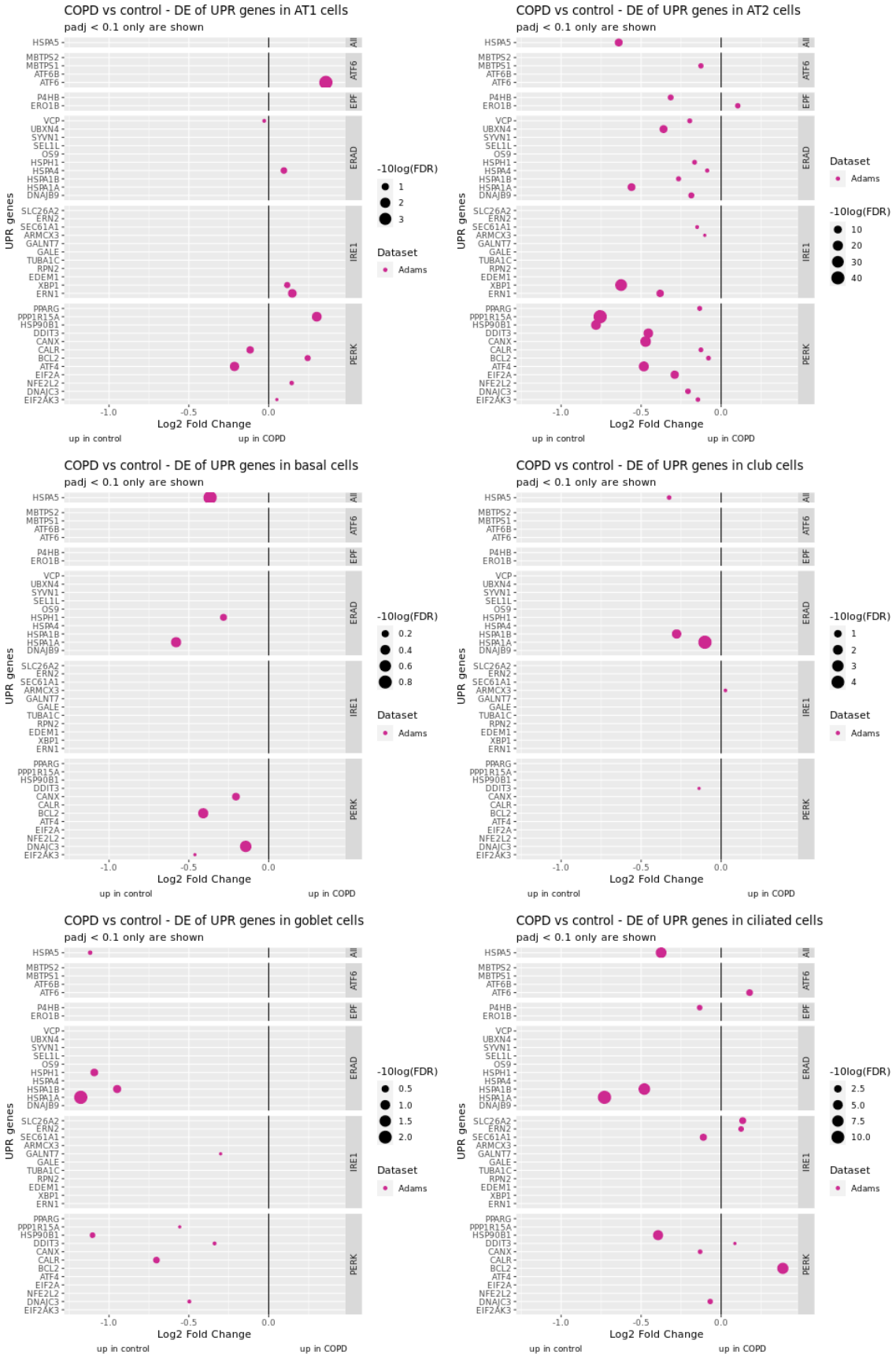

Figure 5.7: DE analysis of UPR genes for COPD vs control for the epithelial cell subtypes AT1, AT2, basal, club, goblet and ciliated annotated in data by Adams.

and to lack of significantly regulated genes, no further investigation of COPD was performed.

5.2 *In vitro* experiments - gene readouts

The purpose with the first experiment was to establish protocols for gene and protein readouts of UPR. Gene readouts from the experiment can be seen in Figure 5.8, where the fold-change has been calculated as the ratio of stimulated cells compared to non-stimulated controls. The cells were stimulated with 0.01-0.5 µg/ml tunicamycin and concentrations transformed to a logarithmic scale with base 10.

An upregulation of gene expression can be observed for almost all genes. Maximum response was observed at 6 h stimulation, with high fold-changes for BiP, CHOP and sXBP1. However, the highest fold-change for sXBP1 seen in Figure 5.8 occurs after 3 h stimulation, which is not a result of transcriptional activation, but due to the fast kinetics of sXBP1 which results in a fast upregulation as a direct effect of IRE1 activation.

When comparing non-spliced and spliced XBP1, expressed in the IRE1 pathway, a compensatory effect can be observed. This could be due to that splicing of XBP1 depletes the stores of full-length XBP1 mRNA, which is restored after 24 h when gene transcription is induced. An upregulation can also be seen for ATF4 and PPP1R15A that represent the PERK pathway, for all concentrations except PPP1R15A at 24 h. Finally, there is almost no regulation of ATF6 at any of the three timepoints.

Cell morphology was imaged at the different timepoints. Tunicamycin inhibits protein glycosylation and is known to induce changes in cell morphology, which was observed in the experiment since the cell viability was reflected by the dose of tunicamycin, with the highest cell density for non-stimulated cells and lowest cell density at 0.5 µg/ml. No visual effect was observed after 3 h stimulation, but a gradual decrease of live cells can be observed at 6 h or first at 24 h. The decrease in cell viability most likely contribute to the decreased gene expression observed at house-keeping gene level at 24 h. See Figure A.4 in Appendix A for cell images.

Thereafter, it was explored how H_2O_2 as an external source of oxidative stress could impair redox balance in the cells and thereby induce ER stress and UPR. The cells were stimulated with 50-1000 µM H_2O_2. Gene readouts of BiP, CHOP, ATF4, ATF6, XBP1, sXBP1 and PPP1R15A from the experiment is visualised in Figure 5.9, where fold-changes were calculated as the ratio of stimulated cells compared to non-stimulated controls.

There is almost no gene regulation after stimulation with H_2O_2 apart from some upregulation that can be seen for CHOP and sXBP1 after 6 h stimulation, and some downregulation of the genes after 24 h stimulation. Cell images were taken after 6 and 24 h stimulation with H_2O_2 and a high cell death can be observed at 600-1000 µM H_2O_2. See Figure A.5 in Appendix A for cell images from experiment

2.

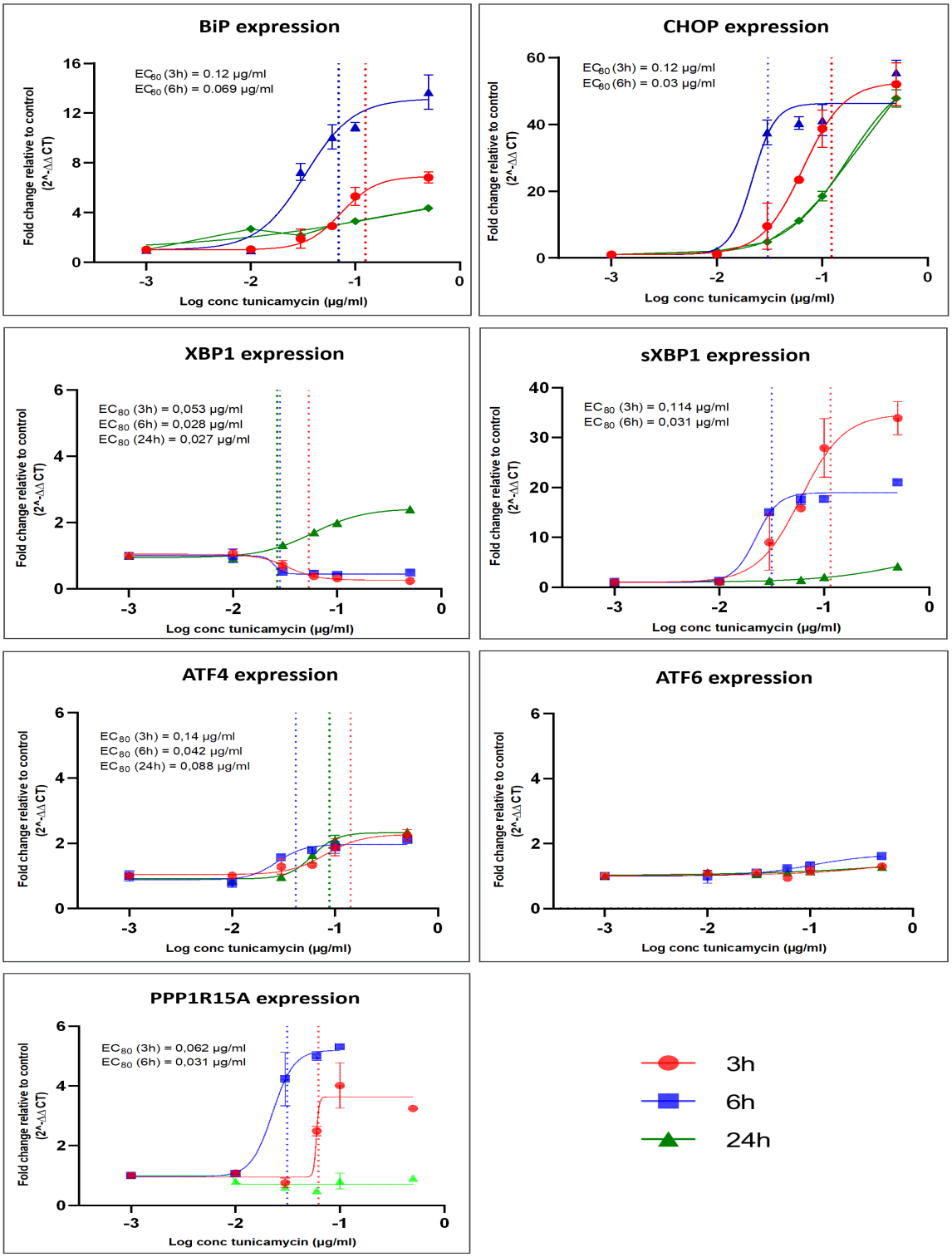

Figure 5.8: Dose-response of UPR genes in healthy HBECs after 3, 6 and 24 h stimulation with 0.01-0.5 µg/ml tunicamycin, normalized against non-stimulated control at each timepoint. The fold-change of stimulated cells relative to non-stimulated controls are visualized and tunicamycin concentrations are shown in a logarithmic scale with base 10. At the EC_{80}-values, a dose-response around 0.1 µg/ml can be observed for the majority of genes after 3 h stimulation and around 0.03 µg/ml after 6 h stimulation

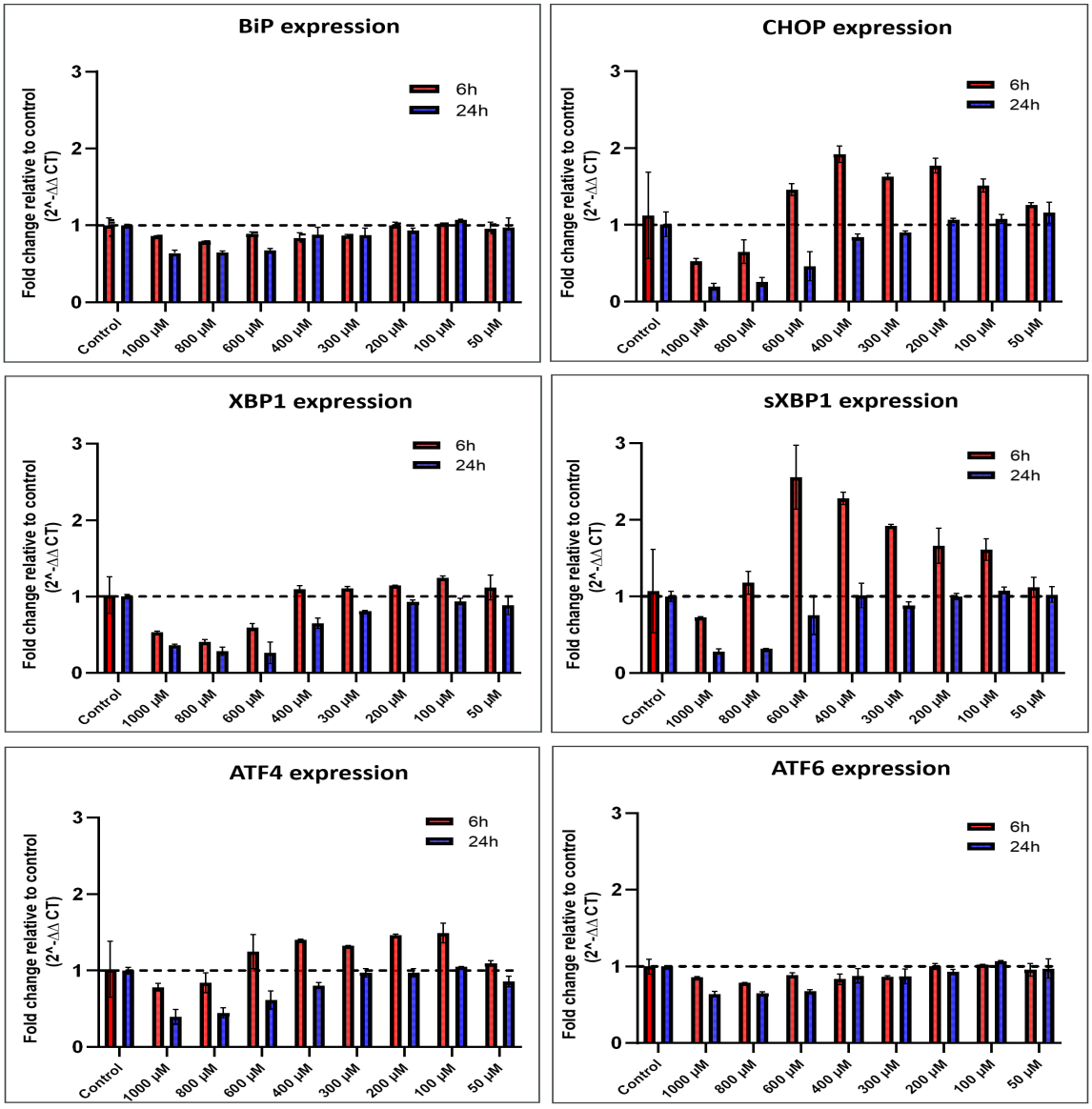

Figure 5.9: UPR gene expression in healthy HBECs after stimulation with 50-1000 µM H_2O_2, normalized against non-stimulated control at each time point. The fold-change of stimulated cells relative to non-stimulated controls are visualized.

In a final experiment, multiple stimuli were used to induce ER stress and compare how stimuli with different modes of action modulates key UPR gene expression and to confirm the results from the previous experiment with H_2O_2. The fold-change was calculated as the ratio of stimulated cells compared to non-stimulated controls at each separate timepoint. Gene readouts of BiP, CHOP, ATF4, ATF6, XBP1, sXBP1, PPP1R15A and HSP90B1 were established for the different stimuli used.

The gene readouts for H_2O_2 can be seen in Figure 5.10. As in previous experiment, very little UPR gene induction was observed after H_2O_2 challenge.

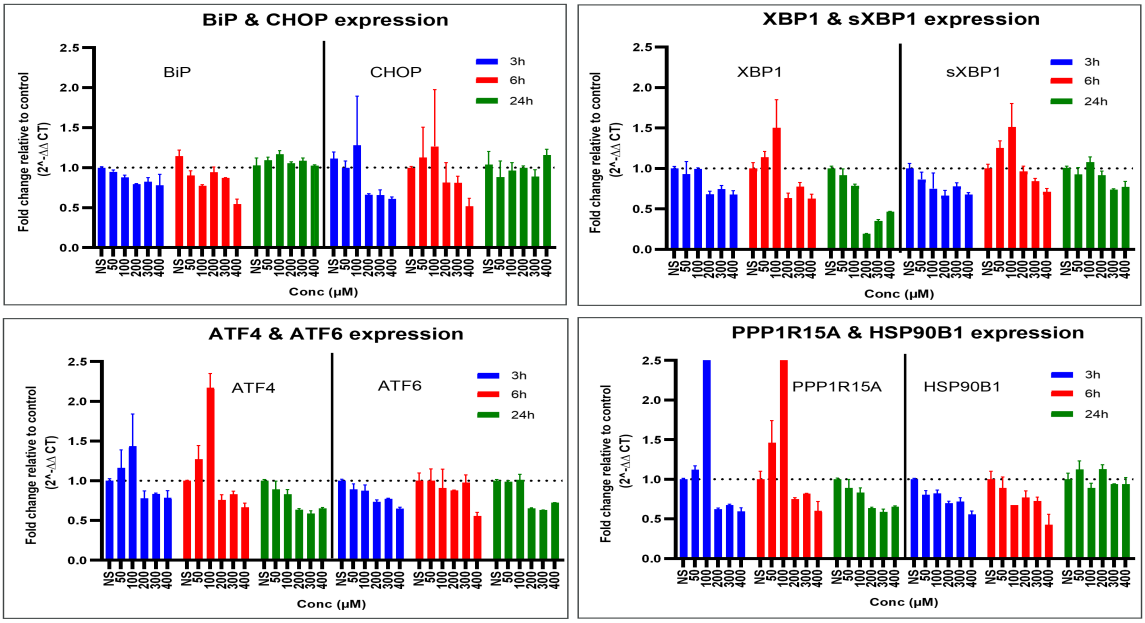

Figure 5.10: UPR gene expression in healthy HBECs after stimulation with H_2O_2, normalized against non-stimulated control at each time point. The fold-change of stimulated cells relative to non-stimulated controls are visualized. The concentrations used were 50, 100, 200, 300 and 400 µM.

UPR gene induction after tunicamycin challenge was similar to that observed in previous experiment. The gene readouts for tunicamycin in the final experiment can be seen in Figure 5.11.

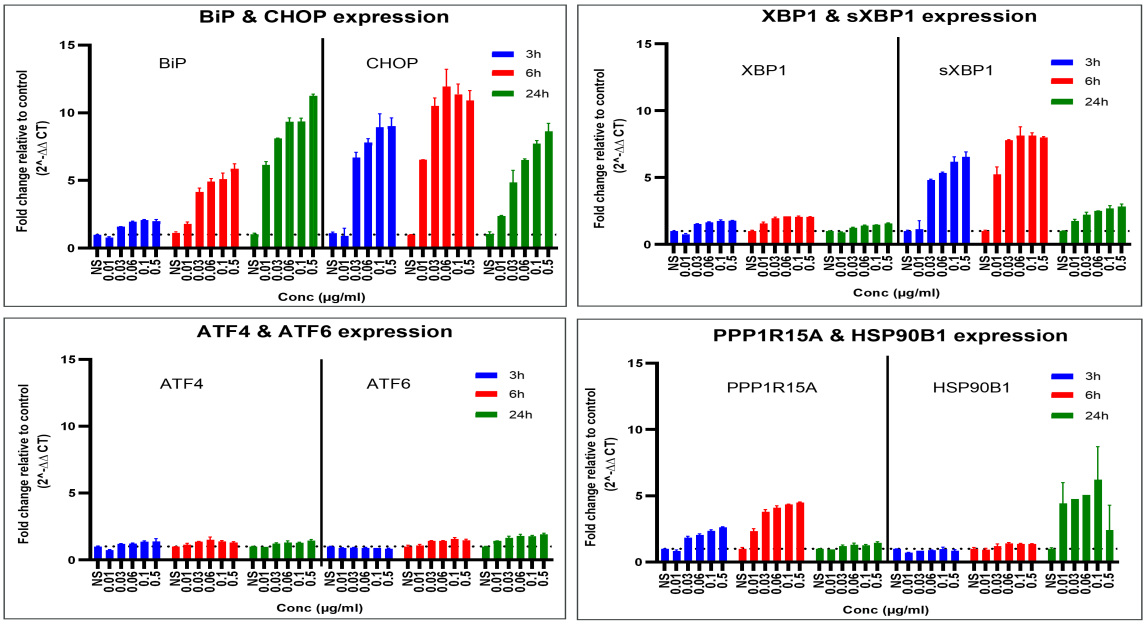

Figure 5.11: UPR gene expression in healthy HBECs after stimulation with tunicamycin, normalized against non-stimulated control at each time point. The fold-change of stimulated cells relative to non-stimulated controls are visualized. The concentrations used were 0.01, 0.03, 0.06, 0.1 and 0.5 µg/ml.

For stimulation with tunicamycin, there is a 10-fold increase of BiP, CHOP and sXBP1 and a 5 fold-increase for PPP1R15A and HSP90B1. For BiP, a dose response can be seen at 6 and 24 h and for CHOP there is a dose response at 24 h. Regarding sXBP1, PPP1R15A and HSP90B1, there is only a small difference in gene expression for 0.03 µg/ml and higher doses, which is the same concentration as for the EC_{80}-values for PPP1R15A and sXBP1 in experiment 1. However, the fold-changes for XBP1, ATF4 and ATF6 are only around two as max.

The gene readouts for thapsigargin can be seen in Figure 5.12. Regarding stimulation with thapsigargin, the same genes are regulated as for tunicamycin, namely CHOP, BiP, sXBP1, PPP1R15A and HSP90B1, and at the same timepoints. However, the upregulation of genes are higher and there is no clear dose-response between concentrations. What can be observed is an increase until 10 or 50 nM for most of the regulated genes. For BiP, the fold-change is highest after 24 h stimulation while CHOP shows a similar regulation at all three time points. For sXBP1 and PPP1R15A, the fold-change is largest at 3 and 6 h stimulation, while HSP90B1 only is upregulated after 24 h stimulation.

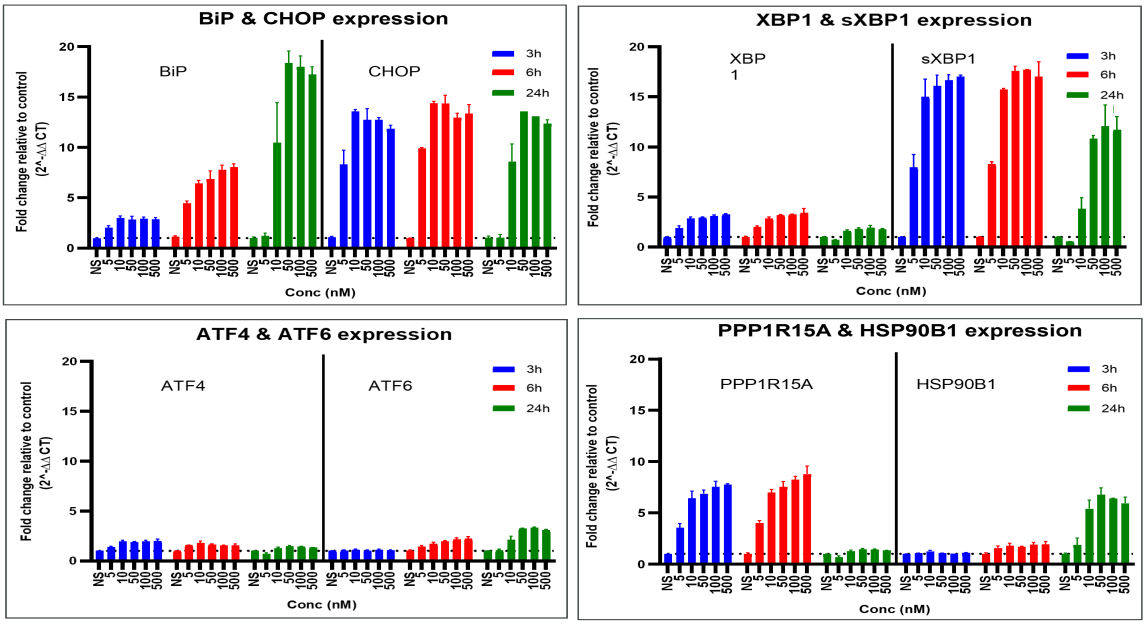

Figure 5.12: UPR gene expression in healthy HBECs after stimulation with thapsigargin, normalized against non-stimulated control at each time point. The fold-change of stimulated cells relative to non-stimulated controls are visualized. The concentrations used were 5, 10, 50, 100 and 500 nM.

The gene readouts for cigarette smoke extract (CSE) can be seen in Figure 5.13. The stimulation with CSE resulted in upregulation of CHOP, sXBP1 and PPP1R15A. All three genes are upregulated at 3 and 6 h stimulation, while only CHOP and sXBP1 is up-regulated after 24 h. Furthermore, the highest upregulation of these genes are around 5 and 7.5 % CSE stimulation.

Finally, the housekeeping genes used in all experiments, to normalize the Ct-values

for the UPR genes, were GAPDH and RPLP0. Gene expressions for both genes are stable at 3 and 6 h gene readouts in all experiments. However, after 24 h stimulation, the Ct-values differed several cycles in the first experiment, for 600-1000 µM H_2O_2 in the second and in the third experiment together with CSE. The reason for deviating Ct-values at 24 h could be due to lower cell viability at this time point. A larger amount of apoptotic or dead cells can be observed at 24 h for higher doses of H_2O_2 and CSE, which contribute to decreased amount of expression in the housekeeping genes used and to higher Ct-values. The Ct-values for GAPDH and RPLP0 for all three experiments can be found in Figure A.7 in Appendix A.

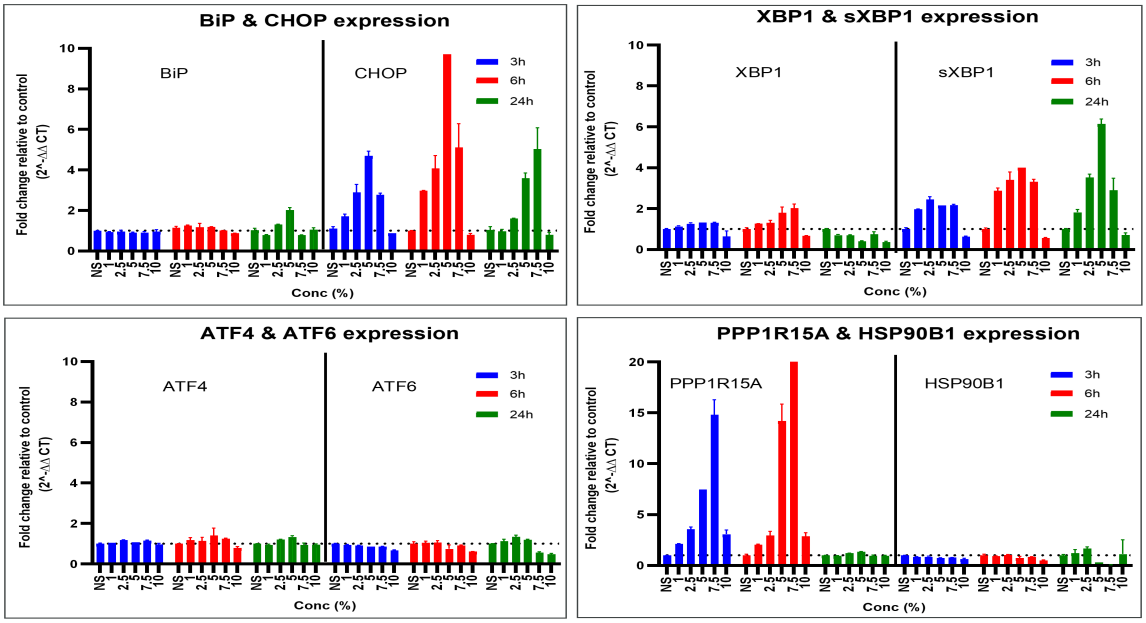

Figure 5.13: UPR gene expression in healthy HBECs after stimulation with cigarette smoke extract (CSE), normalized against non-stimulated control at each time point. The fold-change of stimulated cells relative to non-stimulated controls are visualized. The CSE consisted of smoke from five cigarettes and the percentages of CSE used for stimulation was 1, 2.5, 5, 7.5 and 10 %.

5.3 *In vitro* experiments - protein readouts

The protein readouts from the first experiment is presented in this section. Protein readouts were established to compare the expression at gene and protein level and measure pathway activation of UPR, since gene and protein expression do not always correlate. Protein expressions from automated Western blot with JESS is presented first, followed by the results from traditional Western blot.

5.3.1 Automated Western blot

The automated Western blot, performed with JESS, was first performed on the same samples from an in-house experiment, with cells stimulated with tunicamycin in absence or presence of a commercially available PERK inhibitor. The run was

performed to validate antibodies and find an optimal antibody concentrations that allows for quantitative measurements of protein expressions. Antibodies for protein expressions of eIF2α, phosphorylated eIF2α (phospho-eIF2α), ATF6 and CHOP were tested. Each of the proteins were tested with 1.56 µg protein content and with two different concentrations if primary antibodies, as can be seen in Table 4.6 in section 4.4. The JESS run resulted in protein readouts for eIF2α and phospho-eIF2α, but no readouts for ATF6 and CHOP. See Figure A.8 in Appendix A for visualization of the run and figures of protein expression normalized against β-actin.

Selected antibody concentration for phospho-eIF2α (1:5 dilution) was used to investigate the expression of phospho-eIF2α in the time-course samples. This expression was investigated together with readout of a fifth protein, BiP (1:100 dilution), which optimal antibody concentration previously had been established at AstraZeneca. The ratio of the proteins of interest against β-actin can be seen for the three timepoints for both BiP and phospho-eIF2α in Figure 5.14. For images of the runs, see Figure A.9 in Appendix A.

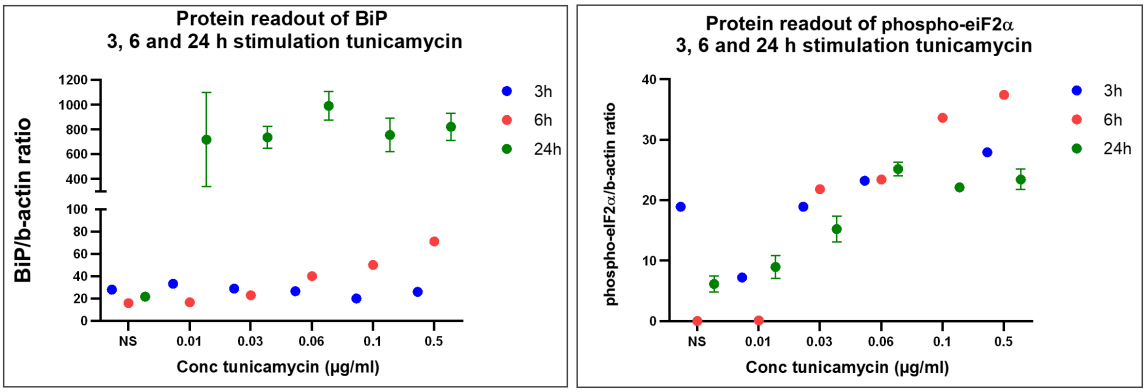

Figure 5.14: UPR protein expression readouts with automated Western blot for BiP and phospho-eIF2α after 3, 6 and 24 h stimulation with tunicamycin.

The protein expressions of BiP after 3 h stimulation are similar for all concentrations of tunicamycin. Furthermore, a dose-response can be seen for BiP and phospho-eIF2α after 6 h stimulation where the protein expression increases with increased concentration of stimulus. At 24 h, BiP was strongly induced at all concentrations of tunicamycin. Phosphorylation of eIF2α displayed a similar dose-response relationship to tunicamycin at all timepoints investigated. When comparing the 24 h readout of phospho-eIF2α from cells stimulated with 0.1 µg/ml in Figure A.8 in Appendix A, the same level of protein expression can be observed, demonstrating that the result is reproducible.

5.3.2 Traditional Western blot

The traditional Western blot was performed to further investigate protein readouts of ATF6 and CHOP, normalized to β-actin, since these readouts could not be established with automated Western blot. Western blot was first performed on cells stimulated with tunicamycin in absence or presence of a commercially available

PERK inhibitor. The experiment resulted in protein readout of ATF6 and cleaved ATF6, which can be seen in Figure 5.15, while no readout of CHOP could be established.

For the PBS control, only bands for non-cleaved ATF6 could be observed. For the tunicamycin stimulated samples, both non-cleaved and cleaved ATF6 could be observed, with a higher expression of cleaved ATF6 for the samples with no inhibitor added.

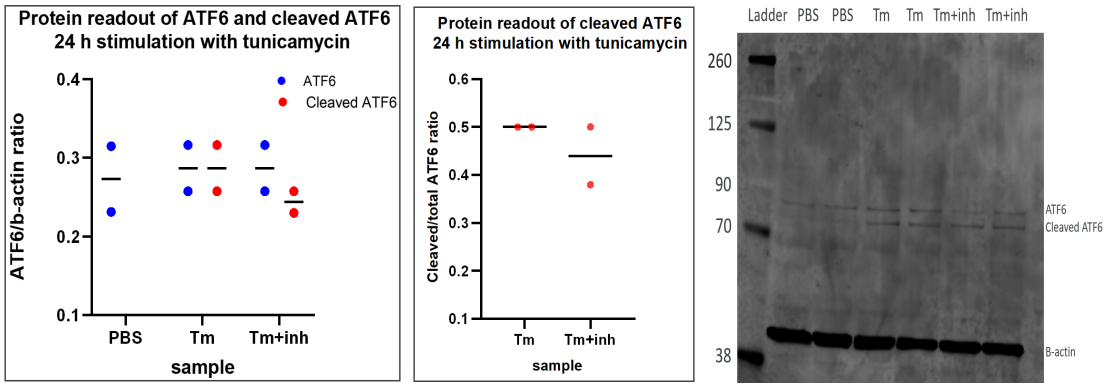

Figure 5.15: Detection of cleaved ATF6 with traditional Western blot after 24 h stimulation with tunicamycin. Protein expression in non-stimulated cells with PBS and tunicamycin stimulated cells with and without inhibitor of the PERK pathway can be seen.

Western blot was thereafter performed with the same antibodies to investigate protein readout for CHOP with the protein extractions from experiment 1. However, the experiment only resulted in protein readout of β-actin, which could be due to an unsuccessful transfer of proteins from gel to membrane since almost no proteins were visualized when staining with Ponceau S. An unsuccessful transfer could depend on the small size of CHOP (27 kDa) or to space between gel and membrane during transfer, disabling transfer of proteins.

5.4 Cell viability

Cell viability over a 48 h time-course stimulation of H_2O_2, thapsigargin and tunicamycin was performed in the final experiment. The cells were stained with IncuCyte NucLight Rapid Red Reagent, that stains nuclei in cells, and with IncuCyte Caspase 3/7 Green Reagent, that detects apoptosis in live cells by cleavage of green fluorescent protein when caspase-3/7 is activated, resulting in staining of nuclei DNA. The fluorescence was measured every third hour and the resulting percentages of live and apoptotic cells can be seen in Figure 5.16.

Regarding live cells, there is a large drop in cell viability for the three highest doses of H_2O_2 compared to the lower doses that had a constant viability throughout the run. For thapsigargin, the constant percentage of live cells seen indicates that there is no or a very small toxic effect of this stimulus, except for 100 nM and 500 nM

where a drop in live cells can be observed. there is a linear decrease of live cells for tunicamycin, indicating that it has a toxic effect on the cells at the chosen concentrations. Regarding the results from the apoptotic stain, the percentage of apoptotic cells did not correlate with live cells for H_2O_2, which could be due to unsuccessful staining or that the programs used to target these cells did not find the correct cells. However, the apoptotic staining did correlate with tunicamycin and thapsigargin, indicating that the staining did work.

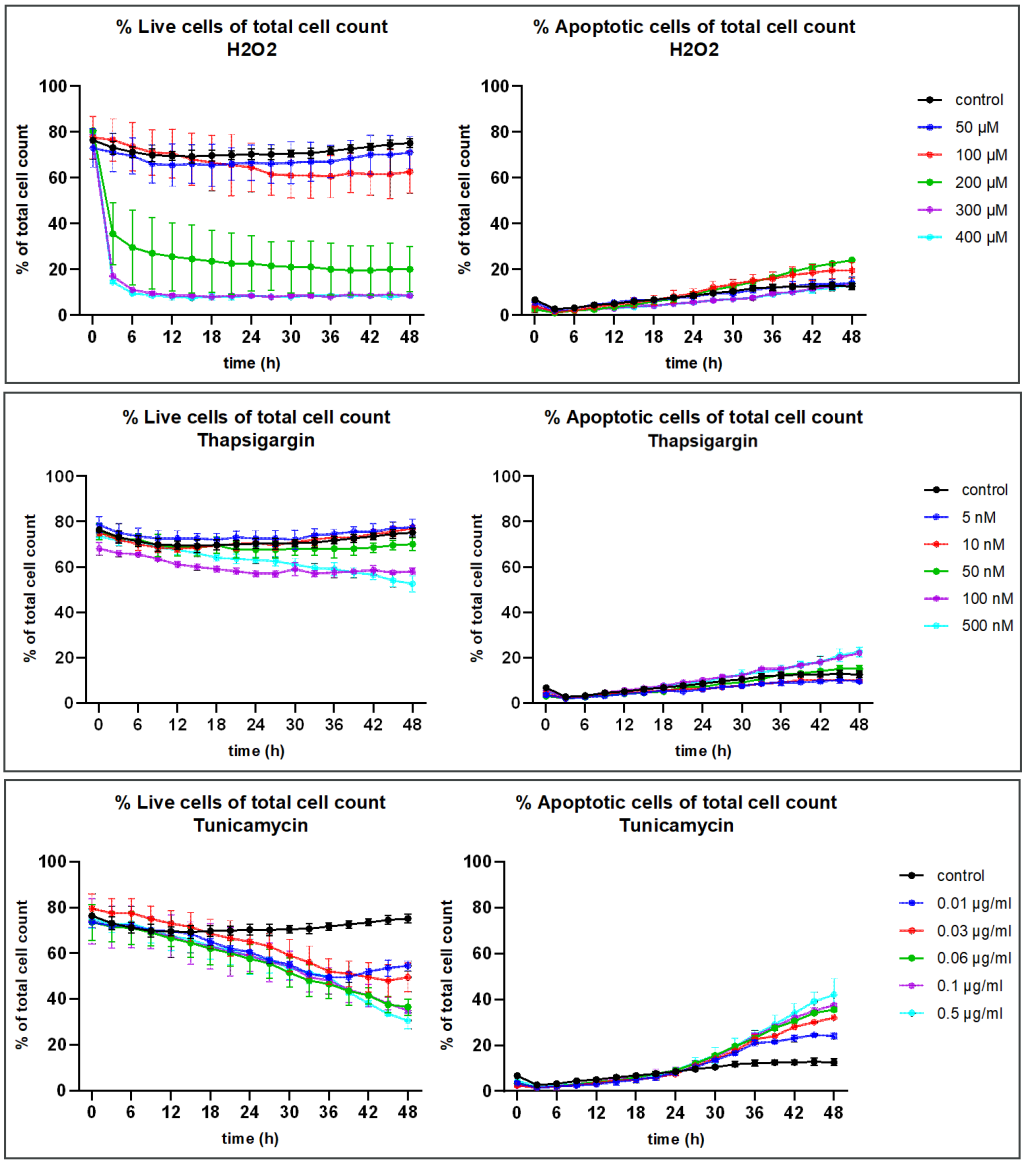

Figure 5.16: The percentage of live and apoptotic cells when stimulated with different concentrations of H_2O_2, thapsigargin and tunicamycin during a 48 h time-course experiment.

6
Discussion

6.1 Differential expression of IPF

To comprehensively understand cell diversity and central mechanisms that drive pathological epithelial remodelling in disease, *in silico* models can provide valuable insights. Both bulk RNA and scRNA can be utilized to gain important knowledge, but due to the variety of functionally distinct cell types that exist in the lung, scRNA-seq provides insight not available from bulk RNA-seq analysis. It enables potential identification of key mechanistic mediators as well as highlighting cell subtypes with key role in disease pathologies. Analysis of the scRNA-seq datasets utilized in this study has enabled identification of epithelial cell subtypes that differs significantly in UPR at the transcriptional level and in distribution between disease and control, especially for IPF.

The DE analyses in IPF identified gene regulations that were consistent across the three independently performed studies, which strengthens the reliability and robustness of the result and the analysis of the UPR pathway. From the DE analysis and the cell distribution between IPF and control several observations were made. For example, all three studies demonstrated a loss of AT2 and AT1 cells in IPF, which indicates that lung fibrosis is promoted by disrupted alveolar epithelial homeostasis, even if this hypothesis is in need of further investigation.

Regarding AT2 cells, their low abundance in IPF is consistent with previous literature. AT2 cells differentiate into AT1 cells when required as a repair mechanism to restore alveolar homeostasis. However, in IPF, exposure to intrinsic and extrinsic factors have been correlated to dysfunctional AT2 cells, which lead to AT2 hyperplasia, damaged capacity of differentiating into AT1 cells, induction of senescence programs, apoptotic AT2 cells, fibroblast activation and tissue remodelling [54], [55]. However, opposite to the hypothesis about upregulation of UPR, no UPR induction could be observed in AT2 cells of IPF patients compared to control. Previously performed studies have shown induction of UPR in AT2 cells in *in vivo* models of pulmonary fibrosis, for example through disrupted EMC3 [56], knockout of BiP [57] or mutations in SFTPC [27]. The downregulation seen in this study could be due to that gene expression do not necessarily correspond to protein expression or to a deficiency in inducing functional UPR in AT2 cells, which could lead to reduced cell numbers in disease. The analysed control lungs, which at least in Habermann were declined for organ donation, could have been derived from non-healthy subjects with

increased UPR.

In IPF tissue, the number of AT1 cells were scarce but those identified had an up-regulated UPR expression profile. This may suggest that the few remaining AT1 cells were under stress. The low abundance of AT1 cells is consistent with previous findings where it has been confirmed that there is a loss of AT1 cells in IPF [55]. It can be explained by their regulatory role in the thin air-blood barrier together with their morphology, ie thin and stretched, and their large coverage of the alveolar area (90-95 %) which makes them very vulnerable in IPF where ECM accumulates and the lungs become rigid. However, dissociation biases between different studies during cell isolation contribute to different cellular proportions than those found *in vivo* due to variable cellular sensitivities and AT1 cells has previously been found to be under-represented in scRNA-seq data [58].

The observed abundance of club cells in IPF are consistent with previous studies. Club cells are the main secretory cells in human small airways and their secretory products SCGB1A1, MUC5B and SCGB3A2 have been observed to be significantly increased in epithelial lining fluid of patients with IPF in a scRNA-seq study by Zuo et al. [59]. Previous correlations between club cells and IPF and the fact that UPR is important for functional secretory cells, contribute to the conclusion that the induction of UPR seen in the annotated club cell types might contribute to the abundance of club cells and to the pathogenesis of IPF [60].

Finally, the large abundance of ciliated cells in IPF could be due to that these cells normally are found throughout the airways and therefore might have replaced alveolar units in IPF. The hypothesis is strengthened by a scRNA-seq study from Xu et al. where cell markers of ciliated, goblet and basal cells but not alveolar cells were observed in IPF and a loss of regional specification and gene expression was suggested [61]. The lack of UPR gene regulation seen in these cells compared to the upregulation seen in more secretory cells could also be due to a smaller need for protein folding, contributing to that UPR is of less significance in ciliated cells.

6.2 Aberrant basaloid cells

The heatmaps of the 12 signature genes from Habermann applied on epithelial cells in Habermann and Adams confirmed that these genes distinguish the aberrant basaloid cells from other epithelial cells and that the annotated subgroups have similar characteristics. The existence of this newly found subgroup was further confirmed by the created subgroup in Reyfman, which showed similar characteristics as the already annotated subgroups. Both by similar heatmap results and since the majority of cells were from IPF and annotated as AT2 cells, which complies with the slingshot-based pseudo-time trajectories made of cells from pulmonary fibrosis in Habermann [10]. From the trajectories, the authors hypothesised that the aberrant basaloid cells might be derived from transitional AT2 cells. Finally, the comparison of biological processes between the annotated subgroups and the created subgroup in Reyfman by GO enrichment strengthened the reproducibility of the aberrant basa-

loid cells since similar GO enrichment as described by Habermann and Adams were achieved.

Altogether, above mentioned findings in IPF patients across the three scRNA-seq datasets investigated confirm the existence of aberrant basaloid cells since similar characteristics were discovered and the cells were shown to be reproducible across datasets.

Regarding the UPR induction and abundance of aberrant basaloid cells in IPF, it can be stated that UPR in these cells might be important for development of IPF. One explanation to the observed upregulation of UPR genes could be the secretory phenotype of these cells which increases the need for protein folding. However, further studies are required to enable such conclusions. The impact of UPR is further strengthened by upregulated T-UPR genes since they are specific for maladaptive UPR signalling. In a study by Auyeung et al. [53], it was suggested that IRE1 is activated during conversion into these cells and that gene expression changes caused by IRE1 activation during lung epithelial injury are part of terminal UPR [53]. The upregulation of IRE1 and T-UPR genes in this study confirms these previous findings and the connection of UPR to IPF.

Unfortunately, regulation of the UPR could not be seen for the created aberrant basaloid subgroup in Reyfman, which could be due to that Reyfman had fewer aberrant basaloid cells annotated than Habermann and Adams and thereby small and insignificant regulations of UPR genes. The fact that aberrant basaloid cells are a rare cell type could also contribute to the difficulties of correctly annotating the cell subgroup across datasets.

6.3 Differential expression of COPD

Regarding the DE analysis of COPD, there might have been too few cells for a significant comparison of COPD and control, which is important to take into account since the number of cells have a large impact on the validity of the results. When comparing the regulation of genes for the epithelial subtypes in this study with literature, there is one article where the distribution of AT2 cells was similar and where the authors concluded the importance of AT2 cells in COPD [58]. Thereby, the impact of AT2 cells in COPD can be confirmed, but not the role of UPR.

6.4 Comparison of datasets

When comparing different datasets it is important to consider the differences in methods and technologies used. For scRNA-seq datasets specifically, multiple factors are likely to influence the result; the number of cells and samples, cell types, method used for annotating cell subtypes, quality control cut-offs and cell extraction methods.

Age distribution was one clear differentiating factor between IPF and control tissue that is likely to have an impact on the analysis.For example, if the difference in gene regulation between controls and diseased subjects mainly are due to disease state, age differences or other age-related changes not related to disease. Another crucial parameter was the epithelial cell numbers, which enables in depth investigation of epithelial cell subtypes. Lung tissue processing most likely influence the extraction of different cell types and the number of cells extracted. It involves both enzymatic and mechanical processing that can injure cells, leading to cell death and loss of specific cell populations. The enzymes chosen for digestion differed between the datasets and their different digestion capacities effect the efficiency and digestion rates. Furthermore, different digestion efficiencies between IPF and control lungs due to the accumulated ECM in IPF can also introduce biases.

6.5 *In silico* future analyses

A comparison interesting to make is to integrate the three datasets and analyse the whole set of epithelial cells. To create an integrated epithelial cell atlas could help capturing the variability seen concerning age, number of cells and samples, and the risk of wrongly annotate cell subtypes. Furthermore, a cell atlas could be used to better distinguish between disease-specific changes and natural variations in the datasets. Biological heterogeneity could be captured more accurately, which is useful when analysing rare cell types such as the newly discovered aberrant basaloid cells.

To further explore the role of UPR in epithelial cell subtypes, a similar process as for the aberrant basaloid cells could be implemented for the MUC5B+ subgroup from Habermann by applying biomarkers of MUC5B+ cells on epithelial cells in Reyfman and Adams. The analyse would be of interest since the MUC5B+ cells showed similar characteristics as the aberrant basaloid cells and could be important for the disease pathology. An UPR dependent genetic link, MUC5B promoter polymorphism, between these cells and IPF has previously been confirmed by Chen et al. via IRE1β and spliced XBP1 in the IRE1 pathway [28]. How MUC5B contribute to IPF initiation and progression is intriguing and it might be a future targeting molecule.

It would also be of interest to apply all three secretory (club) subtypes from Habermann on the data from Adams and perform a trajectory analysis. That would allow the hypothesis made by Habermann to be explored about SCGB3A2+ being able to differentiate into AT1 cells via transitional AT2 cells. A similar cell trajectory analysis has been observed in a scRNA-seq study on mice, where both club cells and AT2 cells could transition into AT1 cells [62].

6.6 Gene readouts

The upregulation of BiP, CHOP and spliced XBP1 seen in HBECs when challenged with tunicamycin confirm the induction of UPR by activation of the IRE1 and PERK pathways at transcriptional level. However, the ATF6 pathway was only activated at protein level, which demonstrates the importance of investigating expression at both transcription level and protein level since observations on mRNA level do not necessarily reflect the response at protein level.

Regarding H_2O_2 challenged cells, PERK-mediated upregulation of CHOP has been observed in previous studies [40], but almost no regulation was seen in any of the two experiments performed with H_2O_2 in this study at gene expression level. This could be due to cell death and that cells changed into a senescent phenotype at higher concentrations of H_2O_2, which affects the transcriptional machinery. However, the ER is overall not sensitive to H_2O_2 stress since its oxidizing environment contribute to that the majority of cysteines in the ER already are oxidized and forms disulfide bonds, with H_2O_2 constantly created as byproduct when the disulfide bridges are created.

The upregulation observed when stimulating cells with CSE indicates that UPR can be induced through oxidative stress and that CSE is more efficient than H_2O_2 in inducing UPR. In previous studies, expression of CHOP and ATF4 were upregulated in primary small epithelial cells and HBECs during acute induction of ER stress, while long-term chronic CSE exposure led to decreased expression of CHOP. The regulation of CHOP in this study is coherent with the acute model and may be important in early induction of lung disease. However, ER stress levels changes over time, which makes it important to establish chronic ER stress models as well [63].

The stimulation with tunicamycin in the final experiment resulted in similar but lower upregulations than in previous experiment, which indicates that the choice of cell media has a large impact on cell growth and can be optimized to best fit the experiment to be performed. The ExPlus media is known to result in higher proliferation of cells than BEGM media and their usage resulted in different morphologies, which can affect transcriptional regulations. Similar results were achieved by stimulation with thapsigargin as for tunicamycin and a conclusion can thereby be drawn that these stimulus can be used to set up robust mechanistic models in primary lung cells for stimulation of UPR.

6.7 Protein readouts

The establishment of protein readouts for eIF2α and phospho-eIF2α indicates that the primary antibodies used to target these proteins worked as well as the secondary antibodies used since signals were detected. Since the signal was very strong for the higher concentration of eIF2α primary antibody, the lower antibody dose is preferred. Regarding phospho-eIF2α, the signal was low and the higher dose of

primary antibody preferred.

No protein expression could be detected for CHOP or ATF6, which might be due to that the primary antibodies chosen did not work with JESS. However, the antibody for ATF6 led to signals of ATF6 and cleaved ATF6 in traditional Western blot, which implies that the ATF6 arm is activated on protein level. Since the sample with inhibitor of the PERK pathway resulted in lower expression of cleaved ATF6, it is possible that the PERK and ATF6 pathways interact. Signaling of UPR is known to be complex with cross-talk and compensations between transcription response from the three arms [53]. Together with the fact that the UPR arms might be expressed at different levels, monitoring both gene and protein level is important to capture the complexity of UPR.

The actual investigation of protein expression with JESS from the time-course experiment performed in this study resulted in readouts for phospho-eIF2α and BiP. Thereby, both BiP and phospho-eIF2α can be used to establish protein readouts of UPR, but preferably at different timepoints and with different antibody concentrations since maximal expression levels were achieved at different timepoints and the detected signals differed in strength between proteins.

6.8 *In vitro* future analyses

To further investigate the use of thapsigargin for UPR induction, an experiment similar to the first one performed in this study could be executed to see if the levels of gene expression would increase even more, as could be seen for tunicamycin when comparing cultivation in ExPlus and BEGM media. Furthermore, a comparison of the two donors could be made, to see if the difference in UPR mostly depends on chosen donor or cell media. To examine UPR in COPD or IPF donors is also of interest since the response in healthy vs diseased cells most likely differs. Thereby, it is important to, besides from mechanistic models with healthy cells, also develop *in vitro* disease models with diseased cells for ER stress and UPR.

To continue with thapsigargin or tunicamycin, that has shown to induce UPR, different inhibitors could be tested to examine if the inhibition of certain genes in the UPR pathway contributes to decreased levels of gene expression and to higher cell viability. A suggestion for tunicamycin is then to continue with the ExPlus media and 6 h stimulation with 0.03 µg/ml since the EC_{80}-value for majority of genes and the largest fold-changes were found at this concentration and timepoint, using ExPlus media.

Finally, it is of interest to investigate chronic vs acute ER stress and UPR since cells are exposed to chronic ER stress in COPD and IPF. Chronic models could be developed by challenging cells with lower doses of stimuli during a longer time-course and then compare with that of more acute ER stress.

6.9 Comparison *in silico* and *in vitro* models

To directly compare results from *in silico* and *in vitro* models is a challenge, but the results from one of these domains can be used to confirm or further validate the results in the other. One of the challenges with *in vitro* experiments is to capture the complexity of the cells in their natural environment. When investigating chronic diseases, such as COPD and IPF, it is important to both look at acute and chronic responses and to capture the cellular response over time. One example from this study is the gene expression of XBP1 and sXBP1 in the IRE1 pathway, where there is a shift in response between 6 and 24 h stimulation with tunicamycin that indicate a cellular adaption to the stress exposure where unspliced XBP1 has been refilled at the 24 h timepoint. The *in silico* analysis can be of great use to investigate such changes over time and to capture the complexity of respiratory diseases by enabling analysis of diseased tissue at single cell level.

The results from single cell analysis could be used to create hypotheses about biological processes in the diseased state that then can be further examined *in vitro* and *in vivo*. One example from this study is the expression of PPP1R15A, which was downregulated in the DE analysis for the majority of epithelial cell subtypes investigated and then analysed in the *in vitro* experiments to validate UPR upon induction of ER stress with different stimuli. However, it was upregulated at gene level *in vitro*, while previous literature has resulted in downregulation of PPP1R15A, similar to the *in silico* result in this study [16]. The difference might be due to that the complexity of cellular processes *in vivo* is not always reflected with *in vitro* models. For example, interactions between different cell types are complex and hard to capture *in vitro*. Furthermore, a challenge with in *in silico* models is that extracted tissues are from sever disease state developed during a long time, which do not directly reflect the early phase of disease, where the possibility of reversing the disease state is maximal.

7
Conclusion

Based on the experimental results of UPR gene regulations with different stimuli, thapsigargin and tunicamycin could be used to create robust mechanistic models for modulation of ER stress and UPR, especially for CHOP, BiP, sXBP1 and PPP1R15A. However, oxidative stress caused by H_2O_2 and CSE lead to small induction of the UPR pathway or none at all and are therefore, based on this study, not optimal to use to set up models for modulation of UPR.

Regarding protein readouts and the use of different Western blot methods, several conclusions can be made. First, that tunicamycin can be used to induce the expression of proteins in the UPR, which was accomplished for eIF2α, phospho-eIF2α, BiP and ATF6. Second, that automated Western blot with JESS can be used with the tested primary antibodies for eIF2α, phospho-eIF2α and BiP, but not for CHOP and ATF6. Third, that traditional Western blot can be used with the tested primary antibody for ATF6, indicating that different methods of Western blot can be useful depending on tested primary antibodies.

Since club cells and aberrant basaloid cells were upregulated in IPF and the majority of cells from these subgroups were IPF cells, they might be important in the development of IPF and be affected by ER stress and induction of UPR in patients. Furthermore, AT1 and AT2 cells were mostly found in healthy patients, which is consistent with previous studies, and the regeneration of these could be used to improve the health of IPF patients.

Regarding the regeneration of the aberrant basaloid cells in Reyfman, the created subgroup did show similar characteristics, strengthening the existence of the subgroup and its reproducibility across datasets.

Both *in vitro* and *in silico* models are of high importance when investigating biological processes of disease. Together, they can be used to understand disease mechanisms in specific cellular subsets and enable identification of new drug targets. scRNA-seq data can be used to establish hypothesis around biological pathways important in disease that thereafter can be further investigated with *in vitro* models.

Bibliography

[1] Han M. Labaki W. "Chronic respiratory diseases: a global view". In: *The Lancet Respiratory Medicine* 8.6 (June 2020), pp. 531–533. ISSN: 2213-2600. DOI: https://doi.org/10.1016/S2213-2600(20)30157-0.

[2] Li X. et al. "Single cell RNA sequencing identifies IGFBP5 and QKI as ciliated epithelial cell genes associated with severe COPD". In: *Respiratory research* 22.1 (Dec. 2021). ISSN: 1465-993X. DOI: 10.1186/S12931-021-01675-2.

[3] World Health Organisation. *Chronic obstructive pulmonary disease (COPD)*. June 2021. DOI: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd). (Visited on 01/30/2022).

[4] Bradley K. L. et al. *Role of unfolded proteins in lung disease*. Jan. 2021. DOI: https://doi.org/10.1136/thoraxjnl-2019-213738.

[5] Chen A. C.H., Burr L., and McGuckin M. A. "Oxidative and endoplasmic reticulum stress in respiratory disease". In: *Clinical & Translational Immunology* 7.6 (2018). ISSN: 20500068. DOI: https://doi.org/10.1002/CTI2.1019.

[6] Burman A., Tanjore H., and Blackwell T. "Endoplasmic reticulum stress in pulmonary fibrosis". In: *Matrix Biology* 68-69 (Aug. 2018), pp. 355–365. ISSN: 0945-053X. DOI: https://doi.org/10.1016/j.matbio.2018.03.015.

[7] Hewitt R.J. and Lloyd C.M. "Regulation of immune responses by the airway epithelial cell landscape". In: *Nature Reviews* 21.6 (Jan. 2021), pp. 347–362. ISSN: 1474-1741. DOI: https://doi.org/10.1038/s41577-020-00477-9.

[8] Davis J. and Wypych T. "Cellular and functional heterogeneity of the airway epithelium". In: *Mucosal Immunology* 14.5 (Feb. 2021), pp. 978–990. ISSN: 1935-3456. DOI: https://doi.org/10.1038/s41385-020-00370-7.

[9] Yaqub N. et al. "Recent advances in human respiratory epithelium models for drug discovery". In: *Biotechnology Advances* (Aug. 2021). ISSN: 0734-9750. DOI: https://doi.org/10.1016/j.biotechadv.2021.107832.

[10] Habermann A.C. et al. "Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis". In: *Science Advances* 6.28 (June 2020), p. 1972. ISSN: 23752548. DOI: 10.1126/sciadv.aba1972.

[11] Wang Y., Tang Z., and Huang H. "Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.10 (Jan. 2018), pp. 2407–2412. ISSN: 10916490. DOI: https://doi.org/10.1073/pnas.1719474115.

[12] Pluquet O., Pourtier A., and Abbadie C. "The unfolded protein response and cellular senescence. A Review in the Theme: Cellular Mechanisms of Endoplasmic Reticulum Stress Signaling in Health and Disease". In: *American Journal of Physiology - Cell Physiology* 308.6 (Mar. 2015), pp. 415–425. ISSN: 15221563. DOI: https://doi.org/10.1152/ajpcell.00334.2014.

[13] Cao S.S. and Kaufman R.J. "Endoplasmic Reticulum Stress and Oxidative Stress in Cell Fate Decision and Human Disease". In: *antioxidants and redox signaling* 21.3 (July 2014), pp. 396–413. DOI: https://doi.org/10.1089/ars.2014.5851.

[14] Dickens J. et al. "Pulmonary endoplasmic reticulum stress-scars, smoke, and suffocation". In: *The FEBS Journal* 286.2 (Jan. 2019), pp. 322–341. ISSN: 1742-4658. DOI: https://doi.org/10.1111/febs.14381.

[15] Cao S. and Kaufman R. "Unfolded protein response". In: *Current Biology* 22.16 (Aug. 2012), R622–R626. ISSN: 0960-9822. DOI: https://doi.org/10.1016/J.CUB.2012.07.004.

[16] Monkley S. et al. "Sensitization of the UPR by loss of PPP1R15A promotes fibrosis and senescence in IPF". In: *Scientific Reports* 11.1 (June 2021). ISSN: 2045-2322. DOI: https://doi.org/10.1038/s41598-021-00769-7.

[17] Donnely N. et al. "The eIF2 kinases: their structures and functions". In: *Cellular and Molecular Life Sciences* 70 (Jan. 2013), pp. 3493–3511. DOI: https://doi.org/10.1007/s00018-012-1252-6.

[18] Maurel M. et al. "Getting RIDD of RNA: IRE1 in cell fate regulation". In: *ScienceDirect* 39.5 (June 2014), pp. 245–254. ISSN: 0968-0004. DOI: https://doi.org/10.1016/j.tibs.2014.02.008.

[19] Rodrigues S., Cunha C., and Soares G. "Mechanisms, Pathophysiology and Currently Proposed Treatments of Chronic Obstructive Pulmonary Disease". In: *Pharmaceuticals* 14.10 (Sept. 2021). ISSN: 1424-8247. DOI: https://doi.org/10.3390/ph14100979.

[20] Aghaei M. et al. "The ER stress/UPR Axis in Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis". In: *Life* 11.1 (Dec. 2020). ISSN: 23752548. DOI: http://doi.org/10.3390/life11010001.

[21] Hassan T., Carroll T.P., and Buckley P.G. "miR-199A-5P silencing regulates the unfolded protein response in chronic obstructive pulmonary disease and 1-antitrypsin deficiency". In: *Am J Respir Crit Care Med* 189.3 (Nov. 2014), pp. 263–273. DOI: https://doi.org/10.1164/rccm.201306-1151OC.

[22] Becker E. et al. “Bronchial gene expression signature associated with rate of subsequent FEV1 decline in individuals with and at risk of COPD”. In: *Thorax* 77.1 (June 2021), pp. 31–39. ISSN: 0040-6376. DOI: http://10.1136/thoraxjnl-2019-214476.

[23] Kaunisto J., Salomaa E., and Hodgson U. “Demographics and survival of patients with idiopathic pulmonary fibrosis in the FinnishIPF registry”. In: *ERJ Open Research* 5.3 (July 2019). ISSN: 00170-2018. DOI: https://doi.org/10.1183/23120541.00170-2018.

[24] Lederer D. and Martinez F. “Idiopathic Pulmonary Fibrosis”. In: *The new england journal of medicine* 378.19 (June 2018), pp. 1811–1823. ISSN: 0028-4793. DOI: https://doi.org/10.1056/NEJMra1705751.

[25] Salton F., Volpe MC., and Confalonieri M. “Epithelial–Mesenchymal Transition in the Pathogenesis of Idiopathic Pulmonary Fibrosis”. In: *Medicina* 55.4 (Apr. 2019). DOI: https://10.3390/medicina55040083.

[26] Martinez F. et al. “Idiopathic pulmonary fibrosis”. In: *Nature reviews disease primers* 3.17074 (Oct. 2017). DOI: https://doi.org/10.1038/nrdp.2017.74.

[27] Katzen R. et al. “An SFTPC BRICHOS mutant links epithelial ER stress and spontaneous lung fibrosis”. In: *Journal of Clinical Investigation* 4.6 (Feb. 2019). ISSN: 0021-9738. DOI: https://doi.org/10.1172/jci.insight.126125.

[28] Chen G. et al. “XBP1S Regulates MUC5B in a Promoter Variant–Dependent Pathway in Idiopathic Pulmonary Fibrosis Airway Epithelia”. In: *American Journal of Respiratory and Critical Care Medicine* 200.2 (Apr. 2019), pp. 220–234. DOI: https://doi.org/10.1164/rccm.201810-1972OC.

[29] Korfei M., Von der Beck D., and Henneke I. “Comparative proteome analysis of lung tissue from patients with idiopathic pulmonary fibrosis (IPF), non-specific interstitial pneumonia (NSIP) and organ donors”. In: *Journal of proteomics* 85 (June 2013), pp. 109–128. DOI: https://doi.org/10.1016/j.jprot.2013.04.033.

[30] Zhong Q., Zhou B., and Ann D. “Role of endoplasmic reticulum stress in epithelial-mesenchymal transition of alveolar epithelial cells: effects of misfolded surfactant protein”. In: *Resp Cell Mol* 45.3 (Dec. 2011), pp. 498–509. DOI: https://doi.org/10.1165/rcmb.2010-0347OC.

[31] Mo X-T., Zhou W-C., and Cui W-H. “Inositol-requiring protein 1 - X-box-binding protein 1 pathway promotes epithelial-mesenchymal transition via mediating snail expression in pulmonary fibrosis”. In: *Int J Biochem Cell Biol* 153.3 (Dec. 2018), pp. 1221–1228. DOI: https://doi.org/10.1016/j.chest.2017.07.034.

[32] Kolodziejczyk A. et al. “The Technology and Biology of Single-Cell RNA Sequencing”. In: *Molecular Cell* 58.4 (May 2015), pp. 610–620. DOI: https://doi.org/10.1016/j.molcel.2015.04.005.

[33] Whang B., Lee J.H., and Bang D. "Single-cell RNA sequencing technologies and bioinformatics pipelines". In: *Experimental and Molecular Medicine* 50.8 (Aug. 2018), pp. 1–14. DOI: https://doi.org/10.1038/s12276-018-0071-8.

[34] Kolodziejczyk A. et al. "The Technology and Biology of Single-Cell RNA Sequencing". In: *Molecular Cell* 58.4 (May 2015), pp. 610–620. ISSN: 1097-2765. DOI: https://doi.org/10.1016/j.molcel.2015.04.005.

[35] Adams T. et al. "Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis". In: *Science Advances* 6.28 (July 2020). ISSN: 23752548. DOI: https://doi.org/10.1126/sciadv.aba1983.

[36] Van den Berge K., Roux de Bézieux H., and Street K. "Trajectory-based differential expression analysis for single-cell sequencing data". In: *Nature Communications* 11.1201 (Mar. 2020). DOI: https://doi.org/10.1038/s41467-020-14766-3.

[37] Sigmaaldrich. *Primary Cell Culture Basics*. 2020. URL: https://www.sigmaaldrich.com/SE/en/technical-documents/technical-article/cell-culture-and-cell-culture-analysis/primary-cell-culture/primary-cell-culture (visited on 04/13/2022).

[38] Abdullah A. et al. "Modeling Acute ER stress in vivo and in vitro". In: *Shock* 47.4 (Apr. 2018), pp. 506–513. DOI: https://10.1097/SHK.0000000000000759.

[39] Oslowski C. and Urano F. "Measuring ER stress and the unfolded protein response using mammalian tissue culture system". In: *Methods Enzymology* 490 (July 2013), pp. 71–92. DOI: https://10.1016/B978-0-12-385114-7.00004-0.

[40] Matos L., Gouveia A., and Almeida H. "ER Stress Response in Human Cellular Models of Senescence". In: *National Library of Medicine* 70.8 (Aug. 2014), pp. 924–935. DOI: https://10.1093/gerona/glu129.

[41] Tagawa Y., Hiramatsu N., and Kato H. "Induction of CCAAT/enhancer-binding protein-homologous protein by cigarette smoke through the superoxide anion-triggered PERK-eIF2 pathway". In: *Toxicology* 287.1-3 (Sept. 2011), pp. 105–112. DOI: https://doi.org/10.1016/j.tox.2011.06.005.

[42] Jorgensen E. et al. "Cigarette smoke induces endoplasmic reticulum stress and the unfolded protein response in normal and malignant human lung cells". In: *BMC cancer* 11.8 (Aug. 2008), p. 229. DOI: https://10.1186/1471-2407-8-229.

[43] ThermoFisher Scientific. *Real-time PCR basics*. Nov. 2019. URL: https://www.thermofisher.com/se/en/home/life-science/pcr/real-time-pcr/real-time-pcr-learning-center/real-time-pcr-basics/how-taqman-assays-work.html (visited on 02/05/2022).

[44] ThermoFisher Scientific. *Overview of western blotting*. URL: https://www.thermofisher.com/se/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-western-blotting.html (visited on 03/19/2022).

[45] Vuong H. et al. “Venice: A New Algorithm for Finding Marker Genes in Single-Cell Transcriptomic Data”. In: (Nov. 2020). DOI: https://doi.org/10.1101/2020.11.16.384479.

[46] Le T et al. “BBrowser: Making single-cell data easily accessible”. In: *BioRxiv* (Jan. 2020). DOI: https://doi.org/10.1101/2020.12.11.414136.

[47] Reyfman P. et al. “Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis”. In: *American Journal of Respiratory and Critical Care Medicine* 199.12 (June 2019), pp. 1517–1536. DOI: https://doi.org/10.1164/rccm.201712-2410OConDecember15.

[48] Krämer A. et al. “Causal analysis approaches in Ingenuity Pathway Analysis”. In: *Bioinformatics* 30.4 (Dec. 2013), pp. 523–530. ISSN: 1367-4803. DOI: https://doi.org/10.1093/bioinformatics/btt703.

[49] Dennis G. et al. “DAVID: Database for Annotation, Visualization, and Integrated Discovery”. In: *Genome Biology* 4.9 (Aug. 2003). DOI: https://doi.org/10.1186/gb-2003-4-9-r60.

[50] Schadewijk A., Wout E., and Stolk J. “A quantitative method for detection of spliced X-box binding protein-1 (XBP1) mRNA as a measure of endoplasmic reticulum (ER) stress”. In: *Cell stress and chaperones* 17 (Oct. 2011), pp. 275–279. DOI: https://doi.org/10.1007/s12192-011-0306-2.

[51] Kemp P., Sugar R., and Jackson A. “Nucleotide-mediated mucin secretion from differentiated human bronchial epithelial cells”. In: *Respiratory Cell Molecular Biology* 31.4 (Oct. 2004), pp. 446–455. DOI: https://doi.org/10.1165/rcmb.2003-0211OC.

[52] *Jess Your Complete Protein Analysis Solution*. URL: https://www.proteinsimple.com/jess.html (visited on 04/10/2022).

[53] Aeuyung V. et al. “IRE1 drives lung epithelial progenitor dysfunction to establish a niche for pulmonary fibrosis”. In: *American journal of physiology* 322.4 (Apr. 2022), pp. L564–L580. DOI: https://doi.org/10.1152/ajplung.00408.2021.

[54] Parimon T. et al. “Alveolar Epithelial Type II Cells as Drivers of Lung Fibrosis in Idiopathic Pulmonary Fibrosis”. In: 21.7 (Apr. 2020), p. 2269. DOI: https://10.3390/ijms21072269.

[55] Katzen J. and Beers M. “Contributions of alveolar epithelial cell quality control to pulmonary fibrosis”. In: *The Journal of Clinical Investigation* 130.10 (Sept. 2020), pp. 5088–5099. DOI: https://doi.org/10.1172/JCI139519.

[56] Tang X. et al. “EMC3 coordinates surfactant protein and lipid homeostasis required for respiration”. In: *The Journal of Clinical Investigation* 127.12 (Oct. 2017), pp. 4314–4325. DOI: https://doi.org/10.1172/JCI94152.

[57] Borok Z. et al. “Grp78 Loss in Epithelial Progenitors Reveals an Age-linked Role for Endoplasmic Reticulum Stress in Pulmonary Fibrosis”. In: *American Journal of Respiratory and Critical* 201.2 (Nov. 2019). DOI: https://doi.org/10.1164/rccm.201902-0451OC.

[58] Sauler M., McDonough J., and Rosas I. “Characterization of the COPD alveolar niche using single-cell RNA sequencing”. In: *Nature communications* 13.494 (Jan. 2022). DOI: https://doi.org/10.1101/705244.

[59] Zuo W. et al. “Dysregulation of club cell biology in idiopathic pulmonary fibrosis”. In: *PLoS One* 15.9 (Sept. 2020). DOI: https://doi.org/10.1371/journal.pone.0237529.

[60] Moore K. and Hollien J. “The Unfolded Protein Response in Secretory Cell Function”. In: *Annual Review of Genetics* 46 (Dec. 2012), pp. 165–183. DOI: https://doi.org/10.1146/annurev-genet-110711-155644.

[61] Xu Y. et al. “Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis”. In: *The Journal of Clinical Investigation* 1.20 (Dec. 2016). DOI: https://doi.org/10.1172/jci.insight.90558.

[62] Strunz M. et al. “Longitudinal single cell transcriptomics reveals Krt8+ alveolar epithelial progenitors in lung regeneration”. In: *bioRxiv* (July 2019). DOI: https://doi.org/10.1101/705244.

[63] Geragthy P., Wallace A., and M D’Armiento J. “Induction of the unfolded protein response by cigarette smoke is primarily an activating transcription factor 4-C/EBP homologous protein mediated process”. In: *Int J Chron Obstruct Pulmon Dis* 6 (June 2011), pp. 309–319. DOI: https://doi.org/10.2147/COPD.S19599.

A

Appendix - Results

In this chapter, figures and tables supporting the results in the result chapter of the report can be found. Results from the bioninformatics part is presented first, followed by results from the experimental part of the project.

A.1 Unfolded protein response gene signature

The UPR gene signature collected from IPA can be seen in Table A.1.

Table A.1: UPR gene signature from IPA.

Gene symbol	UPR pathway
All	HSPA5 (BiP)
PERK	EIF2AK3, DNAJC3, NFE2L2, EIF2A, ATF4, BCL2, CALR, CANX, DDIT3 (CHOP), HSP90B1, PPP1R15A, PPARG
ATF6	ATF6, ATF6B, MBTPS1, MBTPS2
IRE1	ERN1, ERN2, XBP1, EDEM1
ERAD	DNAJB9, HSPA1A, HSPA1B, HSPA4, HSPH1 OS9, SEL1L, SYVN1, UBXN4, VCP
ER protein folding	ERO1B, P4HB

A.2 Cell distribution across epithelial cell subtypes

The analysed epithelial cell subtypes in the three datasets and the number of cells belonging to diseased and control state was compared and can be seen in Table A.2.

Table A.2: Analysed epithelial cell subtypes and number of IPF and control cells from the three datasets by Habermann, Reyfman and Adams.

	Habermann		Reyfman		Adams		
Cell subtype	**IPF**	**control**	**IPF**	**control**	**IPF**	**control**	**COPD**
Abberant basaloid					448	33	
AT1	96	472	96	1013	176	375	502
AT2	3225	4393	2201	20607	496	810	2655
Basal	1786	65			1472	101	114
Ciliated	11857	2119	416	512	7464	912	1144
Club			372	757	1855	470	226
Goblet					1044	132	101
Ionocyte					23		2
KRT5-/KRT17+	374	13					
Mesothelial					233	288	46
MUC5B+	1940	281					
MUCAC+	75	20					
Neuroendocrine					12	33	17
Proliferating epithelial	222	79					
SCGB3A2+	2824	97					
SCGB3A2+/SCGB1A1	854	222					
Transitional AT2	581	443					

A.3 T-SNE images of aberrant basaloid cells

T-SNE images of the aberrant basaloid cell gene signature applied on epithelial cells from the three datasets can be seen in this section. The expression of these genes in the cells are visualised in the first image, red indicating higher expression and lighter color indicating lower expression. The cells with highest expression of these genes were analysed.

The selected aberrant basaloid gene signature applied on the data from Habermann is visualised with T-SNE images in Figure A.1. The selection of cells resulted in a sub-cluster of 497 cells where 97 % are from IPF and 70 % of the cells are from the aberrant basaloid cell subgroup annotated in Habermann.

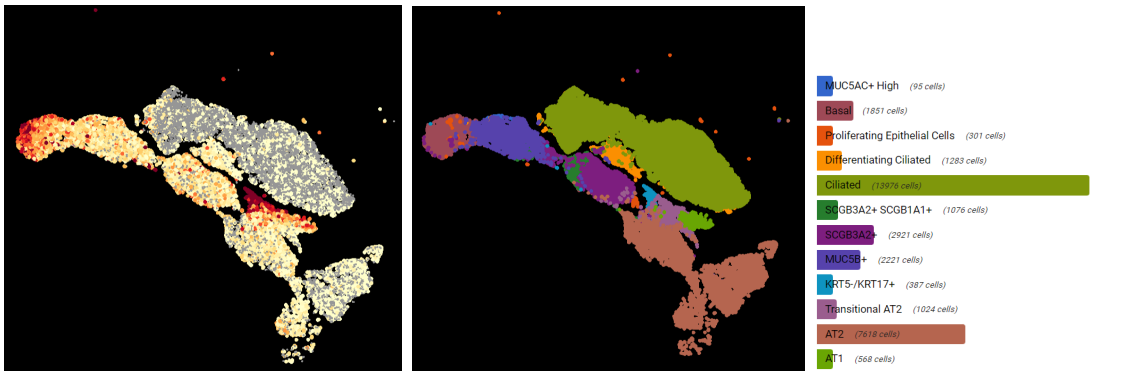

Figure A.1: T-SNE images of epithelial cells from the Habermann dataset, coloured by expression (red to yellow) and annotated cell subgroups. From [46].

The selected aberrant basaloid gene signature applied on the data from Adams is visualised with T-SNE images in Figure A.2. The selection of cells resulted in a sub-cluster where 95 % of the cells are from IPF and 80 % are annotated as aberrant basaloid cells or basal cells.

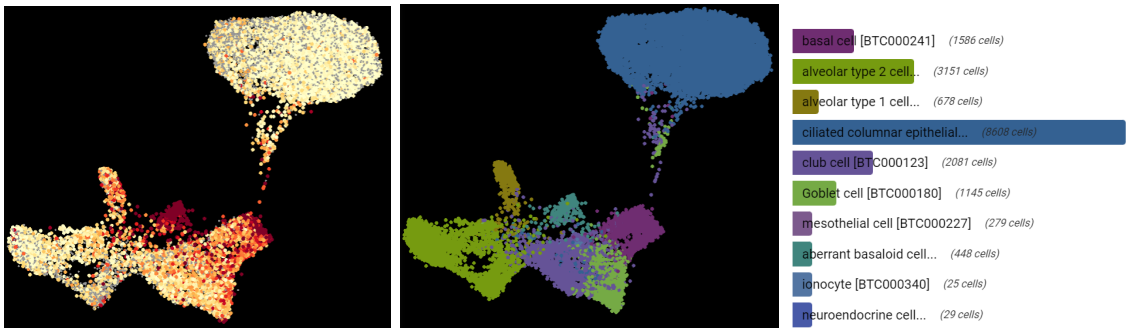

Figure A.2: T-SNE images of epithelial cells from the data by Adams, coloured by expression (red to yellow) and annotated cell subtype. From [46].

The created aberrant basaloid subgroup in Reyfman is visualised with T-SNE images in Figure A.3. The selection of cells resulted in 209 cells being selected where 82 % of the cells was from IPF and the majority was defined as AT1 or AT2 cells.

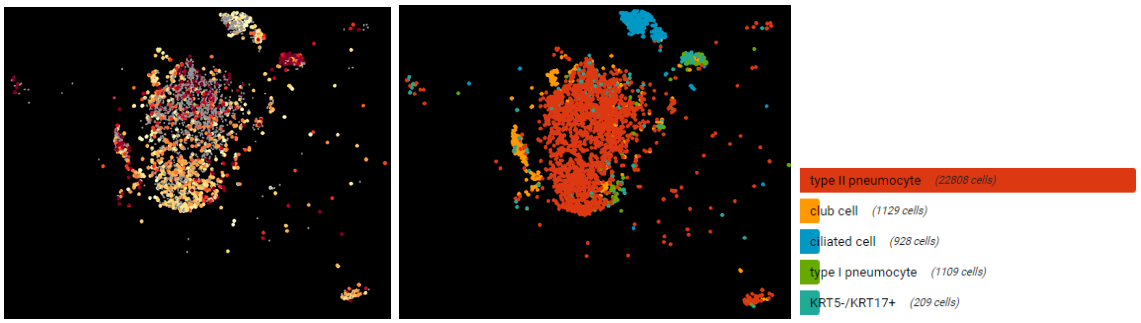

Figure A.3: T-SNE images of epithelial cells from the Reyfman dataset, coloured by expression (red to yellow) and annotated cell subtype. From [46].

A.4 Marker genes for the aberrant basaloid cells

In Table A.3, 28 of the most upregulated genes for the aberrant basaloid cells in Habermann and Adams, and the created aberrant basaloid subgroup in Reyfman can be observed.

Table A.3: Log2 Fold changes of marker genes for the aberrant basaloid cells in Habermann and Adams, and the created aberrant basaloid cell subgroup in Reyfman.

Gene	Habermann	Reyfman	Adams
CALD1	2.00	0.58	2.91
CCND2	1.86	1.00	2.46
CDH2	1.375	-	2.19
CDKN2A	1.65	-	2.26
COL1A1	2.78	0.82	2.68
DST	1.46	0.80	2.72
FN1	3.03	0.89	3.30
GDF15	2.83	-	2.49
IFI27	1.73	1.35	2.85
IL32	2.54	1.25	3.39
ITGB6	2.24	0.61	2.55
KRT17	3.43	1.24	3.43
KRT18	1.41	0.92	2.62
KRT19	-	0.89	2.28
KRT7	2.89	1.63	3.82
KRT8	-	0.71	2.61
MDK	2.03	-	2.56
MMP7	2.16	0.78	3.54
PMEPA1	1.67	0.66	2.40
PTGS2	3.32	1.00	2.22
S100A10	1.80	1.25	2.52
SFN	2.20	0.54	2.22
SOX4	2.22	0.80	2.16
TM4SF1	2.65	1.35	2.74
TMSB10	2.22	1.07	3.33
TMSB4X	2.41	1.06	3.83
TPM1	2.30	0.84	2.51
TRAM1	1.82	0.85	2.36

A.5 Cell images after 3, 6 and 24 h stimulation

Cell images from cells stimulated with tunicamycin in experiment 1 can be seen in Figure A.4.

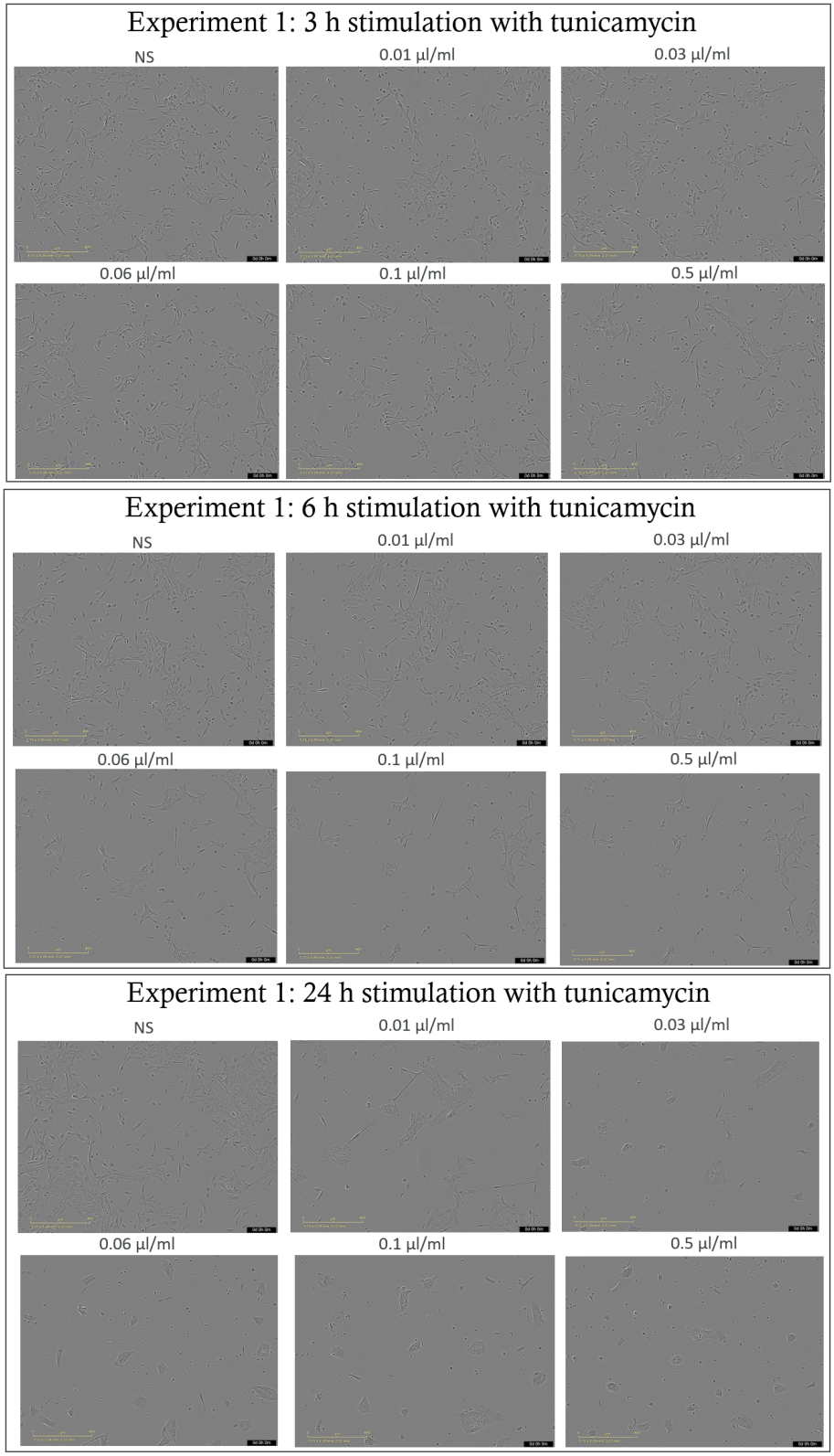

Figure A.4: Cell images of cells stimulated with tunicamycin in experiment 1 after 3, 6 and 24 h stimulation.

Cell images from cells stimulated with H_2O_2 in experiment 2 can be seen in Figure A.5.

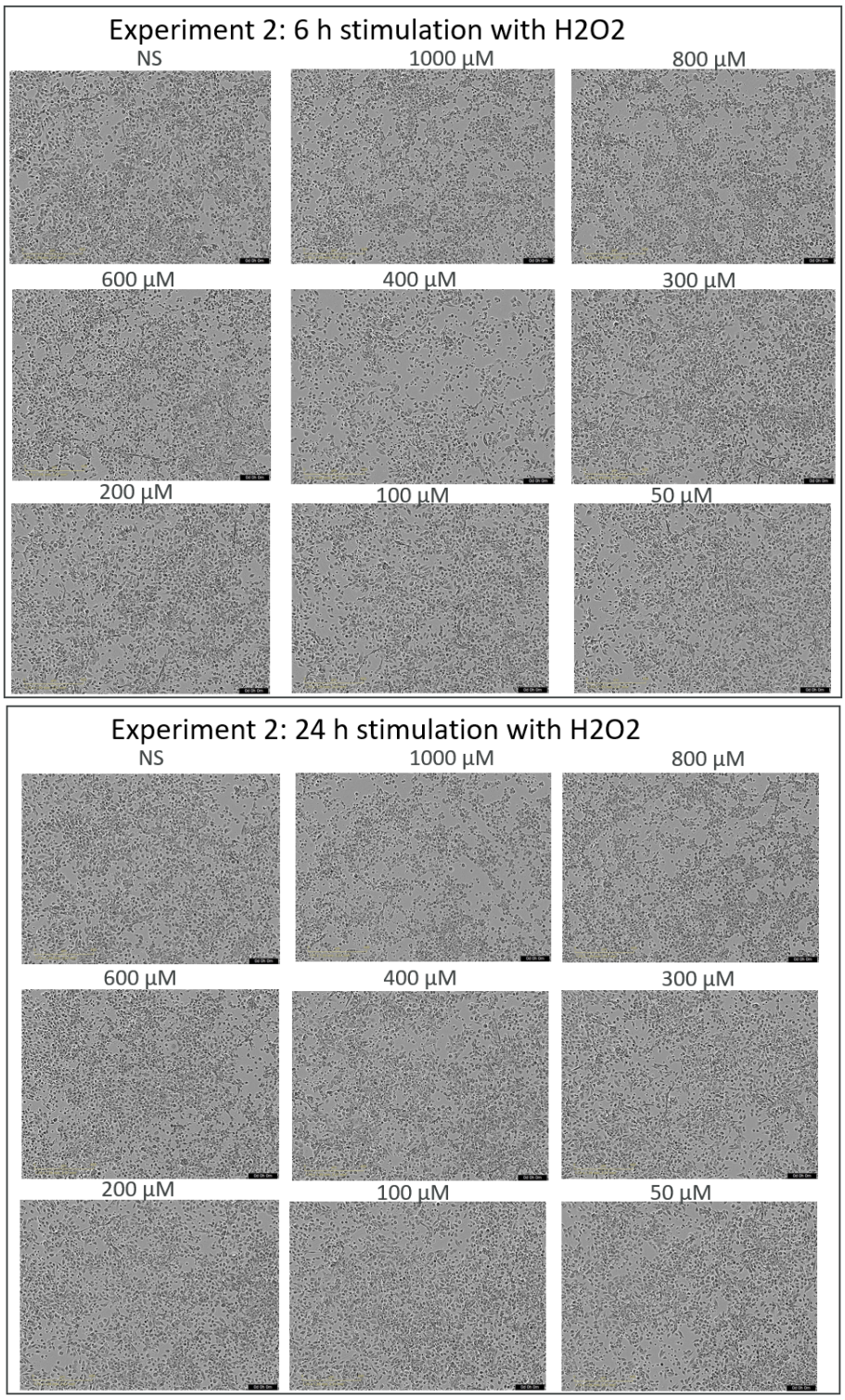

Figure A.5: Cell images of cells stimulated with H_2O_2 in experiment 2 after 6 and 24 h stimulation.

Cell images from cells stimulated with H_2O_2, thapsigargin, tunicamycin and CSE in experiment 3 can be seen in Figure A.6.

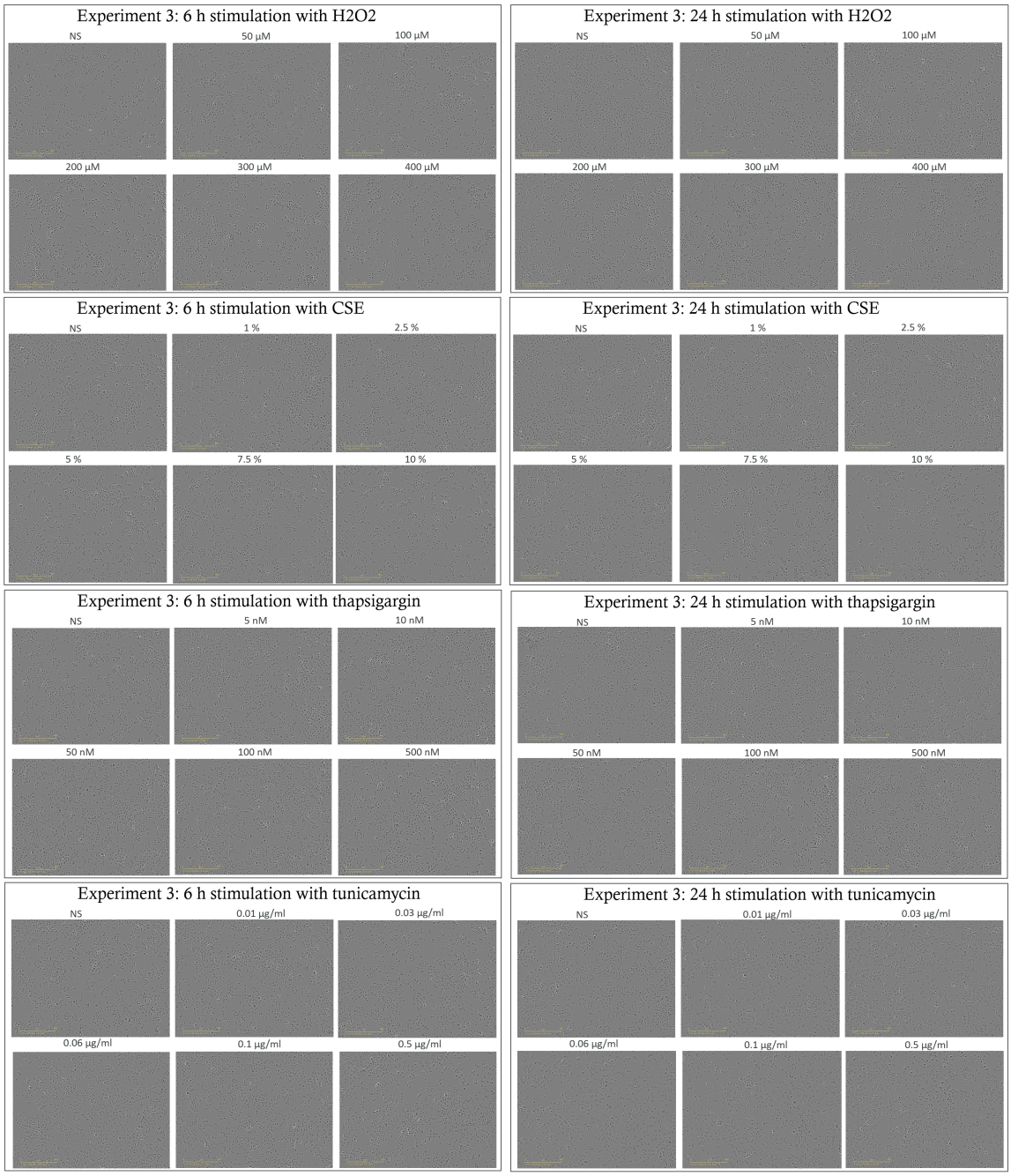

Figure A.6: Cell images of cells stimulated with H_2O_2, CSE, thapsigargin and tunicamycin in experiment 3 after 6 and 24 h stimulation.

A.6 Ct-values for housekeeping genes

The Ct-values for the housekeeping genes GAPDH and RPLP0 from the three experiments can be seen in Figure A.7. The Ct-values for experiment 1 and 2 can be observed in the first row, while Ct-values for for experiment 3 can be seen in the second row.

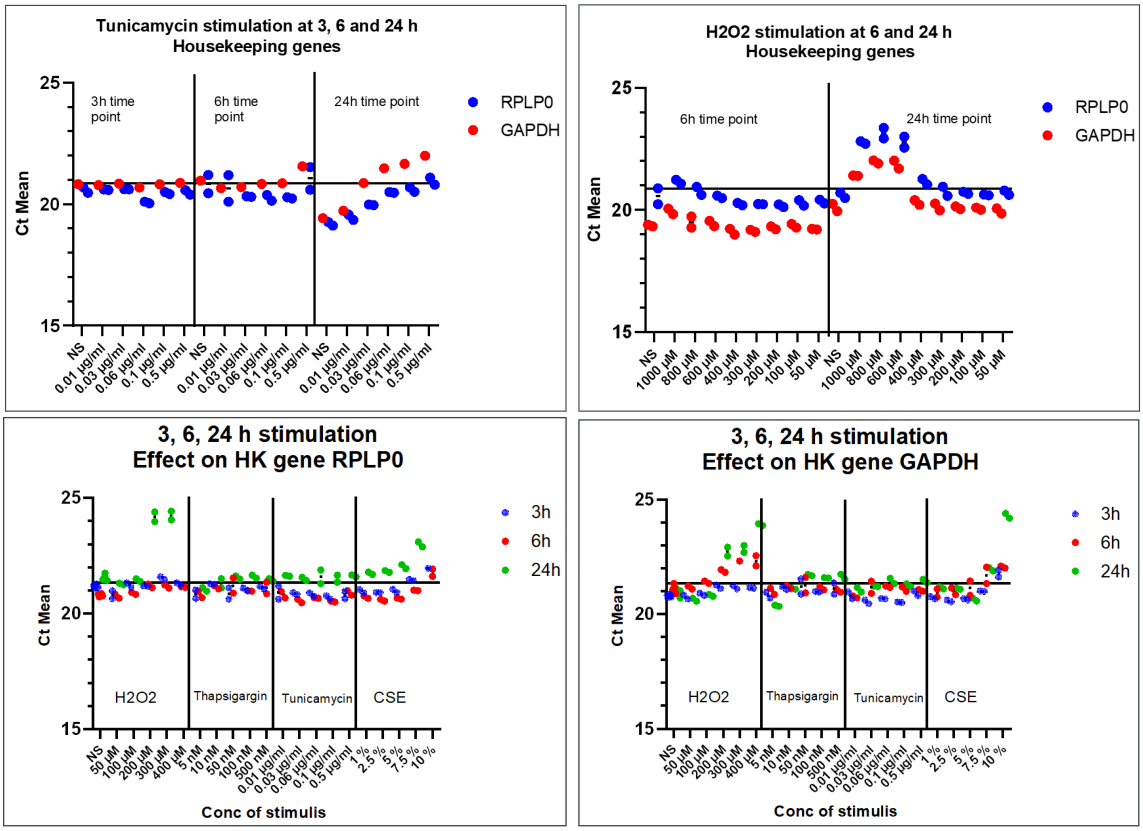

Figure A.7: Ct-values of housekeeping genes from the three performed experiments. Experiment 1 with tunicamycin and experiment 2 with H_2O_2 in the first row, and experiment 3 with H_2O_2, thapsigargin, tunicamycin and cigarette smoke extract (CSE) in the second row.

A.7 Western blot

Images of the result from the automated Western blot runs can be seen in Figure A.8.

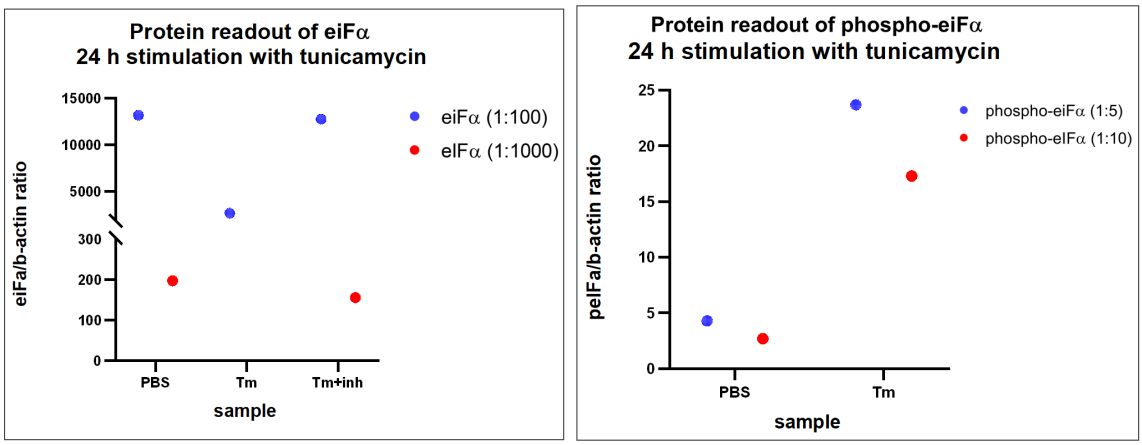

Figure A.8: UPR protein expression readouts with automated Western blot for different concentrations of eIF2α and phospho-eIF2α primary antibodies. Readouts for non-stimulated cells with PBS and tunicamycin stimulated cells with and without inhibitor of PERK can be seen.

Protein readouts with automated Western blot for different concentrations of eIF2α and phospho-eIF2α primary antibodies can be seen in Figure A.9. Samples with cells stimulated with tunicamycin with or without an inhibitor of the PERK pathway were used.

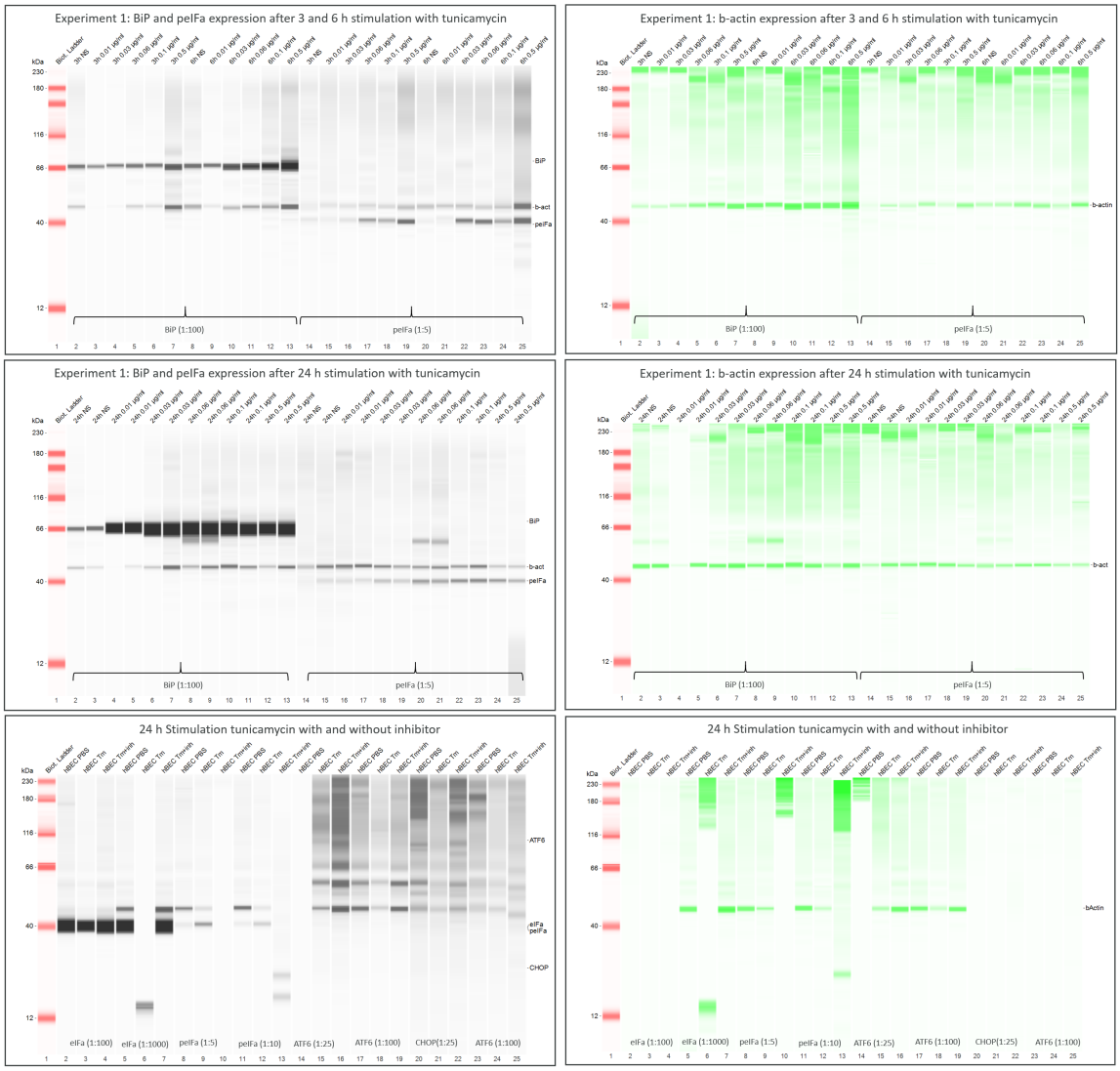

Figure A.9: Protein expression of UPR proteins after 3, 6 and 24 h stimulation. Protein readouts for BiP, phospho-eIF2α and using samples from experiment 1 can be seen in the top four figures and protein readouts for eIF2α and phospho-eIF2α using samples from another experiment can be seen in the last figure together with an attempt to measure ATF6 and CHOP, which did not succeed.

The protein readout of β-actin derived from rabbit after unsuccessful Western blot run for CHOP, using samples with cells stimulated for 24 h with and without inhibitor of the PERK pathway, can be seen in Figure A.10.

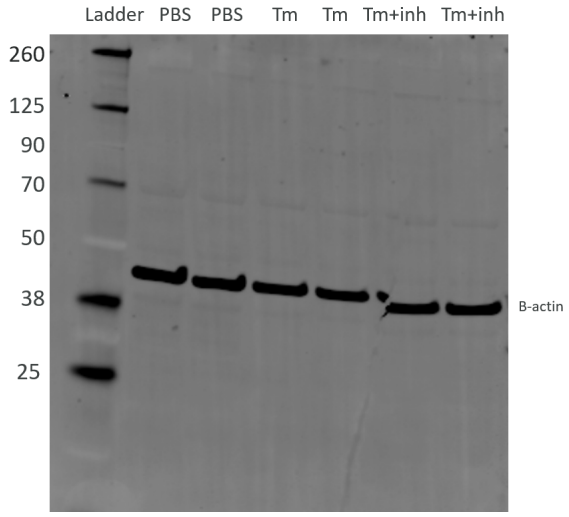

Figure A.10: Detection of β-actin rabbit after unsuccesful traditional Western blot run for detection of CHOP. Samples are from cells stimulated with tunicamycin for 24 h with and without inhibitor of the PERK pathway.

B

Appendix - Protocols

Protocols for the experimental part of the project are presented in this chapter. Protocols regarding cell cultivation, RNA purification, cDNA synthesis, qPCR, protein quantification, traditional Western blot and automated Western blot are provided.

B.1 Cell culture

The following cell culture protocol was used in experiment 1. For experiment 2 and 3, the same protocol was followed as for the 96 well plates explained below, but with different experimental layouts and stimuli, as can be seen in Table 4.2 in the materials and methods chapter.

Day 1

1. Heat medium in 37.5 °C prior to usage.
2. Thaw vials with cells in 37 °C water bath.
3. Add cells to 2x10 mL pre-warmed medium.
4. Centrifuge at 300 x g for 10 min.
5. Resuspend in 8 mL medium.
6. Add 200 µL to eppendorf-tube and count cells in Nucleocounter.
7. Dilute cells in medium to $1*10^5$ cells/mL.
8. Add 100 µL to three 96 well plates (10 000 cells/well, 12 wells in total) and 300 µL to three 48 well plates (30 000 cells/well, 12 wells in total).
9. Add 200 µL PBS to surrounding wells.
10. Incubate cells over night at 37.5 °C and with 5 % CO_2.

Day 2 - 96 well plates

1. Remove medium and add 100 µL fresh medium.
2. Add 10 µL tunicamycin of each concentration (final concentrations: 0.01, 0.03, 0.06, 0.1 and 0.5 µg/mL) in duplicates.
3. Harvest plates at 3 h, 6 h, 24 h by the following steps:

- Aspirate the media carefully.
- Wash cells in 200 µL of PBS and lyse cells in 150 µL of RLT Lysis Plus buffer.
- Quickly transfer cell-lysate plates to -80 °C freezer awaiting sample processing.

Day 2 - 48 well plates

1. Remove medium and add 200 µL fresh medium.
2. Add 20 µL tunicamycin of each concentration (final concentrations: 0.01, 0.03, 0.06, 0.1 and 0.5 µg/mL) in duplicates.
3. Harvest plates at 3 h, 6 h, 24 h by the following steps:
   - Aspirate the media carefully.
   - Wash cells in 300 µL of PBS two times and add 25 µL lysis buffer/well.
   - Keep plates on ice for 5 min and scrape cells with a pipette tip.
   - Shake plates for 30 min in a cold room
   - Transfer plates to -80 °C freezer awaiting sample processing.

The lysis mixture used for the lysis of cells in the 48 well plates was prepared by adding 850 µL M-PER Lysis Buffer, 50 µL DNAse I (50 µg/mL) and 100 µL of a mixture with tablets of 10X cOmplete Mini protease inhibitor (Roche, 11836170001) and 10X PhosSTOP (Roche, 4906837001) dissolved in dH_2O.

B.2 Gene expression analysis

B.2.1 RNA purification

The RNA purification was performed with the RNeasy 96 kit and on 150 µL lysates.

1. Place an RNeasy 96 plate on top of an S-Block. Mark the plate for later identification.
2. Add 1 volume (150 µL) of 70 % ethanol to each well of the assay plate. Mix well by pipetting up and down 3X.
3. Transfer the samples (300 µL) to the wells of the RNeasy 96 plate.
4. Seal the RNeasy 96 plate with an AirPore tape sheet
5. Place plate in vacuum pump before centrifugation to empty wells.
6. Centrifuge at 6000 rpm (5600 x g) for 4 min at 20-25 °C.
7. Empty the S-block and remove the AirPore tape sheet.
8. Add 800 µL Buffer RW1 to each well of the RNeasy 96 plate, and seal the plate with a new AirPore tape sheet.
9. Place plate in vacuum pump before centrifugation to empty wells.

10. Centrifuge at 6000 rpm (5600 x g) for 4 min at 20-25 °C.
11. Empty the S-block and remove the AirPore tape sheet.
12. Add 800 µL Buffer RPE to each well of the RNeasy 96 plate, and seal the plate with a new AirPore tape sheet.
13. Place plate in vacuum pump before centrifugation to empty wells.
14. Centrifuge at 6000 rpm (5600 x g) for 4 min at 20-25 °C.
15. Empty the S-block and remove the AirPore tape sheet.
16. Add 800 µL Buffer RPE to each well of the RNeasy 96 plate, and seal the plate with a new AirPore tape sheet.
17. Place plate in vacuum pump before centrifugation to empty wells.
18. Centrifuge at 6000 rpm (5600 x g) for 4 min at 20-25 °C.
19. Remove the AirPore tape sheet. Place the RNeasy 96 plate on top of a rack of Elution Microtubes CL.
20. Add 50 µL RNase-free water to each well, and seal the plate with a new AirPore tape sheet.
21. Incubate for 1 min at room temperature (15-25 °C).
22. Centrifuge at 6000 rpm (5600 x g) for 4 min at 20-25°C to elute the RNA.

RNA concentrations were measured using NanoDrop ND-1000 spectrophotometer (ThermoFisher Scientific) at 260 nm.

B.2.2 cDNA synthesis

The components in the master mix from the High Capacity cDNA reverse Transcription kit added to each well for the cDNA synthesis can be seen in Table B.1.

Table B.1: Master mix used for each sample in the cDNA synthesis.

Component	**Volume (µL)**
10X RT buffer	4
25X dNTP	1.6
10X Random Primer	4
RNase-free Water	8.4
MultiScribe Reverse Transcriptase	2
Total	**20**

The cDNA protocol used in the Veriti 96w Thermal Cykler (Applied Biosystems) can be seen in Table B.2. The cover temperature used prior to start was 105 °C.

Table B.2: cDNA protocol used for amplification in the Thermal Cykler.

Temperature (°C)	Time
25	10 min
37	2 h
85	5 min
4	∞

B.2.3 qPCR

1. Thaw cDNA plate on ice.
2. Mix and spinn the cDNA for 30 s respectively.
3. Dilute cDNA in nuclease-free water to 3X by adding 80 µL water to each well.
4. Use TaqMan Pipetting BioMek NX robot to add 7 µL master mix and 3 µL cDNA in triplicates to a 384 well plate.
5. Centrifuge at 700 rpm for 1 min.
6. Run the qPCR on the Quantstudio™ 7 Flex Real-Time PCR system (Applied Biosystems) for comparative CT method.

The components in the master mix for the qPCR added to each well can be seen in Table B.3.

Table B.3: Master mix used for each sample in the qPCR run.

Component	Volume (µL)
TaqMan Master Mix	5
TaqMan probe	0.5
nuclease-free water	1.5
Total	**7**

The qPCR protocol used with the Quantstudio Real-Time PCR system can be seen in Figure B.1.

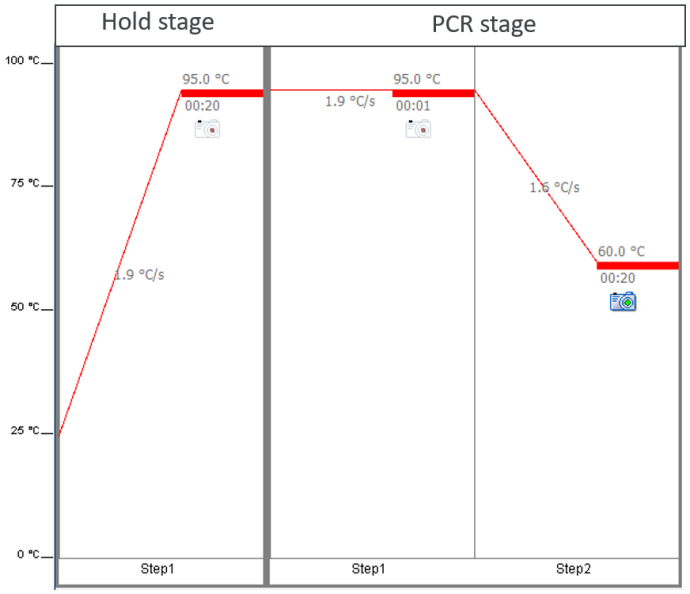

Figure B.1: Program for qPCR run with Quantstudio Real-time PCR system.

B.3 Protein quantification

Protein quantification was performed with Pierce BCA Protein Assay kit by measuring protein concentrations in the supernatants from the cell lysates. All equal samples were pooled and tested as singlets.

1. Thaw lysates on ice and transfer to eppendorf tubes.
2. Spin lysates for 10 minutes at 14 000 rcf at 4 °C and save supernatant.
3. Dilute lysates 1:5 directly in plate by adding 5 µL lysate and 20 µL M-PER buffer.
4. Dilute BSA ampule into clean vials with same diluent as samples (M-PER).
5. Add 25 µg/mL of standards in duplicates into plate.
6. Prepare working reagent by mixing 11 mL working reagent A and 220 µL working reagent B and shake.
7. Add 200 µL working reagent mixture to each well.
8. Cover and incubate at room temperature for 10-20 min (read when enough colour).
9. Uncover the plate and measure OD at 562 nm with SpectraMax 190 (Molecular Devices).
10. Calculate standard curves from BSA standards by plotting BSA concentrations against measured OD-values.

11. Calculate protein concentrations from measured ODs and the slope (k) and y-intercept (m) of the standard curve by using equation B.1.

$$\text{Concentration} = \frac{\text{OD}_{\text{sample}} - \text{m}_{\text{value}}}{\text{k}_{\text{value}}} \tag{B.1}$$

BSA standard dilutions can be observed in Table B.4.

Table B.4: BSA standard dilutions used to determine protein concentrations.

Conc (µg/mL)	Volume BSA (µL)	Volume M-PER (µL)
2000	0	300
1500	125	375
1000	325	325
750	175	175
500	325	325
250	325	325
125	325	325
25	400	100
0	400	0

Standard curves calculated from the OD measurements of BSA standard dilutions can be seen in Figure B.2.

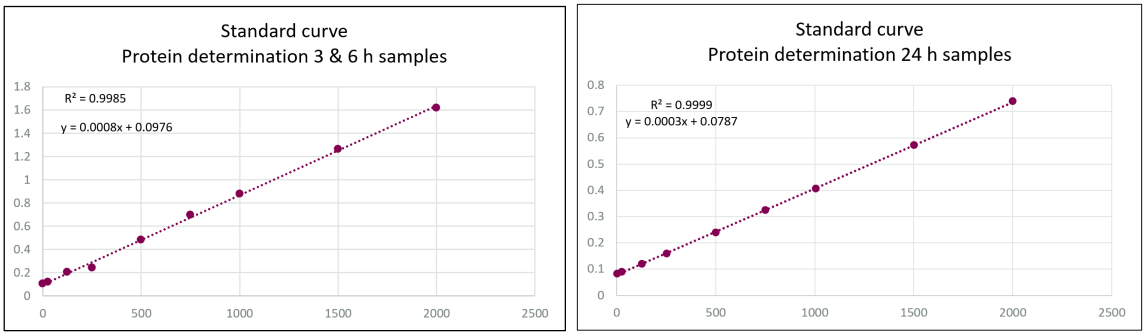

Figure B.2: Standard curves for protein determination of the samples from experiment 1. The samples from 3 and 6 h stimulation with tunicamycin to the left and samples from 24 h stimulation with tunicamycin to the right.

B.4 Traditional Western blot

1. Samples were pre-diluted with water
2. NuPAGE LDS Sample Buffer (4x) and NuPAGE Sample Reducing Agent (10x) were added to each sample at working concentration.
3. Samples were vortexed, spinned and heated to 70 °C for > 10 min.

4. Box for SDS-run was prepared by setting the gel in the box and adding MOPS or MES buffer in the box, covering the wells.
5. 2.5 µL of chameleon Duo ladder was loaded into its designated wells.
6. Samples were loaded into their designated wells according to the defined setup.
7. 0.5 mL NuPAGE Antioxidant was added to the 200 mL of MOPS/MES SDS running buffer in the upper (cathode) buffer chamber.
8. The gels were run with a constant voltage of 150 V and with 1 A as maximal ampere level for 1 h with MES running buffer or 90 min with MOPS running buffer.
9. Prepare concentrated methanol solution and transfer buffer solution for the membrane.
10. Immerse the PVDF membrane for 1 min in concentrated methanol solution and then transfer to transfer buffer.
11. Put the gel after the SDS-run in transfer buffer
12. Prepare the transfer by putting together sponges, membrane protection, gel, membrane (transfer sandwich) och setting it in the running-box.
13. Fill middle of box with blocking buffer and the rest with distilled water.
14. Transfer proteins from the gels to PVDF membranes at room temperature for 105 minutes at 80 V and with max 1 A.
15. Block membranes in LI-COR Intercept TBST blocking buffer for 60 minutes at Room temperature.
16. Incubate with primary antibodies at 4°C overnight.
17. Wash at room temperature with TBS-Tween for 3x10 min.
18. Incubate with secondary antibodies at room temperature for 2 h.
19. Washing at room temperature with TBS-Tween for 3x10 min.
20. Detect fluorescence signals on Odyssey reader.

All antibodies were diluted in LI-COR Intercept TBST blocking buffer.

B.5 Automated western blot

1. Thaw samples in ice.
2. Prepare standard pack reagents by piercing the fold lids and add as follows:
   - Add 40 µL deoinized water to the DTT reagent and pipette up and down
   - Add 20 µL of the diluted DTT solution and 20 µL 10X sample buffer to the 5X master mix reagent.
   - Add 20 µL deoinized water to the ladder.

3. Prepare 0.1X sample buffer by mixing 10 µL 10X SB stock and 990 µL deoinized water.
4. Dilute lysates in 0.1X sample buffer if needed, use 0.5 mL tubes.
5. combine 1 part flourescent 5X master mix with 4 parts diluted lysate.
6. Spinn down and boil lysates at 70 °C for 10 min, spin down.
7. Dilute antibodies in Milk-free antibody diluent.
8. Prepare chemiluminescent substrate by mixing 200 µL Luminol-S and 200 µL Peroxidase.
9. Peel away upper pat of folio on the JESS plate and add 5 µL ladder, 3.9 µL sample, 10 mL antibody diluent, 10 µL streptadivin HRP, 10 µL primary antibody, 10 µL secondary antibody and 15 µL HRP substrate to the designated wells by pipetting to wall of wells, avoiding bubbles.
10. Cover (dedicated Jess cover and an adhesive foil) and spin at 1000 x g for 5 min.
11. Add 520 µL buffer/wash buffer well.
12. Clean inside of Jess door with 70 % ethanol.
13. Peel away lower foil and place plate in the JESS machine.
14. Insert capillaries, blue light when inserted correctly.
15. Open software and modify standard protocol by changing separation time to 30 min and primary antibody time to 60 min.
16. Start the JESS run and run for approximately 4 hrs.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY