Single Transcriptional Unit Gene-Editing for Inheritable Lung Disease

Ifeanyi Enekwa

PhD

Single Transcriptional Unit Gene-Editing for Inheritable Lung Disease

Ifeanyi Enekwa

Thesis was submitted in partial fulfilment of the requirement of the degree of Doctor of Philosophy at University of Northumbria at Newcastle

Faculty of Health and Life Sciences

June 2022

Abstract

Genetic lung diseases such as Cystic fibrosis (CF) and idiopathic pulmonary fibrosis (IPF) are two of the most common examples of obstructive and restrictive lung diseases respectively, which pose an increasing challenge to health care systems. The Δ F508 mutation in cystic fibrosis transmembrane regulatory protein (*CFTR*) and rs35705950 single nucleotide polymorphism (SNP) in the Mucin 5 B (*MUC5B*) promoter region, are the two most common genetic conditions which are causative for CF and associated with IPF, respectively. Furthermore, a putative promoter associated non-coding (panc) RNA (AC061979.1) overlapping rs35705950 is predicted to be transcribed from this locus, positing a functional role in the development of IPF. Such mutations are prime candidates for gene-editing technologies aiming at disease treatment or prevention. Attempts to make this molecular tool more clinically relevant seek to enhance specificity, focus activity, or reduce the duration of action and thus risk of off-target gene editing events.

In this thesis, such an approach was taken by designing a single-transcriptional unit (STU) gene-editing cassette encoding the minimum components needed for genome editing. A set of functional *CFTR*-targeting gRNAs were designed, then tested in Cystic fibrosis bronchial epithelial (CFBE41o-), human alveolar basal epithelial (A549) and HeLa cells, at varying success rates as determined by T7 endonuclease 1 (T7E1) assay, with no statistically significant negative impact of cellular viability. The gRNA found to be most efficacious in CFBE41o- cells was then cloned into a STU gene-editing cassette which was found to express by RT-qPCR, but not translate *in cellulo* as determined by western blotting analysis. Troubleshooting eventually uncovered the lack of expression of the gene-editing endonuclease, present within the wild-type plasmid backbone used for cloning this STU gene-editing cassette.

After 3rd party next generation sequencing transcriptome data analysis, reverse-transcription quantitative polymerase chain reaction (RT-qPCR) was used to validate the expression of AC061979.1 pancRNA in lung epithelia, albeit at low copy number. Attempts to genome edit the rs35705950 locus proved unsuccessful, potentially because of the highly repetitive nature of this genomic locus.

Hence, this work validated the expression of a newly described pancRNA molecule *in cellulo*, which had only previously been predicted *in silico* to be expressed, and not explicitly reported in 3rd party next generation sequencing datasets. This thesis further contributes to the body of work attempting to improve production and design of clinically-relevant gene-editing approaches relevant to lung disease.

Declaration

I certify that the work in my thesis has not been submitted for any other prize and is entirely my own. The sources are correctly recognised where ideas and opinions were imported from other studies.

Word count: 38,758

Acknowledgements

First and foremost, I wish to express my gratitude to my supervisor, Dr Sterghios Moschos, for the opportunity to be part of his group. Thanks to his steadfast advice and support through difficult times, I was able to complete this project. Working with him was a great experience since he is a real mentor. I'd like to express my gratitude to my second supervisor Prof Gary Black for his support for the duration of this project.

A special thanks to Dr Ciaran Kelly for mentoring me through the cloning part of this project and Dr Antony Antoniou for his unconditional kindness and the reagents gifted to me throughout the course of my project.

I would also like to thank Dr Manfred Frick and Dr Giorgio Fois from the University of Ulm for sharing their data relevant to the fourth chapter of this project.

Without genuine friends, no one can go through a PhD. As a result, I will be eternally grateful to Dinh Nguyen for showing me the ropes with gene-editing assays, PCR, every day lab instruments and for helping me troubleshoot most failed experiments. I am very grateful to John Henderson for being a good friend to me throughout my project and providing guidance in writing this PhD thesis. A special thanks to Theodora Mantso, Gina Abdelaal, Effie Kritikaki, Nengi Ogbanga for their help with Western blot, cell viability, qPCR assays and statistical analysis respectively. Off course the mental support and scientific advice provided by my others friends namely, Olga Cugaj, Andrea Bonicelli, Nicola Cowley, Joe O'Sullivan, Migkena Artemis, Emma Riley and John Allen shall not be forgotten.

Last but not least, I owe a huge debt of gratitude to my wonderful parents. Despite the vast distances that divides us, they have shown a great deal of patience and provided the resources and emotional support for the summit of this education mountain. I am grateful for all of this and much more.

Table of Contents

Table of Figures

List of Tables

Abbreviations

A549 cells	Human alveolar basal epithelial cell line
AARS	tRNA synthase
AAT	α-1 antitrypsin
AAV	Adeno-associated virus
AAVS1	Adeno-associated virus integration site 1
ADE1	Purine biosynthetic protein
ANOVA	Analysis of variance
AdV	Adenovirus
AdV-CFTR	CFTR packaged into AdV serotype 5
AS	Allele-specific
ATP	Adenosine triphosphate
B2M	Beta-2-Microglobulin
BCA	Bicinchoninic assay
Bdp1	B double prime
BGH	Bovine growth hormone
Bp	Base pairs
Brf1	B-related factor 1
BSA	Bovine serum albumin
Cas	CRISPR-associated protein
cDNA	Complementary deoxyribonucleic acid
CF	Cystic fibrosis
CFBE41o-	Cystic fibrosis bronchial epithelial cell line
CFTE	Cystic fibrosis tracheal epithelial
CFTR	Cystic fibrosis transmembrane conductance regulator
CF gRNA	CFTR-targeting gRNA
COPD	Chronic obstructive pulmonary disease
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CPPs	Cationic polypeptides
CSF2RA	Colony Stimulating Factor 2 Receptor Subunit Alpha
E. coli	Escherichia coli

dCas9	Catalytically dead Cas9
Δ F508	Phenylalanine 508 deletion
Δ I507	Isoleucine 508 deletion
dBPK4410	Digested BPK4410
D. melanogaster	Drosophila melanogaster
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNase	Deoxyribonuclease
dNTP	Deoxynucleoside triphosphate
DRP	DNase-resistant particles
DSB	Double-stranded break
dsDNA	Double-stranded DNA
dsODN	Double-stranded
FEV_1	Forced expiration volume
ECL	Enhanced chemiluminescence
EP	Endo-porter peptide
eSpCas9	Enhanced specificity S. pyogenes Cas9
FBS	Foetal bovine serum
FOXA2	Forehead box protein A2
GalNac	N-acetylgalactosamine
GBA	Shingolipid glucosylceramidase
Gblock	Gene block
GFP	Green fluorescent protein
GM-CSF	Granulocyte-macrophage colony-stimulating factor
gRNA	Guide RNA
HAEpCs	Human airway epithelial primary cells
HBE	Human bronchial epithelial
HC-AdV	High-capacity AdV
hCFTR	Human cystic fibrosis transmembrane conductance regulator
HD-AdV	Helper-dependent AdV
HDF	Human dermal fibroblasts
HDR	Homology-directed repair
HDV	Hepatitis delta virus
HEK293	Human embryonic kidney 293

hESC	Human embryonic stem cells
HH	Hammerhead
hiPSC	Human induced pluripotent stem cells
HLMVEC	Primary human lung microvascular cells
Hrs	Hours
HR	Homologous recombination
HSV	Herpes Simplex Virus
Hypa Cas9	Hyperactive Cas9
ICP	Infected Cell Protein
IFNB1	Interferon B1
ILA	Interstitial lung abnormalities
ILD	Interstitial lung diseases
Indel	Insertion and deletion
IPF	Idiopathic pulmonary fibrosis
iPSC	Induced pluripotent stem cells
ISCs	Intestinal stem cells
ITR	Inverted terminal repeat
IVT RNA	In vitro-transcribed RNA
Kb	Kilo bases
KLD	Kinase-Ligase-DnpI
Kv	Kilo volts
LAMP-1	lysosome-associated membrane protein-1
LAP	Latency-active promoters
LB	Luria Broth
LMO-2	LIM domain only 2
LNA	Locked nucleic acid
LNC	Lipid nanocapsules
lncRNA	Long non-coding RNA
LVLPs	Lentivirus-like particles
Mbp	Mega base pairs
MDA5	Melanoma differentiation–associated gene 5
MEM	Minimum Essential Medium
MESC	Mice embryonic stem cells
µg	Microgram

M gRNA	MUC5B-trageting gRNA
MHHi002-A cells	Human CF-iPSCs
miRNA	microRNA
µM	Micromolar
µmol	Micromoles
mM	Millimolar
mmol	Millimoles
MOI	Multiplicity of infection
mRNA	Messenger RNA
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MUC5AC	Mucin 5 AC
MUC5B	Mucin 5 B
MYBPC3	Cardiac myosin binding protein C
NC	Negative control
NGS	Next generation sequencing
NHEJ	non homologous end joining
ng	Nanogram
nmol	Nanomoles
nM	Nanomolar
N/P ratio	Nitrogen/phosphorus ratio
NSCLC	Non small cell lung cancer
OC BPK4410	Open-circle BPK4410
OE	Overlap-extension
OLD	Obstructive lung disease
Oligo	Oligonucleotide
OPEN	Publicly available oligomerised pool engineering
OsPDS	rice phytoene desaturase
PAM	Protospacer adjacent motif
PAMPs	Pathogen-associated molecular patterns
pancRNA	promoter-associated long-noncoding RNA
pBAC	PiggyBac
PBS	phosphate buffered saline
PC	Positive control
PCR	Polymerase chain reaction

PEG	Polyethylene glycol
pegRNA	prime-editing gRNA
PEI	Polyethylenimine
Pen/Strep	Penicillin/streptomycin
pg	picogram
pH	power of hydrogen
PIC	Pre-initiation complex
pmol	Picomoles
pM	picomolar
Poly(A)	Poly adenosine nucleotides
PPABLG	PEGylated nanoparticles based on the α-helical polypeptide
PPP1R12C	Intron 1 of the protein phosphatase 1 regulatory subunit 12C
pre-crRNA	Precursor crRNA
Pre-tRNA	Precursor tRNA
P. aeruginosa	Pseudomonas aeruginosa
qPCR	quantitative PCR
RDN5	5S rRNA gene
REC	Recombinant
Rep gene	Replication gene
RIG-I	Retinoic acid-inducible gene I
RIPA buffer	radio-immunoprecipitation buffer
RNA	Ribonucleic acid
RNA pol	RNA polymerase
RNP	Ribonucleoprotein
RT-qPCR	Reverse transcription quantitative PCR
R.U	Virus to host DNA ratio
SBE	SMAD-Binding elements
SCID-CB17	Severe combined immunodeficiency mice with a CB17 genetic background
SD	Standard deviation
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel
SFTPB	Surfactant protein B
sgRNA	Single-guide RNA
siRNA	Short interfering RNA
SMAD	Suppressor of Mothers against Decapentaplegic transcription factors

SOC media	Super Optimal broth with Catabolite repression media
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
SNP	Single nucleotide polymorphism
ssODN	Single-stranded oligodeoxynucleotide
SPDEF	SAM Pointed Domain Containing ETS Transcription Factor
S. pyogenes	Streptococcus pyogenes
SpCas9-HF1	High fidelity S. pyogenes Cas9
STU	Single-transcriptional unit
T1-4	HDR donor templates 1-4
T7E1	T7 endonuclease 1
TALE	Transcription activator-like effector
TALENs	Transcription activator-like effector endonucleases
TAT	Trans-Activator of Transcription
Th2 cells	Type II helper T cells
tracrRNA	Trans-activating crRNA
TBP	TATA-binding protein
TBS	tris buffered saline
TBST	tris buffered saline with tween
TEMED	Tetramethylethylenediamine
TERC	Telomerase RNA Component
TERT	Telomerase reverse transcriptase
TIDE	Tracking of Indels by Decomposition
TMEM	Transmembrane channel proteins
Tris	Tris(hydroxymethyl)aminomethane
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
tRNA	Transfer RNA
TNF-α	Tumour necrosis factor alpha
U2OS	Osteosarcoma epithelial cell line
Vero cells	Monkey kidney epithelial cell line
v/v	Volume by volume
WT	Wild type
w/v	Weight by volume
YFP	Yellow fluorescent protein

ZF	Zinc fingers
ZFN	Zinc finger nucleases

Chapter 1. Introduction

1.1 Diseases of the Lung

Respiratory diseases present an enormous socioeconomic and individual burden world-wide, causing significant disease-related morbidity and mortality (UK cystic fibrosis registry, 2021). Lung diseases can be categorised into four broad groups namely: (i) acute, (ii) chronic, (iii) occupation-related and (iv) other parenchymal diseases including immune-related diseases, as well as pulmonary malignancies (Speizer *et al.*, 2006). Chronic lung diseases can be classed into two broad groups namely, restrictive and obstructive lung diseases (Kim *et al.*, 2016).

1.1.1 Obstructive Lung Diseases

Obstructive lung diseases (OLDs) are characterised by complete or partial impediment of airflow, at any level, from the trachea to the distal bronchioles (Martinez-Pitre, Sabbula and Cascella, 2022). These diseases are associated with serious respiratory morbidity, respiratory failure, and high risk of mortality. Common causes of OLDs include asthma, emphysema, bronchiectasis, chronic obstructive lung diseases (COPD) and Cystic fibrosis (CF), which is the most common inherited genetic disorder (Kim *et al.*, 2016). Of note, OLDs account for approximately eighty percent of respiratory syndromes (Mangera, Panesar and Makker, 2012).

Emphysema and COPD are diseases linked with a mutation in α-1 antitrypsin gene *AAT*, encoding the serine-protease α-1 antitrypsin (AAT), which results in AAT deficiency. This AAT deficiency results in protease to anti-protease disequilibrium in the lungs, leading to inflammation (Kim *et al.*, 2016). A combination of this monogenic trait plus environmental triggers such as cigarette smoking and acute infection, often result in the development of emphysema and COPD (Kim *et al.*, 2016). Similar to COPD, asthma is caused by a combination of genetic predisposition and environmental factors. Asthma is most commonly an allergic disorder induced by uncontrolled synthesis of cytokines, released by allergen-specific type 2 helper T cells (Th2 cells), which leads to immunoglobulin E synthesis and mast cell infiltration (Kim *et al.*, 2016). Mast cells induce histamine, leukotriene and prostaglandin production which encourage vesicular permeability, smooth muscle contraction and mucus production, while chemokines produced recruit macrophages, Th2 cells and basophils to the airway, resulting in localised inflammation and tissue damage (Maes, Tournoy and Joos, 2011).

1.1.2 Cystic Fibrosis

CF is a life-limiting autosomal recessive disorder with a life-expectancy just under 40 years, affecting up to 70,000 people world-wide, particularly those of European descent (UK cystic fibrosis registry, 2021). Due to the untreatable nature of Class I CF mutations (~12% CF patients) and the constant need

to improve efficacy of current CF modulators discussed below, there is a push to develop a universal strategy to tackle this disease.

In CF there is an imbalance in the cross-membrane ionic regulation leading to a depletion of water content present in the airway surface liquid. This fall in water content in the airway surface liquid results in decreased pH of the epithelial lining which in turn causes the viscosity of mucus to increase, subsequently decreasing its mobility by ciliary action (Henke et al., 2004). The pile up of mucus restricts motion and leads to the blockage of airways, inflammation and bronchiectasis (Henke et al., 2004).

CF impacts significantly on patients health and wellbeing. The accumulation of mucus along the airway provides a fertile breeding ground for opportunistic bacteria. One of the most common pathogens in the CF airway, *P. aeruginosa* colonises the airway, recruiting inflammatory cells which tend to damage the airway and parenchymal epithelial cells, hinder lung function and cause mortality via respiratory failure (Wilschanski *et al*., 2003; Liou *et al.,* 2007). Therefore obstructive lung disease is classed as the primary cause-of-death in CF patients, accounting for ~ 60% of all registered deaths (UK cystic fibrosis registry, 2021). However, the pathology of CF is not only restricted to the lung. Other harmful effects of CF include, increased infertility rate, decreased pancreatic function, intestinal malabsorption causing weight loss and impaired growth. Furthermore, due to the increasing CF life-expectancy on account of improved treatment availability over the past fifteen years, novel disease-related complications are being observed such as CF-related diabetes, metabolic bone disease, multidrug resistant pathogenic infections, gastrointestinal and pulmonary malignancies (Than *et al.*, 2016). Hence non-pulmonary causes of death include, malignancies, liver disease, cepacian syndrome, pneumothorax, and hemoptysis, amongst others (Martin *et al.*, 2016).

CF-causative mutations are recessive, requiring two copies of mutant allele to result in disease. CF causative mutations can be grouped into six classes, which affect the cystic fibrosis transmembrane regulator chloride ion channel (*CFTR*) gene. Class I mutations such as G542X, W1282X, R553X are characterised by an absence of *CFTR* protein production due to dysfunctional translation caused by a mutation that results in a premature stop codon. Drugs targeting these mutations are termed 'read-through agents', they rescue protein synthesis by overcoming the premature stop codons (King *et al.,* 2022). Class II mutations such as Δ F508, Δ 507, G85E, are characterised by aberrant trafficking of channel protein to the cellular membrane as a result of mutations that cause translated *CFTR* protein to be misfolded and subsequently degraded in the endoplasmic reticulum (Cheng *et al.,* 1990). Corrector drugs such as lumacaftor, tezacaftor and elexacaftor, aid in reversing misfolding errors, and improving protein trafficking (Miah, Hyde and Gill, 2019). Trikaftor, a combination of three CF

modulator drugs (two correctors tezacaftor and elexacaftor, and a potentiator Ivacaftor) is the most effective therapy prescribed to patients with atleast one copy of the Δ F508 allele (ClinicalTrials.gov, 2020). Class III mutations such as G551D, G178R, G551S, affect channel gating, due to mutations that affect CFTR ATP binding domains, resulting in diminished ATP binding and hydrolysis, necessary for channel closed-dimer and open-dimer conformations, respectively (King *et al.,* 2022). Potentiators such as ivacaftor, potentiate the channel in the open conformation, improving channel function (King et al., 2022). Class IV mutations such as R334W, R117H, R347P, result in reduced channel anionic conductance. Some of these mutations can be remedied using potentiators such as ivacaftor. Class V mutations such as 3272-25A>G, A455E, D565G, are splicing mutations resulting in production of aberrantly spliced transcripts which hinders the amount of CFTR transcripts and subsequent protein production (Amaral *et al.,* 2001). Both potentiators and amplifier drugs in the developmental stages are used against this class of mutations. Class VI mutations such as 1811+1.6kb A>G, result in the reduction in CFTR protein stability on the cellular membrane. Theoretical stabiliser modulator drugs that increase cell surface protein stability would be applicable for this mutation (King *et al.,* 2022). The most common CF mutation accounting for up to 60% of cases globally is the CFTR class II Δ F508 mutation (figure 1.1.2.1), hence why it shall be a major focus of this study. Here, in addition to degradation of the aberrantly misfolded protein, there is a reduction in normal channel functioning (figure 1.1.2.2) (Molinski *et al.,* 2012).

Table 1.1.2.1 Summaries the CF mutations discussed above, their effect and therapeutic approach.

Mutation Class	CFTR Defect	CFTR Function	CFTR Expression	Examples of Mutations	Potential Therapy
I	Defective channel production	No	No	G542X, W1282X, R553X	Read through agents
II	Impaired folding/ processing	No	No	Δ F508, Δ 507, N1303K, G85E	Correctors + Potentiators
III	Impaired channel gating	No	Yes	G551D, G178R, G551S,	Potentiators
IV	Defective conductance	Reduced	Yes	R334W, R117H, RR347P	Potentiators

V	Aberrant splicing, reduced amount	Reduced	Reduced	3272-25A>G A455E, D565G	Splicing modulators, amplifiers
VI	Reduced stability at plasma membrane	Reduced	Reduced	1811+1.6kb A>G	Stabilisers

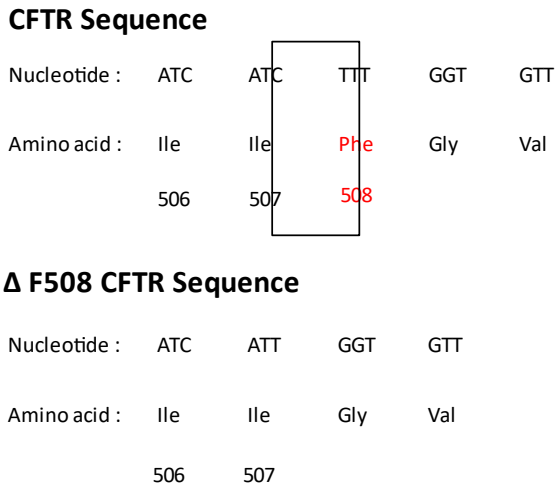


Figure 1.1.2.1 A 'CTT' deletion at codon 508 in *CFTR* results in a frameshift mutation which causes the deletion of Phe at the same position.

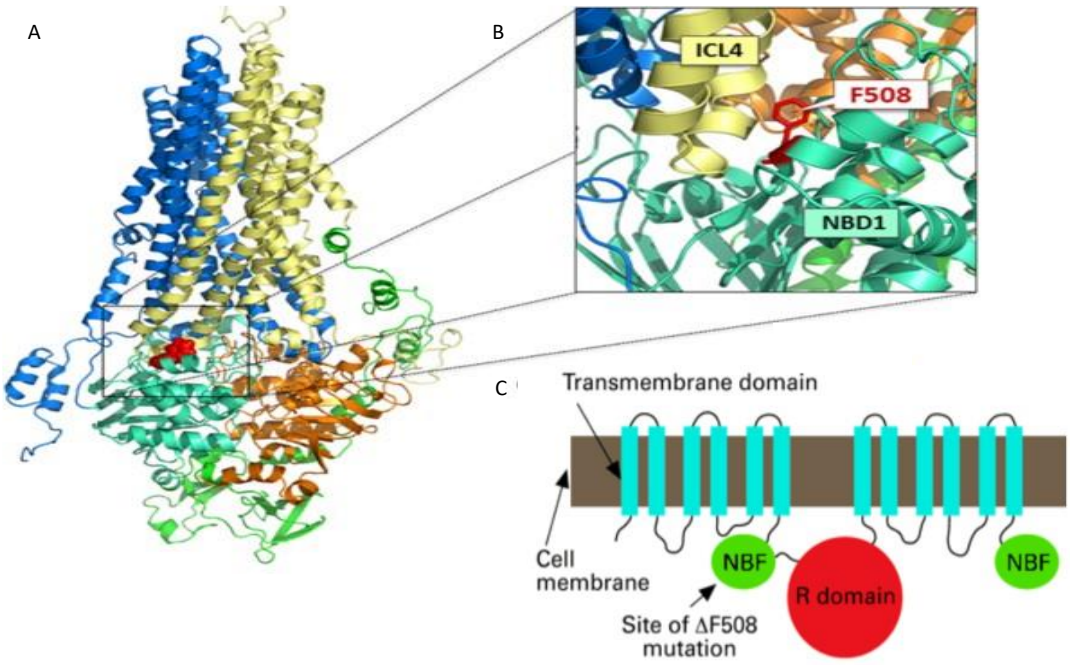


Figure 1.1.2.2 Visualisation of the CFTR channel protein. (A) Full-length ribbon diagrams of CFTR; membrane spanning domains (MSD1, blue); MSD2, yellow; Nucleotide binding domains/ Nucleotide binding folds (NBD1/ NBF1, cyan); NBD2/NBF2, orange; Regulatory (R domain, green); Δ F508, (red). (B) position of Δ F508 at the intercellular loop 4 (ICL4): NBD1/NBF1 interface. (C) Δ F508 mutation negatively impacts upon NBF1 stability, and subsequent CFTR channel folding, as well as NBF1/NBF2 heterodimerisation which impacts channel gating. Taken and adapted from (Molinski *et al*., 2012).

Current treatment options for CF patients exist, however treatment interventions are presented with numerous challenges. Precision *CFTR* modulator listed above, while effective, only benefit CF patients with a specific type of *CFTR* mutation and cannot be universally applied to all CF patients. To tackle CF-related pulmonary infection, administration of antibiotic therapy is often applied (Wilschanski *et al*., 2003). However, *P. aeruginosa* populations are capable of segregating and evolving differently according to the lung microenvironment, leading to microbial resistance (Wilschanski *et al*., 2003). Read-through agents such as the antibiotic gentamicin capable of benefiting Class I CF patients have very limited efficacy with toxic effects associated with long-term use (Wilschanski *et al*., 2003). Hence, class I *CFTR* mutations are the most severe, remaining untreatable, with palliative care rendered to such patients. Non-pharmacological treatments, such as aerobic exercise or physiotherapy, improve patient outcomes in the short-term. However, once pulmonary functional loss has progressed to respiratory failure, then a lung transplant inevitably becomes a last resort. Unfortunately, lung transplantation often leads to multiple complications down the line, only serving to extend mean life-expectancy from 12.9% (10.3-15.5%) for patients on the transplantation waiting list for 12 months, to 67.8% (59.9-76.8%) for patients having received a lung transplant over the same time period (Thabut

et al., 2013). Primary causes of death post-lung transplant include *bronchiolitis obliterans* syndrome, graft failure, technical causes (surgical complications, hemothorax, suture problems), and infection (Martin *et al.*, 2016).

Some approaches to tackling CF disease focus on some alternate ion channels that work in concert with CFTR to regulate homeostasis. These include inhibition of the sodium channel ENaC and ATP12A proton pump, which would reduce airway surface liquid dehydration and acid levels respectively, as well as the activation of transmembrane (TMEM) channel proteins, an alternate calcium-activated chloride channel, with the aim to compensate for deficient CFTR-mediated Cl- transport into cells (Shei *et al.*, 2018). In the context of precision medicine, a commonly used treatments for Δ F508 CF is Orkambi™, which is a combination of two drugs: Ivacaftor, a potentiator which extends the duration mutant *CFTR* remains in the open confirmation on the cellular membrane hence improving chloride conductance, and Lumacaftor, which functions as a chaperon aiding in the folding of mutant chloride channel (Connett, 2019). Orkambi™ for Δ F508 CF patients improves lung function and nutritional status, however, this treatment improved FEV_1 and normalised sweat chloride levels in only about 45% of CF patients (Accurso *et al.*, 2010). In another phase III trial, administration of 150mg Trikafta in Δ F508 CF patients over 24 weeks was equally shown to improve FEV_1 by 14% in these patients, but this treatment is only effective with this specific CF mutation . The use of Ivacaftor for class III G551D *CFTR* mutation, the most common gating mutation, improved FEV_1 in only 5-7% of CF patients, hence there is a need for a more universal approach to treating CF (Edgeworth, 2015). These precise pharmaceutical treatments are also very expensive, costing up to $250,000 per patient per year indicating the need to develop more economical but effective universal interventions against CF (Seidner, 2020).

1.1.3 Restrictive Lung Diseases

Restrictive lung diseases are often associated with a reduction in elasticity of the lung, negatively affecting lung expansion, reducing total lung volume and capacity (Borie *et al.*, 2019). Restrictive lung diseases could have intrinsic origins, often resulting from the damage of distal lung parenchyma caused by inflammatory toxins and responses. This destruction negatively affects the alveolar interstitium, potentially damaging the peripheral bronchial structure in the process. Interstitial lung diseases (ILDs) are a collective name for a group of diseases affecting the distal lung in such manner. Examples of ILDs include acute interstitial pneumonia, sarcoidosis, cryptogenic organizing pneumonia, non-specific pneumonia, and idiopathic pulmonary fibrosis (IPF), which is the most common and severe form of the disease (Martinez-Pitre, Sabbula and Cascella, 2022). Of note, OLDs account for approximately twenty percent of respiratory syndromes (Mangera, Panesar and Makker, 2012).

Acute interstitial pneumonia is a swift progressive idiopathic pulmonary disease frequently resulting in severe respiratory failure and acute respiratory distress syndrome (Mrad and Huda, 2022). The distinctive feature of this disease is its rapid onset. Although the aetiology widely-remains mysterious, acute interstitial pneumonitis is believed to be induced by insult to the alveolar epithelium leading to inflammation and downstream fibrosis (Mrad and Huda, 2022). Sarcoidosis is a heterogeneous group of diseases, which are the most frequently occurring ILDs. Although its causes are unknown, sarcoidosis is primarily characterized by the presence of non-caseating granulomas capable of residing in any organ, albeit primarily within the vessel walls, pleura and interstitium of the pulmonary system (Saketkoo *et al.*, 2021; Martinez-Pitre, Sabbula and Cascella, 2022). Cryptogenic organising pneumonia is a diffuse interstitial disease of the respiratory system that primarily affects the alveolar walls, but equally affects the bronchioles and alveolar ducts (King and Lee, 2022). This disease is characterised by hyperproliferation of connective tissue within small airways and alveolar ducts, as well as blockage of the bronchioles and surrounding alveoli (Martinez-Pitre, Sabbula and Cascella, 2022). Non-specific interstitial pneumonia is a chronic interstitial pneumonia characterised by the uniform distribution of interstitial fibrosis and inflammation (Nayfeh, Chippa and Moore, 2022). It is non-specific due to alterations in the lung structure that vary from other forms of pulmonary interstitial diseases. Nonspecific interstitial pneumonia histopathological patterns can be grouped into three. A cellular pattern, composed of mild to moderate interstitial inflammatory infiltrates including lymphocytes as well as plasma cells, a fibrotic pattern as a result of fibrotic cellular elements, and a mixed pattern comprised of a mixture of the two previously mentioned patterns (Martinez-Pitre, Sabbula and Cascella, 2022).

1.1.4. Idiopathic Pulmonary Fibrosis

IPF is a devastating disease of low incidence (4.6-16.3 cases per 100, 000 persons) and unknown cause characterised by interstitial fibrosis and increased risk of mortality (Noth *et al.*, 2013). The median survival of IPF patients post-diagnosis is between two to five years, yet some patients live much longer (Idiopathic Pulmonary Fibrosis Clinical Research *et al.*, 2012; Fernández Fabrellas *et al.*, 2018). Although IPF prognosis is generally poor, decline could be slow, rapid, or mixed, with periods of relative stability quickly followed by periods of rapid decline. Mortality is frequently caused by respiratory failure at the end stages of disease progression. The disease has numerous phases namely, the asymptomatic sub-clinical phase in which disease is only identifiable via radiography, and symptomatic pre-diagnosis and post diagnosis clinical phases (Ley, Collard and King, 2011) .

Treatments for IPF can be grouped into pharmacological and non-pharmacological. Nonpharmacological treatments include long-term oxygen therapy, pulmonary rehabilitation, and finally a lung transplantation when disease has progressed to an advanced stage (Fujimoto, Kobayashi

and Azuma, 2016). A pharmacological approach for treatments of this disease mainly comprises of glucocorticoids or immunosuppression regimens, applied as a single drug (N-acetylcysteine), a combination of two drugs (azathioprine plus prednisone) or three drugs (azathioprine, prednisone, and N-acetylcysteine) (Idiopathic Pulmonary Fibrosis Clinical Research *et al.*, 2012). Unfortunately, drawbacks in the use of these immunosuppressive treatments include an association with an increase in hospitalisation and mortality rate amongst IPF sufferers (Idiopathic Pulmonary Fibrosis Clinical Research *et al.*, 2012).

A few risk factors are linked to the onset of IPF including an age over 66 years, with older patients displaying poorer prognosis ('Idiopathic Pulmonary Fibrosis: Diagnosis and Treatment,' 2000; Fernández Fabrellas *et al.*, 2018). In the past twenty years, male sex was identified as another risk factor, accounting for up to 70% of cases worldwide, attributable to male sex hormones (Ley, Collard and King, 2011). This was initially uncovered in male C57BL/6 mice that displayed higher reduction in static compliance than female mice after 1-2mg/kg bleomycin treatment, an antibiotic used to induce lung injury (Voltz *et al.*, 2008). Ethnicity is another known risk factor, with people from Caucasian and Hispanic backgrounds being at greater risk (Fernández Fabrellas *et al.*, 2018). The lung microbiome could be a risk factor, with IPF patients having greater bacterial burden in bronchial alveolar lavage which could serve as a stimulus for microinjury thereby elevating risk of disease onset (Zaman and Lee, 2018). Several genetic variations are linked to an increased risk of developing IPF, including mutations, the most common being the G/T rs35705950 SNP located at the promoter region of *MUC5B* gene (Ley, Collard and King, 2011). This rs35705950 SNP is implicated is the overexpression of *MUC5B* in the distal airway, resulting in enhanced mucus production and downstream mucociliary clearance issues which could lead to the development of IPF disease (Hancock *et al.*, 2018). Other risk factors include smoking, increased dyspnoea, increased BMI and other comorbidities such as pulmonary hypertension and emphysema (Ley, Collard and King, 2011).

Numerous genetic variations are linked to an increased risk of developing IPF. Mutations in telomerase genes, *TERT* or *TERC,* result in telomere shortening which is linked to both familial and non-familial IPF (Alder *et al.*, 2008). Mutations in *SFTPA2* and *SFTPC* encoding surfactant protein A2 and surfactant protein C respectively, cause protein instability and ER stress in the lung, resulting in IPF (Maitra *et al.*, 2010). The G/T rs35705950 SNP on mucin 5B (*MUC5B*) on chromosome 11p15.5 is associated with the highest risk of developing IPF (Chen *et al.*, 2019; Wang *et al.*, 2014a; Hunninghake *et al.*, 2013; Hobbs *et al.*, 2019). In one study, genome-wide genotyping data for 1,699 people with interstitial lung abnormalities (ILA) including IPF, and 10,274 healthy people, demonstrated that ILAs were strongly correlated with the *MUC5B* promoter variation rs35705950 (P value of 2.6×10^{-27}) (Hobbs *et al.*, 2019). By comparison, the next SNP with the strongest relation to ILA was rs6886640,

with a p value of 3.8 x 10^{-8}, showing rs35705950 had a much stronger association this disease (Hobbs *et al.*, 2019). In another study, the minor IPF-risk allele had a prevalence of 3.33% in Chinese IPF patients as opposed to 0.66% in health controls, demonstrating its strong link to IPF (Wang *et al.*, 2014a). Equally, in the same study, the minor IPF-risk allele had a prevalence of 10% in Caucasian populations, which was > 10-fold greater than the prevalence reported in a Chinese population, further highlighting ethnicity as a risk factor for disease development. Another study demonstrated that the rs35705950 minor allele increased odds of ILA 2.8-fold greater than in healthy controls (Hunninghake *et al.*, 2013).

This rs35705950 polymorphism is implicated in the over transcription and overexpression of *MUC5B* resulting in elevated ER stress through protein recycling, misfolded protein responses, and other quality control mechanisms (Chen *et al.*, 2019). Equally, the overexpression of *MUC5B* results in elevated mucus production, which perturbs mucociliary clearance. Increased mucus present in the airway leads to greater retention of inhaled substances such as air pollutants, cigarette smoke and inflammatory debris, which results in long-term microscopic scaring and fibrosis in the lung, eventually leading to IPF disease (Hancock *et al.*, 2018).

A series of transcription factors, proteins, lncRNAs and perturbed DNA methylation, are all believed to play a role in the SNP-induced overexpression of *MUC5B* (figure 4.1.1). DNA methylation normally occurs on CpG islands, and it is widely accepted that this epigenetic modification controls gene expression by attracting repressive methyl-binding proteins implicated in gene repression and by preventing transcription factor(s) from binding to DNA promoters (Moore, Le and Fan, 2013). However, hypermethylated CpG regions and upregulated gene expression have been identified in prostate cancer, suggesting an undefined, more diverse mechanistic effect of this epigenetic modification on gene expression and regulation (Rauluseviciute, Drabløs and Rye, 2020); alternatively, hypermethylation in these contexts are the epigenetic footprint of cellular homeostatic mechanisms attempting, and failing, to suppress transcription. The rs35705950 SNP is believed to perturb a 25 CpG motif which is hypermethylated in diseased patients, non-canonically enhancing the association of Forehead box protein A2 (FOXA2) transcription factor, to its DNA-binding site which is situated 32 bp downstream of rs35705950 SNP (Helling *et al.*, 2017). Of note, FOXA2 is known to unwind and displace histones from chromatin with its DNA-binding domain, enabling recruitment of transcriptional cofactors to the promoter, hence promoting gene transcription (Choi, Choe and Lau, 2020; Helling *et al.*, 2017). Equally, this transcription factor is involved in lung development and homeostasis, *MUC5B* expression and IPF (Helling *et al.*, 2017). Therefore, considering the potential role of rs35705950 in FOXA2 transcriptional regulation, it was hypothesised that this SNP is implicated in *MUC5B* expression and its overproduction in the case of IPF patients with the rs35705950 risk variant.

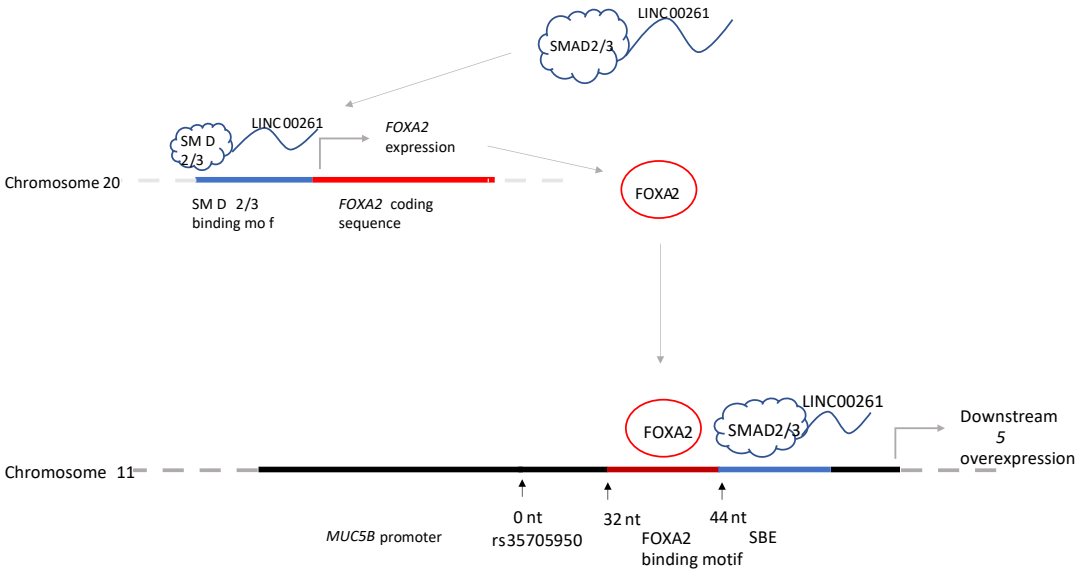


Figure 1.1.4.1. The rs35705950 SNP results in overexpression of MUC5B. LINC00261, an endoderm-associated lncRNA recruits SMAD2/3, to induce expression of FOXA2 (red line) on chromosome 20 (dotted light grey lines). The risk variant of rs35705950 SNP increases the DNA-binding affinity of FOXA2 transcription factor for its binding motif (dark red line), 32 nucleotides downstream of rs35705950, situated on chromosome 11 (dotted dark grey line). SMAD-binding elements (SBE) are located 44 nucleotides downstream of rs35705950 (blue line), permitting the binding of SMAD2/3-LINC00261, hence driving subsequent MUC5B (black line) overexpression in conjunction with FOXA2, in the context of IPF.

1.2 Gene-Editing: An Overview

Traditionally, genetic research was depended on finding and analysing spontaneous mutations. However, Muller and Auerbach proved in the mid-twentieth century that radiation or chemical treatment could speed up mutagenesis (Muller, 1927; Auerbach, Robson and Carr, 1947). Later approaches depended on transposon insertions, which could be induced in some animals; however, these procedures, like radiation and chemical mutagenesis, resulted in alterations to the genome at random locations. In the 1970s and 1980s, the first targeted genetic alterations were made in yeast and mice (Thomas, Folger and Capecchi, 1986; Rothstein, 1983; Scherer and Davis, 1979). The technique of homologous recombination, which was highly accurate yet inefficient, especially in mouse cells, was subsequently used to target genes, although the recovery of the desired products necessitated extensive selection and characterisation (Mansour, Thomas and Capecchi, 1988). Of note, gene-targeting was not easily transferrable to other species due to the low frequency and lack of culturable embryonic stem cells in mammals other than mice (Carroll, 2011).

Targeted gene editing is the creation of localised nucleotide changes, involving the replacement of an *in vivo* sequence with one designed *in silico,* containing the desired changes; typically, these changes can involve intentional insertions/deletions, or nucleotide substitutions. There are substantial challenges to genome editing, namely the delivery of genome editing complexes into target cells, the prevention of activity on genomic loci other than that intended (off-target), and the efficient selection of gene-edited cells (Carroll, 2011). For successful genome editing to take place, there is a requirement for DNA target to be damaged at the locus of interest, in order to increase recombination frequency, hence activating a previously inert and intact DNA target. Presently, there are three platforms for generating this DNA damage such as double-stranded breaks (DSBs) and single-stranded breaks, for gene-editing purposes which are subsequently discussed below.

1.3 Zinc Finger Nucleases

Zinc finger nucleases (ZFNs) were first observed by Chandrasegaran, while taking notice that *Fok*I restriction enzyme had physically separable binding and cleavage activities (Li, Wu and Chandrasegaran, 1992). The FokI cleavage domain does not confer any sequence specificity, however a target sequence could be specified by the recognition domain. The most common of these are the Cys_2His_2 zinc fingers where each unit of about thirty amino acids is bound to a zinc atom and each zinc finger contacts three base pairs of target DNA, whose sequence is defined by the zinc finger amino acid sequence (Pavletich and Pabo, 1991). Therefore, by assembly of multiple different combinations of zinc fingers multiple unique sequences can be used for inducing DSB within a genome (figure 1.3.1.1). One advantage of ZFNs is the requirement for dimerization prior to cleavage which increases the overall specificity of this gene-editing technique by reducing the probability that a pair of zinc fingers (ZFs) bind to the same off-target region. Although dimerization is a requirement for the FokI domain to cut DNA (Bitinaite *et al.*, 1998), in some instances weak dimer formation occurs preventing DNA-cleavage activity. To tackle this dimerization issue, monomeric cleavage domains can be used which would involve attachment of two sets of ZFs on opposite strands to catalyse DNA cleavage.

1.3.1 Homology-Directed Repair

When a DNA donor template is omitted while performing gene-editing, it triggers the default non homologous end joining (NHEJ) repair pathway which characteristically generates a high frequency of mutant DNA insertions and deletions (indels) (Aravind and Koonin, 2001). However, to precisely create desired genetic alterations, homology-directed repair (HDR) can be employed (figure 1.3.1.1) (Rouet et al., 1994). Repair by HDR proceeds by a mechanism known as synthesis dependent strand annealing (Kurkulos *et al.*, 1994). Here, the ends of the target break are resected, leaving a 3’ overhang which hybridizes to the donor template and is subsequently elongated by DNA polymerase until it

makes contact and anneals to the opposite end of the break (Kurkulos *et al.*, 1994). The length of donor sequences incorporated also known as conversion tracts varies depending on the organism, hence in *Drosophila Melanogaster* they can be several kilobases (Nassif *et al.*, 1994), albeit at decreasing frequency at further distances from the break, while in mammalian cells shorter repair tracts of 100 to 200 bases exists (Elliott *et al.*, 1998).

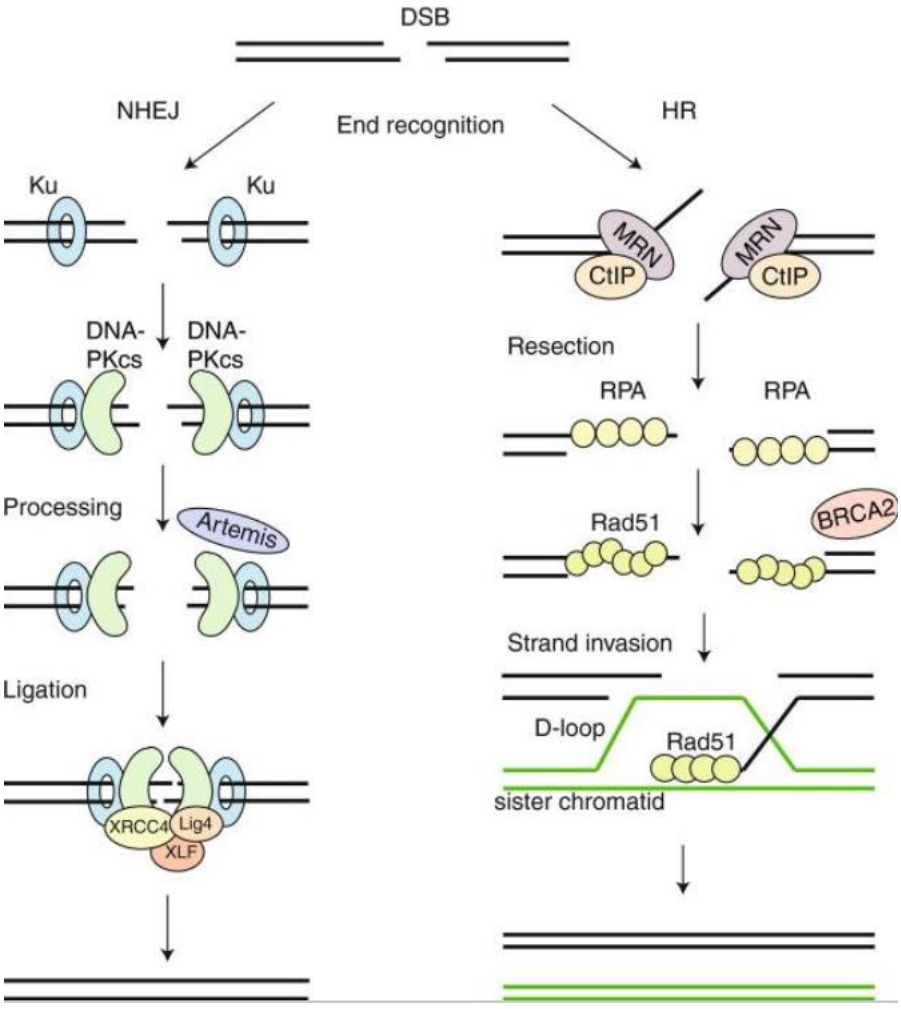


Figure 1.3.1.1 Two pathways for DNA DSB repair following genome editing. HDR or homologous recombination (HR), is a major pathway for the error-free repair of DSBs, which principally occurs during S/G2 phases of the cell cycle. An early determinant of DSB repair pathway choice is DNA end-controlled resection facilitated by Mre11, Rad50 and Xrs2 (MRN)-CtIP complex, which processes DNA ends to generate 3' single strands. RPA coats the single-stranded DNA, then a defining step of HDR, which is controlled by BRCA2, is the formation of RAD51 filaments on the 3' single strands, followed by strand invasion into a homologous DNA, typically the sister chromatid, priming repair DNA synthesis. The 3' end primes DNA synthesis from the homologous DNA; using the sister chromatid, therefore the repair can be precise restoring the original sequence prior to damage. To complete HR, the newly synthesized strand can dissociate to anneal to the other end. Other outcomes are also possible, for example, in which Holliday junctions are formed and either dissolved or resolved. DSBs can also be

repaired by error-prone DNA repair pathways. Canonical nonhomologous end joining (NHEJ) can occur during any cell cycle phase and is considered the other major DSB repair pathway. In NHEJ, DNA ends are protected from end resection but often undergo some processing before joining, which results in small deletions or insertions. Ku70/Ku80 heterodimers are recruited to both ends of the DSB functioning as a docking site for other NHEJ proteins. Protein kinase C (PKC) phosphorylates Artemis which unwinds the DNA double helix and holds opposite ends of the DSB withing proximity. Finally, XRCC4 and ligase IV proteins are recruited to permit the ligation of adjacent ends of the DNA molecule. Taken and adapted from (Brandsma and Gent, 2012).

1.3.2 ZFNs for Gene-Editing

ZFNs were first created as a chimeric restriction enzyme followed by subsequent validation of their activity *in vitro* (Kim, Cha and Chandrasegaran, 1996). The first scenario were ZFNs were designed for the purpose of *in vivo* targeted gene-editing was in *D. melanogaster* (Bibikova *et al.*, 2002). In this instance, targeting ZFNs to the *D. melanogaster yellow* gene locus was achieved by altering seven amino acids in the alpha-helix of each ZF DNA binding motif (Lee *et al.*, 2012a). In many organisms, including *D. melanogaster*, repair by homologous recombination proceeds by a mechanism known as synthesis dependent strand annealing (see section 1.3.2). These ZFNs can be introduced *in cellulo* via non-viral delivery or by viral means. Initial delivery strategies in *D. melanogaster* involved the delivery of ZFN together with a repair template via P-element mediated transformation, a technique that utilises transposases to achieve stable expression of recombinant genes in transformed cells (Bibikova *et al.*, 2002; Bibikova *et al.*, 2003). One drawback to this technique was the requirement for lengthy recombinant gene-editing plasmid construction, therefore a welcome breakthrough occurred when it was demonstrated that efficient homologous recombination could be achieved by injection of ZFN messenger RNA (mRNA) and donor template DNA into embryos (Beumer *et al.*, 2008). ZFN nucleases have since been adopted for the induction of complex genomic alterations in mammalian cells. These include large deletions induced by introduction of two DSBs flanking up to 15 mega base pairs (Mbps) of DNA target intended to be deleted (Lee, Kim and Kim, 2010). Translocations have been induced by introducing two DSBs on different chromosomes (Brunet *et al.*, 2009; Simsek *et al.*, 2011) and inversions on chromosomal sequences induced by two DSBs on the same chromosome (Lee *et al.*, 2012b).

1.3.3 ZFNs to Target Chronic Genetic Lung Disease

Monogenic lung diseases, including cystic fibrosis (CF), are a viable candidate to be targeted using genetherapy approaches due to uncomplicated nature of their underlying mutations. CF-causative mutations are recessive, requiring two copies of mutant allele to result in disease. CF causative mutations can be grouped into six classes, which affect the cystic fibrosis transmembrane regulator chloride ion channel (CFTR) gene. Class I mutations are characterised by an absence of CFTR protein

production due to dysfunctional translation caused by a premature stop codon. Drugs targeting these mutations are termed 'read-through agents', they rescue protein synthesis by overcoming the premature stop codons (King et al., 2022). Class II mutants are characterised by aberrant trafficking of channel protein to the cellular membrane as a result of mutations that cause translated CFTR protein to be misfolded and subsequently degraded in the endoplasmic reticulum (Cheng et al., 1990). Corrector drugs such as lumacaftor, tezacaftor and elexacaftor, aid in reversing misfolding errors, improving protein trafficking (Miah, Hyde and Gill, 2019). Class III mutations affect channel gating, due to mutations that affect CFTR ATP binding domains, resulting in diminished ATP binding and hydrolysis, necessary for channel closed-dimer and open-dimer conformations, respectively (Molinski et al., 2012). Potentiators such as ivacaftor, potentiate the channel in the open conformation, improving channel function (King et al., 2022). Class IV mutations result in reduced channel anionic conductance. Some of these mutations can be remedied using potentiators such as ivacaftor. Class V mutations hinder CFTR transcription and protein production. Both potentiators and amplifier drugs in the developmental stages as used against these sorts of mutations. Class VI mutations which result in the reduction in CFTR protein stability on the cellular membrane. Theoretical stabiliser modulator drugs that increase cell surface protein stability would be applicable for this mutation (King et al., 2022). The most common CF mutation accounting for up to 60% of cases globally is the Δ F508 mutation, which is classed as a type II mutation (UK cystic fibrosis registry, 2021). Here, in addition to degradation of the aberrantly misfolded protein, there is a reduction in the residency time of the minor fraction of CFTR that gets trafficked to the cellular membrane. Furthermore, the Δ F508 mutation caused by a 'CTT' deletion results in a synonymous change in codon 507 which encodes isoleucine; this alters the secondary structure of CF mutant mRNA, leading to a decrease in the translation efficiency (Lazrak et al., 2013).

ZFNs have been specifically applied to study and treat cystic fibrosis and lung disease on numerous occasions. In one instance, ZFNs were utilised in successfully targeting the intergenic regions of human *CFTR* (*hCFTR*) gene. In this instance, DSB induction was up to 203 base pairs upstream of Δ F508 (Lee *et al.*, 2012a). Here, a Pair ZFNs encoded in two separate plasmids were co-transfected into human bronchial epithelial (BE) and cystic fibrosis tracheal epithelial (CFTE) cell lines homozygous for Δ F508. Low insertion-deletion (indel) frequency (~8%) and lower HDR efficiency (<1%) were achieved when codelivered with a donor plasmid, as determined by allele-specific PCR and restriction digestion reactions, respectively (Lee *et al.*, 2012a). Of note, this was the first study to report HDR taking place at a lengthy distance from the ZFN DSB, albeit at low gene-editing efficiency which can be explained by the edit-to-target distance and donor template heterology to wild-type genomic DNA. Also, in this

study the NHEJ efficiency reported was nearly seven-fold greater than that reported in a similar study where ZFNs were utilised to target *hCFTR* (Maeder *et al.*, 2009).

In another study, ZFNs were used to target endogenous mutant *CFTR* in CF patient-derived induced pluripotent stem cells (iPSCs). The iPSCs used here contained compound heterozygous Δ F508 and I507del mutations in exon 11. The ZFNs pairs in this study were co-delivered with a *CFTR* donor template, demonstrating preferential restoration of the more homologous allele over the more heterologous allele, which was sufficient to revert the CF phenotype (Crane *et al.*, 2015). In summary, these gene-edited iPSCs displayed mainly monoallelic corrections, had no occurrence of off-target events in predicted sites, retained pluripotency, retained similar gene expression patterns and displayed restored mutant *CFTR* function (Crane *et al.*, 2015).

Previously mentioned *CFTR* correction strategies focused on specific CF-causative mutations. However, a wider-spectrum approach to treating a majority of CF-causing mutations was performed on the Δ F508 CF patient-derived CFBE41o- cell line. Here, ZFNs co-delivered with a *CFTR* super exon donor template, which encodes *CFTR* exon 11-27 sequences, achieved an integration efficiency of up to 10% when the optimum equimolar ZFN-to-donor template ratio was adopted during transfection (Bednarski *et al.*, 2016). HDR events were mainly monoallelic confirmed by allele-specific (AS) PCR, and Ussing chamber assays confirmed functional chloride channels in corrected clones (Bednarski *et al.*, 2016). This study was significant because it achieved donor template integration into the endogenous *CFTR* site which enabled physiological regulation by the endogenous promoter and also permitted the correction of a wider range of CF-causing mutations. However a substantial prevalence of donor off-target integration was observed, possibly due to the high rate of NHEJ of CFBE41o cells or the inadequate specificity of ZFNs used (Bednarski *et al.*, 2016).

Another attempt at gene-editing was performed using ZFNs involved the use of the publicly available oligomerised pool engineering (OPEN) protocol for making customized ZF arrays to design ZFNs that target a variety of human genes including *hCFTR*, achieving ~ 1.2% targeting efficacy in this locus (Maeder *et al.*, 2008). Another approach made use of ZFNs co-delivered with integration deficient lentiviruses to achieve genetic modification in up to 50% of transduced cell population (Lombardo *et al.*, 2007). Equally, F s were used to generate DSBs in 'safe harbour' genomic locations such as adeno-associated virus integration site 1 (AAVS1), which permitted targeted integration and expression of a drug-inducible transgene into this site, while reducing interference from neighbouring gene expression (Hockemeyer *et al.*, 2009).

The use of ZFNs for the purpose of gene-editing comes with challenges including the requirement for the dimerization of the FokI domains of the ZFN pairs, which can be a complex process resulting in

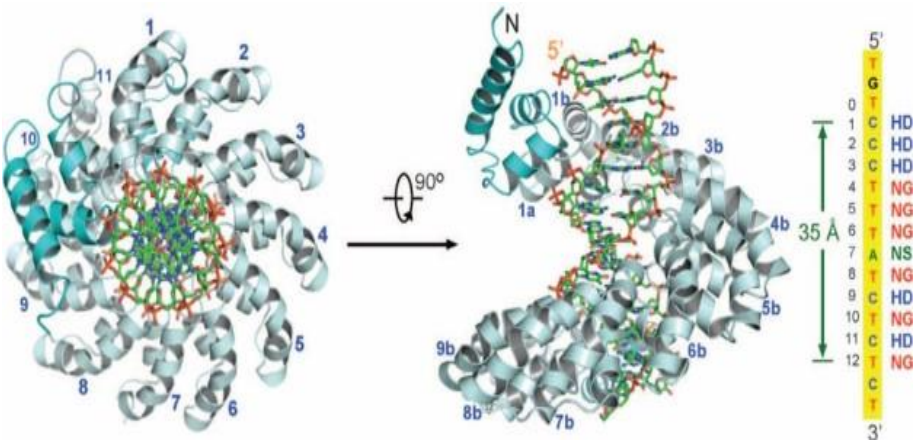
lower editing efficiencies (Bednarski *et al.*, 2016). Equally specificity issues result in some zinc fingers binding with equal affinity to off-target triplet codons, which in most instances is caused by a single member of the ZFN pair (Beumer *et al.*, 2006). To remedy this specificity concern, addition of extra zinc fingers and separation of zinc finger modules pairs with a short linker has been proposed and shown to improve specificity and reduce off-target effects by up to a hundred fold (Moore, Klug and Choo, 2001). Equally introducing substitution mutations in the dimer interface of the cleavage domain which reduces homodimersiation by up to 40-fold but permits heterodimersation has been shown to improve targeting specificity. Here off-target frequency was determined by measuring phosphorylated histone DNA damage biomarkers, which declined from 18.9% in wild-type to 1.7% in modified ZFNs (Miller *et al.*, 2007; Lee *et al.*, 2012a). Despite the modest success of ZFN-based strategies in the context of gene-editing, the constant drive to improves these strategies invariably lead to the emergence of Transcription activator-like effector (TALE) nucleases gene-editing technology.

1.4 Transcription Activator-Like Effector Endonucleases

Another strategy to achieve gene-editing makes use of the TALE protein family, key virulence factors secreted by *Xanthomonas* bacteria that function primarily to alter host cell characteristics by mimicking hosts eukaryotic transcriptional factors (Gu *et al.*, 2005; Kay *et al.*, 2007; Kay and Bonas, 2009). Consequences of TALE gene induction includes developmental changes in hosts such as cell enlargement and division, which contribute to disease symptoms (Kay and Bonas, 2009).

TALE nucleases (TALENs) comprise of a non-specific FokI nuclease fused to a customisable DNA binding domain. The DNA binding domain comprises of highly conserved repeats derived from TALEs (Joung and Sander, 2013). TALEs possess a central domain of tandem repeats, nuclear localisation signals, and an acidic transcriptional activation domain. All TALE family members are highly conserved, varying usually in the amino acid sequence and repeat number, hence the number and pattern of repeats in TALEs underlines their specificity (Herbers, Conradsstrauch and Bonas, 1992).

A



B

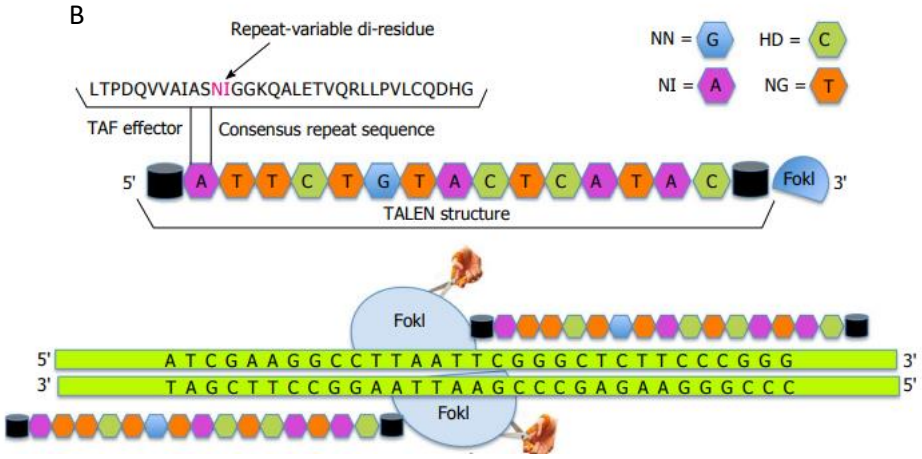


Figure 1.4.1 Transcription activator-like effector nucleases. (A) The general structure of DNA-bound TALENs. DNA's major groove is bound by the super helical structure of TALEN repeats. Each repeat is organized into short (a) and long (b) α helices connected by a short loop where the two hypervariable di-residues are located. The DNA sequence of the sense strand (highlighted in yellow), the repeat numbers (numbers on the right), and the associated hypervariable di-residues (blue, red and green-coloured amino acid di-residues) in the TALE repeats are displayed on the right. The repeat domains of TALEs are flanked by cyan-coloured N- and C-terminal helices. Two perpendicular views are presented and the DNA duplex is depicted as sticks in both. (B) The FokI nuclease effector domains are fused to the TALE DNA binding domains to form TALENs. The DNA binding domains comprises of repeat sequences and hypervariable di-residues which determine target DNA nucleotide specificity. Since FokI can only function as a dimer, two TALENs are built to place the FokI nuclease domains close to the genomic target regions. TALEN dimerization causes DSBs that are repaired by either an error-prone non-homologous end joining route or homology-directed repair, if a donor template is supplied. Taken and adapted from (Deng et al., 2012; Raikwar, Chaurasia and Mohan, 2016).

The repeat region of TALEs consists of a 34 amino acid repeat which is almost identical between TALEs except for the residues at position 12 and 13, commonly referred to as the hypervariable residues, responsible for recognition of a single DNA base (table 1.4.1) (Boch *et al.*, 2009). The final repeat at the C-terminal end shares sequence similarity only in the first 20 amino acids with other repeats, therefore, it is commonly referred to as a half-repeat.

Computational analysis of the binding specificities of naturally occurring TALEs revealed that the number of repeats correlates to the number of base pairs in the target sequence, hence one repeat unit binds one DNA Watson-Crick base pair (Moscou and Bogdanove, 2009). Most TALE repeat arrays used in previous reports use NN, NI, HD and NG dipeptides for the recognition of G, A, C, T DNA nucleotides respectively (Joung and Sander, 2013), however NK has been reported to be more specific at recognising G than NN repeats, which also recognise A (Miller *et al.*, 2011). The base specificity of the hypervariable residues is summarised in table 1.4.1.

Table 1.4.1. The DNA nucleotide specificity of hypervariable amino acid residues within TALE repeats

Hypervariable residues	Base specificity
HD	C
NG	T
NI	A
NS	A, C, G, T
NN	A, G
IG	T

TALENs had generated much interest and excitement due to their ease of design which makes use of a simple protein-DNA code that correlates individual DNA binding TALE repeat domains to specific DNA nucleotides. To this end they have commonly been deployed as a gene-editing tool in the modification of human cells (Cermak *et al.*, 2011; Reyon *et al.*, 2012) and pluripotent stem cells (Hockemeyer *et al.*, 2009).

1.4.1 TALENs to Target Chronic Genetic Lung Disease.

There have been a few instances where TALENs have been deployed as a gene-editing tool in the context of lung disease, including CF (see chapter 3). In one instance TALENs were utilised to enhance

short DNA fragment-mediated correction of Δ F508 CFTR in CF-iPSCs. Here, a 491bp homologous template was co-delivered paired with TALEN-encoding plasmids, demonstrating high HDR efficiency of 10%. Off-target sites were screened by southern blot were undetectable, confirmed by Sanger sequencing of parental and corrected clones. Finally, corrected iPSCs co-cultured with human airway epithelial cells to promote trans-differentiation, displayed restored chloride channel function, as well as successful differentiation into cells with inflammatory cell features, similar to airway epithelial differentiation (Suzuki *et al.*, 2016). The significance of this study was the use of seamless gene correction strategy in iPSCs without the use of an integrating reprogramming vehicle, as well as the successful differentiation of iPSCs into inflammatory cells to permit the development of cell therapies for regenerating an intact haematopoietic system in CF patients (Suzuki *et al.*, 2016; Fleischer *et al.*, 2020).

TALEN-containing plasmids and ssODN HDR templates were transfected into MHHi002-A cells, which are human CF-iPSCs. Heterozygous-targeted clones were observed, hence one corrected clone was chosen for further characterisation, confirming mono-allelic correction of Δ F508. Most importantly, corrected MHHi002 cells maintained pluripotency, as indicated by the expression of OCT4, NANOG, SSEA-4 and TRA-1-60 cellular markers. Hence, results from these two studies focusing on the use of gene-editing to correct CF-iPSCs suggests that these corrected iPSC that have the potential to be redifferentiated into any cell-type of choice, can be used to regenerate systems affected by CF.

Other gene-editing attempts on the *CFTR* locus using TALENs include their delivery in viral vectors such as helper-dependent adenoviral vector, in an attempt to target wild-type *CFTR* repair template to the AAVS1 site (Xia *et al.*, 2019). This achieved ~ 5% editing efficiency in a human bronchial epithelia cell line. In another instance, TALENs were utilised to achieve seamless correction of *CFTR* by subsequent excision of the selection cassette, post genome editing event (Fleischer *et al.*, 2020). This strategy made use of a piggyBac (pBAC) transposase system to excise the selection cassette post-gene-editing, achieving HDR efficiencies of up to 10% determined by Sanger sequencing of recombinant clones (Fleischer *et al.*, 2020). TALENs have also been used to generate reporter CF iPSCs cell lines to study cellular differentiation and lung disease (Hawkins *et al.*, 2017; Mou *et al.*, 2012).

A broader approach to treating a wide range of CF disease-causing mutations has been attempted. Here TALEN-mediated gene-editing was used to correct the *CSF2RA*-mediated form of hereditary pulmonary alveolar proteinosis. The complete *CSF2RA* gene sequence, driven by a CMV early enhancer and chicken beta actin promoter, was integrated into intron 1 of the protein phosphatase 1 regulatory subunit 12C (*PPP1R12C*) gene (aka AAVS1) of iPSCs. Transfection efficiency reached 35% for transfected iPSCs and after puromycin selection, and 5% of clones had bi-allelic modification with

the rest being mono-allelic. Finally gene-edited iPSCs showed classical macrophage marker profiles and restored function (Kuhn *et al.*, 2017). Although the AAVS1 site has been well-characterised, nonvariegated transgene expression and even transgene silencing within this same locus caused by DNA methylation has been described (Ordovás *et al.*, 2015), hence a non-integrating vector approach remains the safest option for genome editing. Highlighted in this study was the potential for physiological transgene expression under control of endogenous promoter through integration of a transgene immediately downstream of a wild-type gene promoter (Kuhn *et al.*, 2017). Also demonstrated was the use of a single wild-type *CSF2A* transgene as a less cumbersome strategy than targeting specific mutations.

Another study demonstrated that bacterial delivery was a viable strategy for delivering TALEN cargo. Here TALEN pairs were delivered using a *P. aeruginosa* type III secretion system into murine embryonic stem cells (MESCs), human ESCs (hESCs), and human iPSCs (hiPSCs) (Jia *et al.*, 2015).

Altogether the use of TALENs for the purpose of gene-editing bears some advantages over ZFNs, including a significantly shorter time to manufacture a functional nuclease, as well as a significant improvement in specificity and toxicity (Shrock and Güell, 2017). However, while TALENs are a substantial improvement over ZFNs, their modular design makes construction difficult. Equally, the multiple repetitive sequences required to be tailored for targeting can be challenging to clone (Shrock and Güell, 2017). In summary, TALENs are a promising gene-editing technique, however, more research is needed to further advance this technology. Hence, development of the CRISPR-Cas9 geneediting platform was a welcome advancement.

1.5 CRISPR-Cas System

Prior to the discovery of CRISPR-Cas9 gene-editing strategy, earlier gene-editing approaches included use of ZFNs, TALENs, and other gene-modification approaches.

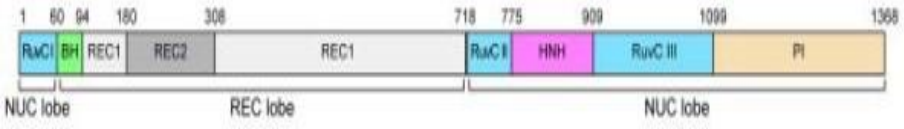
The clustered regularly interspaced short palindromic repeats (CRISPR)- associated proteins (Cas) system was initially identified when repeat sequences in *Escherichia coli* (*E. coli*) separated by nonrepeating sequences in a nearly palindromic manner were described in 1987 (Ishino *et al.*, 1987). Over the following years, similar repeat sequences were described in *Haloarcula* and *Haloferax* archaea (Mojica, Juez and Rodriguez-Valera, 1993). Although the function of these adjacent repeats was initially unknown, they were subsequently confirmed in 40% of bacteria and 90% of archaea (Mojica *et al.*, 2005). The CRISPR acronym was then proposed for the description of such sequences (Lino *et al.*, 2018). In 2005 three separate studies all independently reported the presence of foreign DNA, particularly of bacteriophage origin, present in the CRISPR non-repeat regions, subsequently

hypothesised to play a role in adaptive immunity within prokaryotes (Mojica *et al.*, 2005; Pourcel, Salvignol and Vergnaud, 2005; Bolotin *et al.*, 2005).

To date, over 3000 genes have been shown to be associated with disease causing mutations in humans (Cox, Platt and Zhang, 2015). Hence various applications of CRISPR-Cas9 system are changing the landscape of medical research in many areas including the identification and understanding of genetic disease mechanisms, validating disease targets and in the development of animal models for studying diseases (Findlay *et al.*, 2014; Gilbert *et al.*, 2014; Wang *et al.*, 2014b; Shalem *et al.*, 2014). The procedure for performing gene-editing using this CRISPR-Cas system shall be explained in the next section.

The Cas9 endonuclease components comprises of nuclease lobes and a recognition lobe (figure 1.5.1). The nuclease lobes contain the HNH and RuvC domains are responsible for nuclease activity, while Cas9 recognition lobe is essential for binding gRNA and DNA. The endonuclease is guided by an RNA duplex consisting of CRISPR (cr) RNA and trans-activating (tracr) RNA) gRNAs consist of a target-specific sequence termed 'the spacer' and a constant scaffold part comprised of distinct conserved motifs namely, the stem, nexus and hairpins. Although a synthetic single guide RNA (sgRNA) chimaera that mimics the crRNA: tracrRNA duplex can be used instead to guide Cas9 endonuclease (Briner et al., 2014; Nishimasu et al., 2014).

A



B

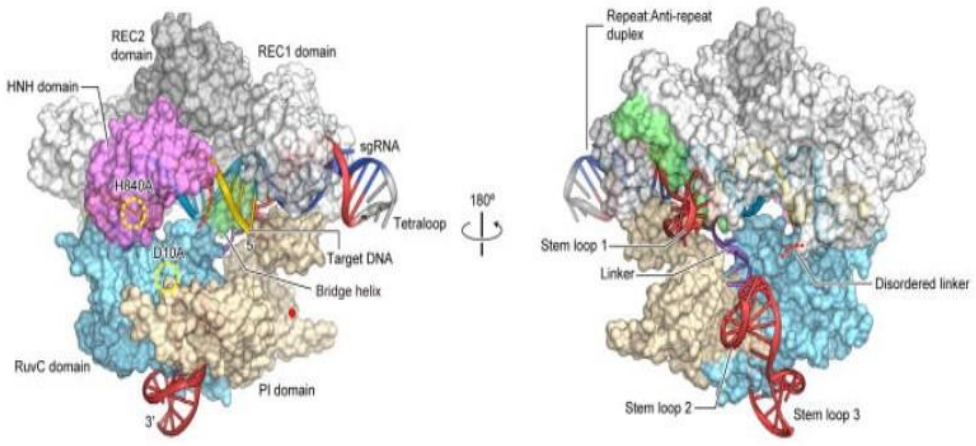


Figure 1.5.1 Structure of Cas9-gRNA-DNA complex. (A) The Cas9 endonuclease components comprises of two nuclease (NUC) lobes and a recognition (REC) lobe. Three sections make up the REC lobe: a lengthy -helix known as the Bridge helix (residues 60–93), the REC1 domain (residues 94–179 and 308–713), and the REC2 domain (residues 180–307). The PAM-interacting (PI) (residues 1099–1368), HNH (residues 775–908), and RuvC (residues 1-59, 718–769, and 909–1098) domains make up the NUC lobe. (B) At the junction of the REC and NUC lobes, the negatively-charged gRNA: target DNA heteroduplex is accommodated in a positively-charged groove. In the NUC lobe, the three separate RuvC motifs (RuvC I–III) that make up the RuvC domain are assembled, where they interact with the PI domain to create a positively charged surface that interacts with the 3' tail of the gRNA. This interaction pemits DSB induction 4 nucleotides upstream of the PAM. Taken and adapted from (Nishimasu et al., 2014).

1.5.1 Mechanisms of Action

The CRISPR-Cas system can be defined as components of a bacterial adaptive immune system used in self-defence against foreign DNA (Wang *et al.*, 2016). In CRISPR-array DNA sequences, repeats are normally interspaced by non-repeat DNA sequences called spacers, commonly found in bacteria and archaea (Adli, 2018). After its initial discovery, the CRISPR Cas9 system was subsequently deployed as a tool for gene-editing which revolutionised the field of molecular biology, as a result of its potential for editing the human genome and treating diseases, while maintaining simplicity and high specificity.

The CRISPR-Ca9 system functions as genomic memory designed to convey resistance against invading foreign phage DNA, to acquire genetic memory of invading phage genomic material (figure 1.5.1.1). This same memory is transcribed into RNA which is then used by the Cas endonuclease to impose DSBs on the invading DNA. The mechanism of action of CRISPR-Cas systems is usually divided into three stages: (1) adaptation or spacer integration, (2) processing of the CRISPR locus' primary transcript (precrRNA) to produce a mature RNA molecule containing the spacer and variable regions corresponding to 5' and 3' fragments of CRISPR repeats, and (3) DNA (or RNA) interference (Makarova and Koonin, 2015). The insertion of spacers into CRISPR cassettes requires only two proteins, Cas1 and Cas2, which are found in the vast majority of known CRISPR-Cas systems (Makarova and Koonin, 2015). Cas1's endonuclease activity is necessary for spacer integration, while Cas2 appears to have a nonenzymatic role (Makarova and Koonin, 2015; Nuñez *et al.*, 2014). The Cas1-Cas2 complex is a highly conserved CRISPR-Cas "information processing" module that appears to be semi-autonomous from the rest of the system. The second stage involves expression of the CRISPR-locus, initiated from an upstream leader sequence which contains promoter elements, followed by the subsequent conversion of precrRNA transcripts to mature guide crRNAs, which is carried out either by a specialised RNA endonuclease complex or by an alternate mechanism involving bacterial RNase III and

other RNA species (Deltcheva *et al.*, 2011). Cas proteins then bind to the mature crRNA to form the effector complex, which targets the cognate DNA or RNA (Makarova and Koonin, 2015).

CRISPR-Cas systems can be generally classified into six types and two classes (Wright, Nunez and Doudna, 2016). The type I and type III systems require multi-complex proteins directed by CRISPR-RNA (crRNA) to induce cleavage of a DNA target, whereas the type II system only requires a single Cas9 protein. This additional operational simplicity makes the type II system a more attractive choice for gene-editing applications.

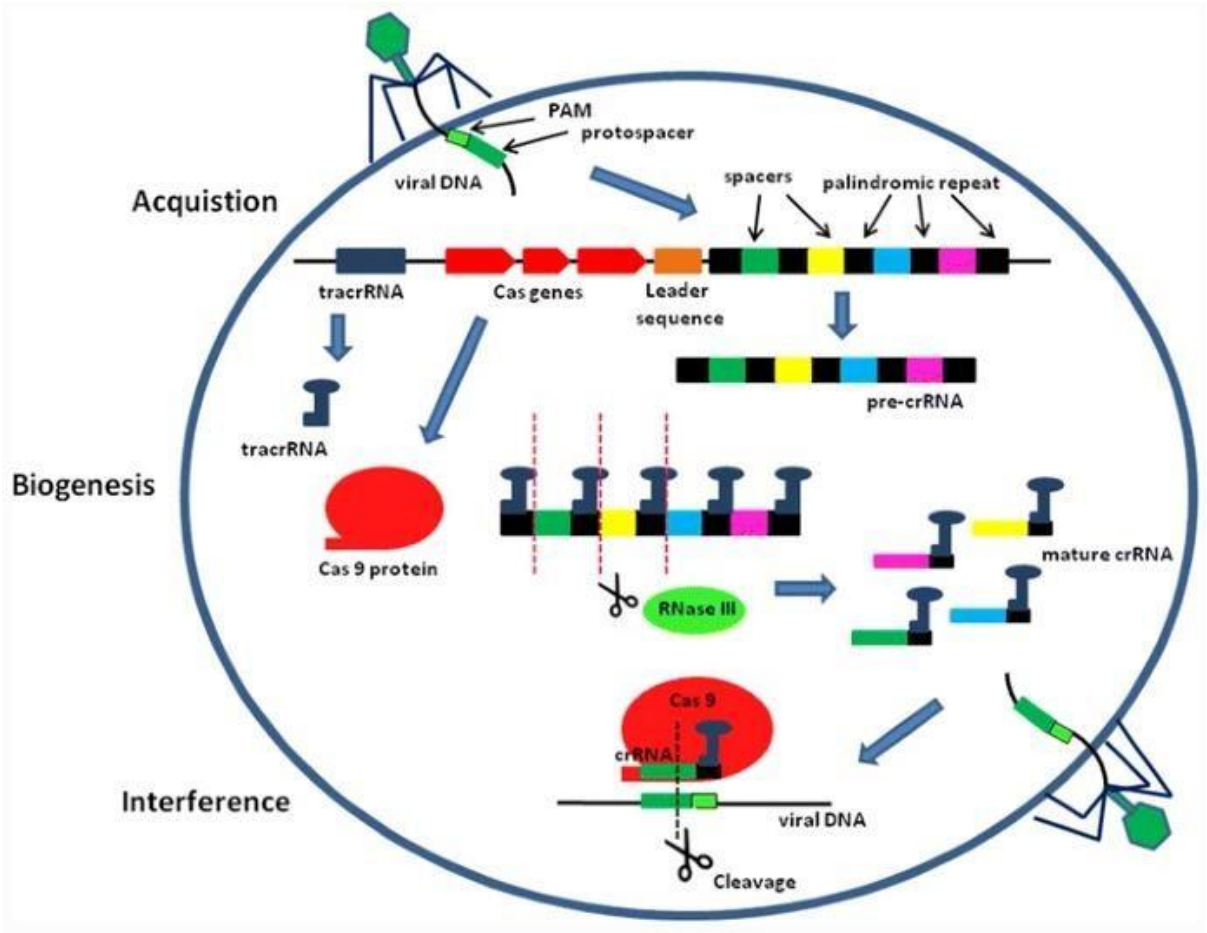


Figure 1.5.1.1 The CRISPR-Cas9 bacterial adaptive immune system functions in three sequential stages. In the acquisition stage, invading viral DNA is incorporated into bacterial genomic CRISPR arrays. Biogenesis proceeds with the transcription of Cas9 protein component as well as RNA components including pre-crRNA and tracrRNA. RNase III and tracrRNA bound to pre-crRNA result in the processing of this component into a mature crRNA version. Mature crRNA hybridised to tracrRNA binds Cas9, and together this ribonucleoprotein complex mediates DSB induction of foreign DNA in a process commonly referred to as Interference. Taken and adapted from (Hryhorowicz *et al.*, 2017).

The type II system comprises of three components, Cas9, trans-activating RNA (tracrRNA) and CRISPR array sequences. Following invasion and incorporation of foreign DNA into CRISPR arrays, pre-crRNA is transcribed from the same array locus, then tracrRNA binds to pre-crRNA repeat sequences, finally cleavage by RNase III forms the mature crRNA product (Deltcheva *et al.*, 2011). It is this mature crRNAtracrRNA complex (gRNA) that loads onto Cas9 endonuclease and guides the Cas9 protein to its DNA target sequence (figure 1.5.1.1). The gRNA-Cas9 ribonucleoprotein complex recognises a triple-base pair protospacer adjacent motif (PAM) located on the target DNA sequence, resulting in subsequent gRNA hybridisation, which then enables Cas9-mediated cleavage of double-stranded DNA (dsDNA) (Wang *et al.*, 2016; Dever *et al.*, 2016). The resulting DSB initiate one of two conserved DNA repair mechanisms namely NHEJ or HDR (figure 1.3.1.1) (Dever *et al.*, 2016). A requirement for gene-editing by HDR is the co-delivery of a DNA donor template which cellular recombination machinery utilises to fix the DSBs, hence enabling precise nucleotide changes within a target genomic DNA locus (Dever *et al.*, 2016).

Due to its simplicity, the type II CRISPR-Cas9 system is the most commonly used system for genomeediting in research consisting of *Streptococcus pyogenes* Cas9 endonuclease and gRNA. (Wang *et al.*, 2016). Initial gene-editing attempts using the type II CRISPR-Cas9 system that occurred in 2012, aimed to introduce DSBs and to develop a truncated, chimeric version of the crRNA-tracrRNA duo, termed a single-guide RNA (sgRNA) still capable of inducing efficient DNA cleavage (Jinek *et al.*, 2012). This further affirmed the convenience of this CRISPR-Cas9 gene-editing system. Subsequently a few pioneering studies (Cong *et al.*, 2013; Mali *et al.*, 2013a; Cho *et al.*, 2013) helped lay the foundation for CRISPR-Cas9 as a global tool for gene-editing. For example Mali et al. 2013 revealed that tethering transcriptional activation domains to either TALENs, null Cas9 or gRNAs were capable of inducing transcription in tdTomato reporter construct. It was equally shown that targeting specificity is sgRNA dependent and that sgRNA-Cas9 complexes can tolerate 1-3 mismatches. Finally the usage of off-set nicking using Cas9 nickase was demonstrated, resulting in higher indel and HDR rates when generating a 5’ overhang, as opposed to a 3’ overhang (Mali *et al.*, 2013a).

In the use of CRISPR-Cas9 system for the purpose of gene-editing, it is necessary to generate precise genomic alterations using a DNA donor template by HDR. Therefore, strategies employed to facilitate this objective includes the design of a single stranded oligodeoxynucleotide (ssODNs) HDR templates which complementary base-pairs to the non-target strand, or that are the same sequence as the target strand. It is equally important to include an asymmetric 3’ P M-proximal overhang by creating a shorter PAM-distal overhang (Richardson *et al.*, 2016; Okamoto *et al.*, 2019). A major challenge to HDR efficiency is the re-cutting of donor template which could be remedied by the introduction of a PAM-altering silent mutation at PAM-proximal sites, leaving PAM-distal sites unaltered to avoid

detrimental effects (Okamoto *et al.*, 2019). Optimal ssODN lengths have been reported to range from 75 to 136 bases (Okamoto *et al.*, 2019; Richardson *et al.*, 2016), with at least 30 to 35 bases of perfectly-matched homology arms on either side (Okamoto *et al.*, 2019). Some other strategies to improve HDR include use of an RNP platform in place of plasmid transfection (Okamoto *et al.*, 2019), and use of cell cycle arrest reagents such as nocodazole to arrest cell cycle in the S and G2 phase where HDR occurs (Lin *et al.*, 2014). Equally, the use of reagents that inhibit NHEJ and enhance HDR such as brefeldin has also been reported (Yu *et al.*, 2015).

1.5.2 Detecting Genome-Editing

Following genome editing, the efficiency of this event can be determined by a variety of assays including T7E1, Sanger sequencing and next-generation sequencing (NGS). T7E1 assay is derived from T7 bacteriophage which detects heteroduplex mismatches in DNA and cleaves at that point (figure 1.5.2.1) (Sentmanat *et al.*, 2018). This assay has the advantage of convenience however questions have arisen about its accuracy. T7E1 assay has a tendency to underestimate cleavage efficiency compared to NGS by up to six fold in some instances (Sentmanat *et al.*, 2018). This can be explained by the reliance of T7E1 efficiency on the variety of indel species present as opposed to the total number of indels present at the DSB locus. Hence, T7E1 was proven to be unreliable for gRNAs with efficiencies lower than 30% (Sentmanat *et al.*, 2018).

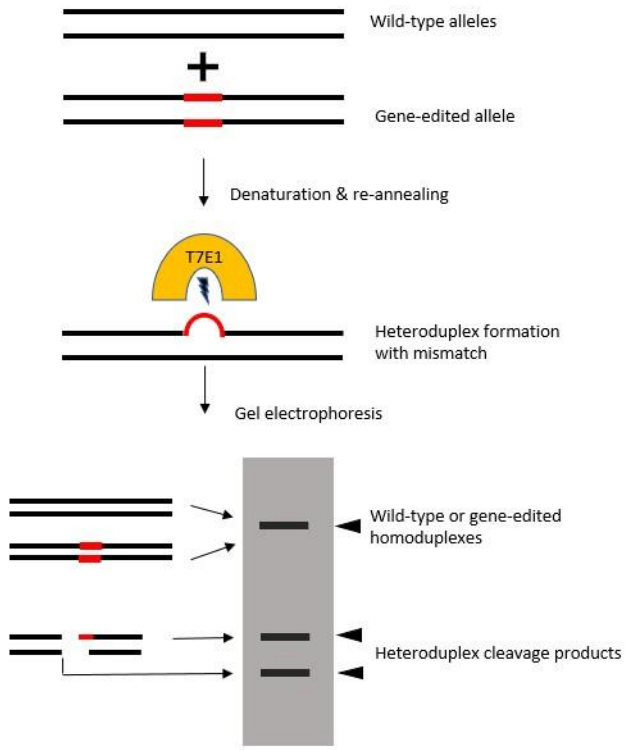


Figure 1.5.2.1. Overview of the T7E1 assay used to detect CRISPR-Cas9 mutations in gene-edited cell lines. Wild-type and gene-edited PCR amplicons are denatured and then re-annealed to form heteroduplexes which

T7E1 identifies and cleaves at mismatch sites (red loop), leaving homoduplexes intact. Cleavage fragments of gene-edited alleles can then be observed following agarose gel electrophoresis.

Sanger sequencing can also be applied to check for generic presence of indels, but if there is need for a more specific identification of indel pools present, then PCR amplicons of gene-edited DNA can be cloned into plasmid vectors and; transformed into bacteria which will enable single colonies to be sequenced (Sentmanat *et al.*, 2018; Maurisse *et al.*, 2010). One drawback to this technique is the labour intensity of picking individual clones to identify edited colonies in the case of low editing efficiency and the need for a large sample size to get an accurate representation of the edited pool present. An alternative to sequencing single clones is the use of NGS which is faster and less time consuming than sequencing single clones. NGS and single clone Sanger sequencing were found to give similar frequencies of the different indels identified, indicating both techniques are suitable for this purpose (Sentmanat *et al.*, 2018).

1.5.3 Off-Target Events

Despite its widespread use and countless benefits, CRISPR editing has several disadvantages. Off target effects are one of the most important factors to consider when planning a CRISPR project (Zhang *et al.*, 2015). Hence, gRNA hybridisation to partially homologous genomic areas must be minimised or eliminated by rational gRNA sequence design (Doench *et al.*, 2016). Many bioinformatics methods have been created to predict off-target sites *in silico* and to facilitate on-target gRNA activity, however, not all off-target sites are predictable. A large variety of experimental investigations have been conducted to investigate Cas9 specificity and the spectrum of its off-target activity in various experimental setups, as well as the factors that influence specificity.

Cas9 can tolerate imperfect base complementarity. However, the amount and location of mismatches throughout the 20mer gRNA sequence are important determinants of specificity. Single base mismatches are frequently ignored by Cas9, resulting in off-targeting, according to genome-wide studies. Two mismatches are tolerable, but they limit cleavage efficiency, whereas three mismatches (and beyond) almost entirely eliminate off-target cleavage. When (many) mismatches, whether dispersed or connected to one another, are found in the P M proximal area (12 nucleotide 'seed' region upstream of the PAM), they have a greater impact on complementarity (Hsu *et al.*, 2013; Anderson *et al.*, 2015). Cas9 then still hybridises to off-target sites, but it does not induce DSBs, since it needs a high degree of pairing to attain the necessary 3D conformation to dock onto target DNA and cleave it (Sternberg *et al.*, 2015).

1.6 CRISPR-Cas9 Gene-Editing Platforms

1.6.1 mRNA Plus gRNA

To perform gene-editing using the Crispr-Cas9 system, it is possible to introduce the Cas9 protein component in the form of mRNA, alongside the gRNA component. The sequence encoding Cas9 mRNA used in this platform includes fifty to two hundred poly-adenine residues at the 3' terminal. The Cas9 mRNA is then subsequently generated *in vitro* using a promoter such as the T7 promoter (Redel *et al.*, 2018). This mRNA-based platform is a promising approach to performing genome-editing, due to its transient effects. This characteristic transient nature ensures that there is a decreased probability of off-target mutagenesis when compared to plasmid Cas9-mediated gene-editing (Liu *et al.*, 2019). One challenge with mRNA-based gene-editing is the inadequate permeability of the cell membrane to mRNA, which is a significant obstacle (Liu *et al.*, 2019). This is attributable to the significant discrepancies in the structure as well as the stability of mRNA and DNA, which hinders the permeability of the former through the cellular membrane (Liu *et al.*, 2019). Finally, *in vitro*-transcribed (IVT) mRNA is capable of inducing a potent type I interferon and tumour necrosis factor alpha (TNF-α) innate immune reaction; the immunogenic potential of these therapeutics could have potential negative impacts on target cell viability and must be carefully considered during their design (Wienert *et al.*, 2018). Attempts to mitigate this IVT RNA-induced innate immune response have been described using chemically modified-IVT RNA with 2'-O-methylatyed nucleosides,

demonstrated up to 10-fold reduced TNF-α expression from dendritic cells measured by an enzymelinked immunosorbent assay (Karikó *et al.*, 2005). This reduction in TNF- α expression was attributable to the blockage of innate immune response-inducing Toll-like receptor activation by pathogenassociated molecular patterns (PAMPs) (Karikó *et al.*, 2005).

1.6.2 Ribonucleoprotein

Here the Cas9 protein is generated recombinantly and then complexed with a gRNA *in vitro*, before being delivered as a ribonucleoprotein (RNP) complex into a target cell. The direct delivery of a CRISPRCas9 and gRNA RNP has numerous advantages such as the rapid onset of gene-editing by removing the need for transcription and translation (Zhang *et al.*, 2021). Another advantage of this approach is its transient nature, which reduces the incidences of off-target events and insertional mutagenesis, while promoting high editing efficiency (Zhang *et al.*, 2021). Equally, RNP gene-editing permits the genomic alteration of cells with intrinsically low transcription and translational activity, enabling progress in the gene-editing of ESCs and iPSCs (Zhang *et al.*, 2021). The CRISPR-Cas9 RNP platform has been reported to outperform mRNA- and plasmid-encoded versions of the endonuclease in the context of gene-editing, attributable to rapid activity and higher local concentrations of Cas9

and gRNA (Kouranova *et al.*, 2016). This RNP platform has the most reliable activity in cell lines, embryos, as well as *in vitro*, hence it is frequently used for initially validating gRNA designs in *in vitro* assays (Kouranova *et al.*, 2016).

1.6.3 Virus-Encoded Platform

CRISPR-Cas9 gene-editing machinery can be encoded in viral DNA or RNA, then transduced into a target cell by packaging it into viruses such as adenoviruses, adeno-associated viruses, retroviruses, herpesviruses, etc. Recombinant viruses encoding CRISPR-Cas9 gene-editing components are generally produced by introducing up to three separate plasmids into a complementing producer cell line such as human embryonic kidney 293 for virus packaging. The multiple plasmids required for viral transduction include (i) A shuttle plasmid that encodes the CRISPR-Cas9 transgene flanked by terminal repeats to permit the packaging of the recombinant viral DNA into the capsid, (ii) A plasmid that encodes genes required for capsid formation as well as replication and (iii) A viral packaging cell line (Lee *et al.*, 2017a). The components encoded by these plasmids are subsequently assembled into a recombinant virus, then purified from the packaging cell culture supernatant.

The use of CRISPR-Cas9-encoded viral gene-editing has some advantages. For example, Adeno-associated viral vector (AAV) and retroviruses can integrate into the host genomic DNA, which enables long-term gene expression in dividing cells. AAVs are significantly less immunogenic than most other virus. Lentiviruses a type of retrovirus, has potential to be pseudotyped with other viral proteins, allowing for engineering and altering viral cellular tropism. Equally AAV numerous serotypes have been demonstrated to enable tissue-specific targeted gene delivery (e.g. AA6 for murine muscle and AAV9 for murine brain) (Xu et al., 2019). However, in the context of CRISPR-Cas9 gene-editing, host genomic integration could result in unwanted off-target insertional mutagenesis and unwanted long-term genome-editing events. In some cases, more than one viral vector is used for gene-editing due to limited packaging capacity (e.g., AAV has 4.7kb capacity) (Xu et al., 2019).

1.6.4 Plasmid-Encoded Platform

In some instances, the components required for performing gene-editing can be encoded in separate plasmids. This plasmid-encoded gene-editing system is often delivered into host cells with the aid of non-viral nanoparticles, or alternative transfection means. Here, the Cas9 endonuclease component and the gRNA component are expressed from two separate plasmids, meaning the two transcription units are under the control of different promoters and (Mali *et al.*, 2013b). Of note, the Cas9 protein component is generally expressed from a plasmid with an RNA polymerase II promoter, while gRNA expression is driven by a separate plasmid using a more powerful RNA polymerase III promoter suited to non-coding RNA expression (Gao, Herrera-Carrillo and Berkhout, 2018). In general, the usage of

plasmids for the purpose of gene-editing is often advantageous due to the low probability that recombinant plasmid DNA shall be integrated into the host genome. More specifically, the usage of separate plasmids for the expression of gene-editing components has numerous advantages. One such advantage is the experimental flexibility offered by this approach; hence a variety of gRNA expression plasmids could be used in combination with plasmids expressing different Cas9 variants such as wild-type Cas9, Cas9 nickase (single-stranded DNA breakage-inducing) or catalytically-dead Cas9. Equally, this approach permits the facile transfection of cells that already stably express Cas9 with a variety of gRNA-expressing plasmids targeting the same or multiple genomic loci (Mali *et al.*, 2013b). Hence, the use of separate plasmids permits the comparison of gRNA editing efficiencies in cell lines with similarly stable Cas9 expression. One disadvantage, however, of using two separate plasmids in the context of gene-editing is the limited ability to facilitate simultaneous expression of both Cas9 and gRNA components in experiments where this may be desired (Adikusuma, Pfitzner and Thomas, 2017).

In other instances, gene-editing is performed using a single plasmid encoding both Cas9 and gRNA components, under the control of two separate promoters. Using this strategy, the Cas9 protein component is ubiquitously placed under the control of an inducible or constitutive RNA polymerase II promoter, while gRNA expression is driven by an RNA polymerase III promoter (Damdindorj *et al.*, 2014; Gao, Herrera-Carrillo and Berkhout, 2018). An advantage to the use of a single expression plasmid includes the improved control of expression timing of both gene-editing components (Adikusuma, Pfitzner and Thomas, 2017). Equally there appears to be a more even cellular localised concentration of gene-editing components when expressed from one plasmid, in contrast to expression from separate plasmids, where one plasmid could be preferentially localised within the nucleus compared to the other. A consequence of uneven local concentration of both plasmids within the nucleus would invariably lead to nonproportional expression of Cas9 and gRNA components, thereby inhibiting gene-editing efficiency (Zhuo *et al.*, 2021). In the context of multiplexed geneediting, a single plasmid which encodes Cas9 plus two or more gRNAs would be more effective at generating deletions within a single gene or simultaneously targeting separate genes, due to simultaneous expression of both gRNAs (Adikusuma, Pfitzner and Thomas, 2017). There are disadvantages to the inclusion of multiple gene-editing components, in separate transcription units, within a single plasmid. One such drawback would include the increased cargo size of such a plasmid, which would negatively affect cloning and transfection efficiency (Nurmenmedov et al., 2007; Yin et al., 2005).

The use of separate plasmids, as well as separate transcription units within a single plasmid, presents challenges for the production of CRISPR-Cas9 systems. One such issue is the enormous size of gene-

editing cargo components (~4.2 kb for Cas9). Challenges also exist within the gRNA cloning step, where RNA pol II promoters are suitable, while limitations are apparent in the potential and options of RNA pol III promoters which can be used in this process (Tang *et al.*, 2016). To achieve effective, simultaneous expression of Cas9 and gRNA components, attempts have been made to place both components under the control of a single RNA pol II promoter. Here, Cas9 components and gRNA components were interspaced by self-cleaving RNA molecules within a STU. Such self-processing RNA species, including ribozymes and tRNA sequences permit *in cellulo* separation of individual gene-editing components (Gao and Zhao, 2014; Dong *et al.*, 2017). This approach to gene-editing has numerous advantages and has been applied in plant, animal, as well as in human cells, with a modest degree of success (Tang *et al.*, 2016; Tang *et al.*, 2019; Gao and Zhao, 2014; Dong *et al.*, 2017).

1.7 CRISPR-Cas9 Gene-Editing to Target Chronic Genetic Lung Disease

A few studies have explored the potential use of CRISPR-Cas9 system in treating monogenic lung disease. A study in 2015 generated iPSCs from CF patients homozygous for Δ F508 mutation then successfully corrected these cells using gRNA and Cas9 plasmids in combination with a pBAC excisable donor to improve the efficiency while reducing the genetic footprint (Firth *et al.*, 2015). No off-target mutations were confirmed determined by whole exome sequencing and selection cassette-specific PCR. Importantly corrected iPSCs regained normal functional expression of *CFTR*, maintaining their differentiation capacity.

Another study made use of CRISPR-Cas9 gene-editing to correct the *CFTR* locus of intestinal stem cells (ISCs) from a CF paediatric patient homozygous for Δ F508. Cas9 and gRNA plasmids specific for *CFTR* exon 11 were introduced together alongside a donor plasmid encoding WT-*CFTR* and a selection cassette (Schwank *et al.*, 2013). Following gene-editing with two different gRNAs, HDR efficiency was very low ranging from 0.2-0.5%, with mostly heterozygous repairs and functional expression of corrected allele was validated using RT-PCR. Notably only one off-target event was detected for one of the gRNAs confirming the high specificity of this technique in adult stem cells (Schwank *et al.*, 2013).

CRISPR-Cas9 has been applied to correct other monogenic lung defects including in surfactant protein syndromes. Here intra-amniotic *in utero* gene-editing was performed using CRISPR-Cas9 packaged in adenoviruses, delivered into foetal lung of mice, which resulted in successful excision of mutant *SFTPC*^{I73T} gene in pulmonary epithelia, translating into improved survival of adults (Alapati *et al.*, 2019). Another form of the same disease caused by homozygous *SFTPB*^{121ins2} mutation was corrected in infantile iPSCs following the lipofection of Cas9 and gRNA expression plasmids, thus restoring surfactant B production in alveolar epithelial type 2 cells (Jacob *et al.*, 2017).

Another study functionally repaired Δ F508 *CFTR* mutation in intestinal organoids by electroporating prime-editing CRISPR-Cas9 plasmids into strategy, demonstrating increased forskolin-induced organoid swelling compared to CF untreated control organoids. Finally a comparison of editing efficiencies of prime-editing (see chapter 6), base-editing and conventional CRISPR-Cas9 HDR was determined by measuring forskolin-induced swelling sensitive organoids after targeting, revealing an editing efficiency up to 5.7% for prime-edited clones, 9.1% corrected organoids by base-editing and 1.22% by conventional CRISPR-Cas9–mediated HDR (Geurts et al., 2021).

In another study a sheep model of Cystic fibrosis was generated by disruption of *CFTR* gene. Following electroporation of CRISPR-Cas9 constructs into Sheep fetal fibroblasts, then restriction fragment length polymorphism was used to determine targeted *CFTR* disruption, which was achieved in both male and female colonies for exons 2 and 11 at various efficiencies ranging between 6.9% and 34.7% (Fan *et al*., 2018).

Another study analysed repair tracts from a donor plasmid with seven nucleotide variations throughout a 216 bp target area in the human *CFTR* gene using deep sequencing (Hollywood et al., 2016). It was uncovered that there were equally as many uni-directional and bi-directional repair tracts and that 90% of the template-dependent repair tracts were >100 bp in length. Long repair tracts were found to be common, signifying that a single gRNA might be used in combination with several iterations of the same template to generate or correct particular mutations within a 200 bp range, which is around the size of 80% of human exons.

CRISPR-Cas9 gene-editing has been used to restore Alpha-1 antitrypsin (AAT) stable expression in mice by knocking out mutant AAT-encoding gene or knocking in wild-type AAT-encoding gene into the adenovirus safe harbour site (Bjursell *et al.*, 2018; Stephens *et al.*, 2018). CRISPR-Cas9 system has also been deployed for epigenetic modifications to reduce mucus hypersecretion in Chronic obstructive lung disease (COPD). This involved targeting transcription repressors to *SPDEF* promoter using catalytically-dead Cas9 (dCas9), effectively repressing *SPDEF* mRNA and protein expression by up to 50%, which inhibited downstream mucus production-related genes such as *MUC5AC* (Song et al., 2016). Table 1.7.1. summarises the genomic targets of the above-mentioned studies.

Table 1.7.1. summarises the genomic targets of the above-mentioned CRISPR-Cas studies

Target	Delivery Platform	Assayed in	Resource
CFTR$_{\Delta F508}$	Plasmid nucleofection	CF-hiPSCs	(Firth *et al.*, 2015)
CFTR$_{\Delta F508}$	Plasmid lipofection	hISCs	(Schwank *et al.*, 2013)
CFTR KO	Electroporation	Sheep fetal fibroblasts	(Fan *et al.,* 2018)
SFTPB$_{121ins2}$	Plasmid nucleofection	hiPSCs	(Jacob *et al.*, 2017)
SFTPC^{I73T}	Adenovirus intra-amniotic injection	Murine Foetus	(Alapati *et al.*, 2019)
hSERPINA1	Adenovirus intravenous injection	PiZ mice	(Bjursell *et al.*, 2018)
AAT	Adenovirus intravenous injection	C57BI/6J mice	(Stephens *et al.*, 2018)
AAT	Adenovirus intravenous injection	PiZ mouse liver	(Bjursell *et al.*, 2018)

CRISPR-Cas9 has been applied in other diseases including anti-cancer immunotherapy by a mechanism which suppresses proteins that interfere with the immune system such as programmed cell death ligand-1 and 2, expressed on the surface of lymphoma cells (Ju *et al.*, 2019). This same gene-editing technology has been equally deployed as a diagnostics tool, for detection of microRNA biomarkers and for minor allele enrichment in cell-free DNA prior to diagnostic testing in NSCLC patients (Qiu *et al.*, 2018; Aalipour *et al.*, 2018). To perform such enrichment, dCas9 was complexed to a gRNA and used to target mutant DNA strands of low abundance in cell-free DNA samples, permitting a > 20 fold detection of three rare epidermal growth factor receptor mutations present in NSCLC (Aalipour *et al.*, 2018).

1.7.1 Advantages and Challenges to CRISPR-Cas Gene-Editing Usage

The CRISPR-Cas9 system enjoys certain advantages over its predecessors. One such advantage is the use of a short RNA sequence as the specificity –determining element to drive DNA cleavage. This avoids the need for a cumbersome protein engineering process to target a new site in DNA, whereas

in the case of ZFNs and TALENs, nucleases need to be redesigned to target each new DNA target sequence. Thus, the gRNA component reduces time required for CRISPR-Cas9 gene-editing design and implementation. Equally, the requirement for solely gRNA alteration between different genomic targets bears economic cost advantages. Finally, the growing pool of studies adopting the CRISPR-Cas technique, serves to improve the understanding and development of this tool for the purpose of gene-editing.

A continued challenge with CRISPR-Cas system is the risk of off-target events and undesired mutagenesis. In the case of conventional CRISPR-Cas9 system this is potentially increased because an sgRNA binding to an off-target site is sufficient to cause a DSB, whereas with ZFNs and TALENs there is a requirement for two sets of nucleases to bind to the same off-target site to trigger a DSB. To remedy this challenge high-fidelity (HF1) *S. pyogenes* Cas9 (SpCas9-HF1), enhanced-specificity spCas9 (eSpCas9), and hyper-accurate spCas9 (HypaCas9) variants containing substitution mutations within the REC3 and HNH nuclease domains have been developed (Chen *et al.*, 2017). These variants reduce off-target specificity by enhanced proof-reading without compromising on-target activity (Chen *et al.*, 2017). Furthermore, conventional gene-editing is not footprint free (leaves a genomic impression), hence, the selection marker remaining in the genomic DNA after HDR has the potential for unanticipated mutagenesis. However, successful attempts have been made to excise selection markers using pBac transposase to excise a given selection marker post-antibiotic selection of corrected clones (Firth *et al.*, 2015)

A challenge to the introduction of foreign bacterial gene-editing components into mammalian cells is the recognition of such components by pathogen-associated molecular pattern (PAMP)-binding receptors in the cytosol and the subsequent immune responses they trigger. Such receptors include the retinoic acid-inducible gene I (RIG-I) and melanoma differentiation–associated gene 5 (MDA5) receptors (Wienert *et al.*, 2018). The 5' triphosphate ends present in IVT gRNAs via T7 polymerase serve as PAMPs, capable of stimulating an innate immune response. Hence as little as 1pmol of gRNA produced by IVT can increase innate immune gene transcription by thirty to fifty fold (Wienert *et al.*, 2018) while the commonly administered 100 pmols of gRNA per 2.0 x 10^5 human embryonic kidney 293 (HEK293) cells induced an increase in *interferon beta 1 (IFNB1)* expression by up to four thousand fold (Wienert *et al.*, 2018). This observable increase in IFNB1 expression was similar to that observed when viruses such as Sendai virus mRNA (Martínez-Gil *et al.*, 2013) and hepatitis C virus PAMPs (Saito *et al.*, 2008) infect hosts cells. Furthermore, non-transfected cells can sense *IFNB1* produced by other cells, thereby activating downstream defence mechanisms. This is especially relevant for *in cellulo* gene-editing were RNP delivery to a set of lung cells could trigger an innate immune response in

surrounding cells (Wienert *et al.*, 2018). Equally of note, factors such as gRNA spacer sequence also determine the severity of immune response (Mehta and Merkel, 2020)

The Cas9 protein component is equally immunogenic, activating humoral, antibody-mediated and T cell-mediated immunity in mice (Mehta and Merkel, 2020). Transfection technique could also determine the degree of immune activation, hence a method such as lipofection may not be suitable in a clinical setting based on longer-lasting immune response in comparison to nucleofection (Wienert *et al.*, 2018). In summary, it is important to remember that gene-editing components and delivery technique can affect immune response, which could result in a decrease in cell viability post geneediting, potentially compromising the main objective of repairing cellular DNA while limiting damage (Wienert *et al.*, 2018).

1.8 Potential Applications of CRISPR-Cas9 Gene-Editing

Certain type II-A and II-C Cas9 enzymes such as *Staphylococcus aureus* Cas9 are capable of inducing RNA interference by binding and cleaving single-stranded RNA (ssRNA) sequences, requiring only a gRNA, without the need for a PAM sequence on the target RNA (Strutt *et al.*, 2018). This can be harnessed to inhibit gene expression by cleaving mRNA of transcribed genes, which can be useful in conferring moderate resistance against ssRNA phages. RNA interference could have applications in targeting RNA viruses of the lung such as influenza virus and syncytial virus that cause upper and lower respiratory tract infections.

CRISPR-Cas9 gene-editing has been explored as an approach in preventing infection of mammalian cells by viruses such as influenza virus which affect the respiratory system, by knocking-out genes implicated in viral infection (Drews *et al.*, 2019). Cas9 and gRNA-encoding plasmids designed to knockout expression of shingolipid glucosylceramidase (GBA) decreased influenza infection by up to 50% in human embryonic kidney 293 (HEK293) cells and 70% in A549 cells. Interference was by a mechanism which prevents the viral particles from fusing with endosomes thereby inhibiting entry (Drews *et al.*, 2019). This GBA knock-out resistance to viral infection was also shown to be applicable to viruses displaying glycoproteins of Ebola virus, vesicular stomatitis, and measles viruses, underpinning the broader relevance and potential of CRISPR-Cas knock-out-mediated viral immunity. Presumably, therefore, the elimination of one or more such proteins exploited by viral pathogens could be knocked out in concert to increase resilience against viral infection in man, assuming their loss is not in itself deleterious to homeostasis or other natural biological mechanisms, and that anti-viral potency matches the multi-log scale performance of efficacious small molecule antivirals.

1.9 Delivery Vehicles for Gene-Editing Components into Cells

To perform successful gene-editing, it is important to access the available options for the delivery of such a cargo into cells, which would help inform the selection process of a delivery vehicle, in an attempt to achieve maximum delivery efficiency. In brief, gene-editing delivery strategies can be grouped into viral and non-viral delivery methods.

1.9.1 Viral Delivery

Viral delivery vehicles such as adeno-associated virus and retroviruses are considered the most efficient delivery system with equally extended longevity of gene therapy (figure 1.9.1.1). However, use of viruses has certain drawbacks including immunogenic responses, high mutation rates which could render them oncogenic, the increased cost associated with scaled-up manufacturing, limited cargo capacity, and the risk of insertion of viral sequences into the host genome (Ryu *et al.*, 2018; Hryhorowicz *et al.*, 2019). One major drawback to the use of viral delivery, is the ability of some viruses to integrate into the host genomic DNA which could lead to unwanted long-term expression of gene-editing components (Xu et al., 2019).

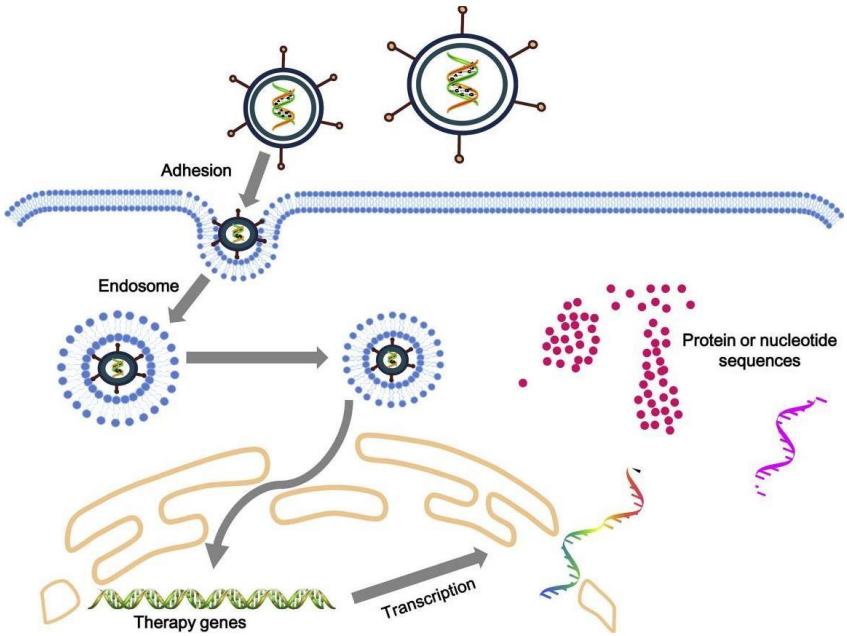


Figure 1.9.1.1 Model of virus-encoded transgene introduction into a cell. Most viruses make use of spike proteins to bind to the plasma membrane receptors, before entering the host cell by clathrin-mediated endocytosis. Within the endosome, low endosomal pH and Ca^{2+} activates viral membrane fusion which leads to the subsequent release of genomic content into the cytoplasm. Transgenes such as CRISPR-Cas9 gene-editing cassettes subsequently replicate or are expressed in the nucleus, followed by exportation of mRNA back to the cytoplasm where translation occurs, to permit the production of a fully functional gene-editing transgene. An

exception to this rule involves RNA replicon-type viruses, where viral genome replication and translation occurs exclusively in the cytosol. Taken and adapted from (Lee et al., 2017a).

1.9.2 Adenovirus Delivery

Adenovirus (AdV) is a non-enveloped linear dsDNA virus, with fifty-seven identified human serotypes, capable of packaging up to 36 kilo bases (kb) of DNA cargo (Lee *et al.*, 2017a). AdV serotypes can be further divided into six subgroups (A-G) based on tropism. The viral capsid delineates tropism amongst the different serotypes, consisting of capsid, core and cement proteins (Lee *et al.*, 2017a). Hence this vast range of viral serotypes can give rise to a wide range of potentially therapeutic viruses. The most commonly used virus is AdV-C5, however the transduction efficiency of AdV-B species (AdV-3, 21 and 14) have frequently outperformed that of AdVs-C5 in human airway cells by two to five fold (Li *et al.*, 2019).

AdV delivery has numerous advantages, including most human cells having the capability to be easily transduced by these viruses due to the expression of AdV receptor and secondary integrin receptors, thereby increasing the probability of higher transgene expression (Crystal, 2014; Lee *et al.*, 2017a). These AdVs serve as a platform for capsid engineering research to enhance targeting specificity of the virus (Lee *et al.*, 2017a). Furthermore, the safety of this delivery method has been improved by development of replication-defective and gutless or helper-dependent AdV vectors, which aids in dampening the host immune response activation. Equally, AdVs are extrachromosomal viruses which predominantly do not integrate their genetic material into the host genome (Ehrke-Schulz *et al.*, 2017). In summary, AdVs have been extensively studied therefore safety as well as risks associated with their usage are well understood.

To establish safety in use, the first-generation AdVs were replication deficient due to a deletion of transcription regulatory genes *E1a* and *E1b*, which resulted in a transgene capacity of 5.2kb (Lee *et al.*, 2017a). In a subsequent attempt to increase cargo capacity and decrease viral replication, secondgeneration AdVs were developed possessing additional deletions of other non-structural genes (*E2, E3, E4*) (Lee *et al.*, 2017a). Although these modifications improved capacity and safety, it still triggered immune responses in transfected cells. Therefore, in order to reduce immunogenic responses, a thirdgeneration of AdVs were developed, also known as high-capacity (HC-AdV), gutless or helperdependent AdV (HD-AdV) (Lee *et al.*, 2017a). These vectors lacked all viral coding sequences, possessing only the 5' to 3' inverted terminal repeats (ITRs) plus a packaging signal. These modifications increased capacity to 36kb while reducing cytotoxicity, yet improving transgene expression (Lee *et al.*, 2017a). Of note, these HD-AdVs lack virus-packaging genetic components, therefore a helper virus is supplied in trans to provide the necessary packaging proteins (Lee *et al.*, 2017a).

AdVs have been deployed in targeting CF-causing mutations within the lung. In one instance the entire 4.5kb protein coding sequence of *CFTR* was packaged into a second-generation replication-deficient AdV serotype 5 (AdV-CFTR), then delivered into Δ F508 CF epithelial cells (CFP C-1 cell line), restoring human *CFTR* mRNA expression and protein function (Rosenfeld *et al.*, 1992). AdV-*CFTR* introduction via *in vivo* intratracheal administration into CF rats, showed expression of wild-type human *CFTR* in lung epithelia after ten days, up to six weeks post-treatment (Rosenfeld *et al.*, 1992; Yang *et al.*, 1994). Induction of expression was observed after eleven to fourteen days, while duration lasted up to six weeks (Rosenfeld *et al.*, 1992). Subsequently third-generation AdV containing human *CFTR* and a helper virus were used to transfect primary human epithelial cell cultures of a CF patient, improving protein expression in 10 - 20% of transduced cells (Fisher *et al.*, 1996). Equally, AdV that possess specific tropism for the lung endothelial over other tissue and vice versa have been developed, hence following intravenous injection of 1 x 10^{11} viral particles into mice, an improvement in targeting efficiency of up to fifteen fold and safety in comparison to conventional HD-AdV control was observed (Lu *et al.*, 2020).

The specific combination of AdV plus CRISPR-Cas9 has also been deployed to correct disease-causing mutations. In one instance, AdV containing a donor template and CRISPR-Cas9 cassette were used to demonstrate stable integration of wild-type *CFTR* DNA into the AAVS1 site. Here, gene-editing by the nuclease was shown to be transient, therefore Cas9 was not detectable at day 7 (Xia *et al.*, 2018). Another study packaged both constitutive and tetracycline-inducible Cas9-gRNA expression cassette into the HC-AdV genome. This was used to target HPV DNA in HeLa cells, dystrophin gene in primary human myoblast cells, and CCR5 co-receptor in A549 cells required for successful HIV infection. Hence, at a multiplicity of infection (MOI) of 10, up to 93% exon deletion in dystrophin gene was observed by PCR, while gene-editing efficiency achieved was ~ 25% in the HPV locus (Ehrke-Schulz *et al.*, 2017). AdV-CRISPR delivered at MOI 2.5 to 80 has also been used in primary human lung fibroblasts and primary human lung epithelial cells to demonstrate up to 40% efficiency in the knockdown of *SMAD3,* which plays a key role in cellular fibrosis (Voets *et al.*, 2017).

1.9.3 Adeno-associated virus delivery

Adeno-associated virus (AAV) is a small, helper-dependent ssDNA virus that is capable of transducing both dividing and non-dividing cells, mainly delivering transgene cargo episomally (Lee *et al.*, 2017a). In comparison to AdVs, AAVs display an improved safety profile because they are naturally nonpathogenic to humans, equally wild-type viruses tend to integrate silently into a specific site on the q arm of human chromosome 19 (Nayerossadat, Maedeh and Ali, 2012; Lee *et al.*, 2017a). Furthermore this ability is lost in AAV vectors due to deletion of *rep* gene (Kay, Glorioso and Naldini, 2001). However, AAV delivery usage is hindered by their relatively small cargo size of about 4.8kb, as

well as their complex production process (Nayerossadat, Maedeh and Ali, 2012). Upon transduction, gene expression is slow perhaps due to the requirement for generation of dsDNA by ssDNA annealing or second-strand synthesis, followed by linkage of vector genome to form concatemers (Kay, Glorioso and Naldini, 2001).

In the context of lung disease, AAV-containing *CFTR* cDNA (AAV-CFTR) has been used in a few phase I dose-escalation trials, delivered into respiratory nasal epithelia in CF patients (Wagner *et al.*, 1999; Flotte *et al.*, 1996; Wagner *et al.*, 1998a). Therapeutic benefits were observed from a dosage of >10,000 viruses to host DNA ratio (R.U), equally wild-type *CFTR* was expressed up to 10 weeks post treatment, with little to no inflammatory or immune responses observed (Wagner *et al.*, 1999; Wagner *et al.*, 1998b). In another study AAV2 vector encoding human *CFTR* was delivered in a dose ranging from 10^{10} to 10^{13} DNase resistant particles, into the lungs of twelve CF patients via aerosols or nebulisation (Aitken *et al.*, 2001). Maximum gene transfer was achieved at 0.6 and 0.1 vector copies per cell after fourteen and thirty days respectively; by 90 days, vector copies were undetectable. Notably, of the two hundred and forty two adverse events recorded, only three were possibly linked to the study viral drug (Aitken *et al.*, 2001), demonstrating the safety of this viral vector. However, a low *CFTR* mRNA expression was observed in these studies, possibly attributable to an inefficient AAVITR promoter driving transgene expression, an increase in lung inflammation, and an increase in AAV2 neutralising antibodies.

Several studies have aimed to improve efficiency of AAV vectors, including the use of other AAV serotypes. Inefficiency is in part due to the reduced number of AAV2 receptor heparin sulphate proteoglycan, present on the apical surface of lung epithelia (Duan *et al.*, 1998). To this end AAV6 vectors have been developed with up to two-fold greater mucus penetrating abilities, determined by yellow fluorescent protein (YFP) bioluminescence measurements, hence improving delivery to airway epithelia by inhalation even in the presence of mucus hypersecretion present in CF (Duncan *et al.*, 2018).

1.9.4 Herpes Simplex Virus

The Herpes Simplex Virus (HSV) is a dsDNA virus with a 152kb genome, capable of delivering a transgene DNA payload of up to 50kb, making it one of the largest viral vectors studied (Lee *et al.*, 2017a). The further deletion of genes *ICP0, ICP4, ICP22* and *ICP47*, involved in viral growth and replication, lead to the generation of vectors with reduced cytotoxicity (Lundstrom, 2004). Similar to AdV, HSV infection is common in the general population therefore pre-existing immunity to viral antigens may affect transduction efficacy, although HSVs are widely able to circumvent host immune detection by inhibiting type I interferon production (Lee *et al.*, 2017a). Equally, similar to AdV, HSV

genomes remain episomal, thereby ensuring transgene expression remains transient (Lee *et al.*, 2017a). However, HSVs can also establish a latent life-cycle translating to a life-long persistence in the host and subsequent long-term transgene expression. Therefore, the engineering of HSV vectors with latency-active promoters LAP1 and LAP2 has been shown to permit long-term transgene expression for over a year (Lundstrom, 2004).

Although HSV vectors display a natural tropism for neuronal cells, the vast majority of studies have explored its application in the context of neuronal diseases and cancer (Marconi *et al.*, 2009). However, HSV has been used to drive transgene expression in human bone osteosarcoma epithelial cells (U2OS), human dermal fibroblasts (HDFs), monkey kidney epithelial cells (Vero cells), human muscle derived stem cells, human pre-adipocytes, and human hepatocytes (Miyagawa *et al.*, 2015). These cells represent bone, skin, kidney, stem cells, adipose tissue, and liver cells respectively, hence highlighting the broad application spectrum of HSV-mediated transduction (Miyagawa *et al.*, 2015).

1.9.5 Retroviral Delivery

Retroviruses are a family of viruses that encode their genetic information in the form of RNA, which is then converted into DNA in transfected cells by a virally-encoded reverse transcriptase (Ciuffi, 2008). Lentiviruses are a retroviral genera whose genome contains three genes for survival namely; *gag* which encodes structural proteins, *pol* which encodes enzymes for reverse transcription and integration into host genome, and *env* which encodes the viral envelope glycoprotein (Milone and O'Doherty, 2018). The life cycle of all retroviruses is similar; for lentiviruses it begins with entry into the host cell by membrane fusion or receptor-mediated endocytosis. Uncoating of virus follows, then reverse transcription of viral RNA into proviral dsDNA, complexing of dsDNA with viral *gag* protein subunits to enable nuclear import, and finally integration into the host cell genome (Milone and O'Doherty, 2018). The integrated proviral genetic material is now dependent on the host cell for transcription and translation of viral proteins required for assembly of new viral particles, which are subsequently released from the host by a process called budding (Milone and O'Doherty, 2018).

Retroviral-based gene therapy has numerous advantages including long-term stable transgene expression, and an increased transgene cargo capacity due to their large genomic size. Also, retroviruses used for research are associated with reduced humoral immune response due to lack of expression of virulence genes. Of note, lentiviruses in particular are capable of integrating into nondividing cells unlike other retroviruses that only integrate into dividing cells (Ciuffi, 2008). However, disadvantages of retroviruses include their inability to integrate into the same location of the host genome, raising consistency and safety concerns (Ciuffi, 2008). To improve retroviral safety for the purpose of gene therapy, third generation vectors were developed with the components

required for viral production being split across multiple plasmids making recombinant virus generation even more unlikely (Milone and O'Doherty, 2018). Furthermore, a deletion in the 3' long terminal repeat which contains enhancer and promoter sequences, is often included in the DNA used to produce viral RNA, hence eliminating transcription of provirus, rendering the virus 'self-inactivating' after integration (Milone and O'Doherty, 2018; ufferey *et al.*, 1998).

To date, lentiviral transduction of gRNA and Cas9 has been used to target up to 18,080 genes, in order to identify genes involved in drug resistance of cancer cells, as well as the cell viability of these mutagenic cells and pluripotent stem cells (Shalem *et al.*, 2014). A knockdown efficiency of up to 93% and indel rates of up to 92% were equally achieved following antibiotic selection, at the *EGFP* locus of an *EGFP*-containing HEK293T cell line, using lentiviral CRISPR-Cas9 constructs (Shalem *et al.*, 2014). Co-delivery of lentivirus-encoding gRNA and adenovirus encoding Cas9 has been used for the transfection of primary human lung microvascular cells (HLMVECs) as a proof of concept that these primary human endothelial cells are capable of being edited (Gong *et al.*, 2017). Up to 90% deletion frequency was achieved which remained stable after five passages, hence there was no need for clonal expansion for selection post gene-editing (Gong *et al.*, 2017). More recently, lentivirus encoding alveolar granulocyte-macrophage colony-stimulating factor (GM-CSF), a protein which is insufficient in pulmonary alveolar proteinosis, were transduced as a single dose into mice. Introduction of 1×10^5 viral particles per mouse successfully resulted in GM-CSF expression up to eleven months posttransduction.

Recent strategies have attempted more transient approaches to deliver Cas9 and gRNA into cells. One strategy attempted to package Cas9 mRNA into lentivirus-like particles (LVLPs) modified with aptamer-binding proteins, using aptamers downstream of Cas9 stop codon (Lu *et al.*, 2019). Another strategy attempted to package Cas9 RNP in LVLPs via the interaction between gRNA modified with aptamers and lentiviruses modified with aptamer binding proteins (Lyu *et al.*, 2019). These strategies were shown to have comparable efficiencies but in some genomic loci, higher gene-editing activity was observed for RNP LVLPs compared to transfection by Cas9 mRNA packaged in LVLPs (Lu *et al.*, 2019). These RNP and mRNA platforms equally displayed less off-target cleavage compared to AAVmediated gene-editing and even standard lentivirus DNA delivery (Lyu *et al.*, 2019).

1.9.6 Non-Viral Delivery

Non-viral delivery techniques could be by physical means, using electric currents such as electroporation-based techniques, or by chemical means using vehicles such as peptides, cationic lipids, and cationic polymers. Delivery by nanoparticles is less challenging in comparison to the usage of viral vectors for the same purpose and much less complex in production, hence lower costs are

incurred. Additionally, they generally provoke less of an immune response and are more easily customisable (Ryu *et al.*, 2018). A challenge of note facing the field of nano medicine, which includes nanoparticle delivery, is that numerous nanoparticles used are non-biodegradable and therefore may never get approved for in ection into humans. Therefore, it's important that nanoparticles >10nm (hence too large to be excreted in urine) are biodegradable (Szwed *et al.*, 2020).

Table 1.9.6.1 A summary of non-viral delivery approaches

Reagent	Description	Pros	Cons
Polyethylenimine	PEI based cationic polymers possess a high concentration of positive charges due to a high concentration of amino groups, which in turn interacts with negatively charged DNA to form a compact polymer-DNA complex (Ryu et al., 2018).	High transfection efficiency with higher branching further improving efficiency, excellent endosomal escape ability.	The positive charge on the surface of PEI can be toxic to cells being transfected
Lipofection	Lipofection is the chemical delivery of DNA and proteins into cells using cationic lipid-mediated transfection (Sharifi Tabar et al., 2015). These lipid nanoparticles are composed of a hydrophilic head and a hydrophobic tail domain, hence once these nanoparticles are in an aqueous environment, the lipids spontaneously self-assemble forming a circular bilayer encapsulating the payload within the core of the hydrophilic tails (Chen, Alphonse and Liu, 2020), yielding so-called lipid nano-capsules (LNCs).	Efficient endosomal escape, cargo protection from lysosomal degradation, and from neutralisation by antibodies (Chen, Alphonse and Liu, 2020).	Inefficient at delivery cargo payload directly into the cellular nucleus.

Cell Penetrating Polypeptide	Cationic polypeptides (CPPs) a technology based on materials that are often used for membrane destabilisation such as oligoarginine and HIV-TAT protein. Theses CPPs are generally comprised of water-soluble α-helical polypeptide bearing a cationic side chain (Wang et al., 2018a). CPPs can be deployed by combining a positively charged amino acid domain essential for DNA binding with an amphipathic membrane transduction domain, such as HIV-TAT (Wang et al., 2018a)..	CPP fused to HIV-TAT actively forms physical pores with diameters in the nanometre range, enabling intercellular entry of gene cargo	CPPs often possess inadequate cationic charge density which is easily neutralised by the negatively-charged DNA cargo , therefore delivery efficiency is reduced severely
Cationic Polysaccharides (CPs)	Oligoamine residues grafted to polysaccharides can be used as effective delivery vehicles. In these nanoparticles, side-chain oligomers are grafted to either a linear or branched hydrophilic polysaccharide backbone which subsequently permits interactions with the anionic surface of cargo DNA (Azzam et al., 2002). These delivery vehicles can be further optimised to suit specific DNA payload by grafting oligoamino side-chains at a certain distance from each	Suitable for *in vivo* transfection because they are biodegradable, water-soluble, and can be transported into cells,	Inadequate cationic charge density which is easily neutralised by the negatively charged DNA cargo, negatively impacting upon cargo delivery, hence the need for grafting oligoamines to CPs

	other, for example, at an interval of two or three monomer subunits (Azzam et al., 2002).		
Hydrodynamic Delivery	Hydrodynamic delivery is an *in vivo* delivery approach which involves introduction of a large volume solution containing gene-editing cargo into the bloodstream of an animal, typically using the tail in mice (Lino et al., 2018). The injection of solution results in increased hydrodynamic pressure in endothelial and parenchymal cells, enabling membrane crossing of cargo where ordinarily this should be impossible (Lino et al., 2018).	Technical simplicity and the nonrequirement of any exogenous delivery components	Traumatic or fatal nature of the technique for the animal model and the low transfection efficiency observed
Electroporation	Electroporation is the most common physical transfection method, involving brief electric pulses applied to cells to increase their permeability to foreign macromolecules (Maurisse et al., 2010; Xu et al., 2018). This short microsecond to millisecond voltage pulse technique causes the formation of aqueous pores in the lipid bilayers, driving the macromolecules through these pores (figure 1.9.12.1)	Applicable to a wide range of commercially available immortalised cell lines. Electroporation enables transfection of hard-to-transfect cells including stem cells and primary human cells isolated from a patient or donor, which bear the most physiological	Need to optimise transfection voltage and length of electrical pulse to minimise cell-death post-transfection.

	(Bolhassani, Khavari and Oraf, 2014).	relevance when studying diseases	
Microinjection	In this technique gene-editing components can be introduced through a plasmid or as an RNP by direct injection into individual cells, using a microscope and a 0.55.0 um diameter needle to pierce the cell membrane (Lino et al., 2018). Most common technique in generation of animal models, through the injection of CRISPR-Cas9 components into the cytoplasm of zygotes (Horii et al., 2014; Lino et al., 2018).	Absence of cargo size limitations and direct cargo delivery to desired target location, circumventing constraints encountered in delivery through the cell membrane, (Lino et al., 2018).	Most suitable for *in vitro,* and *ex vivo* use, as the use of a microscope and needle for injection of individual cells precludes it's use in an *in vivo* setting
Gymnosis	Gymnosis implies the delivery of naked oligonucleotides (oligos) into cells without the need for delivery vehicles. Here, oligos are internalised into cells by clathrin-mediated endocytosis, then accumulate within the endosomes prior to their release (González-Barriga et al., 2017).	Permits cell exposure to a more even concentration of delivered cargo, lack of transfection reagent reduces cost, reduces possible artifacts or false positives that are attributed to the delivery vehicle (González-Barriga et al., 2017).	Slow cellular uptake rates compared to transfection reagents, some artifacts are a result of high oligo concentration and nonspecific surface cell binding events (González-Barriga et al., 2017).

1.10 Aims and Objectives

Chronic heritable lung diseases, such as CF and IPF, are associated with significant disease-related morbidity and mortality, resulting in an increased burden to the global health services. CRISPR-Cas9 targeted gene-editing has been deployed as an investigative and therapeutic tool against these genetic pulmonary conditions, however these approaches are associated with short comings in both therapeutic safety and efficiency.

The aims of this research were: (i) Design of a set of highly efficient gRNAs and HDR templates specific for the *CFTR* and *MUC5B* loci. MUC5B shall be targeted because a G/T SNP in its promoter region is strongly correlated to IPF, hence I shall attempt HDR at this loci as a proof-of-principle gene-editing therapy for IPF (ii) Validate the expression of a MUC5B promoter-associated non-coding RNA implicated as a risk factor for IPF. (iii) Design a STU CRISPR-Cas9 gene-editing system specific for the *CFTR* locus. (iv) Investigate the efficiency of *in vitro* transcribed gene-editing single-transcriptional unit (IVT gene-editing STU) delivered by gymnosis, in the context of gene-editing. Gymnosis would ensure more even concentration of delivered cargo whereas transfection reagents would create an unusually high payload concentration within cells, which makes it more difficult to achieve uniformity *in vivo*. Gymnosis would also serve to reduce possible artifacts or false positives that are attributed to the delivery vehicle and not the therapeutic cargo (Soifer *et al.*, 2012)..

Chapter 2. Materials and methods

2.1 Cell Culture

2.1.1 Cell Recovery

Cell lines used were thawed at 37°C in a water bath, then resuspended in 5 mL complete growth media. Cell suspensions were subsequently centrifuged at room temperature for 3 minutes at 300 x g, then cells were resuspended in full growth medium after the freezing media was aspirated. Cells were then cultured in T25 tissue culture flasks (ThermoFisher, Cramlington, UK). or 6-well plates ThermoFisher).

2.1.2 Cell Propagation

Cells were continuously grown in T25 tissue culture flasks containing their respective growth media. When cells grew to ~ 70% confluency, they were harvested using 1 x trypsin-EDTA, then sub-cultured into fresh flasks or 6-well plates containing complete growth media (growth media supplemented with 10% fetal bovine serum (Sigma-Aldrich, Gillingham, UK), 1% L-Glutamine (Sigma-Aldrich) and 1% Pen/Strep (Sigma-Aldrich). All cells were maintained in humidified 37°C, 5% CO_2 incubators.

CFBE41o- cell line, a kind gift from Professor Uta Griesenbach, was adopted as the primary cell line used to study the effectiveness of designed CF gRNAs, as well as for downstream homology-directed repair experiments in chapter 3 and 5. The CFBE41o- immortalised cell line which was derived from a CF patient homozygous of the Δ F508 CFTR genotype possesses the desired Δ F508 genotypic anomality required for these studies, hence was chosen. CFBE41o- (Sigma-Aldrich) cells were routinely passaged once every five to seven days, after sub-culturing at five to ten-fold serial dilutions. Cells were cultured in Minimum Essential Medium (MEM) (Sigma-Aldrich), then used at passage number <20. CFBE41o- cells were usually seeded down at 10,000 cells/well in 96-well plates and 250,000 cells/well in a 6-well plate during gene-editing and viability experiments, at passage numbers three to ten.

The A549 cell line are an *in vitro* model of type II alveolar epithelial cells first established in 1972. These cells lines form confluent adherent monolayers with a Type II characteristic morphology and distribution of lamellar bodies (Foster *et al.*, 1998). A549 cells were chosen as the primary investigative model in chapter 4 because type II alveolar epithelial cells have been shown to express *MUC5B* and are involved in fibrosis in the context of IPF (Hancock *et al.*, 2018). A549 cells received as a gift from Dr Steve Getting were routinely passaged once every three to five days, after sub-culturing at five to ten-fold serial dilutions. Cells were cultured in Dulbecco's modified Eagles Media (DMEM) (Sigma-Aldrich), then used at passage number <20. A549s were cultivated and seeded down for gene-editing and viability studies under the same circumstances as CFBE41o- cells.

HeLa cells were received as a gift from Dr Antony Antoniou. HeLa cells were routinely passaged once every three to five days, after sub-culturing at five to ten-fold serial dilutions. Cells were cultured in Dulbecco's modified Eagles Media (DMEM) (Sigma-Aldrich), then used at passage number <20. HeLa cells were cultivated and seeded down for gene-editing and viability studies using similar conditions as CFBE41o- cells.

2.1.3 Cell Counting

Cell counts were evaluated by combining 10-20ul cell suspension with 0.4 percent trypan blue (Gibco, ThermoFisher) solution during splitting and seeding of cells for downstream experimental purposes. The cells were subsequently transferred to a haemocytometer (Hawksley, Lancing, Sussex, UK) then viewed under a bright-field microscope using a magnification setting of 10x. During the counting process, all cells were counted except those cells positive for trypan which represented the dead cell population.

2.1.4 Cell Viability

Cell viability was evaluated during splitting and post-gene-editing by performing a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay (Sigma-Aldrich) on cells in 96-well plates. MTT cell-proliferation assay measures the reduction of MTT into an insoluble formazan product by the mitochondria of viable cells. Initially MTT powder was dissolved in PBS (1x) to a final stock concentration of 1mg/mL. Next the growth media was removed from cells prior to addition of 200µl of MTT stock solution per well. Plates were incubated at 37°C for 10-30 minutes, then observed under the microscope for the presence of an intercellular purple tinge. MTT solution in each well was then replaced with an equal volume of dimethyl sulfoxide (DMSO) (Sigma-Aldrich), then absorbance was measured at 550nM using a TECAN plate reader (Biotek, Swindon, UK). Of note, absorbance readings were measured in triplicates for each control and sample. Equally, positive (containing untreated cells) and negative control (containing DMSO only) wells were included in every viability experiment.

2.1.5 Cryopreservation

Subsequent to attaining 70-80% confluency, all cells were frozen down. During cell propagation, the pellets were re-suspended in the appropriate complete culture media at a concentration of roughly 1 million cells/ml and refrigerated on ice after trypsinization for adherent cells. The cells were then put to cryovials then 10% v/v DMSO added, before placing the cryovials in a pre-chilled (4°C) Mr Frosty (ThermoFisher) container at -80°C overnight. cryovials were then transported to a -150°C freezer the following day for long-term preservation.

2.1.6 Cell Transfection

Cas9-gRNA RNP complexes were transfected into cells using Lipofectamine™ CRISPRMAX™ kit (ThermoFisher), following the manufacturers protocol. Experiments using this kit were carried out by 'forward transfection' technique to transfect cells with RNP complexes. Cells were initially seeded at 250,000 cells/well in a 6-well plate or 10,000 cells/well in a 96-well plate, then allowed to adhere overnight. The following day, an initial reaction mix was made containing, 125µl/ 5µl of OptiMEM serum-free media (ThermoFisher) for 6-well/ 96-well plate transfection, 240nM gRNA, 240nM Cas9, 24nM donor template (for HDR experiments) and Cas9 plus reagent (ThermoFisher) (2 x volume of Cas9 nuclease). This initial mix was then left at room temperature for 10 minutes to permit RNP complex formation. Subsequently, a second reaction mix containing, 125µl/ 5µl of OptiMEM serum-free media (ThermoFisher) for 6-well/ 96-well plate transfection and 7.5µl/ 0.3µl of CRISPRMAX™ reagent (ThermoFisher) for 6-well/ 96-well plate transfection, was quickly made ensuring contents were not left at room temperature for longer than 3 minutes. Contents of the first reaction mix were subsequently added to that of the second reaction mix, followed by incubation at room temperature

for 10 minutes. Finally, 250µl/ 10µl of RNP-transfection reagent complex were added dropwise into each well of 6-well/ 96-well plates, replacing the growth media with fresh complete media 4hrs post-transfection, then cells were maintained for 24 hours in a humidified 37°C, 5% CO_2 incubators.

To directly introduce donor template DNA into the nucleus, where Cas9-gRNA RNP exerts it's function, nucleofection was adopted as a strategy instead of lipofection. Cas9-gRNA RNP plus donor template complexes were nucfleofected into CFBE41o- cells using Nucleofector® Kit V (Lonza, Staffordshire, UK) following the manufacturers protocol. Initially, cells from a >70% confluent T25 counted then 1,000,000 cells were transferred to a fresh 1.5mL Eppendorf. Cells were then spun down at 90g for 10 minutes at room temperature, discarding the supernatant, while keeping the pellet. A reaction mix was then prepared containing, 82µl solution V (Lonza, Staffordshire, UK), 18µl supplementary solution (Lonza, Staffordshire, UK), 240nM gRNA, 240nM Cas9, 24nM HDR donor template DNA. As a positive control for the nucleofection reaction, gRNA and Cas9 components were replaced with 2µg of pGFP plasmid (Lonza). The entire reaction mix was then used to resuspend the cell pellet, then cell suspension was transferred into a cuvette supplied with the kit. Next nucfleofection was performed using an Amaxa™ Nucelofector™ I (Lonza) device, with the nucleofection program set to 'T-020'. Once the program was complete, 500µL of complete growth media pre-equilibrated at 37°C was added to the cuvette. Finally, using the kit-supplied dropping pipette, the contents of the cuvette were transferred to a pre-equilibrated 6-well plate, then nucleofected cells were maintained for 24 hours in a humidified 37°C, 5% CO_2 incubators in complete MEM plus 30 µM Alt-R HDR Enhancer V1 (IDT, Leuven, Belgium).

Gene-editing plasmid DNA was transfected into cells using jetPRIME® kit (Polyplus, Illkirch, France), following the manufacturers protocol. Assays using this kit were carried out by 'forward transfection' technique to transfect cells with Plasmid DNA complexes. Cells were initially seeded at 250,000 cells/well in a 6-well plate or 10,000 cells/well in a 96-well plate, then allowed to adhere overnight. The following day, a reaction mix was made containing, 200µl/ 10µl of jetPRIME® buffer (Polyplus) for 6-well/ 96-well plate transfection, 2µg plasmid DNA and 4µl/ 0.2µl of jetPRIME® reagent (Polyplus) for 6-well/ 96-well plate transfection. This mix was vortexed for 10 seconds then incubated at room temperature for 10 minutes to permit plasmid DNA-jetPRIME® reagent complex formation. Finally, the entire plasmid DNA-jetPRIME® reagent complex mix were added dropwise into each well of 6-well/ 96-well plates, replacing the growth media with fresh complete media 4hrs post-transfection, then cells were maintained for 24 hours in a humidified 37°C, 5% CO_2 incubators.

When stated, gene-editing plasmid DNA was transfected into cells using PEI (Polysciences, Warrington, USA), following the manufacturers protocol. Initially PEI powder was dissolved to a final

concentration of 1mg/mL, to be used for subsequent forward transfection experiments. Cells were initially seeded at 250,000 cells/well in a 6-well plate or 10,000 cells/well in a 96-well plate, then allowed to adhere overnight. The following day, an initial reaction mix was made containing, 150µl/ 6µl of PBS (ThermoFisher) for 6-well/ 96-well plate transfection which was used to dilute 2µg plasmid DNA. Subsequently, a second reaction mix containing, 150µl/ 6µl of PBS for 6-well/ 96-well plate transfection and 3mM PEI (ThermoFisher) (2 x volume of Cas9 nuclease), was made. Contents of the first reaction mix were subsequently added to that of the second reaction mix, followed by incubation at room temperature for 10 minutes. Finally, the entire plasmid-PEI transfection reagent complex were added dropwise into each well of 6-well/ 96-well plates, replacing the growth media with fresh complete media 4hrs post-transfection, then cells were maintained for 24 hours in a humidified 37°C, 5% CO_2 incubators.

2.2 Microbial Methods

2.2.1 General Microbiology

Unless otherwise noted, bacteria were cultivated overnight at 37°C in Luria Broth (LB)-miller media (Formedium™, Norfolk, UK) in a shaking incubator set at 250 rpm for general handling and cloning purposes. Aliquots of the initial culture (0.75 ml per 50 ml) were then plated on agar plates treated with suitable selection antibiotics and cultured at 37°C until the following day, when colonies were observed. Single colonies were subcultured in adequate quantities of selection medium for downstream techniques.

2.2.2 Preparation of Electrocompetent TOP 10 *E. coli*

Top 10 *E. coli* were grown on LB agar plates, then incubated at 37°C overnight. The next day a single colony was used to inoculate 3 mL of LB media then incubated overnight at 37°C, shaking at 200 rpm. All 3 mL of starter culture was used to inoculate 30 mL of LB-Miller medium then grown at 37°C until an OD_{600} of 0.4, measured using a spectrophotometer. The culture was then incubated on ice for 20 – 30 minutes, then centrifuged at 3000 x g for 15 minutes at 4°C, discarding the supernatant. The pellet was resuspended in 10 mL ice-cold 100mM $MgCl_2$ then centrifuged at 2000 x g for 10 minutes at 4°C, discarding the supernatant. The pellet was resuspended in 5 mL ice cold $CaCl_2$, incubated on ice for 20 minutes, then centrifuged at 2000 x g for 10 minutes at 4°C, discarding the supernatant. Cells were resuspended in 1.25 mL resuspension solution (85% v/v $CaCl_2$, 15% v/v glycerol). Aliquots of 50 µl were either used immediately for transformation or were stored at -80°C for future use.

2.2.3 Preparation of Chemically Competent TOP 10 *E. coli*

Top 10 *E. coli* were grown on LB agar plates, then incubated at 37°C overnight. The next day a single colony was used to inoculate 3 mL of LB media then incubated overnight at 37°C, shaking at 200 rpm.

All 3 mL of starter culture was used to inoculate 30 mL of LB-Miller medium then grown at 37°C until an OD_{600} of 0.4, measured using a spectrophotometer. The culture was then incubated on ice for 20 – 30 minutes, then centrifuged at 3000 x g for 15 minutes at 4°C, discarding the supernatant. The pellet was resuspended in 10 mL ice-cold 100mM $MgCl_2$ then centrifuged at 2000 x g for 10 minutes at 4°C, discarding the supernatant. The pellet was resuspended in 5 mL ice cold $CaCl_2$, incubated on ice for 20 minutes, then centrifuged at 2000 x g for 10 minutes at 4°C, discarding the supernatant. Cells were resuspended in 1.25 mL resuspension solution (85% v/v $CaCl_2$, 15% v/v glycerol). Aliquots of 50 µl were either used immediately for transformation or were stored at -80°C for future use.

2.2.4 Transformation of TOP 10 Electrocompetent *E. coli*

50 µl of electrocompetent cells were retrieved from -80°C, thawed on ice for 10mins then 50ng of plasmid DNA were added to competent cells, followed by further incubation on ice for a further 2 minutes. The mix was transferred into a pre-chilled sterile 2 mm electroporation cuvette (Bio-Rad, Hertfordshire, UK). The cuvette was placed into a Bio-Rad MicroPulser™ (Bio-Rad, Hertfordshire, UK), then a voltage of 2.5kV was applied across the cuvette for 4 milliseconds. 1 mL of LB media was subsequently added to the cuvette, then the mix was transferred into a sterile 1.5 mL microcentrifuge tube. Electroporated cells were incubated at 37°C for 1 hour, then plated on appropriate antibiotic LB agar plates.

2.2.5 Transformation of TOP 10 chemically competent *E. coli*

After prior cloning steps, recombinant plasmid DNA was transformed and propagated in *Escherichia Coli* TOP 10 as follows. 50 µL of competent TOP10 *E. coli* were retrieved from -80°C, thawed on ice for 10mins then 50ng of plasmid DNA were added to competent cells, followed by further incubation on ice for 30 minutes. A heat-shock was performed at 42°C for 30 second using a water bath, followed by incubation on ice for a further 5 min. 250 µL SOC outgrowth medium (New England Biolabs, Hitchin, UK) was subsequently added to cells followed by incubation in a shaking incubator at 37°C for 1 hour. Following incubation cells were plated on an appropriate antibiotic LB-agar plate overnight at 37°C.

2.2.6 Bacterial Growth

Single colonies of wild-type or transformed TOP 10 *E. coli* were picked and inoculated into 5 mL LB media supplemented without or with 100 µg/mL ampicillin respectively, then incubated at 37°C overnight, shaking at 200 rpm. Plasmid DNA was isolated using the NZYMiniprep purification kit (NZYTech, Lisboa, Portugal), following the manufacturers protocol, storing eluted DNA at -20°C until use.

2.3 Protein Expression

2.3.1 Preparation of Total Protein Extract from Mammalian Cells

Cells contained within 6-well plates were incubated on ice briefly, then washed twice with ice-cold PBS (1x), before addition of 100 µL radio-immunoprecipitation (RIPA) buffer (1x) (10mM Tris-Cl pH 8.0, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton x-100, 140 mM NaCl) to each well. The cells were then detached by scraping from the plate followed by incubation on ice for 30 minutes to promote cell lysis. Next centrifugation of lysates was performed at 12,000 x g for 12 minutes at 4°C, collecting the supernatant in a fresh collection tube while the pellet of cell debris was discarded. Protein amount was determined immediately by a BCA assay (Pierce BCA assay kit, ThermoFisher) or was stored at either -20°C or -80°C for short-term or long-term storage respectively.

2.3.2 SDS Polyacrylamide Gel Electrophoresis

To achieve protein separation from total protein extracts sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed. 10% SDS-PAGE were first prepared using a mini protean gel casting apparatus (Bio-Rad), then sealed using 500 µL isopropanol (Sigma-Aldrich). 50% v/v distilled water, 10% v/v acrylamide (Bio-Rad), 375mM Tris pH 8.8 (ThermoFisher), 0.1% w/v SDS (ThermoFisher), 0.1% w/v ammonium persulfate (ThermoFisher) and 0.04% v/v TEMED (Sigma-Aldrich). Once the resolving gel had set, the stacking gel was prepared comprising of 72% v/v distilled water, 5.1% v/v acrylamide, 130mM Tris pH 6.8, 0.1% w/v SDS, 0.1% w/v ammonium persulfate and 0.01% v/v TEMED. A comb was subsequently inserted then gels were allowed to polymerise.

Polymerised gels were placed into a gel tank (Mini-protean tetra vertical electrophoresis cell; Bio-Rad) then covered in running buffer (1x) (25mM Tris, 190 mM glycine, 0.1% w/v SDS pH 8.3) (Bio-Rad, Hertfordshire, UK). Protein samples were then thawed on ice, SDS-PAGE loading buffer (1x) (60mM Tris-HCl pH 6.8, 50% w/v Glycerol, 10% w/v SDS, 14.4 mM β-Mercaptoethanol, 1% bromophenol blue, distilled water; Bio-Rad) was added, then the mix was incubated at 95°C for 5 minutes in a heat block. Samples and 5µL of precision plus protein standards (Bio-Rad) were then loaded onto the wells of the gel, then electrophoresis was conducted at 100V, for 2 hours at room temperature until the dye front approached the bottom of the resolving gel.

2.3.3 BCA Assay

The concentration of total protein extract was determined by BCA assay (Pierce BCA assay kit, ThermoFisher). This assay makes use of bicinchonic acid (BCA) for the colorimetric quantification of total protein in any given sample. Bovine serum albumin (BSA) standards ranging from 0 mg/ml to 2 mg/mL and samples were made up to a volume of 20 µL in RIPA (1x) buffer then 200 µL of BCA reagent was added to each sample. Samples were transferred to into clear polystyrene 96-well plates (Sigma-

Aldrich), followed by centrifugation at 3,500 x g for 2 minutes at room temperature. Samples were incubated at 37°C for 30 minutes. Protein concentration was then determined using a TECAN spectrophotometer (Biotek, Swindon, UK) set to measure absorbance at 562 nm to estimate protein concentration. Analysis of standards and samples was performed in duplicates, with average absorbance between duplicates being subtracted from the average absorbance of the blank wells containing solely BCA reagents. A standard log-linear curve was generated based on the mean of duplicates per reference concentration and regression fit determination using Sigmaplot software (version 14.0/2018). This standard curve was then used to extrapolate protein concentrations of unknown samples by using the background-substracted absorbance average values of duplicate samples.

2.3.4 Western Blotting

To probe for a specific protein of interest present in total protein extracts, a western blot was performed. After SDS-PAGE separation proteins were transferred for 1 hour at 120V from the polyacrylamid gel onto a 0.45 µm nitrocellulose membrane (Bio-Rad) submerged in transfer buffer (1x) (25mM Tris base, 190mM glycine, 20% v/v methanol) (ThermoFisher) using Trans-blot Turbo transfer system (Bio-Rad). Membranes were then blocked for 1 hour at room temperature in blocking buffer (137 mM NaCl, 2.7 mM KCl, 19 mM Tris base, 0.1% Tween 20 (ThermoFisher) and 5% powdered dry milk (Cell Signalling, Leiden, Netherlands)). Primary antibodies were then diluted in blocking buffer based on manufacturers specification (see table 2.3.4.1), followed by incubation with the membrane overnight at 4°C. Washes were performed 3 times for 5 minutes each in Tris buffered saline-Tween (TBST) (1x) (137 mM NaCl, 2.7 mM KCl, 19 mM Tris base, 0.1% Tween 20) prior to addition of secondary antibody for 1 hour at room temperature. Subsequently 3 washes, 5 minutes each with TBST (1x) were performed on nitrocellulose membrane prior to imaging. Next enhanced chemiluminescence system (ThermoFisher) which uses a luminol-based detection system was added to the membranes then incubated for 10 minutes while being protected from light exposure. Membrane were finally imaged on a Gbox chemiluminescence imager (Chemi G:Box, Syngene, Cambridge, UK), selecting a 1 minute exposure time.

Table 2.3.4.1 List of Antibodies used for western blotting

Antibody Target	Manufacturer (Cat. No)	Concentration
Mouse Anti-GAPDH	Abcam, Cambridge, UK (AB8245)	1 : 3000

Rabbit Anti-Cas9	Abcam, Cambridge, UK (AB189380)	1 : 20000
Anti-Mouse secondary antibody	Cell Signalling, Leiden, Netherlands (7076S)	1 : 3000
Anti-Rabbit secondary antibody	Cell Signalling, Leiden, Netherlands (7074S)	1 : 3000

2.4 Gene-Editing Design

2.4.1 Design of gRNA and Donor Templates

Both CFTR and MUC5B-targetting gRNAs were designed using CRISPOR (Haeussler et al., 2016) and CHOPCHOP (Montague et al., 2014) open-access online software. An in-house B2M-targetting gRNA (gift from Dr Antony Antoniou) was used as a positive control in gene-editing experiments. All gRNA oligonucleotides were purchased from IDT (table 2.4.1.1). Selection criteria was based on gRNA affinity for target loci, the efficiency of DSB induction and the proximity to the desired target loci, three important criteria previously highlighted in the context of gene-editing (Konstantakos et al., 2022; Hollywood et al., 2016). An MIT specificity score of > 50 and a CFD score > 23 were set as the cut-off, as recommended by the tool developers (Haeussler et al., 2016; Montague et al., 2014). The cutting frequency determination (CFD) is a score designed to rank specificity of gRNAs. This prediction algorithm was initially developed for the CRISPR/Cas9 system (Doench et al. 2016) as a weighted function of the mismatch number, position, and nucleotide identity. The CFD score is the percentage activity values provided in a matrix of penalties based on mismatches of each possible kind at each place within the guide RNA sequence (Doench et al. 2016). Increasing numerical value approaching a 100% indicate a higher affinity for DNA target (Doench et al. 2016).

Table 2.4.1.1 List of gRNA *sequences* designed to target CFTR and MUC5B.

gRNA	**Sequence**
CFTR-targeting gRNA 1	ATATTTTCTTTAATGGTGCC
CFTR-targeting gRNA 2	GGGAGAACTGGAGCCTTCAG
CFTR-targeting gRNA 3	AATGGTGCCAGGCATAATCC
CFTR-targeting gRNA 4	TCTGTATCTATATTCATCAT
CFTR-targeting gRNA 5	CAAAGCATGCCAACTAGAAG
Muc5B-targetting gRNA 1	TCAACT GTGAAGAGGT GAAC

Muc5B-targetting gRNA 2	CAGCGCCTTCAACTGTGAAG
Muc5B-targetting gRNA 3	GCAAACAGGC CCGACTCCCA

CFTR donor template 1-4 were custom-designed and purchased from IDT (table 2.4.1.2). CFTR donor template 1 and 2, were purchased as dsDNA, < 200 bp in size, as previously described (Hollywood *et al.*, 2016). CFTR donor template 3 and 4, were purchased as ssDNA, with shorter, asymmetric homology arms, complementary to the non-target DNA strand to optimise annealing, as previously described (Richardson *et al.*, 2016).

Table 2.4.1.2 List of donor template *sequences* designed to target CFTR.

gRNA	Sequence
CFTR donor template 1	GCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCTTGGATTATGTCTGGC
(Containing *AclI*)	ACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATAT AGA TAC
	AGAAGCGTCATCAAAGCATGCCAACTAGAAGAAGTAAGAAACTATGTG
	AAAACGTTTTGATTATGCATATGAACCCTTCACACTACCCAA
CFTR donor template 2	TTCTAATGATGATTATGGGAGAACTCGAGCCTTCAGAAGGTAAAATTAAG
(Containing *TaqI-v2*)	CACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCTTGG ATTATG TCT GGC
	ACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGAT
CFTR donor template 3	TTCTGTATCTATATTCATCATAGGAGACGACAAAGATGATATTTTCTTT
(Containing *BsmBI-v2*)	AATGGTGCCAGGCATAATCCA
CFTR donor template 4	GGCACCATTAAAGAAAATATCATCTTTGGCGTCTCTTATGATGAATATAGA
(Containing *BsmBI-v2*)	TAC AGA AGC GTC ATCAAAGCATGCCAACT

The absent 'CTT/AAG' (highlighted in red) triplet bases in Δ F508 CF genome was introduced in each template. Each donor contained silent mutations that introduced restriction enzyme sites for downstream restriction fragment length polymorphism (RFLP) screening. CRISPR shield mutations/silent mutations (highlighted in green) were included in the ssODN sequences to prevent gRNA binding and reversal of the knock-in allele.

2.4.2 Design 1: tRNA^{Gly}-Flanked gRNAs

The gRNA chosen was self-designed CF-gRNA 1. Rice tRNA^{Gly} obtained from (Dong *et al.*, 2017; Xie, Minkenberg and Yang, 2015; Tang *et al.*, 2016) was chosen to flank gRNAs on each side to permit *in cellulo* maturation of STU components. Two cassettes were subsequently designed, the first comprised of fifty adenosine residues (50 poly A) plus a tRNA^{Gly} sequence (Cfg 1 frag 1), while the second comprised of a gRNA-tRNA^{Gly} sequence (Cfg 1 frag 2). Cassettes were purchased as gBlock gene fragments from (IDT). GBlocks were delivered as 500 ng of dry lyophilized powder, which were then reconstituted in 10 µL IDTE pH 7.5 (IDT). Oligonucleotide sequences used for subsequent cloning steps

are listed (table 2.4.2.1). The $tRNA^{Gly}$ – interspaced STU was cloned into the *PmeI* site of BPK4410, downstream of HypaCas9 sequence, a variant containing six substitution mutations (N692A, M694A, Q695A, H698A) from the wild-type *S. pyogenes* Cas9, which improves specificity without compromising on-target efficiency within human cells (Chen *et al.*, 2017)

Table 2.4.2.1 List of primers used for the amplification of $tRNA^{Gly}$-interspaced gene-editing cassettes and host plasmid vectors

Oligonucleotide	Sequence
Cfg 1 frag 1_F	CATCACCATCACCATTGAGTTTGTTCATCGTCTCATCAAG
Cfg 1 frag 1_R	CTAGTTGGCATGCTTTGCTGCACCAGCCGGGAATCGAAC
Cfg 1 frag 2_F	GGCTGGTGCAGCAAAGCATGCCAACTAGAAG
Cfg 1 frag 2_R	GTCGAGGCTGATCAGCGGGTTTCCTACCAAGCTTCACATG
BPK4410_F	AAACCCGCTGATCAGCCTCG
BPK4410_R	AAACTCAATGGTGATGGTGATG
Cfg 1 frag 1	GTTCATCGTCTCATCAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACG GTACAGACCCGGGTTCGATTCCCGGCTGGTGCA
Cfg 1 frag 2	GCAAAGCATGCCAACTAGAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGT CTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGT GCACATGTGAAGCTTGGTAGG.

2.4.3 Design 2: Ribozyme-Flanked gRNAs

The gRNA sequence used was self-designed CF-gRNA 1. Self-cleaving 5' Hammerhead (HH) and 3' hepatitis delta virus (HDV) ribozyme sequences obtained from (Gao and Zhao, 2014), were chosen to flank the gRNA on either side to permit *in cellulo* maturation of STU components. gRNA-ribozyme cassettes were purchased as gBlock gene fragments (IDT). GBlock was delivered as 500 ng of dry lyophilized powder, which was then reconstituted in 10 µL IDTE pH 7.5 (IDT). Oligonucleotide sequences to amplify insert for subsequent cloning steps are listed below. The ribozyme – interspaced STU was cloned into the *XhoI and PstI* restriction sites of BPK4410 plasmid, downstream of HypaCas9.

Table 2.4.3.1 List of primers used for the amplification of ribozyme interspaced gene-editing cassettes

Oligonucleotide	Sequence
Cfg1_F4	GTTCATCTCGAGATCAAGCTTTG

Cfg1_R4	CCTACCCTGCAGCACATG
Cfg 1 frag 1	GTTCAT*CTCGAG*ATCAAGCTTTGCCTGATGAGTCCGTGAGGACGAAACGAGTAAGC
	TCGTCGCAAAGCATGCCAACTAGAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAAT
	AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGGCCGGCAT
	GGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATG
	GGACCATGTG*CTGCAG*AGTAGG

Black letters, gblock adapter sequences and corresponding amplification primers; Brown nucleotides, gRNA sequence; Green nucleotides, Ribozyme sequences; Italics, *XhoI* and *PstI* type II restriction enzyme recognition sequences used for cloning.

2.5 Gene-Editing Detection

2.5.1 Genomic DNA Extraction

DNA extraction of genomic DNA gene-edited cells was performed using the PureLink™ genomic DNA mini kit (ThermoFisher). Initially gene-edited cells were harvested by washing In PBS (1x) before detaching by addition of Trypsin-EDTA (1x) solution following standard cell culture procedure (section 2.1.2) The cell pellets were then resuspended in 200µL PBS (1x). Next 20 µL of Proteinase K (ThermoFisher) and RNase A (ThermoFisher) were added to each reaction, followed by vortexing before incubating at room temperature for 2 minutes. Next 200 µL of PureLink™ genomic lysis buffer (ThermoFisher) was added to each reaction, vortexed for 5 seconds, then incubated at 55°C for 10 minutes. The contents (640 µL) were then transferred into an PureLink™ spin column (ThermoFisher) then centrifuged at 10,000 x g, for 1 minute at room temperature, discarding the collection tube. The spin column was placed into a fresh collection tube, next 500 µL of wash buffer 1 was added to the spin column, then centrifuged at 10,000 x g, for 1 minute at room temperature, discarding the collection tube. The spin column was placed into a fresh collection tube, then 500 µL of wash buffer 2 was then added to the spin column, then centrifuged at maximum speed, for 3 minutes at room temperature, discarding the collection tube. To elute genomic DNA, the spin column was placed in a fresh 1.5 mL microcentrifuge tube, then 50 µL elution buffer (ThermoFisher) was applied to the centre of the column, then incubated at room temperature for 1 minute, followed by centrifugation at maximum speed, for 1 minutes at room temperature. Eluate containing genomic was stored at -20°C until future use.

2.5.2 Magnetic Purification of DNA Amplicons

Prior to T7E1 assay, DNA amplicons were purified using the Agencourt AMPure XP reagent (Beckman Coulter, Buckinghamshire, UK). Initially DNA amplicons were transferred into a sterile 1.5 mL microcentrifuge tube followed by addition of 1.8 volumes of Agencourt AMPure XP reagent, mixing and incubation at room temperature to permit binding of magnetic beads to DNA amplicons. Reaction tubes were subsequently placed on a magnetic rack for 2 minutes to permit bead capture, then the supernatant was discarded while contents were still on the magnetic rack. Magnetic beads were washed 2 x 30 seconds with 70% molecular-grade ethanol, then allowed to air-dry for 2 minutes to ensure residual ethanol had evaporated. Reaction tubes were removed from the magnetic rack, then 40 µL of elution buffer (ThermoFisher) was used to resuspend the magnetic beads, followed by incubation for 2 minutes at room temperature. The reaction tubes were then placed on a magnetic rack for 1 minute to separate the beads from the eluate. The eluate was transferred into a fresh 1.5 mL Eppendorf and stored at -20°C until future use.

2.5.3 Mutation Detection Assay

To assay for indel mutations post-gene-editing, a T7E1 assay was performed. T7E1 assay (New England Biolabs) was carried out following manufacturer's instructions. A standard 20 µL T7E1 reaction contained NEBuffer 2 (1x), 200 ng DNA substrate and the reaction volume was topped-up to 19 µL using nuclease-free water (New England Biolabs). Of note, a positive control for T7E1 reaction was included comprising of an equimolar ratio of Δ F508 CF and wild-type PCR amplicons generated using *CFTR*_F and R primer pair (table 2.6.4.1). An annealing step ensued in a ProFlex™ PCR system as follows: 95°C for 5 minutes, 95-85°C at a ramp rate of -2°C/second, 85-25°C at a ramp rate of -0.1°C/second, followed by a 4°C infinite hold. Next 1 µL (10 units) of T7E1 was added, then reactions were incubated at 37°C for 1 hr to permit enzymatic activity. Cleavage products were accessed by agarose gel electrophoresis (section 2.6.1). Once gene-editing was detectable by T7E1, initial purified PCR products were outsourced and sequenced by Eurofins Genomics using *CFTR*_F and R primers. TIDE (Tracking of Indels by Decomposition) analysis software was used to analyse the sequencing chromatograms, following the authors protocol (Brinkman *et al.*, 2014).

2.5.4 Densitometry

Densitometry was carried out on the individual DNA bands of interest following gene-editing, when a thorough quantitative examination of these fragments within agarose gels was necessary. When analysing gels with non-specific bands, the DNA ladder was utilised as a reference to determine the bands of interest. Genesys™ software version 1.8.5 (Syngene) was used on SNG files, to perform densitometry analysis. To start, a rectangular box was created manually to cover the sample lanes on

the gels, with each lane being numbered. Once the 'analysis icon' was selected, numeric values were autogenerated, representative of the density of each DNA peak within a given gel lane. The relative density of a desired peak(s) was calculated as a percentage of the densities of all the peaks combined within a given gel lane.

2.5.5 *In vitro* cleavage assay

In vitro screening assay in this context refers to an assay performed *in vitro* (on the lab bench) on purified *CFTR* PCR amplicons generated using primers that flank the custom-designed gRNA sequence to be assayed. The objective here being the validation of gRNA efficiency in a controlled environment, prior to *in cellulo* experiments.

Initially 1uM Cas9 RNP was made by mixing 1uM gRNA and 1uM Cas9 then incubating at room temperature for 10 minutes. A 20uL IVT-cleavage reaction comprised of 1 x Cas9 reaction buffer (IDT), 100nM Cas9 RNP, 10nM DNA (1 : 10 RNP: DNA molar ratio was used unless stated otherwise) and nuclease free water up to 20uL. The reaction mix was incubated at 37°C for 1 hour, followed by digestion of RNP by addition of proteinase K (ThermoFisher) then incubating at 56°C for 10 minutes. Cleavage products were analysed by performing agarose gel electrophoresis (section 2.6.1).

2.6 Cloning

Ribozyme-flanked gBlock was cloned into the *XhoI* and *PstI* sites of BPK4410, then successful gene integration was confirmed by colony PCR of gene of interest. $tRNA^{Gly}$ flanked gRNA was cloned into the *pmeI* site of BPK4410 via Gibson assembly.

2.6.1 Agarose Gel Electrophoresis

To separate DNA fragments based on molecular size, agarose gel electrophoresis was performed. Agarose powder was dissolved in 1 x Tris-acetate-EDTA (Biorad) to a final concentration of 1.6 % w/v. 1 x loading buffer (New England Biolabs) was added to each DNA sample before loading into the wells of a gel. 100bp DNA ladder (New England Biolabs) was run alongside the samples. Electrophoresis was conducted at 100V for 1 hour in 1 x TAE buffer, using an electrophoresis tank (Mini-Sub Cell GT Cell, Biorad) connected to a power supply (PowerPac Basic Power Supply, Biorad). Agarose gels were stained with SYBR™ safe (1x) (ThermoFisher). Agarose gels were visualised on a Syngene Gbox UV transilluminator (Syngene).

Plasmid vectors, alongside a 1 kb Hyperladder (Bioline, London, UK) were run on 1 % w/v agarose gels at 20 mA overnight. Agarose gels were stained with SYBR™ safe (1x) then bands of the correct anticipated size were subsequently excised and purified by gel extraction using the NZYGelpure

purification kit (NZYTech). Agarose gels were visualised on a Syngene GBox UV transilluminator (Syngene).

2.6.2 Gel Extraction

In order to purify linearised plasmid DNA or PCR amplicons of a specific size with an agarose gel a gel extraction was performed. Gel extraction was performed using the NZYGelpure purification kit (NZYTech). Initially DNA fragments were excised from the agarose gels using a sharp scalpel, weighed then, then transferred into a 1.5mL microcentrifuge tube. Next 300 µL of Binding buffer was added per 100 mg of gel mass followed by an incubation step at 55°C for 10 minutes. The fully dissolved homogenous mixture was loaded in an NZYTech spin column placed in a collection tube, followed by centrifugation at 11,000 g for 1 minute at room temperature, discarding the flow-through. 600 µL of wash buffer was added to each column then centrifuged at 11,000 g for 1 minute at room temperature, discarding the flow-through. This step was repeated once more. Next centrifugation of the column at 16,500g for1 minute at room temperature was performed to dry any residual ethanol. Finally elution was performed by adding 50 µL of elution buffer to the center of the NZYTech spin column placed in a 1.5 mL microcentrifuge tube, then centrifuging at 11,000 g for 1 minute at room temperature. The eluate containing purified DNA was stored at -20°C.

2.6.3 Conventional Polymerase Chain Reaction

Conventional polymerase chain reaction (PCR) was set up using the Taq DNA polymerase kit (New Englad Biolabs, #M0273). PCR primers used were synthesised by IDT (IDT). 50µl PCR reactions systems were set up consisting of 200nM forward and reverse primers, standard taq reaction buffer (1x) , 200µM dNTPs, 1.25 units Taq DNA polymerase, 5-1000ng template DNA and made up to 50µl with nuclease-free water (New England Biolabs, # B1500S). The PCR thermal cycling was performed in a PCR thermal cycler as follows; 95°C for 30 seconds, 35 cycles of 95°C for 15 seconds, 55°C for 15 seconds, 68°C for 30 seconds, then final extension at 68°C for 5 minutes, followed by a 4°C infinite hold. PCR products were stored at -20°C for future use.

2.6.4 Overlap-Extension PCR

To fuse two smaller insert fragments into one larger fragment for subsequent cloning steps, an overlap-extension (OE) PCR was performed. The OE-PCR method consists of two primary PCR reactions which add the overlapping region to each fragment and a secondary reaction which fuses the two fragments together (see section 5.4.2). PCR primers used were synthesised by IDT (table 2.6.4.1). For the primary reactions, 25 µl PCR reaction systems were set up consisting of 500nM OE_primer pairs 1 and 2, 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix and 50ng GBlock fragments. The reaction volume was made up to 25 µl with nuclease-free water (New England Biolabs). The PCR thermal cycling

was performed in a ProFlex™ PCR system as follows; 98°C for 30 seconds, 35 cycles of 98°C for 10 seconds, 55°C for 15 seconds, 72°C for 15 seconds, then a final extension at 72°C for 2 minutes, followed by a 4°C infinity hold. PCR products were stored at -20°C for future use.

Primary PCR products were visualised using a 1.6% agarose gel, following electrophoresis. After primary PCR, if a single band of anticipated size was observed then gel purification was unnecessary. If multiple bands were observed, then the band with the correct anticipated size was excised and gel-purified using the NZYGelpure purification kit (NZYTech). The concentration of PCR products was quantified using a Nanodrop One machine (ThermoFisher).

Secondary PCR reaction used 50 ng of each primary PCR product as a template and 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix. The reaction volume was made up to 25 µl with nuclease-free water. The PCR thermal cycling was performed in a ProFlex™ PCR system as follows; 98°C for 30 seconds, 10 cycles of 98°C for 10 seconds, 55°C for 15 seconds, 72°C for 30 seconds, then a final extension at 72°C for 2 minutes. Once reaction was complete then 500nM fragment-flanking forward and reverse primers (OE_F1 and OE_R2) were added then PCR thermal cycling was performed as follows; 98°C for 30 seconds, 25 cycles of 98°C for 10 seconds, 55°C for 15 seconds, 72°C for 30 seconds, then a final extension at 72°C for 2 minutes. The final product contained the full length insert which was subject to agarose gel electrophoresis, gel-purification then finally used for subsequent cloning steps.

Table 2.6.4.1 List of primers used for the amplification in PCR reactions

Oligonucleotide	Sequence
SDM_F (Containing 50 x A nucleotides)	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAACTTTGCCTGATGAGTCCG
SDM_R	CTTGATCTCGAGACTTTCCTC
Cfg 1 frag 1_F	CATCACCATCACCATTGAGTTTGTTCATCGTCTCATCAAG
Cfg 1 frag 1_R	TGCACCAGCCGGGAATCGAAC
Cfg 1 frag 2_F	GTTCGATTCCCGGCTGGTGCAGCAAAGCATGCCAACTAGAAG
Cfg 1 frag 2_R	GTCGAGGCTGATCAGCGGGTTTCCTACCAAGCTTCACATG
OE_F1	CATCACCATCACCATTGAGTTTGTTCATCGTCTCATCAAG
OE_R1	TGCACCAGCCGGGAATCGAAC
OE_F2	GTTCGATTCCCGGCTGGTGCAGCAAAGCATGCCAACTAGAAG
OE_R2	GTCGAGGCTGATCAGCGGGTTTCCTACCAAGCTTCACATG
Colony PCR_F	GTGACGGATCCCCCAAGAAG
Colony PCR_R	TCTTCCCAATCCTCCCCCTT

CFTR_F	CATGTGCCCTTCTCTGTGAACC
CFTR_R	CCATTCACAGTAGCTTACCC

2.6.5 Site-Directed Mutagenesis PCR

PCR was used to add overhangs to both ends of gBlock fragments prior to Gibson assembly and to introduce a poly A tail into pre-cloned gene-editing cassettes to improve mRNA stability. PCR was set up using Q5 DNA polymerase kit (New England Biolabs). PCR primers used were synthesised by IDT (table 2.6.5.1). 25 µl PCR reactions systems were set up consisting of 500nM forward and reverse primers, 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix, 5-1000ng template DNA. The reaction volume was made up to 25 µl with nuclease-free water (New England Biolabs, Hitchin, UK). The PCR thermal cycling was performed in a ProFlex™ PCR system as follows; 98°C for 30 seconds, 25 cycles of 98°C for 10 seconds, 63°C for 30 seconds, 72°C for 3 minutes 49 seconds, then a final extension at 72°C for 2 minutes, followed by a 4°C infinity hold. PCR products were stored at -20°C for future use.

2.6.6 Colony Polymerase Chain Reaction

To confirm successful integration of the recombinant insert into a vector, a colony PCR was performed. PCR was set up using the Taq DNA polymerase kit (New England Biolabs). PCR primers used were synthesised by IDT and are listed in (table 2.6.5.1). Briefly 25µl PCR reactions systems were set up consisting of 200nM forward and reverse primers, standard Taq reaction buffers (1x), 200µM dNTPs, 0.625 units Taq DNA polymerase, a single bacterial colony as a template, then made up to 25µl with nuclease-free water (New England Biolabs). The PCR thermal cycling was performed in a ProFlex™ PCR system as follows; 95°C for 5 minutes, 35 cycles of 95°C for 15 seconds, 55°C for 15 seconds, 68°C for 30 seconds, then a final extension at 68°C for 5 minutes, followed by a 4°C infinite hold. Agarose gel electrophoresis was performed on PCR products immediately or samples were stored at -20°C for future use.

2.6.7 Restriction Digestion

To linearise plasmid DNA or to generate appropriate sticky ends for subsequent cloning, restriction digestion was performed. DNA digestion using restriction enzymes (New England Biolabs) was carried out in their manufacturer-recommended buffer following manufacturer's instructions (New England Biolabs). A standard 50 µL digestion reaction contained 5 µL of 10 x reaction buffer, 1 µg DNA substrate, 1 µL (20 units) of each restriction enzyme and the reaction volume was topped-up to 50 µL using nuclease-free water (New England Biolabs). Digestions were incubated at 37°C for 1 hr or 15 min in the case of time-saver qualified enzymes.

2.6.9 Gibson Assembly

Gibson assembly was adopted as a stratedgy to clone the gene-editing $tRNA^{Gly}$ interspaced STU. Initially a PCR reaction using the Q5 DNA polymerase kit was performed on the two Gblocks to be assembled, using primers with overlapping sequences to add homologous DNA overhangs, then a restriction digest reaction was performed to linearise the plasmid vector. These Gblocks containing overhangs and a linearised plasmid vector would permit the assembly of all three DNA components in a one-step isothermal reaction. This isothermal *in vitro* reaction is completed by three enzymes; a T5 exonuclease which exposes ssDNA with complementary sequences capable of annealing; a polymerase which fills the gaps between hybridised DNA and finally a DNA ligase that covalently seals the DNA backbone (Kalva, Boeke and Mita, 2018).

Gibson assembly 20 µl reactions consisted of 10 µl NEBuilder HiFi DNA assembly master mix (New England Biolabs), 0.5 mM vector DNA, 2.5 mM insert DNA, made up to 20 µl using nuclease-free water. One-step Gibson assembly was performed by incubating reaction tubes at 50°C for 1 hour in a thermal cycler. Samples were used immediately for transformation or were stored at -20°C until use.

2.6.10 Kinase-Ligase-DnpI Digestion Reaction

Following SDM PCR, Mutant DNA product was subject to Kinase-Ligase-*Dnp*I (KLD) digestion reaction. This reaction allowed for efficient phosphorylation, intramolecular ligation or circularisation and template removal in a single reaction. A standard 10 µl KLD reaction contained 50 ng mutant DNA template, 5 µl of 2 x KLD reaction buffer (New England Biolabs), 1 µl of 10 x KLD enzyme mix (New England Biolabs) and nuclease-free-water up to 10 µl. KLD reactions were incubated at room temperature for 10 minutes. KLD products were used in subsequent transformation of *E. coli*.

2.6.11 Dephosphorylation of DNA

To prevent the sticky ends of a vector from self-religating after restriction digestions, a dephosphorylation reaction was performed. A standard DNA phosphorylation reaction was performed in 20 µl reactions containing 1 pmol plasmid DNA, 2 µl Antarctic phosphatase reaction buffer (10x), 1 µl Antarctic phosphatase (New England Biolabs), and nuclease-free water up to a volume of 20 µl. Reaction system was incubated at 30 °C for 30 minutes followed by a de-activation step at 80 °C for 2 minutes. Dephosphorylated DNA was subsequently used for ligation.

2.6.12 Ligation of DNA

To seal the sticky ends of two corresponding DNA fragments, a ligation reaction was performed. A standard DNA ligation reaction was performed in 10 µl reactions comprising of 1 µl T4 DNA ligase buffer (10x), 1 µl of T4 DNA Ligase (New England Biolabs), 50ng insert DNA and 300ng plasmid DNA at

a molar ratio of 5: 1 respectively, and finally nuclease free water up to a volume of 10 µl. Reaction mix was then incubated at 16 °C for 48 hrs.

2.6.13 Plasmid DNA Extraction and Purification

DNA extraction of recombinant plasmid vectors from *E.coli* strains was performed using the NZYMiniprep kit (NZYTech). Briefly 5mL LB liquid cultures were inoculated with a single bacterial colony then incubated at 250rpm, 37°C for 16 hr. The next day liquid cultures were pelleted at 11,000 g for 30 seconds at room temperature, discarding the supernatant. The cell pellets were then lysed by addition of 250 µL of Buffer A1 then vortexing vigorously. Next 250 µL of Buffer A2 was added then mixed by inverting the tube 6-8 times. Next 300 µL of Buffer A3 was added followed by mixing by inverting the tube 6-8 times. To separate the cell debris from the lysate, the components of the tube were then centrifuged at 11,000 g for 10 minutes at room temperature. The supernatant was then transferred into an NZYTech spin column placed in a 2ml collection tube, then centrifuged at 11,000 g for 1 minute at room temperature, discarding the flow-through. 500 µL of Buffer AY wash buffer was then added to the NZYTech spin column followed by centrifugation at 11,000 g for 1 minute at room temperature, discarding the flow-through. 600 µL of Buffer A4 was subsequently added to the spin column followed by centrifugation at 11,000 g for 1 minute at room temperature, discarding the flow-through. To dry the silica membrane the spin column was re-inserted into an fresh collection tube then centrifuged at 11,000 g for 2 minutes at room temperature. Finally the dried spin column was re-inserted into a fresh 1.5 mL Eppendorf, then 50 µL of Buffer AE elution buffer was added followed by centrifugation at 11,000 g for 1 minute at room temperature. Eluted pure plasmid DNA was stored at -20°C long-term until use.

2.6.14 Clean-up of Plasmid and Insert DNA

Prior to subsequent ligation steps, digested DNA fragments were cleaned-up using NZYGelpure purification kit (NZYTech). Initially contents of restriction digest reaction were transferred into a sterile 1.5 mL microcentrifuge tube followed by addition of 5 volumes of Binding Buffer, mixing and a brief centrifugation step to pull contents to the bottom of the microcentrifuge tube. Contents were subsequently transferred in fresh NZYTech spin columns then centrifuged at 11,000 g for 1 minute at room temperature, discarding the flow-through. Next NZYTech spin columns were washed with 600 µl of Wash Buffer followed by centrifugation at 11,000 g for 1 minute at room temperature, discarding the flow-through afterwards. NZYTech spin columns were then centrifuged at 11,000 g for 1 minute at room temperature to remove any residual ethanol and dry the column. Finally, DNA was eluted by addition of 50 µl of elution buffer into the centre of the NZYTech spin column, incubating at room temperature for 1 minute then centrifuging at 11,000 g for 1 minute at room temperature. The

cleaned-up DNA was then used immediately for subsequent ligation reactions or stored at -20 °C for later use.

2.7 Reverse-Transcription Quantitative PCR

2.7.1 RNA Extraction

Before proceeding with RNA extraction procedure, pipettes, bench top, fumehood and gloves were cleaned with RNaseZap (ThermoFisher). RNA extraction was carried out using the miRNeasy mini kit (Qiagen) on cells cultured on a 6-well plate to obtain a sufficient amount of total RNA. Prior to total RNA extraction cells were first harvested by washing In PBS (1x) before detaching by addition of Trypsin-EDTA (1x) solution following standard cell culture procedure (section 2.1.2). The cell pellets were then resuspended in 0.7 mL Qiazol Lysis reagent (Qiagen), vortexed then incubated at room temperature for 5 minutes. 140 µL of chloroform (ThermoFisher) was added, followed by vortexing before incubating at room temperature for 3 minutes. The tubes were then centrifuged at 12,000 x g, for 15 minutes at 4°C. The aqueous upper phase was then carefully transferred into a new microcentrifuge tube, followed by addition of 1.5 volumes of molecular grade 100% ethanol (ThermoFisher), then thorough mixing by pipetting. The contents were then transferred into an miRNeasy mini spin column then centrifuged at 8000 x g, for 15 seconds at room temperature, discarding the flow-through. Next 700 µL of buffer RWT was added to the spin column, then centrifuged at 8000 x g, for 15 seconds at room temperature, discarding the flow-through. 500 µL of buffer RPE was then added to the spin column, then centrifuged at 8000 x g, for 15 seconds at room temperature, discarding the flow-through. Another 500 µL of buffer RPE was added to the spin column, then centrifuged at 8000 x g, for 2 minutes at room temperature, discarding the flow-through. To dry the spin column, it was then place in a fresh 2 mL collection tube, then centrifuged at 16,000 x g, for 1 minute at room temperature, discarding the collection tube. To elute total RNA, the spin column was placed in a fresh 1.5 mL microcentrifuge tube, then RNase free water was applied to the centre of the column followed by centrifugation at 8000 x g, for 1 minutes at room temperature. Eluate containing total RNA was subject to subsequent DNase treatment or was stored at -20°C until future use.

2.7.2 DNase Digestion

DNase treated RNA was converted into complementary DNA (cDNA) using the High-Capacity cDNA reverse transcription kit with RNase Inhibitor (ThermoFisher). Reverse-transcription (RT) reactions comprised of 250 - 2000 ng of DNase-treated RNA template, RT reaction buffer (1x), 4 mM dNTPs (1x), Random primer mix (1x), 1 µL Multiscribe reverse transcriptase, 1 µL RNase Inhibitor and reaction volume was made up to 20 µL using nuclease-free water. Tubes were then incubated in a thermal

cycler (ProFlex™ PCR system, ThermoFisher) at 25°C for 10 minutes, 37°C for 2 hours, 85°C for 5 minutes then held at 4°C indefinitely. The cDNA products were immediately used for quantitative PCR reactions (qPCR) or were stored at -20°C until use.

2.7.3 Complementary DNA Synthesis

DNase treated RNA was converted into complementary DNA (cDNA) using the High-Capacity cDNA reverse transcription kit with RNase Inhibitor (ThermoFisher, Cramlington, UK). Reverse-transcription (RT) reactions comprised of 250 - 2000 ng of DNase-treated RNA template, RT reaction buffer (1x), 4 mM dNTPs (1x), Random primer mix (1x), 1 µL Multiscribe reverse transcriptase, 1 µL RNase Inhibitor and reaction volume was made up to 20 µL using nuclease-free water. Tubes were then incubated in a thermal cycler (ProFlex™ PCR system, ThermoFisher, Cramlington, UK) at 25°C for 10 minutes, 37°C for 2 hours, 85°C for 5 minutes then held at 4°C indefinitely. The cDNA products were immediately used for quantitative PCR reactions (qPCR) or were stored at -20°C until use.

2.7.4 Quantitative PCR

Custom primers and probes were designed with the aid of the PrimerQuestTM tool (IDT) and were purchased from IDT. Quantitative polymerase chain reaction (qPCR) was performed in a 0.2 mL MicroAmp™ Optical PCR tube (ThermoFisher) as 10 µL reactions consisting of 900nM forward primer, 900nM reverse primer, and 250nM probe per reaction, 5uL Taqman Fast Advanced Master Mix (2x) (ThermoFisher), 5 - 100 ng cDNA template and nuclease-free water up to 10 uL, on a StepOnePlus™ real-time PCR system (ThermoFisher). Cycling conditions consisted of a UNG incubation at 50°C for 2 minutes, then initial denaturation at 95°C for 2 minutes was followed by 40 cycles of 95 °C denaturation for 1 second and 60°C anneal-extension for 20 seconds All reactions were run as technical duplicates and biological triplicates. qPCR reactions made use of ROX as a passive reference dye, 6-FAM as a reporter dye and ZEN/ IBFQ as dual quenchers. Gene expression was calculated according to the delta Ct method (Livak and Schmittgen, 2001). The '$2^{-\Delta\Delta Ct}$ method' (Livak and Schmittgen, 2001) was used to analyse the qPCR data for pancRNA transcript levels, as a fold change in expression between high and low confluence conditions and to the 18S rRNA house-keeping gene.

Table 2.7.4.1 List of primers and probes used for the amplification in qPCR reactions

Oligonucleotide	Sequence or Catalogue number
*Muc5B*_Spliced_F	GCTGTGTCCCTTTCCTTCC
*Muc5B*_Spliced_R	CAAGGCCACAGCTATTGAGAC
*Muc5B*_ Spliced_P	CTGTTTTCAGCGCCTTCAACTGTGAAGAGG

*Muc5B*_Holo_F	GCTGTGTCCCTTTCCTTCC
*Muc5B*_Holo_R	CAAGGCCACAGCTATTGAGAC
*Muc5B*_Holo_P	CTGTTTTCAGCGCCTTCAACTGTGAAGAGG
18S assay	Hs99999901_s1
MUC5B control assay	Hs00861595_m1
Mature_STU_F	GCAAAGCATGCCAACTAGAAGG
Mature STU_R	GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG
Mature STU_P	CACCGACTCGGTGCCAC
STU_Holo_F	CTGGAAGGTGCCACTCCCACTG
STU_Holo_R	CTAGTTGGCATGCTTTGCGAC
STU_Holo_P	CTGATGAGTCCGTGAGGACGAAACG
Cas9_F	CTGTTCATCCAGTTAGTACAAACC
Cas9_R	GGTAATTGTGCGATCAGG
Cas9_P	CGAAGGCTATTCTTAGCGCCCGC
bGH_F	CTGTGCCTTCTAGTTGCCAG
bGH_R	CAGAATAGAATGACACCTACTCAGAC
bGH_P	CTGGAAGGTGCCACTCCCACTG

2.8 Statistical Analysis

All data in this thesis is provided as the average +/- standard deviation (SD). The efficiency of gene-editing *in vitro* and *in cellulo* assays, as well as cell viability assays, were normally determined with three biological replicates (as stated in figure legends). The SD reported in this thesis is the variability between the individual experiments. Equally, the mean and SD presented for all *in vitro* and *in cellulo* work, as well as cell viability assays are determined from the means of the separate biological studies (n=3). When comparing two datasets, a student's independent t-test was used to determine statistical significance. T-tests employed in all cases were two-tailed. In instances were three or more datasets were compared, a one-way ANOVA was performed, usually followed by a Dunnett's post hoc test to determine the significance of the control group against experimental groups. Statistical analysis was performed using IBM SPSS statistics software (IBM, Hampshire, UK) and RStudio cloud (Rstudio, Boston, USA).

Chapter 3. Design of gRNAs and Donor templates to target the CFTR locus

3.1 Introduction

CRISPR knock-in strategies have been evaluated in the fight against CF on numerous occasions (see section 1.7) (Bjursell *et al.*, 2018; Jacob *et al.*, 2017; Schwank *et al.*, 2013; Stephens *et al.*, 2018). Geneediting using the CRISPR-Cas9 system utilises both a Cas9 component and a gRNA component. Despite its widespread usage and countless benefits, CRISPR gene-editing has several disadvantages. Offtarget effects are one of the most important factors to consider when planning a CRISPR project (Zhang *et al.*, 2015). Therefore, the first step before commencing gene-editing would be the careful design of highly effective gRNAs which are conceived to maximise on-target efficiency while minimising or eliminating off-target binding to similar genomic sequences (Doench *et al.*, 2016).

To achieve this goal, numerous predictive bioinformatic tools have been deployed to *in silico* predict gRNA on and off-target activity, identifying most off-targets with mutagenesis rates > 0.1%, although not all off-target sites can be anticipated (Haeussler *et al.*, 2016). Numerous gRNA design tools exist, although their effectiveness is cell-specific or species-specific. For example, the CRISPRscan design tool is less effective at predicting efficient gRNAs in mammalian cells, in comparison to zebrafish, the genome in which the prediction model was built. A gRNA design tool referred to as CHOPCHOP, has over 200 genomes available on it's website, permitting the user to provide genomic coordinated or target sequences as inputs (Montague *et al.*, 2014). CHOPCHOP permits gRNA design for Cas9, Cpf1, CasX, C2C2, and TALEN nucleases, as well as a variety of DNA altering-functions including knock-in and knock-out experiments (Liu, Zhang and Zhang, 2020). When compared to other gRNA prediction software, CHOPCHOP-designed gRNA achieved the highest precision compared to other predictive tools (Bradford and Perrin, 2019). The CRISPOR gRNA design tool is a comparison and assimilation of different gRNA predictive scores, integrated into a flexible framework that supports up to 113 genomes (Gao *et al.*, 2020). CRISPOR analyses genomic target sites, outputting off-targets with up to four mismatches, in the genome of interest, while predicting on–target activity by ranking gRNAs according to different scores. One such score is the cutting frequency determination (CFD) which functions to rank gRNAs by their specificity for their DNA target (Doench et al. 2016). CFD score is a function of the mismatch number, position, and nucleotide identity (see section 2.4.1). Hence, CHOPCHOP and CRISPOR were the two gRNA design tools adopted for use in this thesis.

3.2 Aims

This work set out to: (i) design a set of gRNA with high affinity for the *CFTR* Δ F508 locus, for future cloning experiments; (ii) gene-edit the *CFTR* locus by co-delivering custom designed gRNA and recombinant Cas9 protein into CFBE41o-, A549 and HeLa cells; (iii) attempt precise correction of the CF-causative Δ F508 genotype in CFBE41o- CF cell lines through HDR; (iv) assess the effect of gene-editing experiments on cellular viability. CRISPR-gRNA DSB induction rates from 34 % up to 88% have been shown to be achievable at the Δ508 locus in previous studies gRNA (Fan et al., 2018; Ruan et al., 2019). HDR efficiency at the same locus are typically low ~ <2% (Schwank *et al.*, 2013; Geurts et al., 2021), although the highest know rate reported was 16.7% (Firth et al., 2015). It is possible that multiple designs of donor templates shall be made, including dsDNA, ssDNA, asymmetric and non-asymmetric donors as previously described (Hollywood et al., 2016 Richardson *et al.*, 2016; Okamoto *et al.*, 2019), in an attempt to achieve efficient HDR. This indicates the need for more efficient donor template iterations to target *CFTR.* Finally, although highly efficient gRNA exists, this study shall attempt to design original gRNAs to add to the repertoire of existing gRNAs and for future patenting implications.

3.3 Hypothesis

The *CFTR*-targeting gRNA designed shall be capable of targeting the Δ F508 locus with high efficiency in assayed cell lines with minimal impact on cell viability. This high affinity for the *CFTR* locus is expected to permit HDR in CFBE41o- cells. There might a requirement, however to perform some optimisation steps to achieve HDR.

3.4 Results

3.4.1 In silico Design of gRNA

In order to target the region surrounding the Δ F508 locus in human *CFTR* exon 11, a set of five gRNAs were designed using CRISPOR (Haeussler et al., 2016) and CHOPCHOP (Montague et al., 2014) open-access online software. These gRNA's were selected based on three criteria. Predicted efficiency, specificity and proximity to the Δ F508 mutation in *CFTR* exon 11. For predicted specificity and efficiency, an MIT specificity score of > 50 and a cutting frequency determination score (CFD) score of 23 were set as the cut-off, as recommended by the tool developers. The gRNAs selected had the least predicted off-target events and were designed to target within 200 bps of the Δ F508 mutation (figure 3.4.1.1), previously demonstrated to be an optimised targeting window in a cystic fibrosis tracheal epithelial (CFTE) cell line (Hollywood *et al.*, 2016).

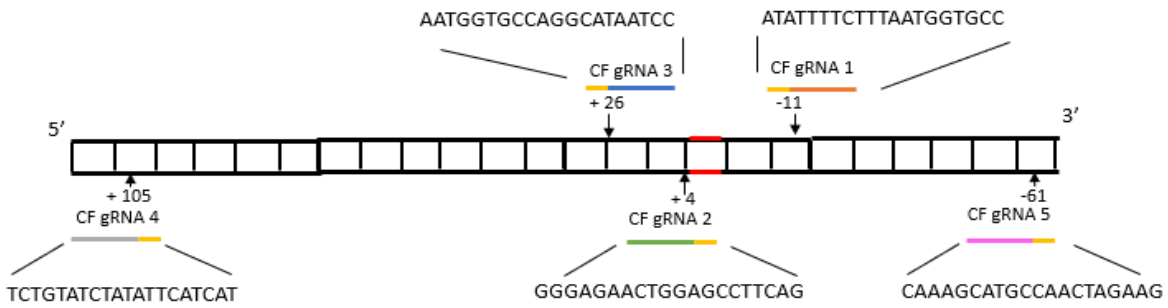


Figure 3.4.1.1 Candidate gRNA mapping coordinates relative to the Δ F508 locus. Five CF gRNAs (coloured lines), their nucleotide sequences (capital letters), corresponding PAM sequences (amber lines) and their cleavage sites (directional arrows) in relation to Δ F508 mutation (in red). All gRNAs were <200bp from the mutation site.

The top five gRNA candidates that were subsequently selected possessed an optimal combination of predicted target specificity, efficiency, and low off-target events, and were assayed *in vitro* (table 3.4.1.1). Of note, a majority of the predicted off-target events were situated within intergenic regions, hence off-target mutagenesis should not have excessive deleterious implications on the protein function.

Table 3.4.1.1 Analysis of the CFTR exon 11 sequence using the CRISPOR CRISPR-Cas9 target predictor.

CF gRNA sequence (number)	MIT specificity score	CFD specificity score	Off-targets 0-1-2-3-4 mismatches	Top 5 Off-target genomic locations, sorted by CFD off-target score (number of mismatches)
ATATTTTCTTTAATGGTGCC (1)	59	73	0-1-2-36-278	Intergenic: *WWOX-RP11-190D6.2* (4) Intron: *ASTN1* (3) Intron: *SLC7A6OS* (4) Intergenic: *LINC01105-AC017053.1* (4) Intergenic: *MIR3660-RP11-61G23.1* (3)
GGGAGAACTGGAGCCTTCAG (2)	61	77	0-0-5-38-248	Intron: *LPCAT3* (3) Intron: *TINCR* (4)

				Intergenic: *RP11-330M19.1-TRABD2B* (4) Intron*: SDK2* (4) Intron*: ADAM2* (4)
AATGGTGCCAGGCATAATCC (3)	76	90	0-0-3-11-136	Intron: *AGBL4* (4) Intergenic: *TRIM60P18-YWHAEP1* (4) Intergenic: *SLC6A17-RP5-1028L10.1* (4) Intron*: CSMD3* (4) Intergenic: *BTBD11-PWP1* (4)
TCTGTATCTATATTCATCAT (4)	30	64	0-1-39-266-336	Intergenic: *OR6K5P-OR6K6* (4) Intergenic: *CTC-463N11.1-RGMB* (3) Intergenic: *UBA6-AS1-ST3GAL1P1/UBA6-AS1* (3) Intergenic: *KRT18P46-SLC4A10* (3) Intron: *PAK3* (3)
CAAAGCATGCCAACTAGAAG (5)	67	85	0-0-3-25-140	Intergenic: *KIAA1009-RP1-90L14.1* (4) Intergenic: *DOCK9-AS2-RPS6P23* (2) Intergenic: *RP11-411H5.1-RP11-556K13.1* (4) Intergenic: *HSPD1P15-CDH18* (4) Intron: *ATP8B4* (4)

The MIT and CFD scores for ranking gRNA specificity range from 0-100, measuring the strength and uniqueness of a guide in the genome. The higher the specificity score, the higher the on-target binding affinity and the lower the incidence of off-target events in the genome, respectively. For each gRNA, the number of off-targets and mismatches are given, as well as the genomic locus for the top five off-target sites.

3.4.2 *In Vitro* Validation of gRNA

To measure the editing efficiency of *in silico* designed gRNAs an *in vitro* screening assay was performed on PCR amplicons generated from genomic DNA of Δ F508-containing CFBE41o- cells. Initially an optimisation step was performed using a positive control gRNA to determine the optimum molar ratio of RNP to DNA to be combined for maximum editing efficiency (figure 3.4.2.1A). RNP to DNA substrate molar ratios of 5 to 1, 10 to 1 and 20 to 1 achieved *in vitro* targeting efficiencies of 38.32 ±0.50%, 38.75 ±0.74%, and 46.37 ±5.27% (n=3) respectively. All molar ratio combinations achieved sufficient and no statistically significant *in vitro* cleavage efficiencies when tested by two-way ANOVA, therefore the manufacturer-recommended RNP to DNA substrate molar ratio of 10 to 1 was adopted for further *in vitro* validation assays to conserve reagents while maximising efficiency.

In vitro cleavage assayed five CF gRNAs alongside a control gRNA targeting the human *B2M* gene for nuclease efficacy validation. Cleavage efficiency of CF gRNA 3 was shown to be the most robust at 76.55 ±3.95% (n=3), as determined by densitometry, highlighting it as a prime candidate for future *in cellulo* validation and cloning experiments. The difference in targeting efficiencies of all gRNAs was statistically significant (two-way ANOVA, p value < 0.001), ranging from 49.41 ±0.54% to 76.55 ±3.95% (table 3.4.2.1), indicating that all gRNAs were good candidates for future *in cellulo* validation as well. Cleavage of all gRNA target amplicons was observed within the expected target region suggested by predicted cleavage fragments of the correct size (figure 3.4.2.1B).

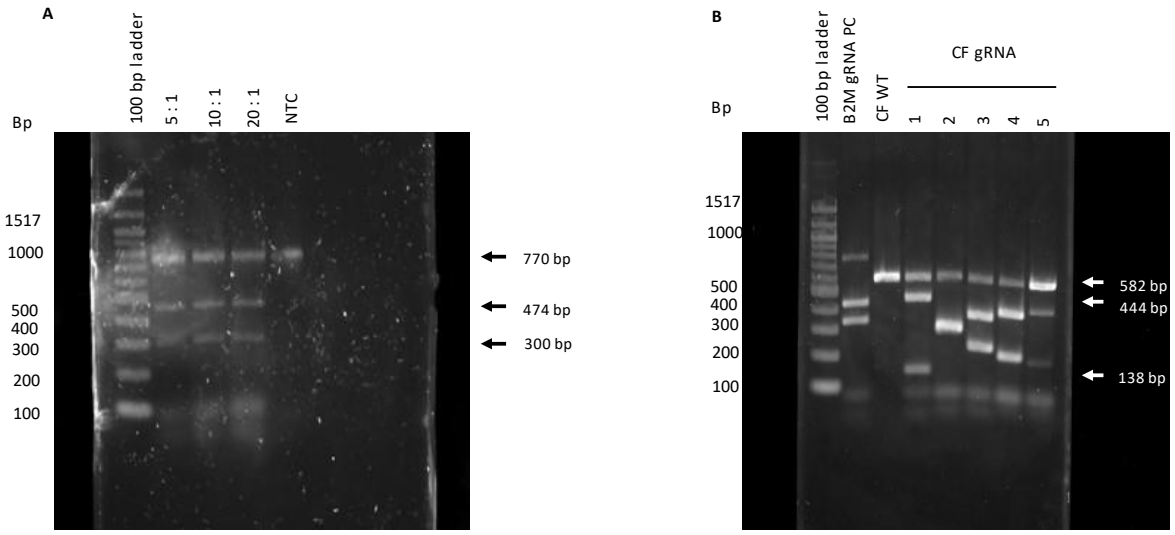


Figure 3.4.2.1. Cleavage assay optimisation and CFTR Δ F508 -targeting gRNA in vitro screening. A. In vitro cleavage assay optimisation using a positive control B2M gRNA using RNP to DNA molar ratios of 5 : 1, 10 : 1 and 20 : 1, while including a non-targeted B2M wild-type amplicon (NTC) as a negative control. **B.** Validation of custom designed CFTR-targeting gRNA (CF gRNA 1-5) activity and potency in vitro on PCR amplicons, while including B2M-targetting positive control gRNA (B2M gRNA PC) and a non-targeted CFTR wild-type amplicon (CF WT) as a negative control. Cleavage products were analysed on a 1.6% agarose gel stained with SYBR™ safe. Arrows indicate the expected fragments upon cleavage by Cas9 endonuclease for B2M gRNA (**A**) and CF gRNA 1 (**B**). Blurry bands can be observed around or below the 100bp mark potentially due to primer dimer formation.

Table 3.4.2.1 Efficiency of CF gRNAs assayed *in vitro*

CF gRNA	Targeting Efficiency (±SD) %
1	64.07 (1.98)
2	64.69 (1.47)
3	76.55 (3.95)
4	73.91 (3.37)
5	49.42 (5.41)

3.4.3 *In Cellulo* Validation of gRNAs

To measure the editing efficiency of CF gRNAs *in cellulo*, CF gRNAs were transfected in CFBE41o-, A549, and HeLa cells by lipofection. The efficiency of each gRNA was determined using the structure-selective T7 endonuclease 1 enzyme assay, which detects structural deformities in heteroduplexed DNA (figure 1.5.2.1).

Densitometry analysis revealed that gRNA efficiency varied between cell lines (figure 3.4.3.1 and table 3.4.3.1), with CF gRNA 1 having the most robust activity in CFBE41o- cells with an efficiency of 45.19 ±3.50% (n=3), while CF gRNA 5 was found to be the most active in HeLa cells with a maximum cleavage efficiency of 26.82 ±2.83% (n=3). Finally in A549 cells, CF gRNA 5 outperformed other gRNAs with a cleavage efficiency of 24.77 ±4.96%. Based on the direct relevance of CFBE41o- genotype to this project, CF gRNA 1 was noted as a prime candidate for future gRNA cloning strategies.

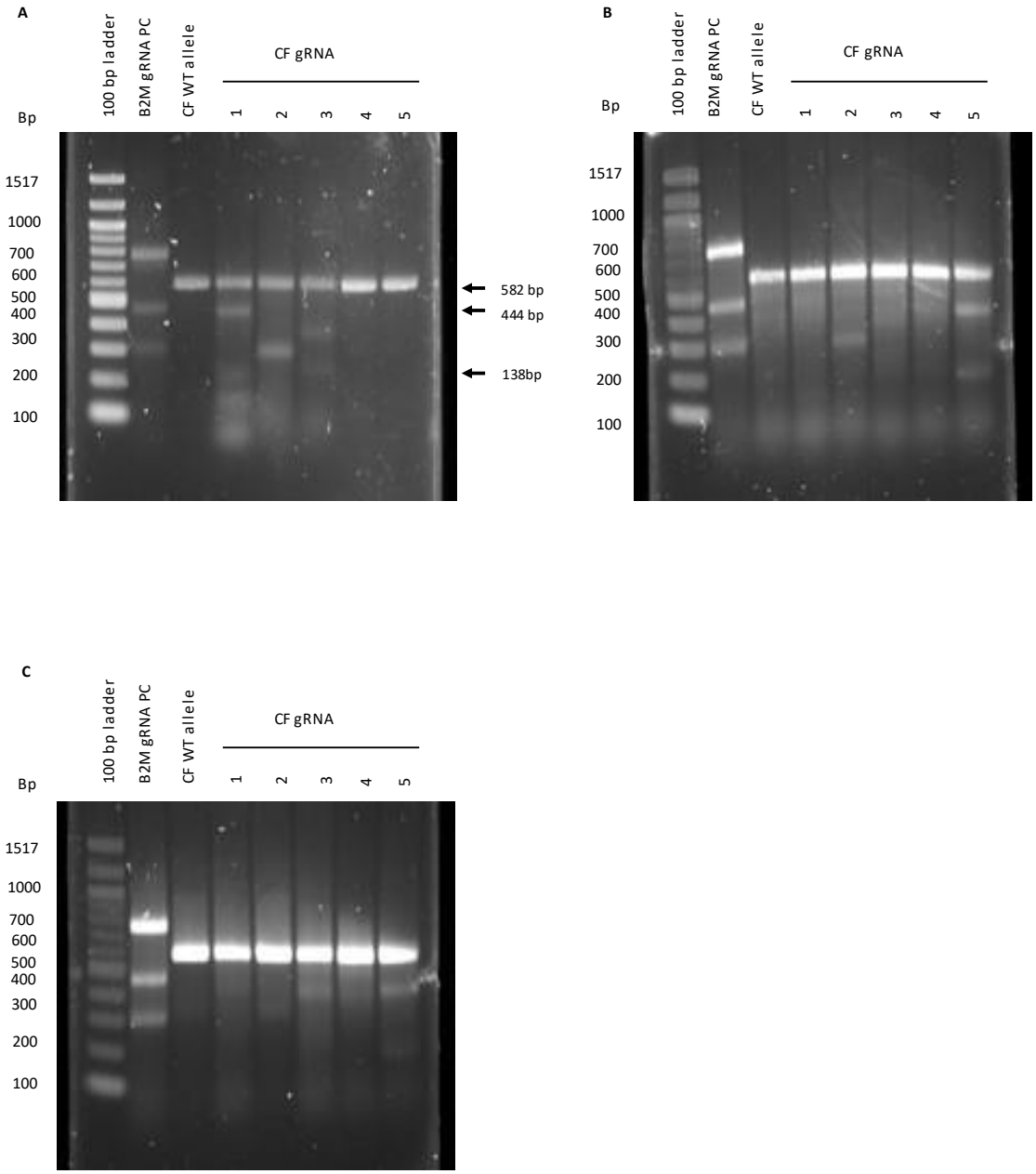
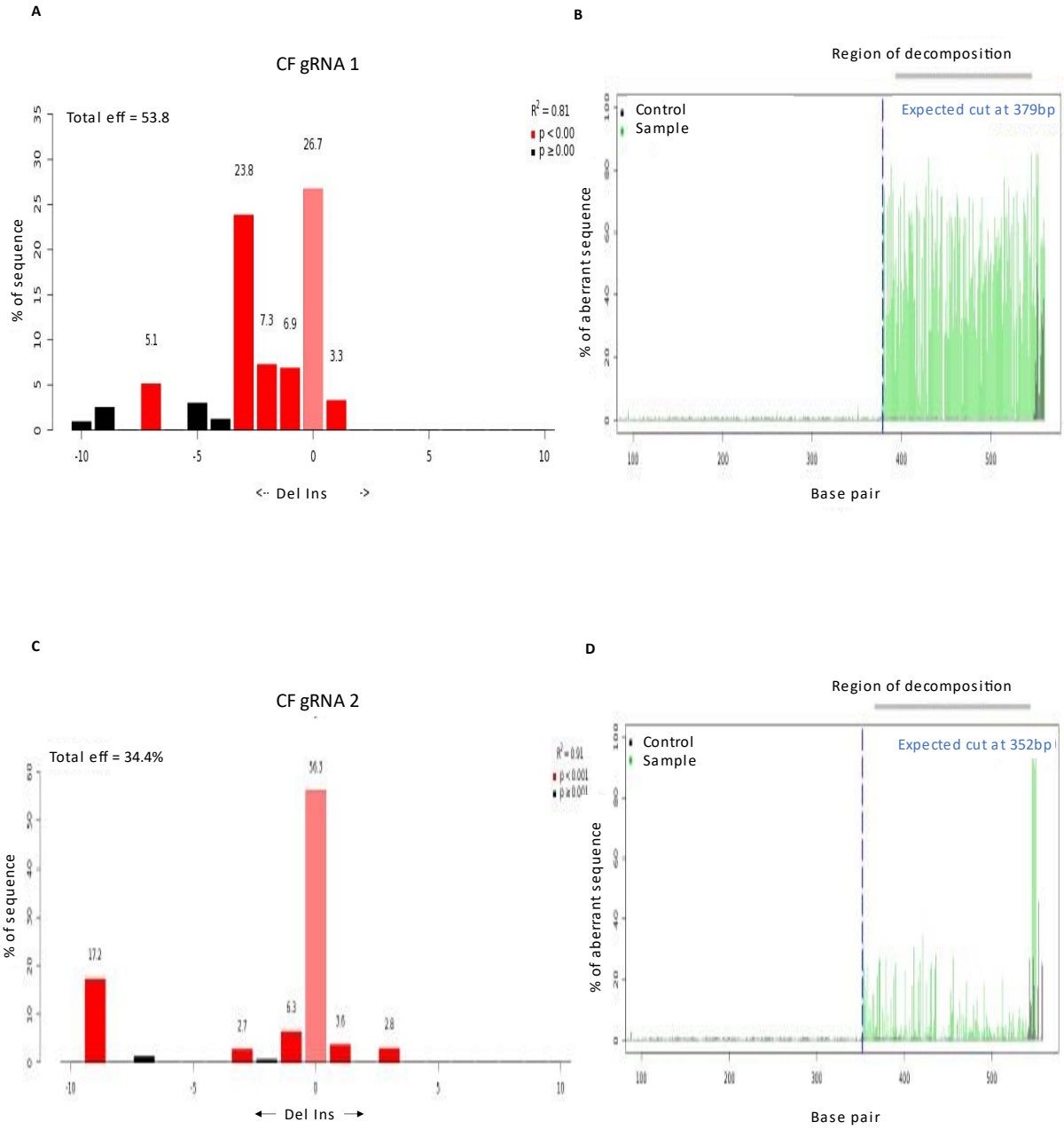


Figure 3.4.3.1 Attempted In cellulo gene-editing *using* CFTR-targetting gRNAs. In CFBE41o- (A), A549 (B) and HeLa (C) cells*,* T7E1 cleavage assay*s* w*ere* performed on DNA targeted by a positive control gRNA (B2M gRNA *PC*); CFTR-targeting gRNAs (CF gRNA 1-*5*). Arrows indicate the expected fragments of CF gRNA 1-targetted DNA, upon cleavage by T7 enzyme. Cleavage products were analysed on a 1.6% agarose gel stained with SYBR™ Safe. Blurry bands can be observed around or below the 100bp mark potentially due to primer dimer formation.

Table 3.4.3.1 Efficiency of CF gRNAs across three different cell lines

CF gRNA Efficiency (+/- SD)%					
Cell line	1	2	3	4	5
CFBE41o-	45.19 (3.50)	24.27 (2.62)	14.41 (1.12)	No edit	No edit
HeLa	18.82 (0.92)	13.77 (1.28)	12.07 (4.35)	7.12 (4.05)	26.82 (2.83)
A549	No edit	9.03 (1.59)	3.27 (3.67)	No edit	24.77 (4.96)

The T7E1 assay used for densitometric estimates is not entirely accurate as it only supplies an estimate of the CF gRNA editing efficiencies, therefore to precisely measure editing effciency, Sanger sequencing was performed on *CFTR* Exon 11 in the CFBE41o- cells, then downstream analysis was performed using TIDE online tool (figure 3.4.3.2) (Brinkman *et al.*, 2014). This tool was used to confirm that CF gRNAs bind and induce double-stranded breaks 3 bp upstream of the PAM. CF gRNA 1, CF gRNA 2 and CF gRNA 3 demonstrated editing efficiencies of 53.80 %, 34.40% and 8.90%, repectively, confirming CF gRNA 1 as a prime candidate for downstream analysis and cloning experiments.


A
CF gRNA 1
Total eff = 53.8
R^2 = 0.81
p < 0.00
p ≥ 0.00
% of sequence
5.1
23.8
7.3
6.9
26.7
3.3
-10
-5
0
5
10
<-- Del Ins -->
B
Region of decomposition
Control
Sample
Expected cut at 379bp
% of aberrant sequence
100
200
300
400
500
Base pair
C
CF gRNA 2
Total eff = 34.4%
% of sequence
17.2
2.7
6.3
36.1
3.6
2.8
Del Ins
D
Region of decomposition
Control
Sample
Expected cut at 352bp
% of aberrant sequence
Base pair

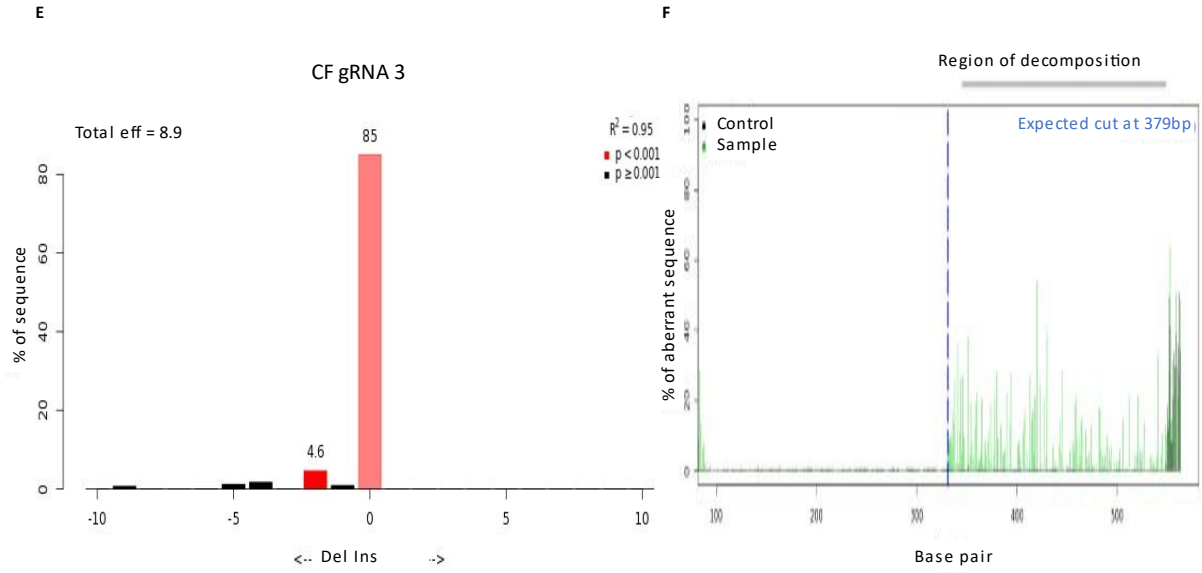


Figure 3.4.3.2 Rank-ordered performance of CF gRNAs by NHEJ-directed indel induction; The three active sgRNAs identified from T7EI experiments in CFBE41o- cells were analysed quantitatively by TIDE following Sanger sequencing of the NHEJ-derived amplicons. Panels A, C, and E present the total efficiency of CF gRNAs 1, 2 and 3, respectively, as the total number of sequences with insertions (right of zero on X axis), or deletions (left of zero on the X axis) with the distance from zero on the X axis identifying the number of indel bp per sequence. Panels B, D, and F represent a visualisation of the aberrant sequence signal in gene-edited DNA (vertical green lines), the DSB location (vertical blue dotted line) and the region of decomposition (grey bar).

3.4.4 *In silico* Design of HDR Templates

To introduce the wild-type allele into CFBE41o- cells which contain the Δ508 genotype, a series of dsDNA HDR templates were designed. Initially, two donor template designed contained 150 and 193 bp of double-stranded CFTR sequence covering CF gRNA 1, 2 and 3 with at least 36 bp of homology arms from the double-stranded break sites of each gRNA (figure 3.4.4.1). Donor templates one and two were designed to introduce 6 and 7 nucleotide changes in comparison to the wild-type sequence including 'CTT' insertion that reverts the allele to wild-type. Silent mutations were introduced at the Cas9 PAM sites preventing the gRNA-Cas9 complex from re-cutting the template. Finally, silent mutations were introduced to donor one and two to generate *Acl*I and *Taq*1v2 restriction enzyme sites respectively for rapid digestion-based validation of edited cells.

HDR donor templates 3 and 4 were equally designed with significantly shorter lengths at 70bp and 80bp respectively, equally possessing shorter mutation-to-cut-site distances (figure 3.4.4.1). These new templates were complementary to the non-target DNA strand and possessed asymmetric

homology arms optimised for annealing, strategies previously described to improve HDR efficiency (Richardson *et al.*, 2016).

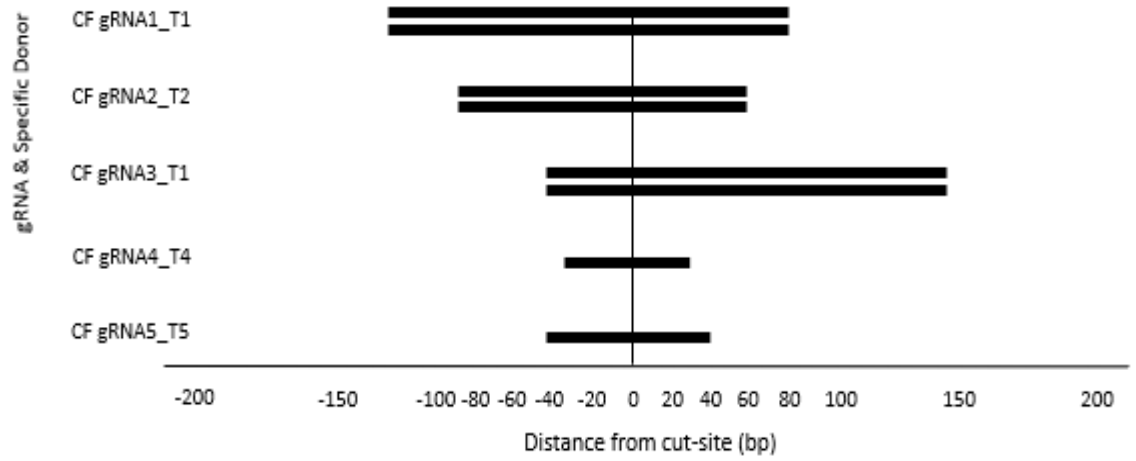


Figure 3.4.4.1 HDR donor template designs and homology arm lengths in relation to DSB site. HDR dsDNA templates (T1 and T2) and HDR ssDNA templates (T3 and T4), the DSB location (vertical black line) of their respective CF gRNAs (CF gRNA 1-3).

3.4.5 HDR Attempted in CFBE41o- Cells

To introduce the donor template directly into the nucleus of cells to permit HDR, CF gRNA, Cas9 and HDR template were nucleofected into CFBE41o- cells, then genomic DNA was extracted from gene-edited and wild-type cells at 24 hours post-nucleofection. HDR events measured by percentage of *Acl*I and *Taq*1v2 digestion efficiency revealed no editing by any gRNA, despite the inclusion of a manufacturer-supplied HDR enhancer in the cell growth media post-nucleofection, intended to boost HDR efficiency (figure 3.4.5.1).

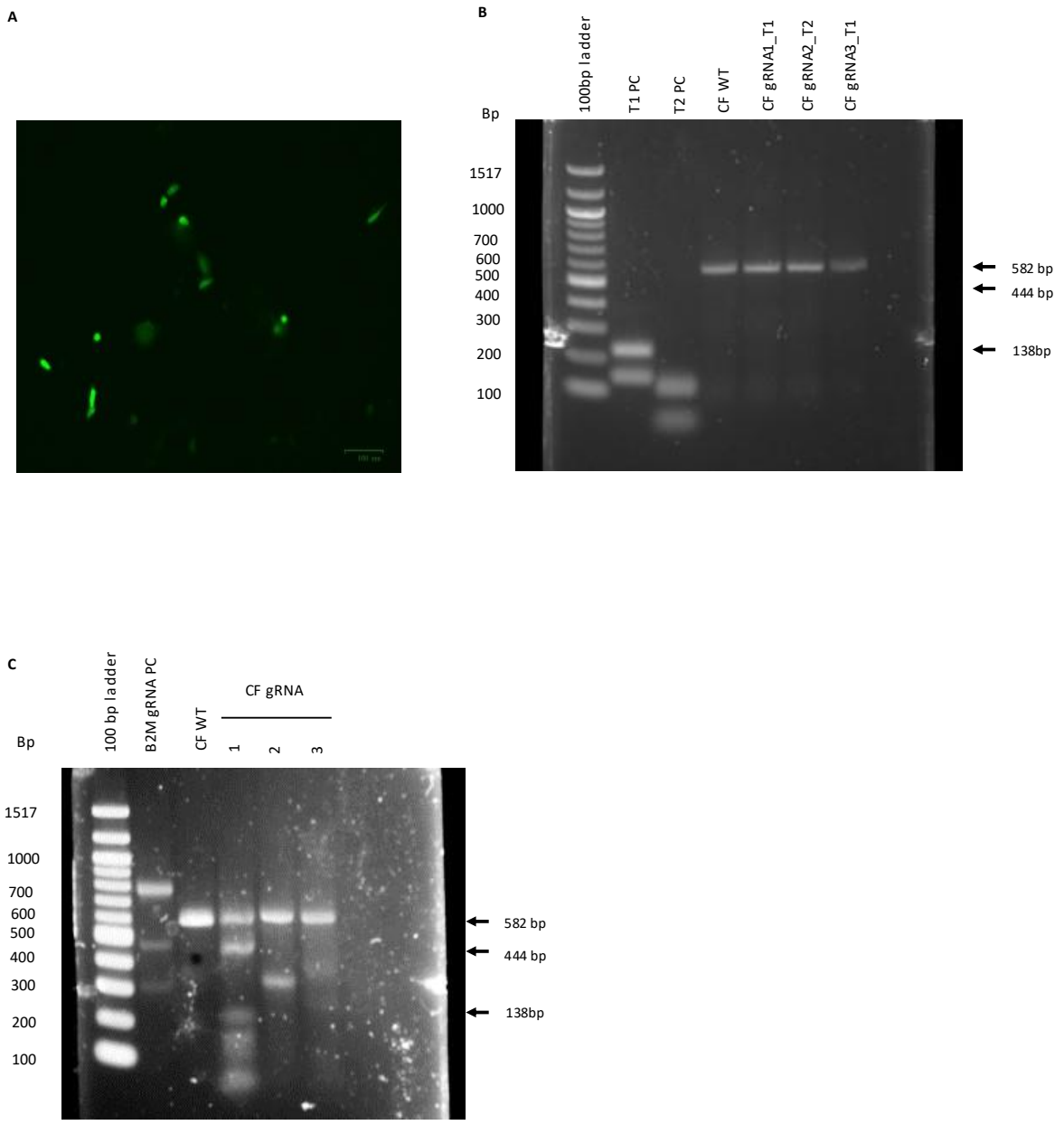


Figure 3.4.5.1 Attempted validation of CF gRNA-specific donor templates *in cellulo*. **A**. Nucleofection of CFBE41o- cells with a GFP encoding plasmid served as a positive control for this transfection technique. **B**. Restriction digestion (**B**) and T7E1 assay (**C**) was performed on DNA targeted by CF gRNA 1 – 3. T1, *HDR* donor template 1; T2, *HDR* donor template 2; CF gRNA _T, CF gRNA co-delivered with respective *HDR* donor template; T PC, HDR donor template positive control for the respective restriction digestion reactions; CF WT, CF wild-type amplicon; B2M gRNA PC, B2M positive control gRNA. Arrows indicate the anticipated fragments for CF gRNA 1-targetted DNA upon restriction enzyme digestion (B) and T7E1 assay (C). Data is representative of three separate biological experiments (n=3). Cleavage products were analysed on a 1.6% agarose gel stained with SYBR^{TM} Safe.

The effect of HDR on cell viability was further investigated using trypan blue staining. As observed HDR had no impact on cell viability assessed after 24 hours post-nucleofection (table 3.4.5.2). Hence one-way ANOVA did not reveal any significant difference between CFBE41o- cells transfected by gRNA plus donor templates and un-transfected control groups. Equally a Dunnett's' post hoc test, did not reveal any significant difference between any CF gRNA plus donor template-transfected groups and non-transfected control groups. Therefore, these results suggest that the manufacturer-recommended amount of 30pmols of donor template used to be sufficiently tolerable to the cells.

In a subsequent attempt to achieve HDR, donor templates 3 and 4 were designed and nucleofected into CFBE41o- cells. These ssODN templates were denatured at 95°C prior to delivery to eliminate complex secondary structures, then nucleofected in CFBE41o- cells with the inclusion of a HDR enhancer, however, no knock-in events were observed (figure 3.4.5.2). A one-way ANOVA was then performed, followed by a Dunnett's post hoc test, revealing that these newly designed templates had no statistically significant effect on cell viability after 24 hours post-nucleofection (table 3.4.5.2).

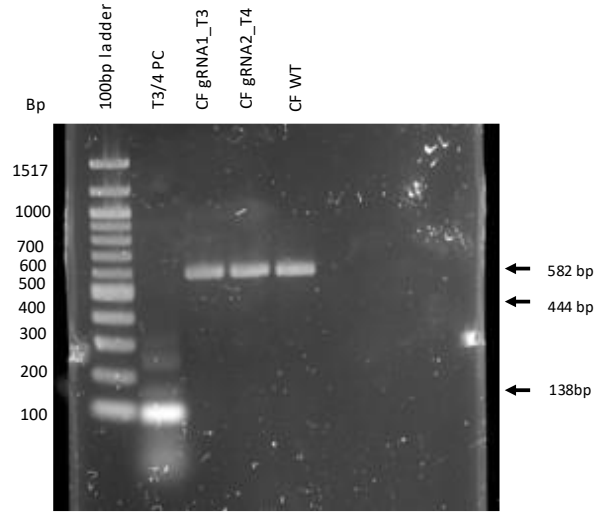


Figure 3.4.5.2 Attempted validation of CF gRNA-specific donor templates in cellulo. Restriction digestion was performed on DNA targeted by CF gRNA 1 & 2. T3, *HDR* donor template 3; T4, *HDR* donor template 4; CF gRNA _T, CF gRNA co-delivered with respective *HDR* donor template; T3/4 PC, positive controls for HDR donor template 3 and 4 restriction digestion reaction; CF WT, CF wild-type amplicon. Arrows indicate the anticipated fragments for CF gRNA 1_T3 targetted-DNA upon digestion by restriction enzymes. Data is representative of three separate biological experiments (n=3). Cleavage products were analysed on a 1.6% agarose gel.

Table 3.4.5.2 Cell survival as a percentage of untransfected controls following attempted HDR (n=3).

CF gRNA_T	Viability (±SD) %
1_1	87.82 (9.21)
2_2	97.88 (5.25)
3_1	91.44 (5.91)
1_3	85.99 (10.58)
2_4	90.67 (1.14)

3.5 Discussion

The shortened life-expectancy of CF patients underpins the need for development of new treatments for the disease. Current approaches to tackling CF disease focus on some alternate ion channels that work in consensus with CFTR in regulating homeostasis, including the inhibition of sodium channel ENaC and ATP12A proton pump and activation of the TMEM protein family, an alternate calcium-activated chloride channel, to compensate for deficient CFTR-mediated Cl- transport into cells (Shei *et al.*, 2018). Other approaches tackle the downstream consequences of the disease; hence antibiotics are employed to treat bacterial infections, and physiotherapy to aid breathing (Lopes-Pacheco, 2016a). In contrast, although precise pharmacological interventions are indeed effective against Δ F508 CF such as Trikafta (comprising of two correctors tezacaftor and elexacaftor and a potentiator Ivacaftor), they come at an elevated cost. Moreover, CF class I mutations which are nonsense mutations resulting in a premature stop codon, remain widely untreatable as there a currently no modulators effective against these mutations (King et al., 2022) and there is increased toxicity associated with long-term use of read through reagents such as gentamicin, commonly used to treat CF class I patients (Wilschanski *et al*., 2003). All these factors justify the need for research into novel gene-editing approaches, as a mechanism to treat CF.

Therefore, this work set out to design and test a set of gRNAs targeting the region surrounding the Δ F508 locus in mutant human *CFTR*, as a model for the development of a future STU-gene-editing approach. This would further add to the pool of effective previously described gRNAs designed to target human *CFTR*. To design these, CHOPCHOP and CRISPOR were all consulted with the later preferentially adopted as the primary design tool because it integrates many gRNA-design tools into one algorithm (Liu, Zhang and Zhang, 2020; Concordet and Haeussler, 2018). The resulting gRNAs had the highest predicted on-target efficacy with minimal off-target events. Of note, a majority of the predicted off-target events were situated within intergenic regions, therefore, minimal off-target effects were anticipated. In general, these intergenic regions are tightly structured due to a high abundance of nucleosomes, hence limiting the accessibility to Cas9 (Horlbeck *et al.*, 2016). However,

in an ideal situation, a gRNA screen should evaluate nucleosome occupancy and mutagenesis activity *in cellulo* and eventually *in vivo* by performing a T7E1 on PCR amplicons generated from these predicted off-target sites. The use of Genome-wide Unbiased Identification of DSBs Enabled by Sequencing (GUIDE-Seq) approach to detecting off-targets could have been deployed in this study. This approach involves the introduction of a dsODN along with RNP when performing gene-editing on cells, permitting dsODN integration into the DSB site in a process similar to NHEJ, then genome-wide NGS is performed on sheared genomic DNA from gene-edited cells using PCR primers specific for the newly-introduced dsODN and the sequencing adapters. This approach has been used to screen for off-targets identifying up to 15 fold more off-targets than *in silico* prediction tools (Tsai *et al.,* 2015). Another approach CIRCLE seq would involve the sequencing of circularised DNA. Here, genomic DNA of gene-edited cells is sheared, then stem loop adapters are ligated to both ends to circularise the DNA molecule, followed by digestion of undesired linear DNA molecules. Circular DNA containing a Cas9 cleavage site is linearised once Cas9 and gRNA are introduced, permitting subsequent adaptor ligation and downstream sequencing, with each pair-end read representing a single off-target (Tsai *et al.,* 2017). This CIRCLE-seq approach has been used to identify over 100 unreported off-targets sites in a single gene (human beta globulin gene) (Tsai *et al.,* 2017).

In this study, gRNA designs tested *in vitro* demonstrated up to >76% activity, which is a robust outcome (Ruan *et al.,* 2019). Equally, gRNA designs demonstrated significant activity *in cellulo*, with the most effective gRNAs in each cell line demonstrating indel frequency > 24%, which is a sufficient level of activity, especially in the context of CF where such a low level of CFTR function restoration is generally considered sufficient to achieve therapeutic outcomes (Duchesneau *et al.,* 2017). Unsurprisingly, the *in vitro* activity of gRNAs far outperformed their respective *in cellulo* activity, due to the ideal conditions being present in the former, such as facile gRNA access to unmodified DNA molecules and optimal buffer salt concentrations to promote endonuclease cleavage activity (Bente, Mittelsten Scheid and Donà, 2020). Interestingly, the effectiveness of individual CF gRNA varied between cell lines, this is not surprising as genomic locus accessibility and transfection efficiency can be quite different amongst different cell lines (Horibe *et al.*, 2014). One possible way to improve gene-editing efficiency would have involved the optimisation of transfection efficiency in the three assayed cell lines. Hence optimisation steps such as, cell seeding density, Cas9 RNP and transfection reagent concentrations and post-transfection incubation time could have been altered to optimise transfection efficiency in a cell-type dependent manner. In the case of nucleofection, the amount of GFP positive cells expressed as a percentage of the total number of cells transfected, could have been used to determine the effect of the above-mentioned optimisation steps on transfection efficiency. TIDE analysis was able to further show these indel species were mostly short insertions or deletions

of 1-10 nucleotides around the double-stranded break which is in line with previously published data (Allen *et al.*, 2019). Furthermore, the gene-editing efficiency values calculated using the TIDE online tool were slightly elevated compared to those calculated by densitometry following T7E1, except in the case of CF gRNA 3, suggesting a general trend where editing efficiency is underestimated by T7E1 analysis, as previously described (Sentmanat *et al.,* 2018). Effectiveness of CF gRNAs designed in A549 and HeLa cells demonstrated their versatility and robust activity *in cellulo*, which would be particularly useful if these gRNAs were used to edit cells of varying physiology.

Nonetheless, it would have been preferable to identify gRNAs of even higher efficiencies. Numerous factors affect gRNA editing efficiency including gRNA sequence composition including GC content, nucleotide composition, gRNA energetic properties such as RNA secondary structure, free energy, and melting temperature, and genetic and epigenetic factors such as, chromatin structure and gene expression (Bruegmann, Deecke and Fladung, 2019). A combination of these confounding factors could have invariably led to the limited editing efficiency. If more time and resources were available, the introduction of chemically modified bases shown to improve Watson-Crick binding of antisense oligonucleotides, to the 5' end of gRNA could improve on-target efficiency and HDR probability (Moschos *et al.*, 2011; Janas *et al.*, 2019).

In this study I aimed to revert the Δ508 mutation present in CFBE41o- cells to the wild-type allele. Careful considerations needed to be made while designing a donor template including the choice of double-stranded or single-stranded donor, sequence length, mutation-to-cut site distance, alterations to the PAM sequences to avoid Cas9 re-cutting and selection criteria. All these factors were considered when designing a first generation of double-stranded donor templates which had alterations to the PAM sequences to avoid Cas9 re-cutting and a cut-to-mutation distance of 70-80 bp. This initial design revealed no detectable HDR events. In an attempt to achieve HDR, a second generation of single-stranded donor templates were designed with new features including shorter cut-to-mutation distances < 20 bp and an edit encoded within the shorter PAM-proximal arm, all factors previously described to improve HDR (Hollywood *et al.*, 2016; Richardson *et al.*, 2016; Okamoto *et al.*, 2019). Surprisingly, these new template design inclusions did not result in any detectable HDR events. There was a possibility that a sizeable population of edited cells were incapable of surviving such a genomic-altering event, hence leading to inadequate gene-editing and HDR events observed (Alvarez *et al.* 2022). However, my cell viability assay results suggested that both conventional gene-editing and HDR attempts had no significant impact on the survival of edited cells. This signifies that off-target mutagenesis did not significantly impact upon genes required for cell survival. The MTT assay used to access viability does have certain drawbacks. MTT-formazan production may decline in tandem with a drop in the levels of D-glucose in the culture medium and mitochondria are still intact during

apoptosis, therefore may be a slight MTT drop in the early stages (Vistica et al., 1991). In future, single cell cloning followed by Sanger sequencing for the confirmation of successful HDR in gene-edited cells could be deployed in place of the less sensitive RFLP which is ineffective at detecting low-efficiency HDR events, typically in the range of 1-3%, but have been reported as high as 40% (Xu *et al*., 2018; Li *et al.,* 2019; Di Stazio *et al.,* 2021). Single-cell cloning would involve the isolation of single cells from a gene-edited population by limited dilution method and growing in a 96-well plates until confluence, followed by genomic DNA extraction, PCR amplification of gene-edited locus and finally Sanger sequencing (Li *et al.,* 2019).

Another potential strategy to achieve HDR would involve blocking NHEJ pathway using small molecules that enhance precision gene-editing directly by promoting HDR pathway (Yu *et al.*, 2015) or indirectly by inhibiting the NHEJ pathway (Frit *et al.*, 2014; Maruyama *et al.*, 2015).

The choice of transfection technique employed for targeted indel generation and introduction of gene-editing cargo was lipofectamine and nucleofection respectively. The levels of gene-editing detected of up to 3 - 45 % was mostly within the normal range of 10 - 90% observable using lipofectamine (Chen, Alphonse and Liu, 2020; Yu *et al.*, 2016; Wang *et al.*, 2018b). Nucleofection as a delivery technique was adopted due to its reported increased HDR efficiency, compared to lipofection (Roche *et al.*, 2018b). This is because nucleofection can deliver up to 70% of gene-editing cargo directly into the cellular nucleus where function is exerted, as opposed to the primarily cytosolic localisation of the lipofectamine cargo, were only about 30% of free donor DNA reaches the nucleus through passive diffusion (Maurisse *et al.*, 2010; Roche *et al.*, 2018b). Hence HDR rates of within 1-20% can be achieved using nucleofection (Sharifi Tabar *et al.*, 2015). However, in this study this targeted repair event was not observable. Of note, viral-mediated delivery of gene cargo is to date the most effective means of achieving HDR (Xiong *et al.*, 2005; Chong, Yeap and Ho, 2021), hence given more time and resources it could have been a viable strategy to test the efficacy of designed CF donor templates. However, use of viral vectors sacrifices the transient nature of the current RNP delivery platform, which could have future safety implications.

In this chapter a set of five CF gRNA's were designed *in silico* then successfully validated in three different cell lines, further adding to the pool of existing gRNA's targeting the CFTR locus. In the CF-relevant CFBE41o- cell line, CF gRNA 1 outperformed other custom-designed gRNAs *in cellulo*, hence a candidate for future cloning experiments. To follow up on these findings, it would be of use to invest time and effort to further design and screen even more gRNA's to improve the probability of successful knock-in using a donor template. In this study knock-in using HDR was not achievable as previously described (Firth et al., 2015; Schwank et al., 2013), however, steps to further optimise donor design

combined with an optimal gRNA design would significantly boost the probability of reverting the ΔF508 allele hence creating a pathway for future studies.

Chapter 4: A SNP involved in Idiopathic Pulmonary Fibrosis.

4.1 Introduction

In the region surrounding the rs35705950 SNP, a promoter-associated long-noncoding RNA (pancRNA) AC061979.1 (gene ID ENSG00000286275) was predicted by the GENCODE project while using computational analysis to identify new genes and new transcript isoforms (Frankish *et al.*, 2019). This novel lncRNA was reported to be subject to splicing, with the rs35705950 SNP situated at the second nucleotide of exon 2, hence potentially perturbing AC061979.1 splicing (Frankish *et al.*, 2019). The possibility of a perturbed splicing event induced by rs35705950 SNP could result in aberrant FOXA2 binding and downstream gene expression. Altogether, the increased understanding of pancRNA involvement in gene expression regulation, together with the location of rs35705950, raised important questions about the role of the AC061979.1 in *MUC5B* gene expression control. Indeed, the pancRNA could be hypothesised to participate in *MUC5B* expression regulation, with rs35705950 perturbing FOXA2 activity directly, or affecting AC061979.1 splicing.

Of significance, there are currently no reports confirming *in cellulo* expression of this pancRNA molecule, nor any functional implication of its' expression in *MUC5B* expression, or IPF-related aberrant MUC5B expression. If the expression of lncRNA AC061791.1 *in cellulo* is confirmed, it would amass evidence that this lncRNA with implications in IPF pathogenesis is expressed. Based on the possible implications of this SNP in the aetiology of IPF, there is an opportunity to attempt the use of gene-editing as a platform to revert the rs35705950 'T' risk-variant to the 'G' non-risk variant in the context of IPF. Hence, we hypothesised it would be possible to target the rs35705950 locus using the CRISPR-Cas9 gene-editing platform in a suitable disease-relevant cell line such as A549 cell lines.

4.2 Aims

This chapter sets out to: (i) validate the *in cellulo* expression of AC061979.1 in A549 and CFBE41o- cells using probe hydrolysis RT-qPCR. (ii) Gene-edit rs35705950 SNP in A549 cells genome using custom-designed gRNA and Cas9 RNP platform. (iii) Assess the effect of gene-editing experiments on cellular viability.

4.3 Hypothesis

AC061979.1 transcript and the spliced variant of this pacnRNA are detectable by RT-qPCR analysis in total RNA extracts of lung epithelial cells. Novel *MUC5B*-targeting gRNAs shall target the *MUC5B* rs35705950 SNP locus in assayed cell lines with minimal impact on cell viability.

4.4 Results

4.4.1 Design and Validation of RT-qPCR Probes for the Detection of the Putative MUC5B pancRNA AC061979.1 Variants

Initially, a set of primers and hydrolysis probes were designed for RT-qPCR detection of the spliced and holo variants of AC061979.1. Subsequently, the amplification efficiency of RT-qPCR primer and probe targetting the spliced version of AC061979.1 was tested against serial dilutions of a commercially procured geneblock used as a PCR template, whereas primers and probes targeting the holotranscript were tested against A549 cell DNA template. The efficiency of primer-probe sets targetting spliced and holo AC061979.1 were determined at 128.8% and 90.6%, respectively (figure 4.4.1.1).

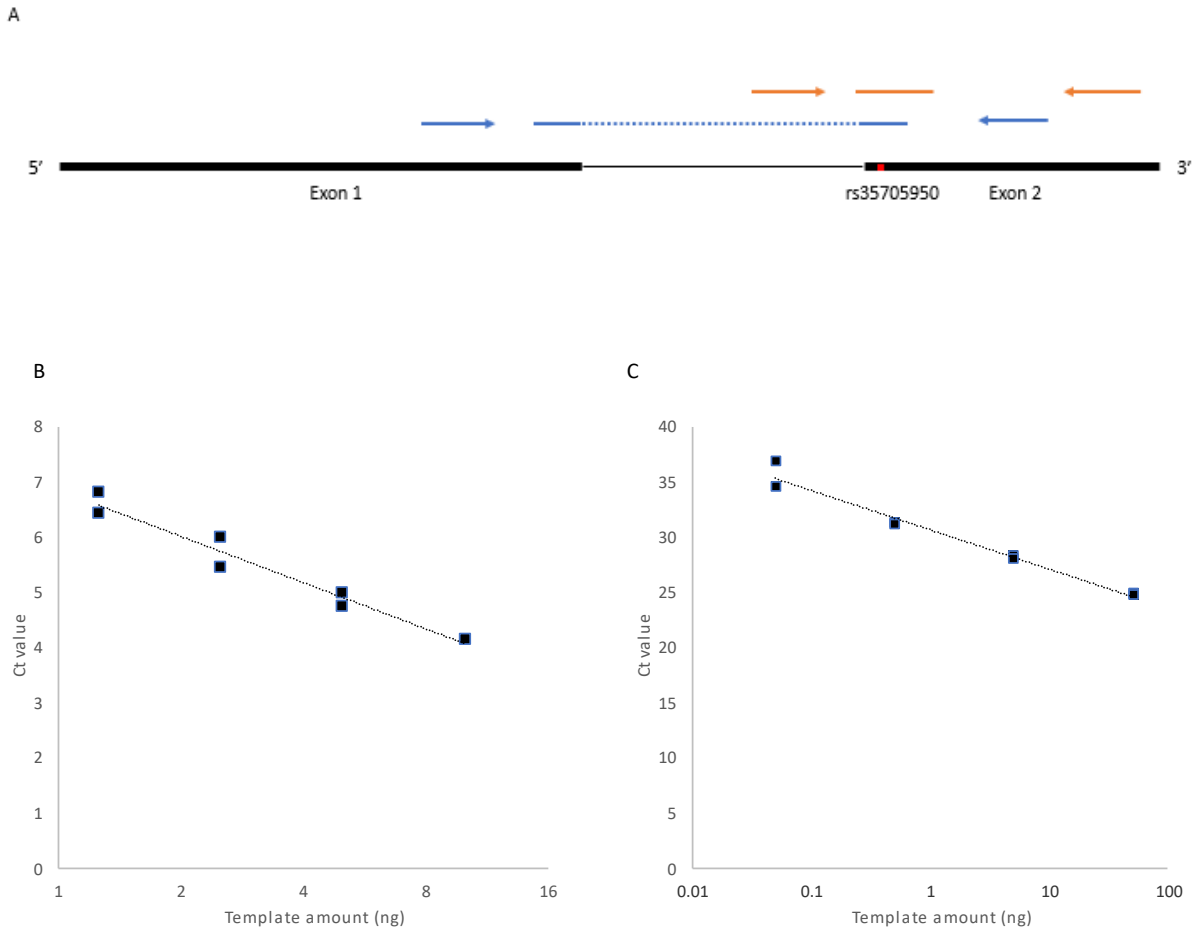


Figure 4.4.1.1 AC061979.1 RT-qPCR primer-probe set validation. (A) The location of the primers (thin arrows) and probes (thin continuous and dotted lines) specific for the splice variant (blue lines and arrows) and holotranscript (orange line and arrows) of AC061979.1 (thick black line), with the spliced region (thin black line) indicated between exons 1&2, and the rs35705950 SNP (thick red dot) located on the second nucleotide of exon

2. Duplicate two-fold serial dilutions of a spliced AC061979.1 Gblock (B) or A549 genomic DNA (C) were analysed by probe-hydrolysis qPCR using custom primer-probe sets.

4.4.2 *MUC5B* pancRNA AC061979.1 Expression *In cellulo*

In order to confirm the cellular expression of AC061979.1, probe hydrolysis RT-qPCR experimental assays were initially performed on total RNA extracts obtained from A549 cells and CFBE41o- cells representing alveolar and bronchial epithelial cells, respectively. Total RNA extracted from cells grown in T25 flasks to 70% confluency was subjected to AC061979.1, MUC5B, and 18S rRNA gene expression analysis by RT-qPCR across serial dilutions of the extracted RNA, with the highly expressed 18S rRNA house keeping gene used as a normalising internal control across replicate RT-qPCR experiments (Kuchipudi *et al.,* 2012). Results indicated that of the two AC061979.1 variants assayed, only the holotranscript was deteactable albeit at a low copy number (figure 4.4.2.1). In A549 cells, the highest concentration of input cDNA per RT-qPCR reactions resulted in AC061979.1 full transcript ΔCt to 18S of 24.12 ±3.16 (n=3). Equally in CFBE41o- cells the AC061979.1 full transcript ΔCt to 18S was 28.59 ±0.81 (n=3), at the highest concentration of input cDNA per RT-qPCR reactions. By contrast, *MUC5B* control gene expression was detected at ΔCt to 18S of 18.29 ±2.72 (n=3) and 24.78 ±2.47 (n=3), in A549 and CFBE41o- cells, respectively.

Importantly, the low expression level of AC061979.1 meant that where RNA extraction resulted in elevated Ct values for 18S > 20, the AC061979.1 full transcript could no longer be detected as its concentration dropped below the detection limit of the RT-PCR assay, suggesting the need for high quality, undegraded total RNA and high cDNA template quantities for its efficient detection. In addition, where no RT control reactions were carried out, no amplification at Ct <40 cycles occurred underlining the RNA nature of the detected material as opposed to carryover DNA contamination.

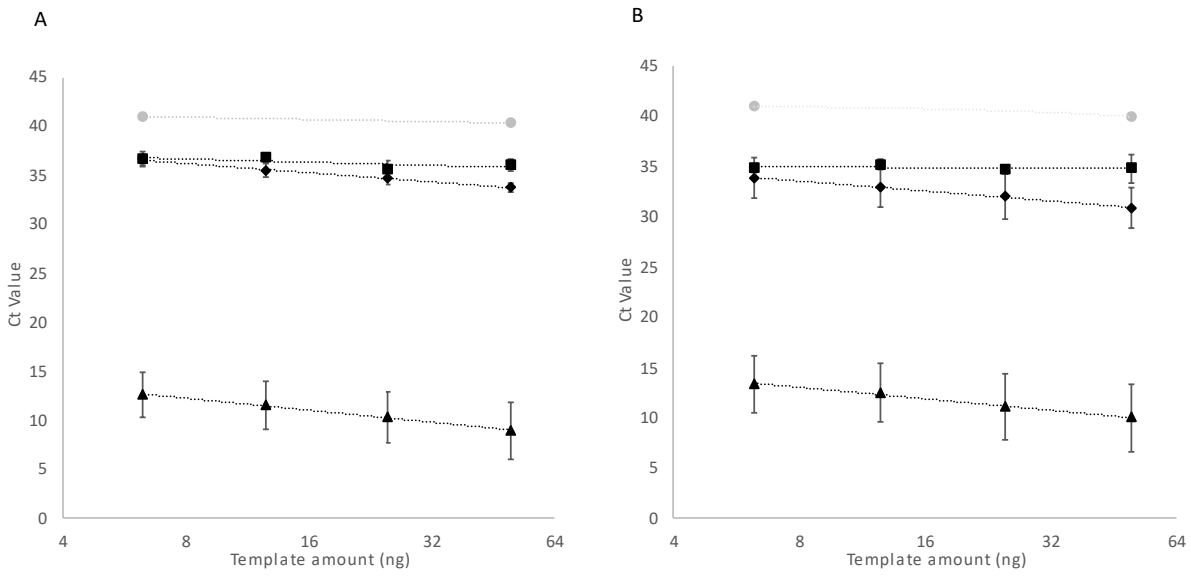


Figure 4.4.2.1 AC061979.1 expression validation *in cellulo*. Two-fold serial dilutions of CFBE41o- (A) and A549 (B) RNA extracts obtained from cells harvested at 70% confluency were subjected to probe hydrolysis RT-qPCR for 18S rRNA as an endogenous control (triangle), MUC5B (diamond), and AC061979.1 pancRNA (square) expression across 2-fold serial dilutions of cDNA derived from total cellular RNA extracts. No RT controls of the AC061979.1 holotranscript reaction were used as negative control (grey circles) to check for contaminating genomic DNA. Data are shown in log2-linear scale, representing 3 independent biological replicates and dual technical replicate points.

4.4.3 Effect of Contact Inhibition on AC061979.1 Expression

Cellular contact inhibition can affect cell growth (Cox et al., 2015) and therefore protein expression. To assess whether low cellular expression of AC061979.1 was as a result of cellular contact inhibition at high confluence, probe hydrolysis RT-qPCR assays were performed on total RNA extracts obtained from A549 cells and CFBE41o- cells harvested at low confluence (figure 4.4.3.1). In this instance, total RNA was extracted from these cells at 30% confluency in a T25 culture flask, to document AC061979.1, *MUC5B*, and 18S rRNA expression levels. Again only the holotranscript was deteactable albeit at a similarly low copy numbers suggesting that contact inhibition did not impact spliced transcript detection.

In A549 cells, the highest concentration of input cDNA per RT-qPCR reactions resulted in AC061979.1 full transcript ΔCt to 18S of 22.44 ±5.94 (n=3). Equally in CFBE41o- cells, the AC061979.1 full transcript ΔCt to 18S was 21.92 ±5.53 (n=3), at the highest concentration of input cDNA per RT-qPCR reactions. Overall, the reduction in contact inhibition resulted in a 3.02 ±0.1 and 101.83 ±9.24-fold increase in the expression of holo AC061979.1 in A549 and CFBE41o- cells respectively. The same trend was

observed for the *MUC5B* gene, where a reduction in contact inhibition resulted in a 3.68 ±0.06 and 10.98 ±3.98-fold increase in its expression in both A549 and CFBE41o- cells respectively.

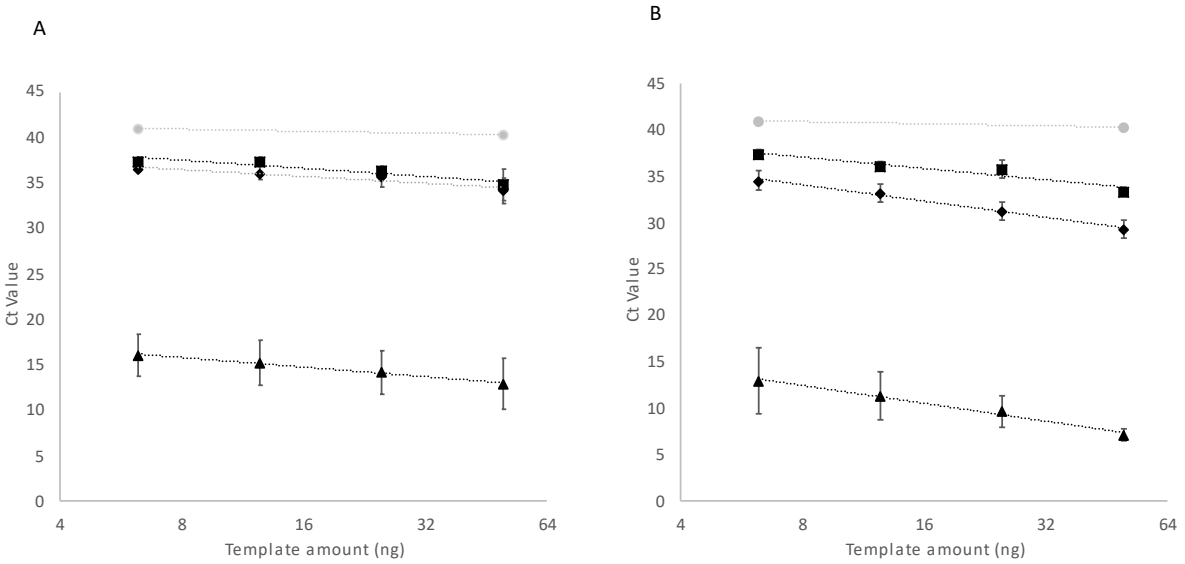


Figure 4.4.3.1 AC061979.1 expression is not affected by cellular contact inhibition. Two-fold serial dilutions of CFBE41o- (A) and A549 (B) RNA extracts obtained from cells harvested at <30% confluency were subjected to probe hydrolysis RT-qPCR. Analysis consisted of 18S rRNA (triangle), MUC5B (diamond) and AC061979.1 paRNA (square) expression across 2-fold serial dilutions of cDNA derived from total cellular RNA extracts, including a no RT of AC061979.1 reaction set as a negative control (grey circles) to check for contaminating genomic DNA. Data are shown in log2-linear scale, representing 3 independent biological replicates and dual technical replicate Ct's points.

4.4.4 Targeting the rs35705950 Locus for Gene-Editing

To begin exploring the significance of the rs35705950 SNP on *MUC5B* expression I first sought out the existence of lung epithelial cell lines with a G/T or T/T genotype. Due to the lack of such cell lines, I next sought to gene-edit the SNP into A549 cells. Such cell lines would be useful tools in understanding *MUC5B* expression control at both the pancRNA and epigenetic level. Furthermore, an effective gene-editing strategy that would restore canonical *MUC5B* expression regulation could conceivably offer future preventative treatment opportunities to individuals with rs35705950 genotype.

To institute the desired mutations at the rs35705950 SNP position, a set of three gRNAs were designed (M gRNA 1-3), ensuring gRNAs induced DSBs within 100 bps of rs35705950 to optimise the chances of incorporation of future edits in this site (figure 4.4.4.1) (Hollywood *et al.*, 2016). The top three candidates were subsequently put forward for *in vitro* screening analysis.

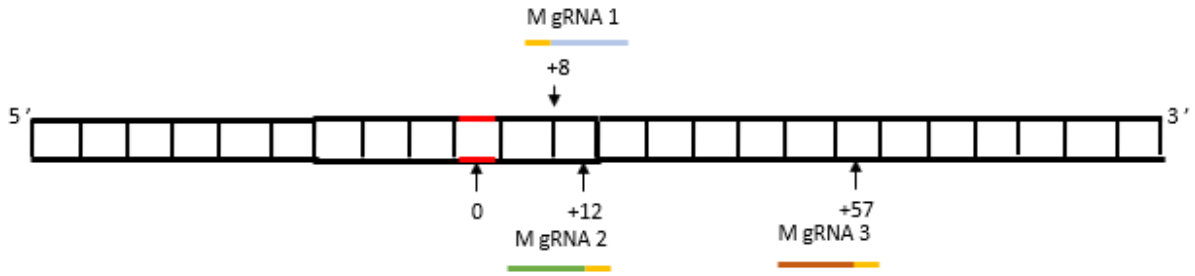


Figure 4.4.4.1 Candidate gRNA mapping coordinates relative to the rs35705950 SNP. Three M gRNAs (coloured lines), corresponding PAM sequences (amber lines) and their cleavage sites (directional arrows) in relation to rs35705950 SNP (in red). All gRNAs were <100bp from the mutation site.

Table 4.4.4.1 Analysis of the *MUC5B* rs35705950 locus using the CRISPOR CRISPR-Cas9 target predictor.

M gRNA sequence (number)	MIT specificity score	CFD specificity score	Off-targets 0-1-2-3-4 mismatches	Top 5 Off-target genomic locations, sorted by CFD off-target score (Number of mismatches)
GTTCACCTCTTCACAGTTGA	70	85	0-0-1-20-133	Intron: RPH3A (3) Intergenic: *PSMB7-GPR144* (4) Intron: *RP11-15E18.4* (3) Intergenic: *PLCB1-AL031655.1* (4) Intron: *ZFAND3* (4) Intergenic: *MIR3660-RP11-61G23.1* (3)
CAGCGCCTTCAACTGTGAAG	52	67	0-0-5-37-266	Intron: *LLGL1* (3) Intergenic: *RP11-838N2.5-RP11-838N2.3* (3) Intergenic: K*CNQ1OT1/KCNQ1-KCNQ1* (4) Intron: *AMOTL1* (4) Intergenic: *RNU6-257P-RP11-42I10.1* (3)
GCAAACAGGCCCGACTCCCA	79	89	0-0-1-10-103	Intron: *RPH3A* (3) Intergenic: *PSMB7-GPR144* (3) Intron: *RP11-15E18.4* (3) Intergenic: *PLCB1-AL031655.1* (4) Intron: *ZFAND3* (4)

4.4.5 *In Vitro* Validation of AC061979.1-Targetting gRNAs

To validate the editing efficiency of the *in silico* designed gRNAs targeting the *MUC5B* promoter an *in vitro* screening experiment was carried out on PCR amplicons generated from genomic DNA extracted from A549 cells. The manufacturer-recommended RNP-to-DNA substrate molar ratio of 10:1 was adopted alongside a control gRNA targeting the human *B2M* gene for nuclease efficacy validation. Cleavage of all gRNAs was observed within the expected target location based on the correct size of the predicted cleavage fragments (figure 4.4.5.1). Of the three *Muc5B*-targetting gRNAs assayed, cleavage efficiency of M gRNA3 was shown to be the most robust at 63.87 ±1.55 % as determined by densitometric analysis, highlighting it's potential as a prime candidate for further *in cellulo* validation. The targeting efficiencies of other gRNAs ranged from 42% to 63% (table 4.4.5.1) which equally highlighted their potential use for future *in cellulo* validation.

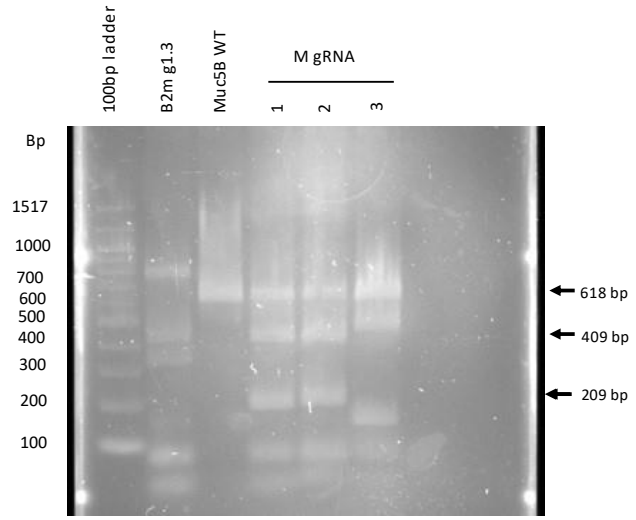


Figure 4.4.5.1 rs35705950-targeting M gRNA validation *in vitro*. Custom-designed M gRNA activity and potency *in vitro* was assayed. Cleavage products were analysed on a 1.6% agarose gel stained with SYBR™ Safe. Red asterisks indicate the expected fragments upon cleavage by Cas9 endonuclease. Arrows indicate the expected fragments upon cleavage by Cas9 endonuclease for M gRNA 1. Blurry bands can be observed around or below the 100bp mark potentially due to primer dimer formation.

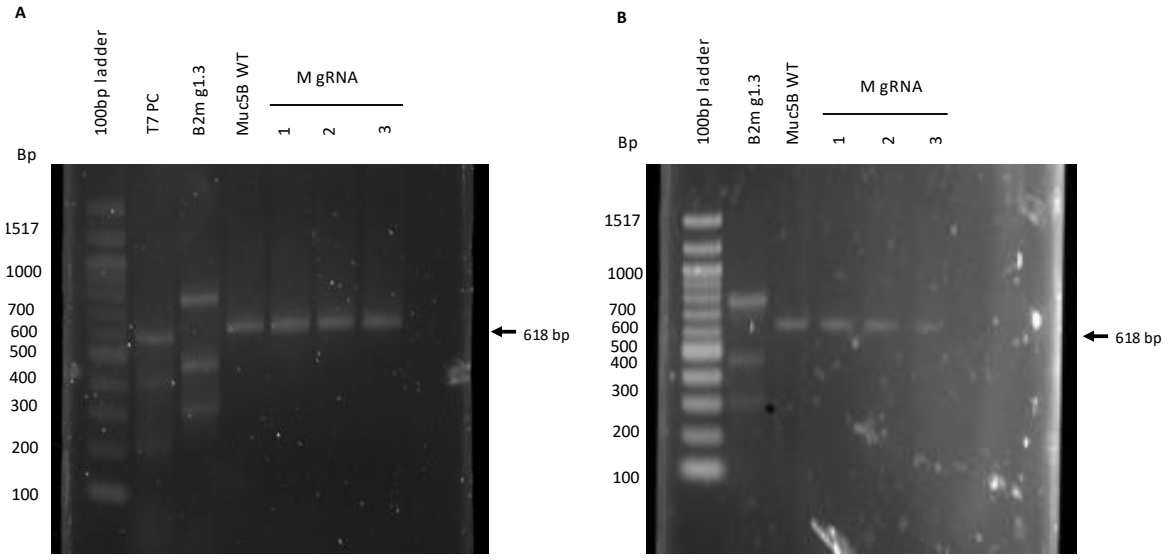
Table 4.4.5.1 The efficiency of *Muc5B*-targetting gRNAs assayed *in vitro*

gRNA	Targeting Efficiency (SD) %
M gRNA 1	62.23 (4.52)
M gRNA 2	42.32 (3.15)

M gRNA 3	63.87 (1.55)

4.4.6 Attempted Gene-Editing of AC061979.1 Genomic Sequence.

To measure the *in cellulo* editing efficiency, the three gRNAs targeting rs35705950, was next transfected into A549 cells, CFBE41o- cells, and HEK293 cells by lipofection and efficiency was determined by T7 endonuclease 1 enzyme assay. Unexpectedly, densitometric analysis revealed that *MUC5B*-targetting gRNAs were not effective irrespective of the cell line used (figure 4.4.6.1). In stark contrast, the B2M positive control gRNA exhibited 49.42 ±3.89%, 35.56 ±2.08% and 25.84 ±4.78% (n=3) cleavage activity in A549, CFBE41o- and HeLa cells respectively, underscoring that the failure to edit the rs35705950 locus was on account of the target site and probably not the candidate gRNAs. Based on these results, no sequencing analysis of the resulting cell lines was attempted, nor could any *MUC5B* gRNA be triaged forward for attempting homology directed repair-mediated introduction of the T allele in rs35705950.



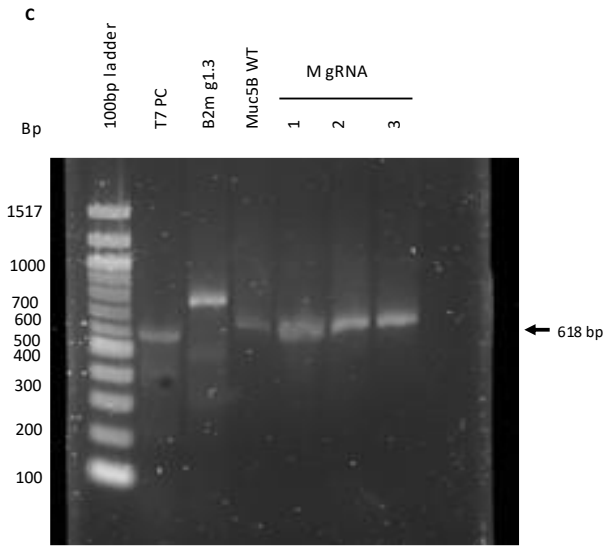
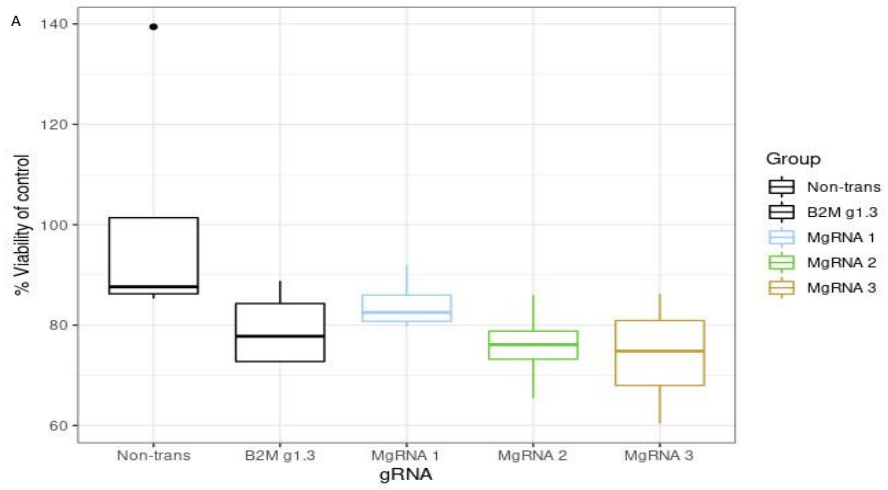


Figure 4.4.6.1 Attempted In cellulo gene-editing using MUC5B-targetting gRNAs. *In A549 (A), CFBE41o- (B) and HeLa (C) cells,* T7E1 cleavage assay*s* w*ere* performed on DNA targeted by a positive control gRNA (B2M g 1.3 *PC*); MUC5B-targeting gRNAs (M gRNA 1-3) and untransfected cells (MUC5B WT) while including a T7 positive control (T7 PC) to validate T7 enzyme functionality. Arrows indicate the expected fragments upon cleavage by Cas9 endonuclease for M gRNA 1.Cleavage products were analysed on a 1.6% agarose gel stained with SYBR™ Safe.



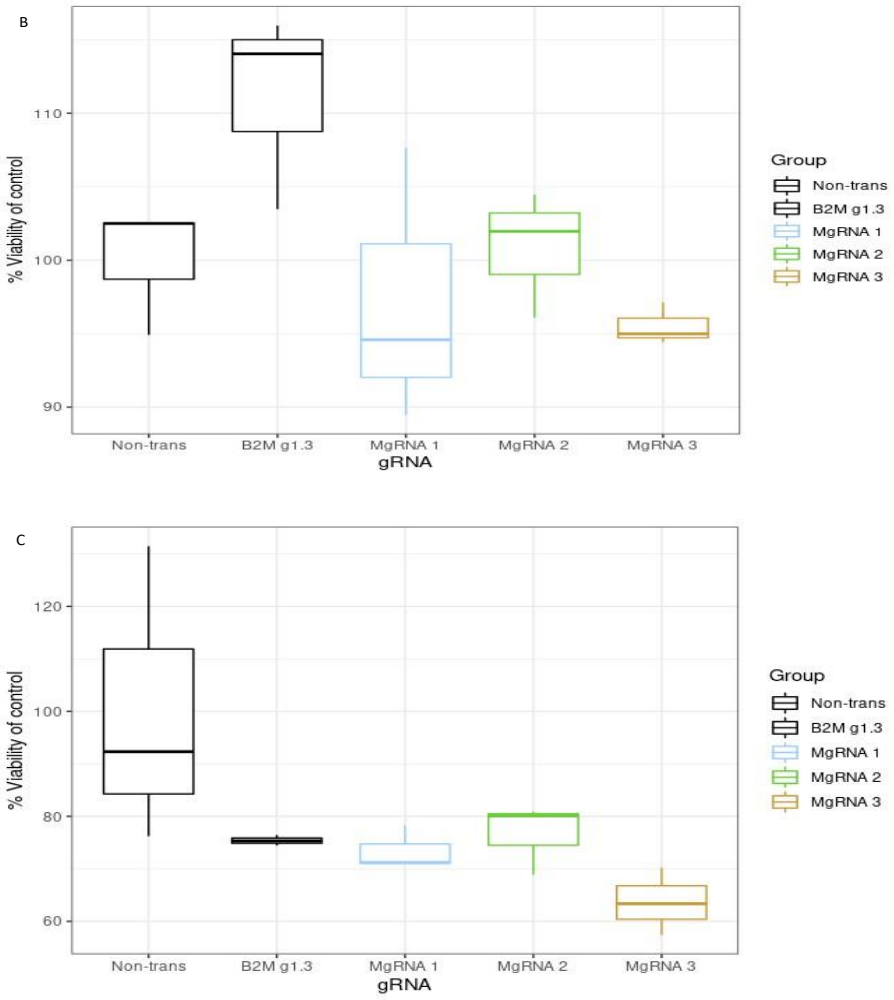


Figure 4.4.6.2 Assessment of cell viability following lipofection with Muc5B-targetting gRNAs. A549 cells (A), CFBE41o- cells (B) and HeLa cells (C) were transfected with M gRNA 1-3 then incubated at 37°C. 24 hours later cell viability was assessed by performing an MTT assay followed by absorbance measurement at 550nM. Data is representated as a percentage of the untransfected control, shown as the mean of three separate biological experiments (n=3). Error bars are representative of +/- SD. Statistical significance was accessed via a one-way ANOVA and a subsequent Dunnett's post hoc test. *p<0.05, **p<0.01, p<0.001.

To decipher whether gene-editing lead to increased cell toxicity, an MTT assay was performed on A549 cells, CFBE41o- cells, and HeLa cells 24 hours post transfection. Treatment with gRNAs did not have impact upon cell viability of A549 cells as a one-way ANOVA did not reveal any significant difference between gRNA transfected and un-transfected control groups (figure 4.4.6.2). For CFBE41o- cells, a one-way ANOVA equally revealed that no significant difference was observed between controls and M gRNA 1-3-transfected cells. In HeLA cells, a one-way ANOVA revealed that no significant difference was observed between controls and M gRNA 1-3-transfected groups.

4.5 Discussion

The G/T rs35705950 SNP found within the AC061979.1 lncRNA encoded in the *MUC5B* gene protomer region is associated with a much-increased risk of developing IPF amongst white, Hispanic, and Asian populations (Seibold *et al.*, 2011; Noth *et al.*, 2013). Individuals that are homozygous for the minor TT allele exhibit higher disease incidence as well as mortality, driven presumably by SNP-induced increase of transcription and translation of *MUC5B* in healthy individuals and IPF patients. Standard treatments for IPF including immunosuppression combination regimens have been associated with potentially harmful side effects in the past, while lung transplantation although effective at extending life, still poses a risk of transplant rejection (Idiopathic Pulmonary Fibrosis Clinical Research *et al.*, 2012).

Although AC061979.1 has been proposed as a lncRNA in humans (Frankish *et al.*, 2019), there was no data available to confirm the *in cellulo* expression of this highly relevant lncRNA molecule. This work is the first to document the existence of AC061791.1, at levels close to the limit of detection of standard RT-PCR methodology, and without any documentable processing into a spliced isoform. Importantly, the existence of this transcript has been successfully validated in a separate laboratory (Neatu *et al.*, 2022), where expression in air liquid interface culture of undifferentiated basal cells and differentiated human airway epithelial primary cells (HAEpCs) were also evaluated. Here, holo AC061979.1 was detectable at Ct 36.19 ± 0.29 and 36.84 ± 1.38 in ALI cultures of basal cells and HAEpCs respectively, further confirming the low copy number *in cellulo* prevalence of this transcript. Equally the spliced version of AC061979.1 was not detectable by RT-qPCR in ALI cultures of basal cells and HAEpCs, which is in line with the results of this thesis and that of an RNA-SEQ-verse survey performed by a research assistant internally, where no explicit splicing signal was observed for the same pancRNA (Neatu *et al.*, 2022).

The rs35705950 SNP on chromosome 11 has previously been linked with an increase in MUC5B expression in the distal airway and respiratory bronchioles (Nakano *et al.*, 2016; Seibold *et al.*, 2011; Noth *et al.*, 2013; Hancock *et al.*, 2018). Hence, RT-qPCR was therefore used on total RNA extracts from A549 and CFBE41o- cells to obtain representative evidence for alveolar and bronchial epithelial cells, respectively. In both cell lines only the full-length AC061979.1 transcript was detectable albeit at very low copy number. However, the low abundance of this transcript is neither unexpected, nor a proxy for its functionality, or capability to induce overexpression of adjacent genes. To emphasise this point with an example, a predicted lncRNA VELUCT was detectable only at 0.1 copies per cell (or 3.5 Ct's lower than DNA levels per extract), present only in the chromatin-associated RNA fraction. However, upon knockdown, loss of VELUCT functionality reduced viability of some lung cancer cell lines by up to 90% (Seiler *et al.*, 2017).

Contact inhibition increases the activity of tumour suppressor genes which in turn inhibits cell cycle progression into S-phase, a phase noted for maximal gene-expression (Küppers *et al.*, 2010; van der Meijden *et al.*, 2002). To test if contact inhibition reduced AC061979.1 expression, the levels of the holo- and spliced transcripts was assessed in both cell lines, using RNA extracts from high confluent and low confluent cells. Unsurprisingly, total RNA extracted from A549 and CFBE41o- cells at lower confluence demonstrated higher full-length AC061979.1 transcript abundance compared to higher confluence RNA extracts, highlighted by a lower ΔCt to 18S for the earlier. This similar trend of increased expression in lower confluent cells was also observed in the *MUC5B* gene included as a control for this RT-qPCR assay. This is in line with previous reports were increased cell-cell contact in higher confluence RNA extracts resulted in the reduction of gene expression (Küppers *et al.*, 2010; van der Meijden *et al.*, 2002); this would suggest that expression of AC061979.1 is potentially necessary for *MUC5B* expression and its knockdown could alleviate aberrant *MUC5B* expression; alternatively, AC061979.1 is only stochastically expressed on account of polymerase engagement with the MUC5B promoter. Of note, contact inhibition affects differently *MUC5B* gene and AC061979.1 pancRNA expression, probably on account of the lower level of *MUC5B* expression expected from an alveolar cell. Furthermore, low cell confluency against high confluency culture conditions resulted in higher levels of AC061979.1 expression in CFBE41o- cells (101.83 ±9.24-fold) compared to A549 cells (3.02 ±0.1-fold), and a similar trend was observed in the *MUC5B* gene, suggesting that these physiologically different cell lines are impacted to varied degrees, by contact inhibition. Hence these data suggest that increased AC061979.1 levels, as well as *MUC5B* expression observed in A549 and CFBE41o- cells, might be elevated by a reduction in cellular stress which was facilitated by reducing cell-cell contact in tissue culture flasks.

The spliced variant of AC061979.1 was not detectable by RT-qPCR in this work. It is possible the splicing of the full-length AC061979.1 is a very rare event requiring much higher levels of total RNA loading in the RT-PCR reaction for efficient detection. This is a distinct possibility because, although RT-qPCR is capable of detecting very low zeptomolar concentrations of RNA (Yao *et al.*, 2009) and even Poisson-level single copy number transcripts per RT-PCR reaction (Shah *et al.*, 2017), there was no evidence for the spliced isoform in these experiments. Another possibility is the erroneous suggestion of a splicing event by the GENCODE study or that splicing might involve different acceptor/donor sites within the full-length AC061979.1 transcript, such as cryptic exon acceptor sites recently described in CFTR (Nissim-Rafinia *et al.*, 2004). Alternatively, the proximity of rs35705950 to the putative splice acceptor site could potentiate aberrant AC061979.1 splicing, which in turn might inhibit or alter this lncRNA's downstream biochemical functionality, such as its association with e.g. the transcription factor FOXA2 (Helling *et al.*, 2017). Such a discovery could give rise to the use of splice-correcting

oligonucleotide therapies in preventing and treating IPF in a rs35705950 G/T or T/T background, similar to marketed Duchenne's muscular dystrophy and spinal muscular atrophy treatments (Kumar and Moschos, 2017). Another approach would involve the use of gene-editing as a strategy to permanently revert the rs35705950 risk variant among those with risk for IPF or at the early stages of disease.

In pursuit of engineering A549 cell line variants with the rs35705950 G/T or T/T genotype, genome editing of the locus was attempted. All three gRNAs designed demonstrated robust activities *in vitro* which made them promising candidates for *in cellulo* assays. However, custom-designed *MUC5B*-targetting gRNAs did not display any observable *in cellulo* activity in any of the cell lines tested. At first glance, the lack of activity among the three *MUC5B*-targetting gRNAs in A549, CFBE41o- and HeLa cells suggests their universal ineffectiveness *in cellulo* which could also translate to other cells of varying physiologies not included in this study. The lack of *in cellulo* gRNA activity could be attributable to numerous factors such as gRNA sequence composition including GC content, and gRNA energetic properties such as RNA secondary structure and free energy. Alternatively, genetic and epigenetic factors such as chromatin structure and gene expression could affect gRNA access to the locus. However, the universal lack of activity of all gRNAs across all cell lines suggests that the observation is not a cell line-dependent event but an event potentially attributable to the genomic locus being edited. Inspection of the rs35705950 locus revealed a highly repetitive, GC-rich region, which could invariably hamper gRNAs from binding to the desired target locus, while promoting non-specific binding to similar local genomic sequences, particularly if mismatches lie outside the gRNA seed region (Wu, Kriz and Sharp, 2014; Neatu *et al.*, 2022). A combination of the genomic locus, cell survivability, as well as the previously mentioned confounding factors could have invariably led to the complete absence of gene-editing observed in this context. Therefore, to uncover gRNAs that are effective in this hard-to-target genomic locus, a future alternate strategy would involve the introduction of chemically modified bases such as LNA, to the 5' end of *MUC5B*-targetting gRNA in an attempt to improve on-target efficiency and prevent non-specific binding (Soifer *et al.*, 2012).

Lipofectamine was chosen as the gene-editing cargo delivery vehicle due to it's low-cost, cargo protection and convenience, demonstrating transfection efficiencies from 25% up to 60% in various cell lines ranging from humans to zebrafish to mice (Yu *et al.,* 2016; Sandbichler, Aschberger and Pelster, 2013; Song, Liu and Liu, 1998). However, the toxicity of the lipofectamine transfection reagent adopted for the gene-editing experiments was observable. The cellular membrane lipid bilayer may break down due to the introduction of cationic or short-chain lipids, or toxicity may arise from the oversaturation of nanoparticle lipids in the membrane (Strachan *et al*., 2020). Indeed, positively charged lipids nanoparticles have been demonstrated to upregulate proinflammatory IFN gamma, TNF

alpha and IL2 cytokines up to 15-25 fold in mice Th2 cells (Kedmi *et al*., 2010). In this study, only indel frequency was determined, and not transfection efficiency, a short-coming which has been acknowledged.

Had targeting of the rs35705950 locus in *MUC5B* present in A549 cells been successful, the next step would have included an attempt to knock in the 'T' IPF risk variant into this locus, to generate a disease model cell line. To achieve this careful design of DNA donor templates would be necessary, hence the choice of double-stranded or single-stranded donor, sequence length, mutation-to-cut site distance, alterations to the PAM sequences to avoid Cas9 re-cutting, as well as other selection criteria all would need to be considered carefully. In an attempt to achieve HDR in future, single-stranded donor templates would be designed with short cut-to-mutation distances and an edit encoded within the shorter PAM-proximal arm, all factors previously described to improve HDR (Hollywood *et al.*, 2016; Richardson *et al.*, 2016; Okamoto *et al.*, 2019). It is however anticipated that the highly repetitive sequence nature around the rs35705950 locus would require an extension of the donor homology arms to enhance specificity. Another potential strategy to improve HDR would involve the blocking of the NHEJ pathway using small molecules which enhance targeted gene-editing directly by promoting HDR pathway (Yu *et al.*, 2015) or indirectly by inhibiting the NHEJ pathway (Frit *et al.*, 2014; Maruyama *et al.*, 2015).

In this chapter *in cellulo* expression of a full-length pancRNA AC061979.1 transcript was validated following previous suggestion of expression by *in silico* analysis. This pancRNA is expressed off a genomic locus which hosts a SNP with one of the strongest correlations with IPF, hence this lncRNA could be integral in the development of this disease. One hypothesis that could explain this lncRNA-disease association is that of aberrant lncRNA splicing, caused by the SNP; however closer analysis did not detect this particular splicing event at least in cell lines with wild type background. As a result of the strong correlation between the rs35705950 risk variant and IPF, identifying a gene-editing solution efficacious *in cellulo* could be utilised as a valuable strategy to revert this locus back to the non-risk variant which, if implemented protectively in the clinic, could prevent IPF onset and potentially IPF outcomes. To this end three *MUC5B*-targetting gRNAs were designed *in silico* which exhibited good activity *in vitro*, however they were unable to induce any gene-editing events in at least three cell lines assayed. Future work could explore chemical modification of gRNAs in the context of locus structure to resolve loss of activity and open a route to their translational use. Achieving knock-in of wild type sequences using HDR however would likely present its own challenges of specificity.

Chapter 5. Design of a Single-Transcriptional Unit Gene-Editing Platform to Target the *CFTR* Locus

5.1 Introduction

There are several challenges in the production of CRISPR-Cas9 systems for the purpose of gene-editing. One of the most impactful, is the size of Cas9 endonuclease which generally does not favour the endonuclease and gRNA being encoded within the same expression systems. This makes it more challenging to achieve co-ordinated expression of both components, because typically the two genes and their associated regulatory sequences must be expressed individually from two separate expression vectors. The Cas9 protein component is often expressed from an RNA pol II promoter. Eukaryotic RNA pol II transcription is initiated by the assembly and recruitment of pre-initiation complex (PIC) comprising of transcription factors TFIIA, B, D, E, F, H, RNA Pol II, and Mediator protein to the promoter (Muniz, Nicolas and Trouche, 2021). PIC causes the DNA to adopt an open conformation, enabling RNA pol II to commence transcription and subsequent elongation of the nascent transcript. Many components including elongation and splicing factors, travel with RNA Pol II during productive transcription elongation and assist coordinate pre-mRNA processing with transcription. Finally, transcriptional termination involves the catch-up of RNA Pol II by the exonuclease Xrn2, caused by a poly(A) site-induced elongation slow-down that favours transcription termination. Xrn2 degrades the RNA downstream of the poly(A) site, permitting subsequent post-transcriptional polyadenylation (Muniz, Nicolas and Trouche, 2021). While the Cas9 gene is usually expressed from a eukaryotic RNA pol II promoter, the gRNA component cannot be expressed from the same type of promoter. This is because RNA pol II transcripts undergo extensive post-transcriptional processing including, splicing, 5’ 7-methyl guanosine capping, and the incorporation of additional bases to the 5’ end of gRNA sequences during transcription which are vital for guiding the Cas9 endonuclease to its DNA target (Gao and Zhao, 2014). Hence, gRNA are usually produced from RNA pol III promoters.

Eukaryotic RNA pol III promoters function to transcribe short, structured RNA molecules, including tRNA and rRNA. RNA pol III transcription is initiated by transcription factors TFIIIA, TFIIIC, and TFIIIB consisting of TATA-binding protein (TBP), B-related factor 1 (Brf1) and B double prime (Bdp1) (Turowski and Tollervey, 2016). TFIIIA is a single protein that is exclusively involved in the transcription of the 5S rRNA gene (*RDN5*). TFIIIB is made up of Brf1, Bdp1, and TBP (TATA-binding protein), which is found in all eukaryotic RNA pol III promoters. TFIIIC is a big, versatile six-subunit complex that recognises all RNA pol III transcription unit promoter sites (Turowski and Tollervey, 2016). During transcription initiation, the internal promoter elements are recognised and bound by TFIIIC, which

then recruits TFIIIB upstream of the transcription start point, leading to the subsequent recruitment of RNA pol III. The RNA pol III transcription termination signal is a stretch of > 4 adenosine residues on the template DNA strand (Turowski and Tollervey, 2016). This short poly(A) stretch on the template strand facilitates polymerase stalling and transcript release, to complete the transcriptional process (Turowski and Tollervey, 2016; Gao, Herrera-Carrillo and Berkhout, 2018). While it is possible to achieve controlled inducible expression using a pol II promoter, challenges arise in the production of a gRNA using pol III promoters. These include: (1) the uncharacterised nature of pol III promoters often resulting in difficulties in choosing suitable promoters and inherently poor promoter performances; (2) the time-consuming nature of cloning multiple gRNAs in multiple cassettes in the context of multiplex gene-editing; (3) The ubiquitous expression profile of these pol III promoters that precludes cell-type specific expression; (4) In the context of *in vitro* transcription pol III promoters are not suitable because RNA pol III is not commercially available; (5) Pol III promoters such as U3 and U6 require an additional 5' 'G' and 'A' respectively which could negatively affect gRNA binding and Cas9 activity (Tang *et al.*, 2016).

An attempt to achieve a facile, compact, and synchronised expression of all components necessary for gene-editing would be to express both Cas9 and gRNA under the control of a single pol II promoter. The use of STU avoids all the challenges associated with usage of conventional dual plasmid-encoded CRISPR-Cas9 expression. These challenges include, production of gRNAs from pol III promoters, lack of spatiotemporal colocalization of gRNA and Cas9 components, as discussed in the previous paragraph. Equally, the usage of an STU package in a plasmid vector, avoids the challenge of packaging a large 160kDa Cas9 RNP in delivery vehicles (Lino *et al.*, 2018). Equally, the usage of STU gene-editing platform (either as an IVT mRNA or encoded in a plasmid) would permit the inclusion of a donor template sequence in the same transcriptional unit to permit spatiotemporal co-localisation of all three gee-editing components, a feature RNP delivery is incapable of achieving.

One means of achieving transcript maturation from such a STU gene-editing precursor involved the use of Csy4 endoribonuclease, an endoribonuclease responsible for *in vivo* CRISPR crRNA processing, by including its recognition sites between the gRNA and Cas9 components (figure 5.1.1). This strategy requires the inclusion of a Csy4 protein coding sequence within the gene-editing STU, resulting in a considerably larger primary transcript (Tang *et al.*, 2019).

Another strategy to achieve this involved the separation of Cas9 and gRNA components by ribozyme cleavage sites (figure 5.1.1) which would permit the processing of the holo-transcript by *cis*-acting ribozymes (Tang *et al.*, 2016; Tang *et al.*, 2019). In one instance the same hammer-head ribozyme sequence was used to interspace gene-editing components (Tang *et al.*, 2019), while in a different

instance two separate ribozymes, namely hammer-head (HH) and hepatitis delta virus (HDV), were used to process the STU (Gao and Zhao, 2014). The hammerhead ribozyme, originally detected in subviral plant pathogens, are tiny endonucleolytic self-cleaving catalytic RNA motifs. It consists of a conserved nucleotide catalytic core surrounded by three helices (stems), two of which establish critical tertiary interactions for rapid self-excision *in vitro* and *in cellulo* (Hammann et al., 2012). HDV ribozymes catalyse their own excision by a transesterification reaction from the HDV concatenated genome, during rolling circle replication (Webb and Lupták, 2011). An advantage of using ribozymes for STU processing is their ability to self-cleave, enabling maturation of STU individual components without reliance on endogenous or exogenous cleavage factors (Gao and Zhao, 2014). Moreover, in the case of hammer-head ribozyme, only the first six-nucleotides of the ribozyme, comprising stem I of the RNA molecule, require alteration to successfully process gRNAs specific for any DNA sequence (Gao and Zhao, 2014).

A third strategy interspaces Cas9 and gRNA components with tRNA sequences, which are dependent on the conserved eukaryotic tRNA-processing system, for STU cleavage and gene-editing precursor maturation (figure 5.1.1) (Tang *et al.*, 2019; Xie, Minkenberg and Yang, 2015). *In cellulo* tRNA are transcribed by RNA pol III as a pre-tRNA molecule, which is then trimmed at the 5' and 3' ends by RNase P and Z endonucleases, to generate a mature functional tRNA molecule (Kikuchi, Sasaki-Tozawa and Suzuki, 1993). In the context of gene-editing STU maturation, the tRNA-processing system has the following advantages: (i) tRNA precursors (pre-tRNAs) are processed at specific, well-defined sites by RNase P and RNase Z, eliminating additional 5' and 3' sequences in the process (Xie, Minkenberg and Yang, 2015); (ii) Because pre-tRNA sequences are highly conserved they are easily identifiable by RNase P and RNase Z regardless of the sequence (Barbezier *et al.*, 2009); (iii) tRNA sequences are widely located in bacterial polycistronic genes, highlighting their evolutionary use for the maturation of polycistronic transcripts (Kruszka *et al.*, 2003); (iv) The ubiquitous cellular expression profile of tRNAs should ensure processing of large quantities of STU transcripts (Xie, Minkenberg and Yang, 2015).

When the efficacy of tRNA, Csy4 and ribozyme interspaced STUs were compared within rice phytoene desaturase (*OsPDS*) gene in transgenic rice plants, Sanger sequencing revealed that editing efficiencies achieved were 88.2, 81.0 and 74.5% respectively, highlighting that all three gene-editing strategies are both comparable and viable (Tang *et al.*, 2019). Ribozyme interspaced STU has been used to target dicot plants such as *Arabidopsis thaliana* Leucine-rich receptor-like protein kinase family protein *(AtBRI1)* gene and *Nicotiana tabacum* pleiotropic drug resistance (*NtPDR6)* gene in tobacco, with efficiencies ranging from 18.1 to 65% (Tang *et al.,* 2016).

Other alternative CRISPR-Cas9 gene-editing approaches are as follows. The use of one construct encoding multiple gRNA sequences interspaced by $tRNA^{Gly}$, have been deployed in human cells, knocking down histone deacetylase gene in HEK293 cells with up to 50% efficiency (Dong et al., 2017). Another promising alternative strategy to the conventional use of CRISPR-Cas9 gene-editing involves the use of base editors. Base-editors compose of a gRNA and Cas9 enzyme attached to a deaminases by a proline-alanine linker for programmable DNA binding, followed by a deamination reaction that converts one nucleotide base into another (Xie *et al.,* 2015). Base-editing and conventional CRISPR-Cas9 HDR editing efficiencies were determined and compared, by measuring forskolin-induced swelling sensitive organoids after targeting, revealing an editing efficiency up to 9.1% corrected organoids by base-editing and 1.22% by conventional CRISPR-Cas9–mediated HDR (Geurts et al., 2021).

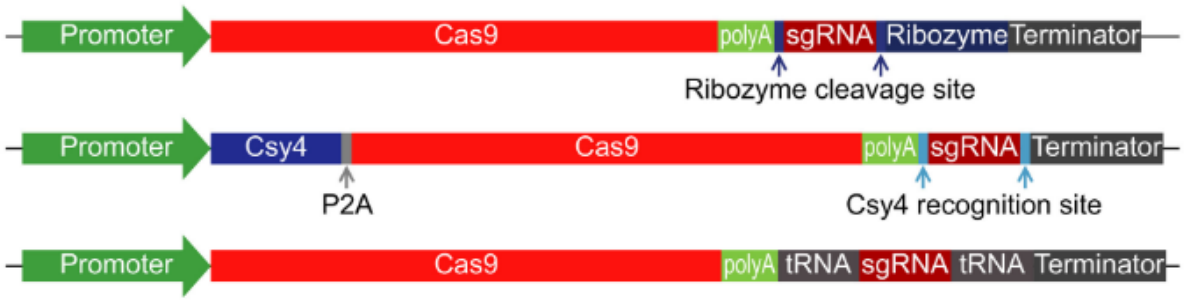


Figure 5.1.1 Three CRISPR-Cas9 STU expression systems. Cas9 coding sequence is followed by a poly(A) tail that improves mRNA stability while all three systems apply different approached to permit specific processivity of gRNA *in cellulo.* Top, STU-Cas9-ribozyme system, ribozyme sequences flank the gRNA sequence. Middle, STU-Cas9-Csy4 system, a P2A self-cleaving peptide interspaces Csy4 and Cas9 endonucleases while Csy4 recognition sites flank gRNA sequence. Bottom, STU-Cas9-tRNA system, tRNA sequences flank the gRNA sequence. Taken and adapted from (Tang *et al.*, 2019)

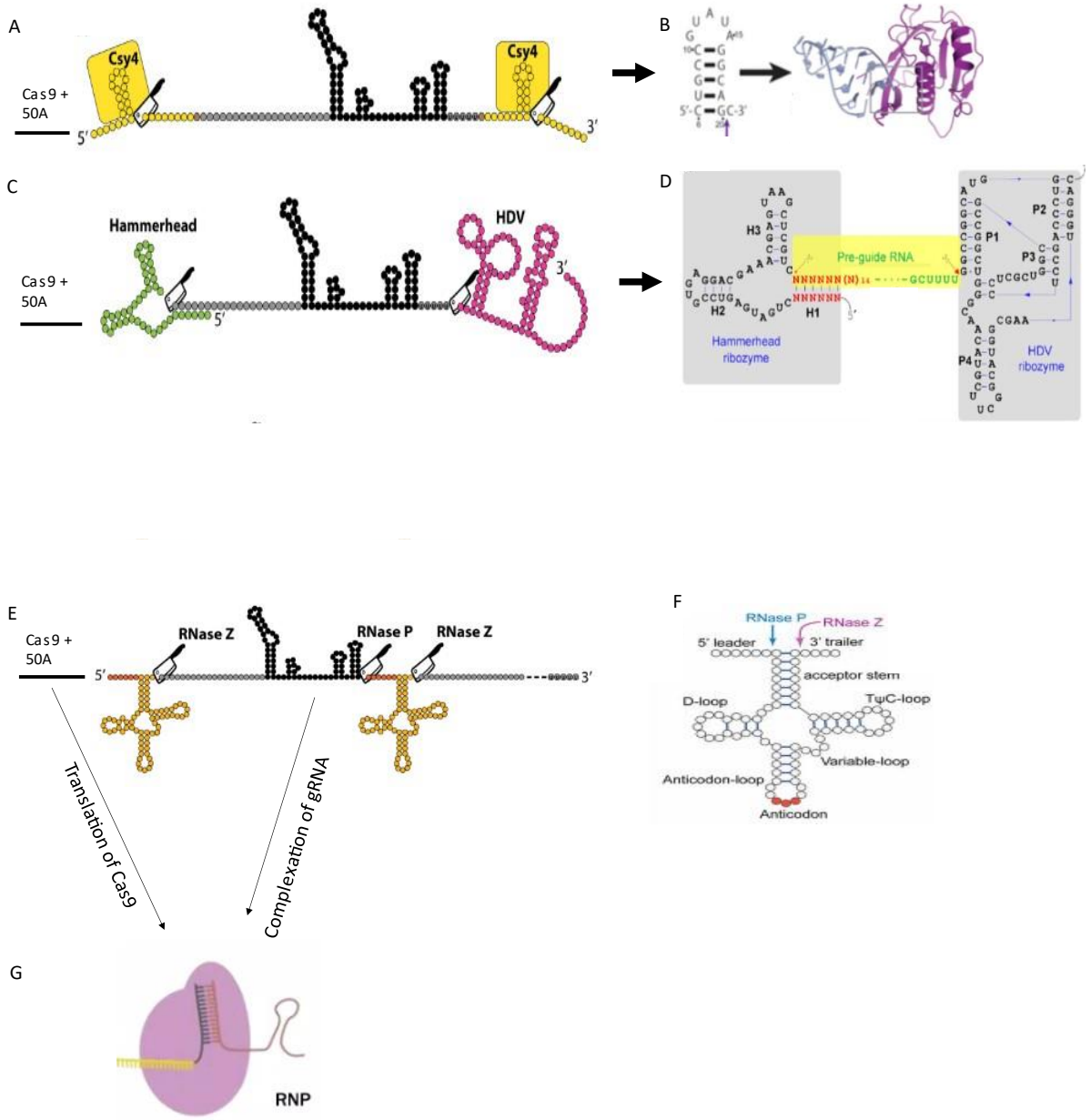


Figure 5.1.2. Three mechanisms of STU maturation. The downstream component of gene-editing (A) 28-nucleotide sequences flanking the gRNA are recognised by RNA endonuclease Csy4. (B) Csy4 which cleaves after the 20th recognition nucleotide (dark purple arrow) while staying attached to the upstream region. Overall structure of Csy4 (dark purple) bound to product RNA (light purple). (C) The ribozyme interspaced STU with the gRNA structure and two flanking ribozymes highlighted, showing the gRNA sequence is flanked by a hammerhead ribozyme at the 5' end, and HDV ribozyme at the 3' end. (D) Both ribozymes' secondary structures are displayed. The Hammerhead ribozyme's hairpin (stem) sections are designated by the letters H1, H2, and H3. P1, P2, P3, and P4 relate to the HDV ribozyme's hairpin sections. The mature gRNA is processed out of the STU by self-catalyzed processing with the aid of both flanking ribozymes. The pregRNA's 5'end is complementary

to the 5' end of Hammerhead ribozyme (shown in red), while the 3' end (shown in green) is universal across all gRNAs'. (E) Polycistronic tRNA-gRNA architecture enables STU to be generated from a single promoter. (F) The 5 leader and 3 trailer sequences are cut by the endogenous RNases, RNaseP and RNase Z at predefined locations, respectively. These three STU manufacturing methods all enable RNA poll II-mediated transcription and termination. (G) Once gRNA component has successfully separated and matured from STU, it localises within the nucleus, until Cas9 protein is translated, then both components as an RNP complex can successfully mediate gene-editing. Taken and adapted from (Nowak *et al.,* 2016; Gao and Chao, 2013; Haurwitz *et al.,* 2012; Xie *et al.,* 2015)

To assay for the *in cellulo* expression of all three gene-editing STU, RT-qPCR assays could be designed specific for the holo transcripts, gRNA and Cas9 components (Livak et al., 2001). In regards to the self-cleaving ribozyme interspaced STU, successful processing of gene-editing components could be determined *in vitro* by performing *in vitro* transcription of STU, then performing agarose gel electrophoresis using an ethidium bromide-stained denaturing urea polyacrylamide gel to separate and visualise the anticipated separated RNA products (Gao *et al.,* 2013). For tRNA interspaced STU, successful processing of gene-editing components could be determined *in vitro* by performing *in vitro* transcription of holo STU, followed by co-incubation of holo-transcript with recombinant RNase P and RNase Z to mediate *in vitro* STU component separation due to cleavage at tRNA 5' leader and 3' trailer sequences, then performing agarose gel electrophoresis using a denaturing urea polyacrylamide gel to separate and visualise the anticipated separated RNA products using (Kikuchi, Sasaki-Tozawa and Suzuki, 1993). For Csy4 interspaced STU, successful processing of gene-editing components could be determined *in vitro* by performing *in vitro* transcription of holo STU, followed by co-incubation of holo-transcript with recombinant Csy4 endoribonuclease to mediate *in vitro* STU component separation due to cleavage at interspacing Csy4 recognition sequences, then performing agarose gel electrophoresis using a denaturing urea polyacrylamide gel to separate and visualise the anticipated separated RNA products (Haurwitz *et al.,* 2011).

Therefore, to circumvent the above stated challenges in the conventional production of CRISPR-Cas9 systems in the context of gene-editing, I assayed the potential for expression of both Cas9 and gRNA components in an STU, driven from a single constitutive cytomegalovirus (CMV) pol II promoter, in human cell lines. This would ensure better co-ordination and spatiotemporal localisation of gene-editing components post-expression. Initially, the tRNA STU-processing system was employed due to its conserved activity in a wide range of cells. Then, the ribozyme STU-processing system was utilised due to ease of construction and cloning.

5.2 Aims

The aims of the work described in this chapter were to: (i) design a STU-gene-editing system with high affinity for the *CFTR* Δ F508 locus; (ii) gene-edit the *CFTR* locus by delivering STU-gene-editing system into CFBE41o-, A549 and HeLa cells; (iii) validate the expression of STU-gene-editing system individual components; (iv) assess the effect of gene-editing experiments on cellular viability.

5.3 Hypothesis

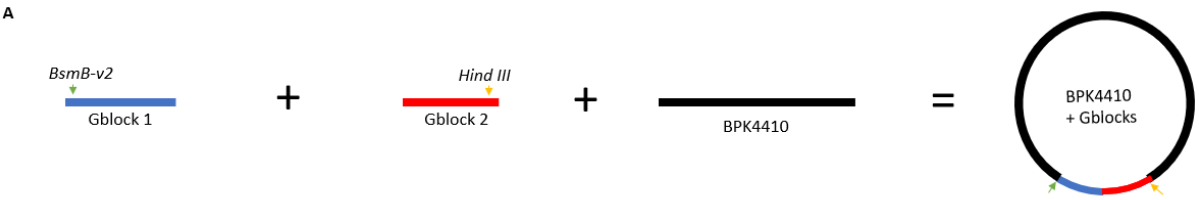
The STU-gene-editing system shall be capable of targeting the Δ F508 locus with high efficiency in assayed cell lines with minimal impact on cell viability. This individual components of STU-gene-editing system shall be expressed and detectable post-transfection, by RT-qPCR and western blotting,

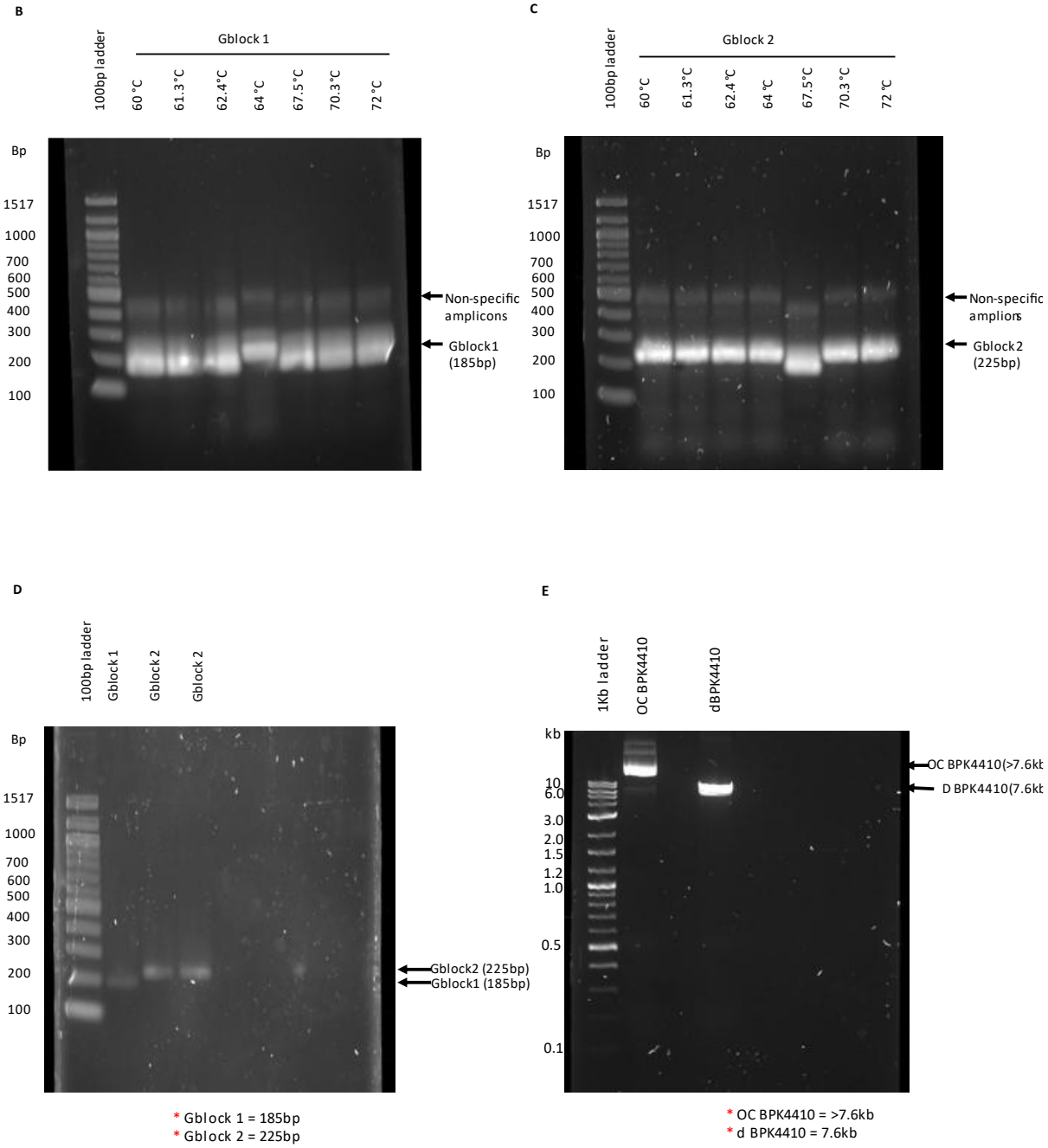
5.4 Results

5.4.1 Engineering a tRNA^{Gly}-Separated STU Gene-Editing Precursor

To synthesise all the individual compoenents for gene-editing in a STU gene-editing precursor, HypaCas9 and a CF gRNA 1 were spliced into one dicistronic gene. The tRNA^{Gly} RNA-processing system was selected for driving the maturation of the precursor within the eukaryotic nucleus because, the tRNAs are amongst the most abundant RNA species within cells, therefore RNA-processing capabilities of this system should be sufficient to achieve a high level of STU-gene editing processing.

Gibson assembly was adopted as a stratedgy to clone the gene-editing tRNA^{Gly} interspaced STU (section 2.6.9). Initial attempts to introduce homologous overhangs and amplify the two Gblock fragments produced non-specific bands even by gradient PCR (figure 5.4.1.1 B, C). To obtain the desired fragments, DNA bands of 185bp and 225bp were excised, respresenting the anticipated size of fragments 1 and 2 respectively (figure 5.4.1.1 D). After gel purification, the presumed gene-blocks were then combined with linearised BPK4410 in a one-step isothermal reaction. Following assembly, a restriction digest using *BsmB*I-v2 and *Hind* III found on fragment one and fragment two respectively (figure 5.4.1.1 A, F), failed to document the desired Gblock fusion products anticipated at 311bp.




B
100bp ladder
Gblock 1
60 °C
61.3 °C
62.4 °C
64 °C
67.5 °C
70.3 °C
72 °C
Bp
1517
1000
700
600
500
400
300
200
100
Non-specific amplicons
Gblock1 (185bp)
C
100bp ladder
Gblock 2
60 °C
61.3 °C
62.4 °C
64 °C
67.5 °C
70.3 °C
72 °C
Bp
1517
1000
700
600
500
400
300
200
100
Non-specific amplions
Gblock2 (225bp)
D
100bp ladder
Gblock 1
Gblock 2
Gblock 2
Bp
1517
1000
700
600
500
400
300
200
100
Gblock2 (225bp)
Gblock1 (185bp)
* Gblock 1 = 185bp
* Gblock 2 = 225bp
E
1Kb ladder
OC BPK4410
dBPK4410
kb
10
6.0
3.0
2.0
1.5
1.2
1.0
0.5
0.1
OC BPK4410(>7.6kb
D BPK4410(7.6kb
* OC BPK4410 = >7.6kb
* d BPK4410 = 7.6kb

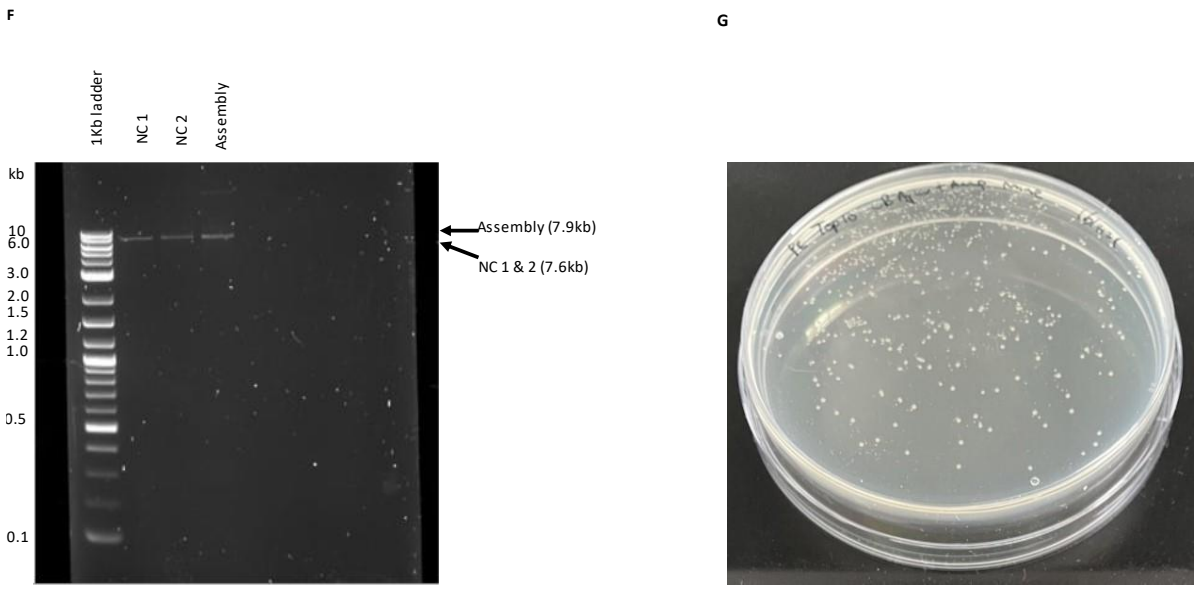


Figure 5.4.1.1 First Cloning Attempt of Gblock 1 & 2 into BPK4410 plasmid. A. Gblock 1 (blue line) containing a *Bsm*B-v2 restriction site (green arrow), Gblock 2 (red line) containing a *Hin*d III restriction site (orange arrow) and linearised BPK4410 were combined in a one-step isothermal Gibson assembly reaction, permitting subsequent confirmation of insert integration by a *BsmB-v2- Hind III* restriction double-digest. **B.** Gradient PCR of Gblock fragment 1 across a range of annealing temperatures to identify optimal PCR conditions. **C.** Gradient PCR of Gblock fragment 2 to identify optimal PCR conditions. **D**. Gblock 1 & 2 PCR amplicons were extracted from A and B respectively, gel purified, and confirmed by agarose gel electrophoresis. **E**. Linearisation of BPK4410 plasmid vector by *Pme*I restriction enzyme digest ahead of Gibson assembly. OC BPK4410, open-circle BPK4410 negative control; dBPK4410, digested BPK4410. **F.** Following Gibson Assembly, a restriction digest was performed to identify the integration of the two Gblocks. NC 1, Gibson assembly negative control containing all three assembly components without the assembly master mix; NC 2, Gibson assembly negative control comprised of solely linearised BPK4410 plasmid; Assembly, Gibson assembly reaction containing all three assembly components plus the assembly master mix. **G**. A manufacturer-supplied Gibson assembly positive control reaction which confers ampicillin resistance to transformants was included, alongside sample reactions, then plated on ampicillin LB agar plates, to demonstrate kit functionality. Arrows indicate the expected size of DNA fragments. PCR productions and restriction digestion products were analysed on a 1.6% agarose gel.

To attempt a different strategy of assembling all three DNA components, a set of primers were designed to simultaneously amplify and linearise the vector backbone, a more efficient approach than relying on restriction digests (Benoit *et al.*, 2006). Following isothermal reaction, a restriction digestion reaction using *Bsm*BI-v2 and *Hin*d III failed to produce the anticipated products (figure 5.4.1.2 A). To apply a more sensitive approach for assembly detection, a PCR was performed using the forward

primer of Gblock one and the reverse primer of Gblock two, but no amplification products at ~311bp were observed suggesting unsuccessful assembly reaction (figure 5.4.1.2 B).

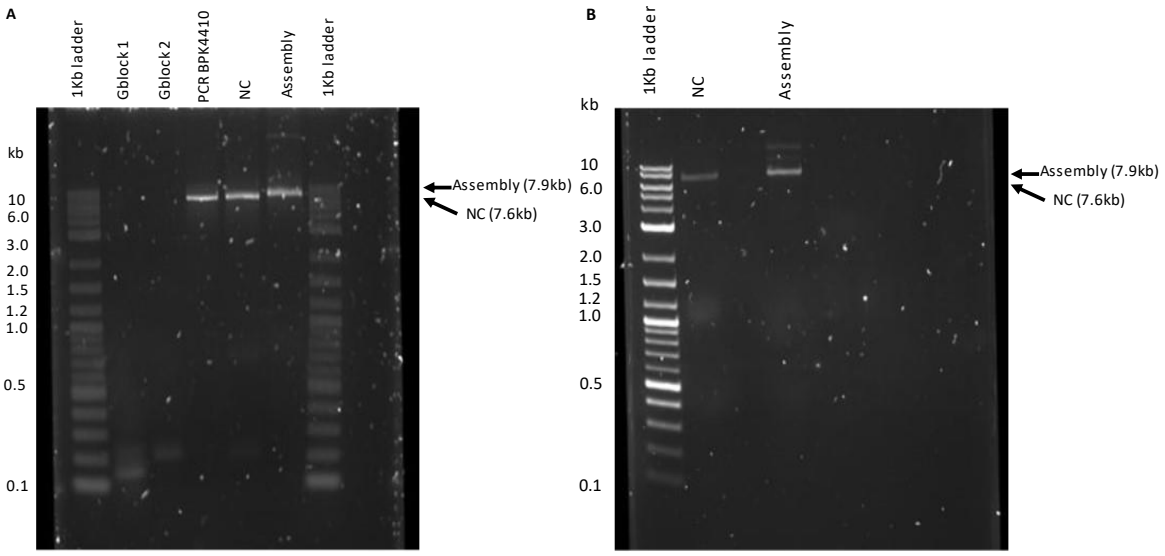


Figure 5.4.1.2 Second Gibson cloning Attempt of Gblock 1 & 2 into BPK4410 plasmid. A. Following Gibson Assembly, a restriction digest was performed to identify the integration of the two Gblocks. PCR BPK4410, BPK4410 plasmid linearised by PCR; NC, Gibson assembly negative control containing all three assembly components without the assembly master mix; Assembly, Gibson assembly reaction containing all three assembly components plus the assembly master mix. **B**. Gibson assembly negative control (from A) and assembly reactions (from B) were used as templates in a PCR reaction, in a more sensitive attempt to detect successful assembly. Arrows indicate the expected size of DNA fragments. PCR productions and restriction digestion products were analysed on a 1.6% agarose gel

5.4.2 Overlap Extension PCR-Based Gibson Assembly

Next, I attempted to assemble Gblock 1 and Gblock 2 together using OE-PCR. The technique is used to splice together two or more DNA fragments that contain complementary sequences (Bryskin and Matsumura, 2010). The OE-PCR method consists of two primary PCR reactions which add a common region to each fragment and to the vector. Next a secondary PCR reaction ensues which fuses the two fragments together using the newly introduced overlapping complementary sequences as primers, to facilitate fusion of multiple DNA fragments (Bryskin and Matsumura, 2010) (figure 5.4.2.1). OE-PCR cloning is advantageous because it is a quick and effective approach to achieve multiple gene fusions, and there is no requirement for restriction enzyme sites to stitch adjacent fragments together .

The OE-PCR method consists of two primary PCR reactions which add a common region to each fragment, and a secondary reaction which fuses the two fragments together (figure 5.4.2.1). For the

primary reaction Gblock1 was amplified in a gradient PCR using a forward primer with overhangs complementary to the BPK4410 plasmid vector and a reverse primer with overhangs complementary to the 5' end of Gblock2. Gblock2 was simultaneously amplified in a separate gradient PCR using a forward primer with overhangs complementary to the 3' end of Gblock1 and a reverse primer with overhangs complementary to the plasmid vector. After both primary PCR reactions, a secondary PCR reaction consisting of the Gblock 1 and Gblock 2 amplicons without any PCR primers was performed. Finally, addition of Gblock 1 forward primer and Gblock 2 reverse primer served to amplify the newly-extended full-length insert.

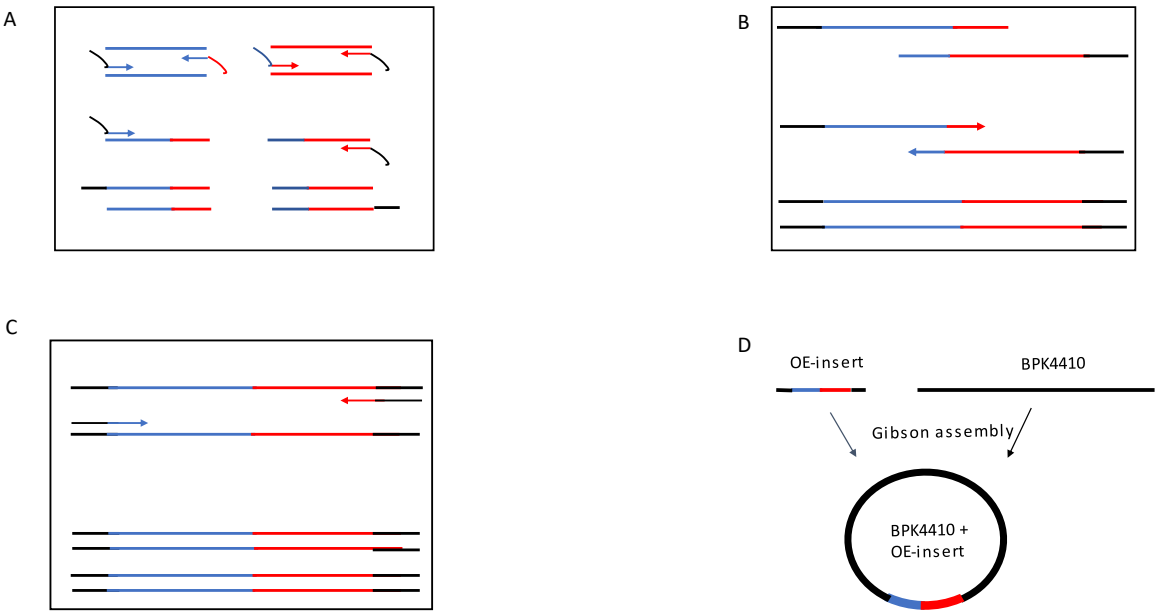


Figure 5.4.2.1 Overlap extension PCR overview. A) The primary PCR reaction produces double stranded smaller Gblock 1 (thick blue line) and Gblock 2 (thick red line) fragments with overlapping ends to each other (thin red and blue overhang lines) and to the vector backbone (thin black overhang lines). B) Secondary PCR reaction commences with the mixing of the primary PCR reaction products followed by amplification for 10 cycles to produce full-length fused Gblocks. C) Secondary PCR reaction is completed by the introduction of insert flanking primers to permit amplification of the newly fused full length Gblock fragment. D) OE-Gblock fragment were combined with linearised BPK4410 in a one-step isothermal Gibson assembly reaction.

Following the secondary PCR reaction using a set of flanking primers, a band ~400 bp corresponding to the Gblock 1 and Gblock 2 fusion was observed (figure 5.4.2.2 C). Unexpectedly, an amplification of a band ~200 bp which could correspond to Gblock 2 was detected. Equally for the overlap-extension samples where only one Gblock was used as a template in the secondary PCR reaction, the original Gblock amplicon was present along with the 400 bp amplicon which could have corresponded to the

desired insert fusion (figure 5.4.2.2 C). This suggested a degree of carryover contamination had occurred.

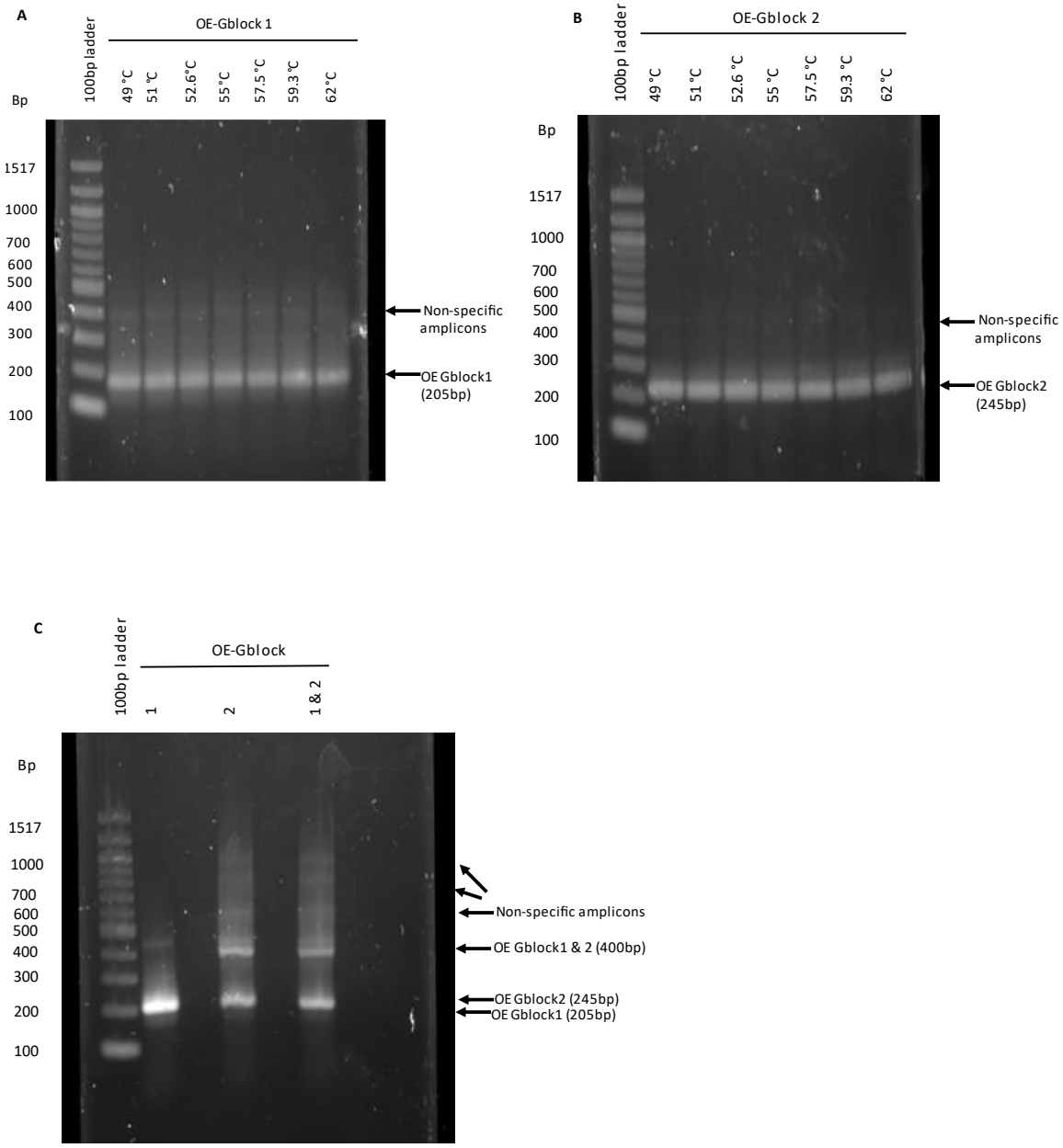


Figure 5.4.2.2 Gblock fusion by OE-PCR. **A**. Gradient PCR of Gblock fragment 1 across a range of annealing temperatures to identify optimal PCR conditions and to introduce overhangs. Amplicons of the desired size were excised across all wells, then gel purified **B.** Gradient PCR of Gblock fragment 1 across a range of annealing temperatures to identify optimal PCR conditions and to introduce overhangs. Amplicons of the desired size were excised across all wells, then gel purified. **C**. Secondary PCR reaction commenced with the mixing of the primary PCR reaction products (OE-Gblock1 & 2), while negative control reactions consisting of isolated individual OE-

Gblocks (Gblocks 1 and Gblock 2) were included. Next amplification for 10 cycles ensued to produce full-length fused Gblocks, then flanking primers were introduced to complete OE-PCR reactions. PCR products using these flanking primers were analysed on a 1.6% agarose gel.

Based on the possibility a population of the overlap-extended products contained the desired fusion products, Gibson assembly was once again attempted. The Assembly products were subsequently used to transform *E. coli* TOP 10 via electroporation, but colony PCRs using primers situated within the vector backbone, both flanking the insert, failed to identify any recombinant clones (figure 5.4.2.3) suggesting a new cloning strategy had to be reconsidered.

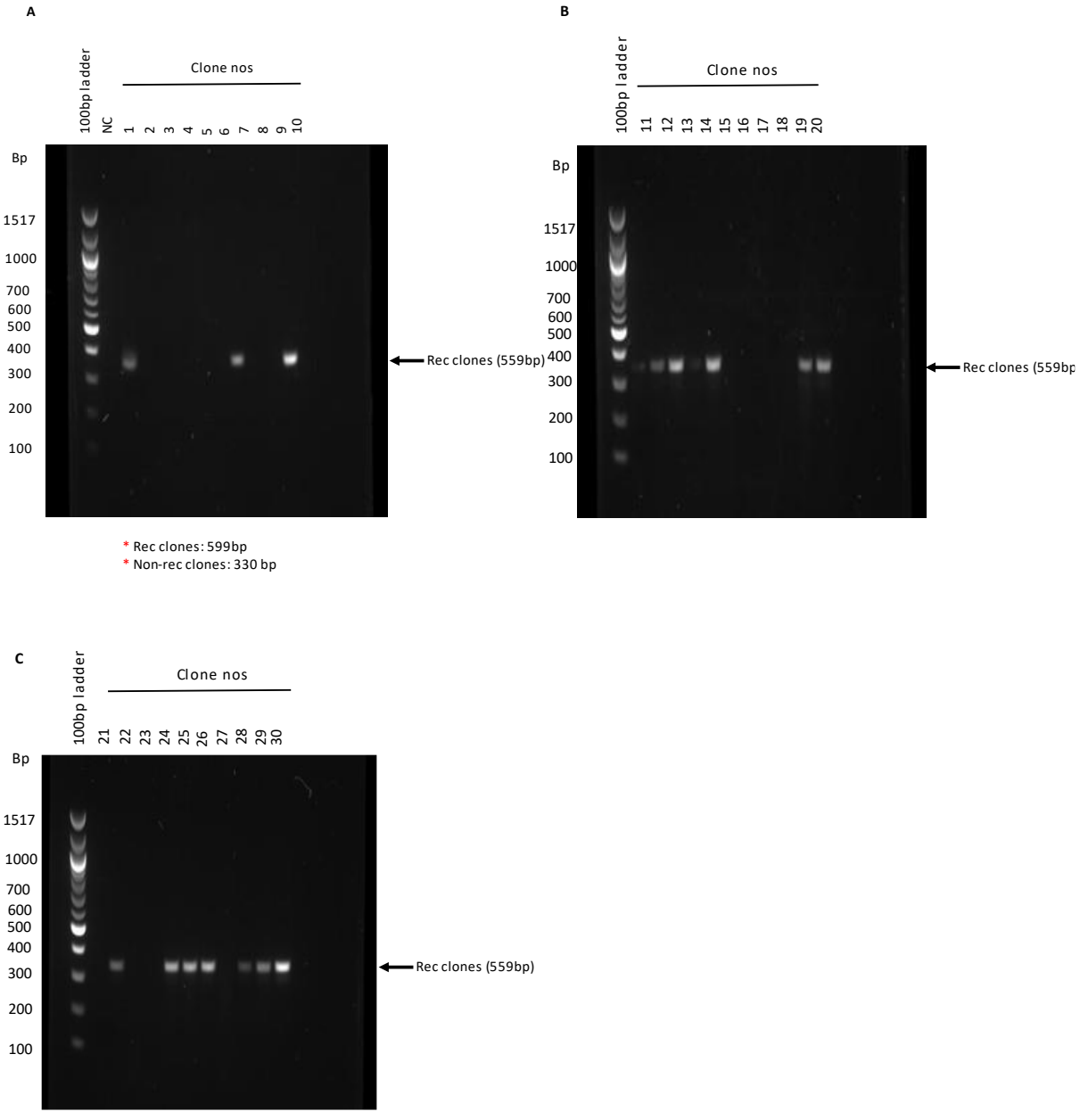


Figure 5.4.2.3. Screening of *E. coli* TOP 10 OE-PCR transformants by colony PCR. **A, B, C**. *E. coli* TOP 10 bacteria were electroporated with *OE*-Gibson assembly products then clones were screened using primers located within the vector backbone that flanked the inserts. Presence of recombinant clones (Rec clones) with the correct insert produced amplicons of 599 bp in size, clones transformed by wild-type plasmid without OE-Gibson assembly products (Non-Rec clones) produced amplicons of 330bp; a negative control (NC) no-DNA reaction was included to check for vector DNA contamination in PCR master mix. Clones untransformed by plasmid vector produced no amplicons. Colony PCR products were analysed on a 1.6% agarose gel.

5.4.3 Engineering a Ribozyme-Separated Gene-Editing STU Precursor

One of the reasons behind the failure to complete the Gibson Assembly could be the use of two $tRNA^{Gly}$ sequences in a single plasmid; cloning of extended sequence duplications in *E. coli* are reportedly problematic. Therefore, an alternative approach using two ribozymes, the Hammerhead (HH) and Hepatitis delta virus (HDV) ribozymes, was selected (figure 5.4.3.1). In this design, the HH ribozyme was cloned downstream of the Hypa Cas9 poly-A tail, followed by the CF gRNA 1 sequence and the HDV ribozyme sequence immediately after, to allow for precise separation and processivity of the 3' end of the gRNA.

Figure 5.4.3.1 Outline of the Ribozyme separated STU cassette. Hypa Cas9 coding sequence is followed by a 50 nucleotide-long poly(A) tail that improves mRNA stability. HH and HDV Ribozymes flank the gRNA sequence on the left and right of the gRNA, respectively, to permit specific processivity of gRNA *in cellulo.*

To generate the desired STU architecture classical restriction digestion cloning was employed as a tool using the Hypa Cas9-containing BPK4410 plasmid vector and a single Gblock (figure 5.4.3.2). This Gblock contained an adaptor sequence possessing an *Xho*I restriction site, the HH ribozyme - CF gRNA 1 sequence - 3' HDV ribozyme sequence in tandem, and finally another adaptor sequence containting a *Pst*I restriction site. The Gblock was initially amplified by PCR and restriction digestion reaction was performed on both the Gblock amplicon and the BPK4410 plasmid vector to introduce compatible sticky-ends to both DNA fragements. The digested BPK4410 vector was then dephosphorylated to prevent self re-ligation. The digested Gblock and dephosphorylated plasmid vector were subsequently ligated together. The ligation products were used for chemical transformation of bacteria.

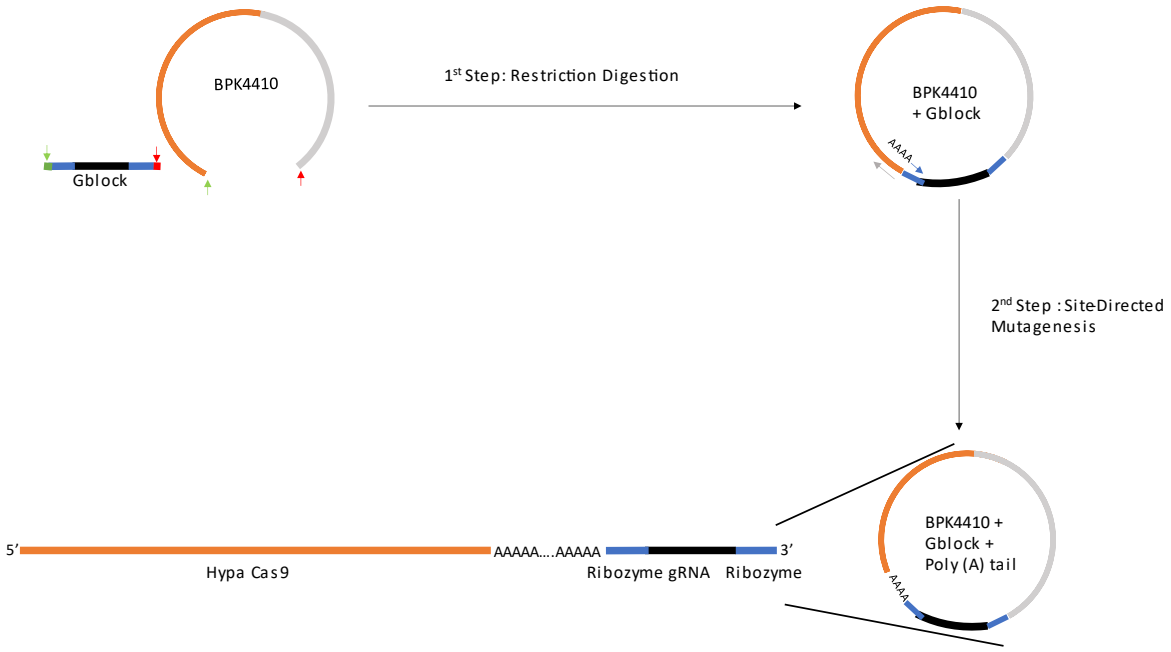


Figure 5.4.3.2 Cloning of Ribozyme-separated STU gene-editing cassette by restriction digestion and site-directed mutagenesis. The BPK4410 plasmid vector encoding Hypa Cas9 and a ribozyme-flanked gRNA Gblock (black line flanked by blue lines) containing adaptor sequences (green and red lines) were double digested by *Xho* I (green arrow) and *Pst*I (red arrow) restriction enzymes, for overhang-mediated ligation cloning. A site-directed mutagenesis PCR was performed to introduce a stretch of 50 adenosine residues between the Hypa Cas9 and HH sequences, by using a primer complementary to HH sequence with a poly(A) tail overhang (AAAA plus blue arrow). The final STU consisted of a Hypa Cas9 coding sequence, a poly(A) tail to improve mRNA stability, and HH and HDV Ribozymes that flanked the gRNA sequence, to permit specific processivity of gRNA *in cellulo.*

Colony PCR on transformed TOP10 *E. coli* revealed an amplicon of the anticipated size suggesting a successful recombination event (figure 5.4.3.3 A). Next, a poly-A tail consisting of 50 adenosine nucleotides was added between the stop codon of Hypa Cas9 and the HH sequence by site-directed mutagenesis to improve mRNA stability, and confirmed by agarose gel electrophoresis (figure 5.4.3.3 B). Site-directed mutageneisis was carried out by performing a PCR using a forward primer containing a 50-nucleotide-long poly-A tail overhang and the recombinant gene-editing cassette as a template, followed by a ligation reaction to circularised the newly-synthesised plasmid DNA containing the poly-A sequence, and finally DpnI digestion to degrade non-ligated linear plasmid DNA (figure 5.4.3.2). Plasmid DNA extracted from these recombinant clones appearing to possess the intact Gblock plus poly-A tail were confirmed to contain an intact STU comprising all the necessary components required for gene-editing by Sanger sequencing.

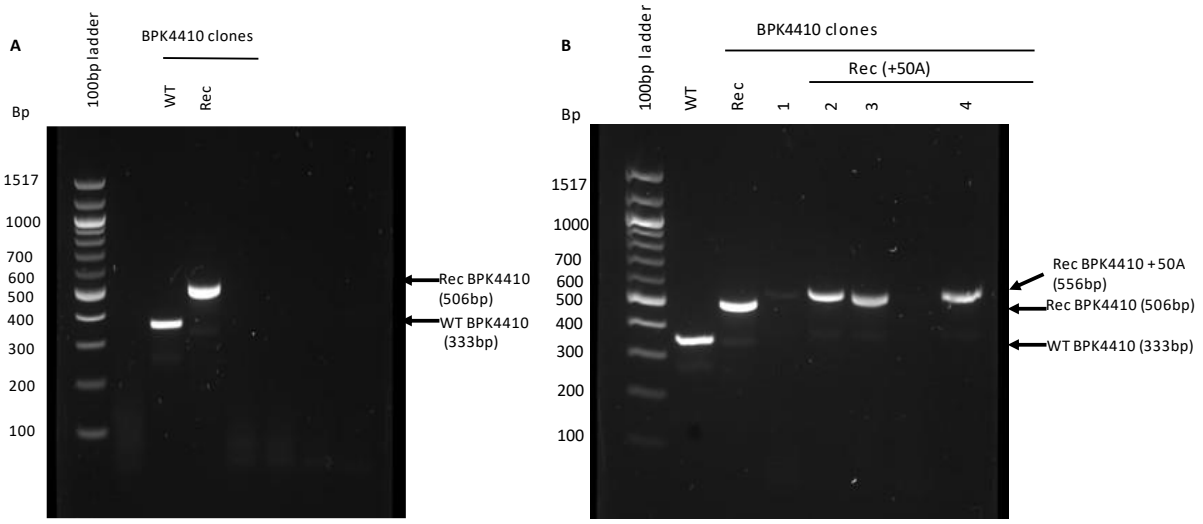


Figure 5.4.3.3 Screening dual ribozyme-flanked gene-editing STU transformants. **A.** Restriction digestion and ligation were performed then bacteria were chemically transformed with either the HypaCas9-encoding BPK4410 vector or recombinant vector encoding a HH ribozyme (HH)-gRNA-HDV ribozyme (HDV) Gblock without a poly-A tail. **B.** Upon Gblock incorporation, site-directed mutagenesis was performed to introduce a poly-A tail between HypaCas9 and the dual ribozyme-flanked gRNA Gblock. Clones were screened using primers located within the vector backbone to produce amplicons of 333bp for Wild-type BPK4410 (WT-BPK4410), 506 bp for BPK4410-HH-gRNA-HDV (RecBPK4410), or 556 bp (BPK4410-A_{50}-HH-gRNA-HDV) transformants.

5.4.4 Attempted Gene-Editing Using Ribozyme-Separated Gene-Editing STU

To determine the editing efficiency of the dual ribozyme-flanked gene-editing STU targeting CFTR exon 11, this plasmid construct was transfected into CFBE41o-, A549 and HeLa cells using lipofection and PEI, while assaying different post-transfection incubation periods. Gene-editing efficiency was determined by T7EI (see section 1.5.2 and figure 5.4.4.1) and quantified densitometrically. Disappointingly, the *CFTR*-targeting STU gene-editing construct was ineffective at generating indels in any cell line tested, irrespective of incubation period, preventing any further attempts to quantify indel induction by TIDE or engineering HDR template inclusion in the STU gene-editing unit. Instead, efforts were focused on troubleshooting the possible reasons why gene-editing was not observed in the desired locus.

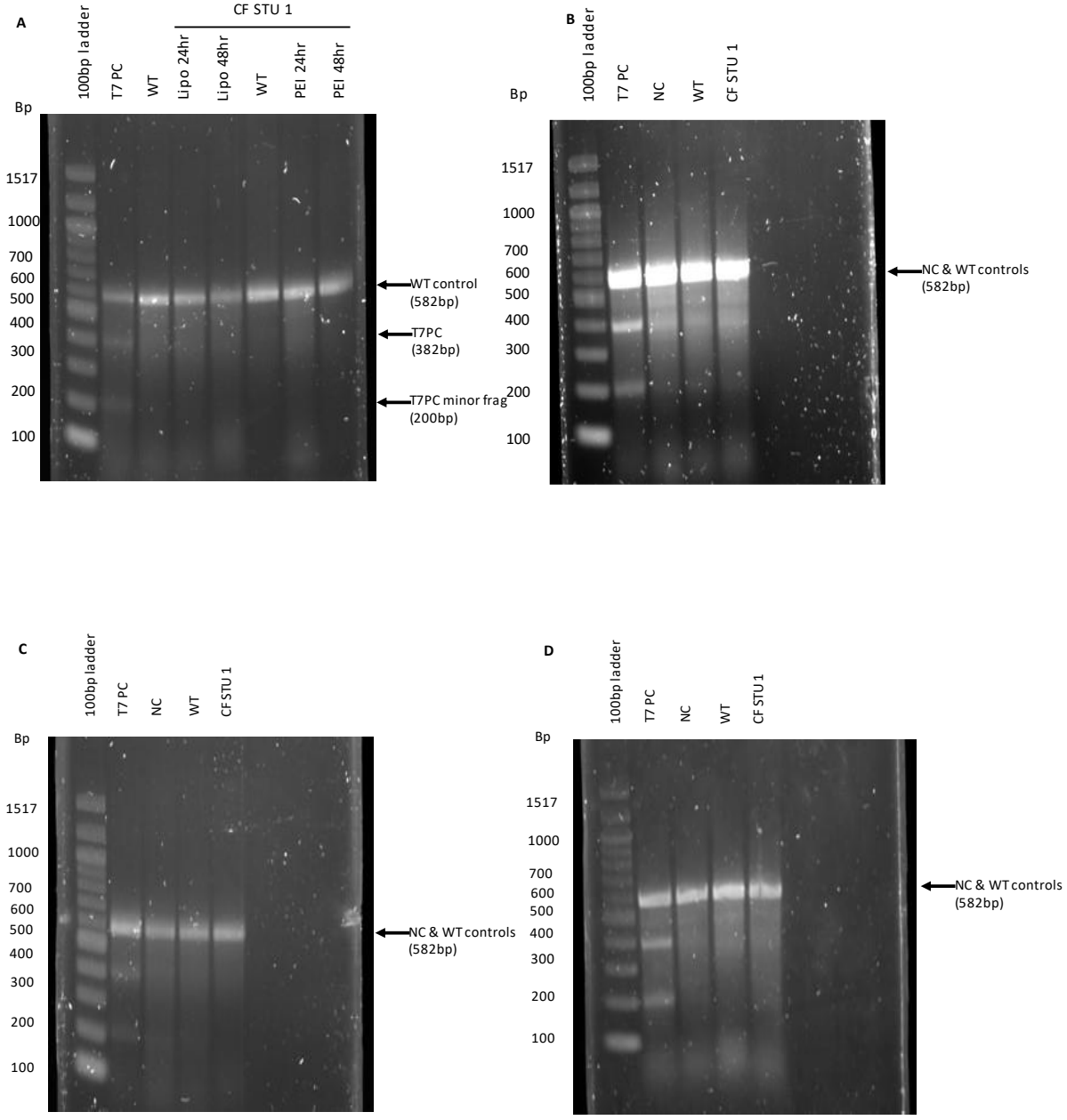
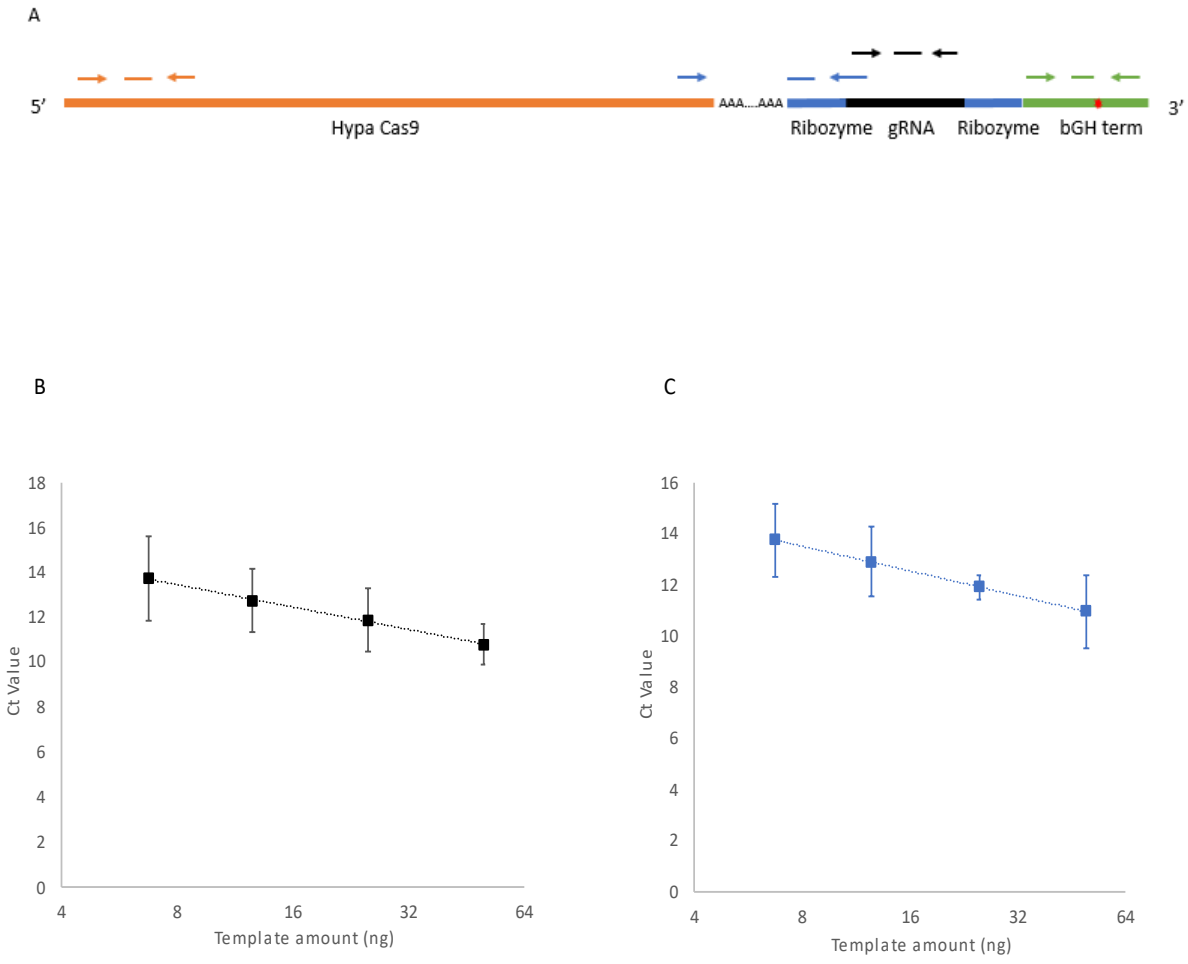


Figure 5.4.4.1 CFTR-targeting STU 1 gene-editing plasmid activity *in vitro*. A. STU gene-editing plasmid transfection with lipofeactamine CRISPRMAX™ (Lipo) or polyethyleneimine (PEI) for 24 or 48 hours does no result in CFTR gene-editing in CFBE41o- cells. In CFBE41o- cells **(B)**; A549 cells **(C)** and HeLa cells **(D)**; T7E1 cleavage assay with a T7 positive control reaction included (T7 PC), was performed on DNA extracted from untransfected (NC), wild-type BPK4410 plasmid (WT)-transfected and ribozyme separated CF gRNA 1 STU (CF STU 1)-transfected cells. Arrows indicate the DNA fragments sizes. Cleavage products were analysed on a 1.6% agarose gel.

5.4.5 Ribozyme-Separated Gene-Editing STU are Transcribed in Mammalian Cells

One of the possible reasons why the ribozyme-based STU gene-editing was inactive, could have been that the components of this gene-editing STU were not transcribed as intended, or suffered differential stability in cells. To achieve this, a set of primers and hydrolysis probes were designed for RT-qPCR validation of individual STU gene-editing component expression, specifically: the STU holotranscript (unprocessed gene-editing precursor), processed STU, and HypaCas9 mRNA, and a sequence downstream of the bGH termination site (bGH term), serving as a control for plasmid DNA carryover during the total RNA extraction step (figure 5.4.5.1 A). Thus, in a single set of reactions, any amplification in excess of background ascribable to carryover plasmid could be attributed to transcriptional activity as opposed to plasmid DNA carryover, a notorious issue in quantifying plasmid-borne gene expression in eukaryotes. The amplification efficienies of RT-qPCR primers and probes specific for various STU-gene-editing components were tested against serial dilutions of recombinant plasmid, which were determined as 96.24%, 102.36%, 98.20% and 87.52% for the processed STU, STU holotranscript, HypaCas9 mRNA and bGH term site aspects respectively (figure 5.4.5.1 B - E).

To quantify the cellular expression of the STU gene-editing components (mature STU, holo-unprocessed STU and Cas9 mRNA), CFBE41o- cells, A549 cells and HeLa cells were transfected with either the recombinant BPK4410 plasmic harbouring the ribozyme *CFTR*-targeting gene-editing STU, BPK4410 plasmid, or a plasmid encoding GFP, and RT-qPCR assays were performed on total RNA extracts at set intervals, using a commercially procured 18S rRNA-specific assay as the endogenous control. All STU gene-editing components were detectable, and expressed to varying degrees in all three cell lines, however statistical significance above background levels was only observable in CFBE41o- cells (figure 5.4.5.2). In terms of component expression intensity, the Cas9 transcript was detected at the highest levels followed by the gRNA component. Moreover, evidence of gRNA excision from the gene-editing STU precursor was documented due to the higher expression of the mature STU transcript compared to the precursor, particularly in CFBE41o- cells, but less so in A549 and HeLa cells. Off note, there was a modest amount of plasmid DNA carry over during the RNA extraction, reflected in the relatively high Ct values for the no-RT control reactions, wherein amplification from RNA templates was presented. Equally, the bGH primer probe set also resulted in modest amplification suggesting plasmid carryover in spite of DNAse treatment of total RNA extracts; reactions loaded with cDNA amplified at Ct values above the respective no-RT controls, suggesting this bGH terminator sequence was also being expressed.


A
5'
Hypa Cas9
AAA....AAA
Ribozyme
gRNA
Ribozyme
bGH term
3'
B
Ct Value
Template amount (ng)
C
Ct Value
Template amount (ng)

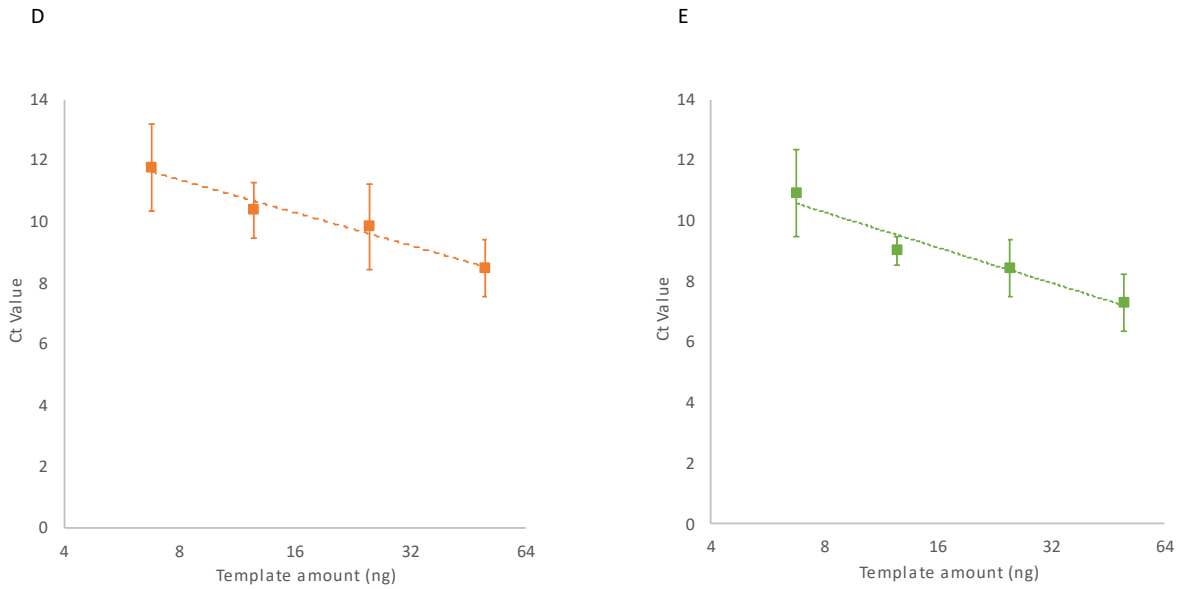


Figure 5.4.5.1 RT-qPCR primer probe set development for STU gene-editing components. The mapping of the RT qPCR primers (arrows) and probes (thin lines) targeting the mature STU (blue), holo-unprocessed STU (black), Hypa Cas9 mRNA (orange) and bGH terminator (green). RT-qPCR primer-probe sets targeting the mature STU (B), holo-unprocessed STU (C), Hypa Cas9 mRNA (D) and bGH terminator (E), were analysed by probe-hydrolysis RT-qPCR against two-fold serial dilution of recombinant ribozyme-separated STU plasmid DNA to determine assay efficiency, according to (Ginzinger, 2002) method.

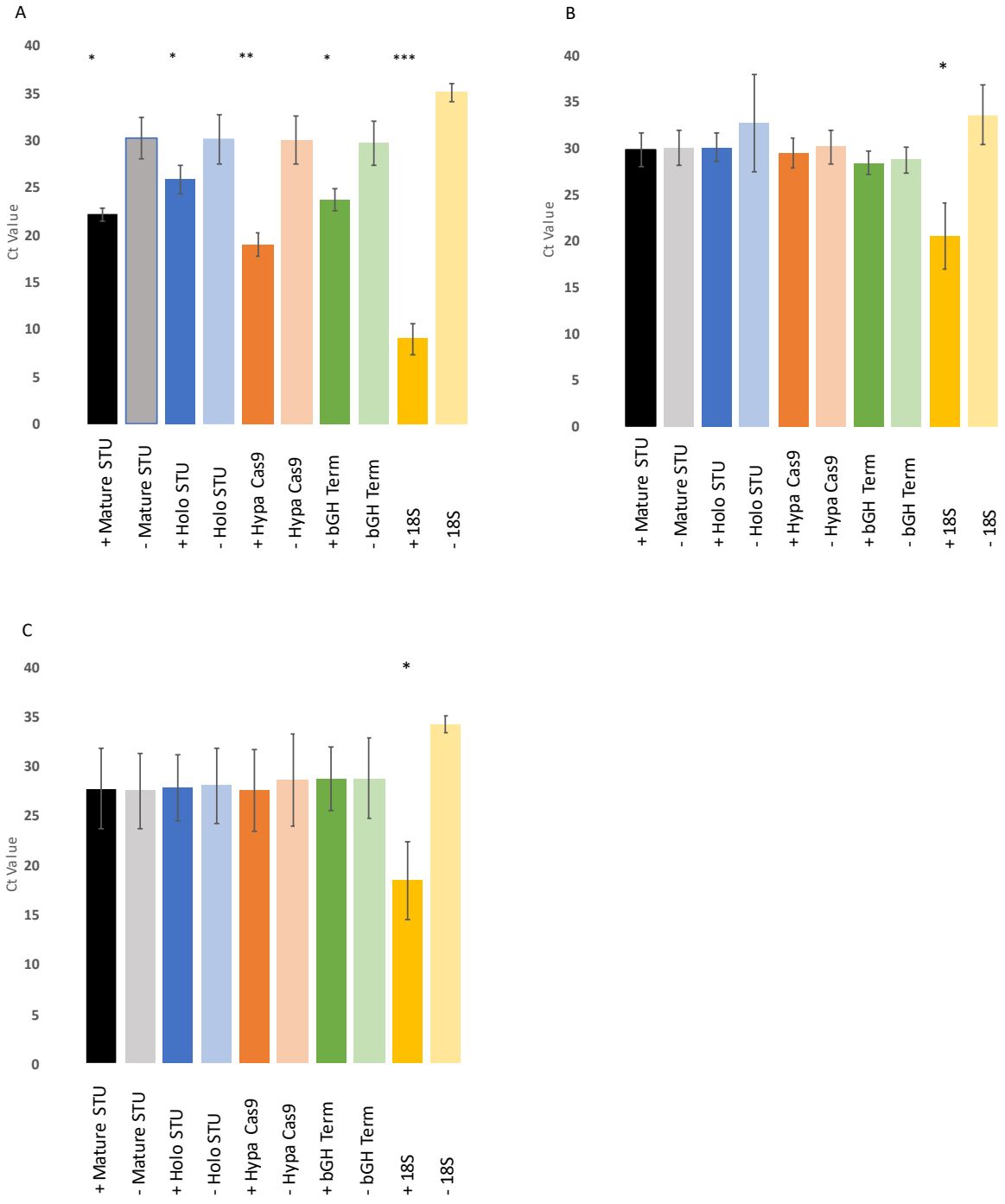


Figure 5.4.5.2. Transcriptional analysis of STU gene-editing component production from a plasmid *in cellulo*. Primer-probe sets targeting the processed STU gene-editing precursor (black bars), unprocessed STU gene-editing precursor (blue bars), Hypa Cas9 mRNA (brown bars) and bGH terminator (green bars), were used to analyse the expression of these individual components in total RNA extracts from CFBE41o- cells (A), A549 cells

(B), and HeLa cells (C) transfected with a gene-editing STU-encoding plasmid. No-RT negative controls were included for each primer-probe set (lighter shades of respective colours). Data is represented as the average of three separate biological experiments (n=3). Error bars represent +/-SD. Statistical significance was assessed by paired t test between primer pairs and corresponding no RT controls. *p<0.05, **p<0.01, ***p<0.001.

5.4.6 In-Efficient STU Gene-Editing Activity is Due to Low Cas9 Translation in Mammalian Cells

These results confirmed that, at least in the disease-relevant CFBE41o- cells, gene-editing STU transcription and post-transcriptional maturation was proceeding resulting into detectable quantities of gRNA, yet no appreciable gene-editing activity was detectable. Furthermore, Sanger sequencing was performed to confirm the sequence integrity of the CMV promoter driving STU transcription with no mutations observed which could have negatively impacted upon the expression of this ribozyme-separated STU. This finding implicated inefficient HypaCas9 translation as one of the possible factors contributing to failure in detecting gene-editing activity. Hence, an antibody specific to HypaCas9 was used to evaluate HypaCas9 expression in CFBE41o- cells, A549 cells and HeLa cells, after transfection with GFP encoding, BPK4410 CF STU 1-encoding or wild-type BPK4410 encoding plasmids (figure 5.4.6.1). Western blotting assays were performed 24 hours post-transfection on total protein extracts of these cells for HypaCas9 in comparison to a GAPDH house-keeping gene control, showing no bands around the anticipated 154kDa mark for Hypa Cas9, despite clear detection of GAPDH. These results suggested that translation of HypaCas9 was not occuring in BPK4410 CF STU 1 or wild-type BPK4410-transfected cells, or that expression of was minimal and below the limit of detection of this highly sensitive chemiluminescent assay.

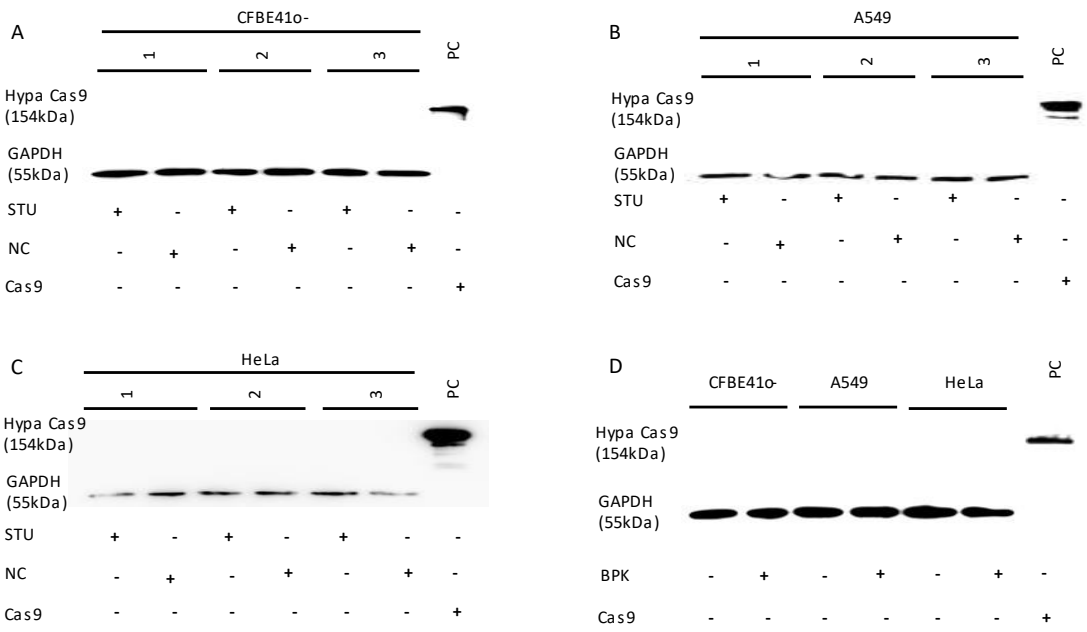


Figure 5.4.6.1 HypaCas9 protein is not efficiently expressed from a STU gene-editing plasmid. The expression of HypaCas9 in cells treated with a *CFTR*-targeting STU gene-editing plasmid (STU), or a GFP-encoding plasmid (NC) was determined by western blot analysis in CFBE41o- cells (A), A549 cells (B), and HeLa cells (C). **D.** The expression of HypaCas9 in cells treated with wild-type BPK4410 plasmid (BPK) was equally assayed in the same three cell lines. GAPDH expression was used as a loading control, while recombinant Cas9 protein (Cas9) was used as an assay positive control (PC) to validate the functionality of the rabbit anti-cas9 antibody. The blot shown presents 3 separate biological experiments (1-3).

To further test this hypothesis of HypaCas9 inactivity, the ability of the wild-type BPK4410 plasmid vector used for cloning CF STU 1 to mediate gene-editing was assayed in CFBE41o-, A549 and HeLa cells. Here, co-transfection of wild-type BPK4410 along with a positive control B2M gRNA, delivered at differential timepoints, was ineffective at generating indels in any cell line tested (figure 5.4.6.2). These results further validated the hypothesis that the failure to induce genome editing was not necessarily due to faulty STU design, but rather failure for HypaCas9 to translate, which remains a mystery why this observation was made.

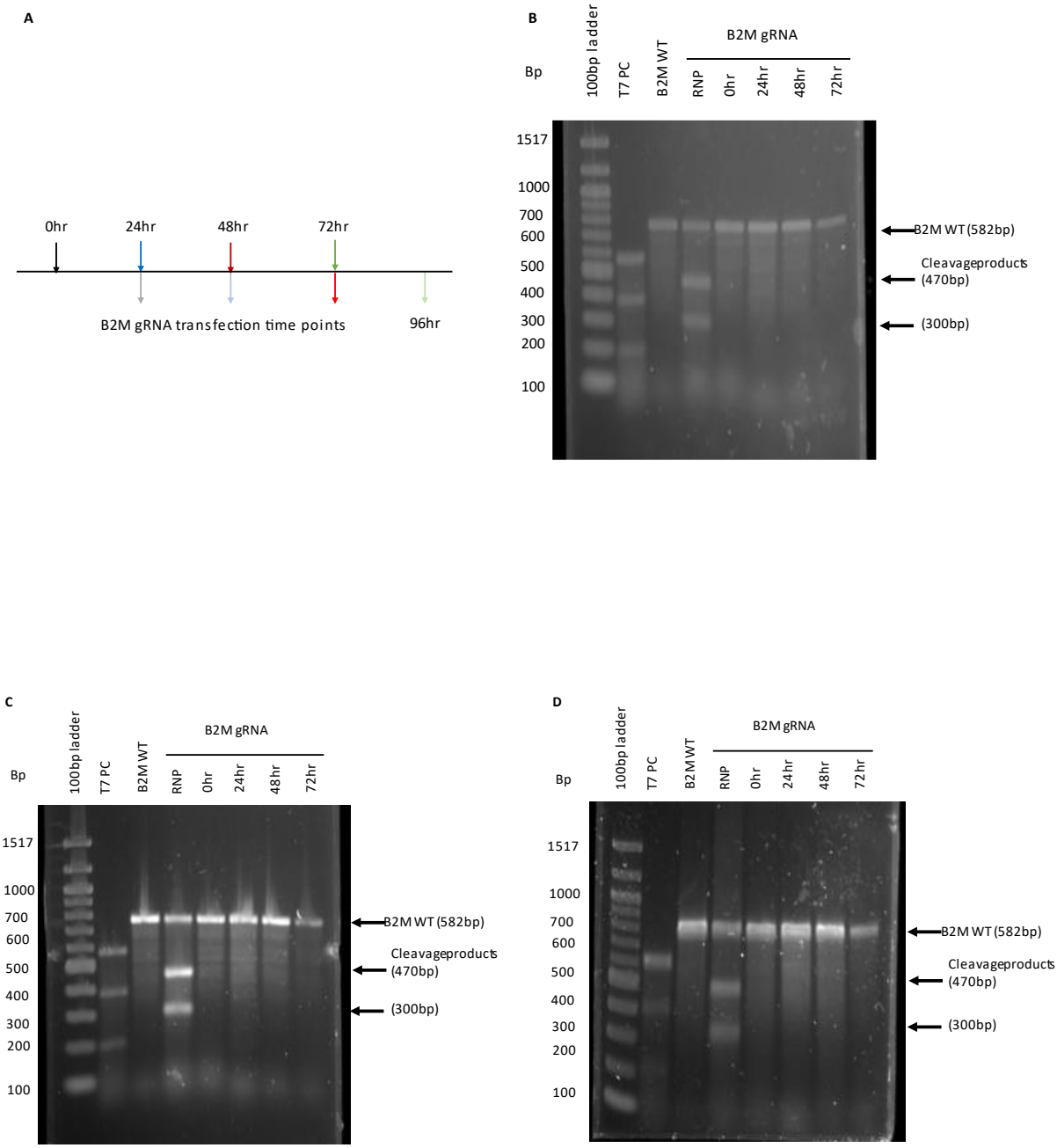


Figure 5.4.6.2. Assaying BPK4410 plasmid activity *in cellulo*. A.BPK4410 gene-editing was lipofected at the 0 hour time point, then B2M gRNA was lipofected at 0 (black arrow), 24 (blue arrow), 48 (red arrow) and 72 (green arrow) hours, with genomic DNA extracted from cells 24 hours post transfection (light shades of respective colours). Differential B2M gRNA transfection does not result in *B2M* gene-editing in CFBE41o- cells **(B)**; A549 cells **(C)** and HeLa cells **(D).** Arrows indicate the expected DNA fragment sizes. Cleavage products were analysed on a 1.6% agarose gel.

5.5 Discussion

There are several challenges in the production of CRISPR-Cas9 system for the purpose of gene-editing including, considerable endonuclease size, variable and lack of coordinated expression of gene-editing components, non-tissue specific nature of pol III promoters, and time-consuming nature of cloning multiple gene-editing components. To resolve these issues, the value of expressing both Cas9 and gRNA components was evaluated in a STU driven from a single constitutive pol II promoter. This would ensure for better co-ordinated expression, as well as nuclear and spatiotemporal localisation of gene-editing components post-transcriptionally, in a single transfection step. To achieve *in cellulo* separation and maturation of Cas9 and gRNA components, a STU precursor of these two components was designed separating these two key gene-editing components with a $tRNA^{Gly}$ sequence previously reported as functional (Dong *et al.*, 2017; Tang *et al.*, 2019; Xie, Minkenberg and Yang, 2015). This strategy relied on the endogenous tRNA-processing system which cleaves both ends of the tRNA precursor, and was expected to permit precise gRNA separation from Cas9.

Cloning such a construct proved challenging for several reasons. Synthetic production of a single Gblock comprising of a poly-A tail and two pre-$tRNA^{Gly}$ repeat sequences interspaced by a CF gRNA 1 sequence proved impossible according to the supplier. This forced a split of the components into two smaller Gblocks, the first containing a poly-A tail and one pre-$tRNA^{Gly}$, while the other comprised of the CF gRNA 1 and the second pre-$tRNA^{Gly}$ repeat. Amplification of these Gblocks proved unsuccessful probably due to the highly repetitive nature of their sequence components which could've invariably lead to polymerase stalling and concatemer generation (Hommelsheim *et al.*, 2014), which negatively impacted downstream Gibson assembly reactions. Switching to OE-PCR to stitch Gblock 1 and Gblock 2 PCR together appeared to generate a heterologous combination of the desired fusion Gblock and DNA fragments of a smaller sizes. However, after Gibson assembly and transformation into bacteria, no recombinant clones were obtained. One possibility is that the pre-$tRNA^{Gly}$ repeats within each Gblock resulted in their cross-hybridisation after each PCR cycle, forming fusion concatemers of hybrid DNA products. Since these challenges were not reported in other studies (Tang *et al.*, 2019; Dong *et al.*, 2017), the issue may not solely be a result of the $tRNA^{Gly}$ repeat sequences, but be amplified by the inclusion of fifty adenosine residues. This synthetic poly (A) tail was introduced to protect HypaCas9 mRNA from post-transcriptional degradation by nucleases once STU components had achieved *in cellulo* maturation (Xu Tang, 2017). Therefore, to achieve successful STU cloning, it was decided that the introduction of poly(A) tail and indeed any repeat sequences present in Gblock cassettes needed to occur in separate cloning steps.

An alternative strategy was therefore employed relying on self-cleaving ribozymes (Gao and Zhao, 2014; Tang *et al.*, 2019). Furthermore, the fifty adenosine residues were added at a later stage using

site-directed mutagenesis to further lower the design complexity of the initial Gblock, concurrently reducing any PCR-inhibition and concatemer formation during amplicon generation steps. Finally, cloning was shifted to simpler, type II restriction enzyme digestion then ligation, which eliminated the complexities involved with designing long Gibson assembly primers capable of self-hybridising, risking the formation of primer dimers, or undesirable complex secondary structures. These changes resulted in the correct STU production as determined by Sanger sequencing.

Yet despite the efficiency of CF gRNA 1, this ribozyme-separated gene-editing STU construct did not display any activity *in cellulo*, as determined by T7EI, across 3 different cell lines including a highly permissive, non-lung cell line. The observation posited transcriptional, processing, half-life, or translation problems for the construct and its products.

To systematically deconvolute the problem, the transcriptional performance of the ribozyme-separated STU gene-editing construct was assessed, using custom probe hydrolysis RT-PCR assays, including an assay sensitive to key junctions of the STU components. The analysis confirmed that all individual components were expressed in CFBE41o- cells, albeit at low expression levels. This was somewhat unexpected given the well-described transcriptionally powerful CMV promoter, one of the strongest constitutive pol II promoters (Damdindorj *et al.*, 2014; Xia *et al.*, 2006), was driving expression of the STU. Yet no mutation in the promoter sequence was observed that could account for the low expression levels or loss of transcription in successfully transfected cell lines other than CFBE41o-. In an alternative cloning strategy, recombinant gene-editing STU cassette could have been introduced into a plasmid vector such as PX330, under the control of a strong constitutive chicken beta-actin hybrid (CBh) promoter, which was demonstrated to drive higher GFP expression in mice hippocampal slices, compared to CMV promoter (Gray et al., 2011).

One possible way to improve expression of gene-editing STU components would have involved the optimisation of transfection efficiency in the assayed cell lines following the introduction of a GFP -encoding plasmid, then calculating the amount of GFP positive cells expressed as a percentage of the total number of cells transfected. Hence, by introducing a GFP-encoding plasmid, then making alterations to, cell seeding density, plasmid and transfection reagent concentrations, it would have been possible to optimise transfection efficiency in a cell-type dependent manner and invariably improve the expression of gene-editing components. After the introduction of GFP encoding-plasmid, the amount of GFP positive cells expressed as a percentage of the total number of cells transfected, could have been used to determine the effect of the above-mentioned optimisation steps on transfection efficiency.

Results also suggested that the gRNA component was less abundant than the Cas9 mRNA, a reflection probably of the lower halflife of the gRNA due to the absence of stabilising sequences. Furthermore, the assay sensitive to the ribozyme cleavage point was showing lower quantities of the precursor vs the processed STU confirming efficient cleavage and gene-editing STU *in cellulo* processing (Gao and Zhao, 2014; Tang *et al.*, 2019). Thus, with adequate Cas9 mRNA and gRNA quantities confirmed in CFBE41o- cells, one would expect some gene-editing activity, subject to efficient Cas9 translation, and assuming the gRNA length and sequence remained as intended for a sufficient amount of time relative to Cas9 translation kinetics and protein halflife.

Interestingly no appreciable expression of the STU was observed in A549 and HeLa cells. This might be attributable to the quality of RNA extracted from these cells lines as implied by the comparatively higher Ct value for 18S in the corresponding extracts, despite the fixed amount of total RNA template per reaction. This drop in RNA quality coupled with the modest plasmid-DNA carry-over during the total RNA extraction step (a known issue in RNA extraction that cannot be overcome by DNase I treatment) evidenced in CFBE41o-'s by the bGH terminator sequence amplification, could have confounded detection in both A549 and HeLa cells. An unexpected observation, however, was bGH terminator sequence expression: this transcriptional read-through, unaccounted for during STU design, could have had implications on transcript secondary structure that could subsequently affect processing of the gRNA component. However, given the evidence of gRNA and Cas9 mRNA separation, one would not expect any impact of bGH readthrough on Cas9 translation.

Strikingly, no translation of HypaCas9 was observed across all cell lines surveyed by Western Blotting, despite the confirmation of STU expression. This observation could be attributable to either one of three possibilities. First, there was an issue with the cellular translation machinery which could have limited or prevented the translation of the recombinant protein component, albeit with the presence of an intact Hypa-Cas9 encoding RNA transcript. Although improbable, it is possible that faulty cellular translation machinery in all three assayed cell lines, affecting components of the protein synthesis machinery, could have hindered cellular protein production. For example, mutations in tRNA synthase (*AARS*) gene could have prevented the tRNA aminoacylation (attachment of amine groups to tRNAs). Indeed, tRNAs are a vital component in the protein synthesis machinery responsible for delivering amino acid residues to the appropriate protein translation site, through Watson-Crick base pairing using endogenous anticodons to bind to codon sequences present on an mRNA (Kapur and Ackerman, 2018; Nakayama *et al.*, 2017). The functional relevance of this synthase was confirmed when the *in vitro* aminoacylation catalytic activity of a truncated mutant AARS was found to decrease by 86%, tested using ^{14}C-labeled alanine with various concentrations of tRNA (Nakayama *et al.*, 2017). Equally mutations in components of ribosomal proteins, could affect the ability of ribosomes to distinguish

aminoacyl tRNAs anticodon sequences, increase stop codon readthrough and prevent ribosomal translocation along the mRNA being translated (Drummond and Wilke, 2009; Kapur and Ackerman, 2018). In one instance, mutations in ribosomal uS12 subunit, responsible for monitoring the complementarity between the mRNA codons being translated and the anti-codons of aminoacyl-tRNAs, increased stop codon readthrough by up to 10 fold in *Saccharomyces cerevisiae* purine biosynthetic protein (*ADE1*) gene, determined using a luciferase reporter assay (Loenarz et al., 2014).

Second, the transcript encoding HypaCas9 might have been incorrectly processed: although qPCR results suggested this transcript was present, neither fragmentation due to RNase activity, nor epitranscriptomic post-transcriptional modifications interfering with translation can be excluded. Third, unanticipated microRNA recognition elements within the Cas9 mRNA cassette, or interactions with other regulatory RNA may have suppressed Cas9 translation. Finally, it is indeed possible that the transcript encoding HypaCas9 had been correctly processed, however, recombinant protein production could remain below the detection limit of this immunoblotting assay. Indeed, previous experiments have demonstrated the detection limitations of HRP-conjugated western blot chemiluminescent assay to be > 3pg of target protein substrate (Ghosh, Gilda and Gomes, 2014). Such low levels of translation, coupled to the arguably low copy number of the gRNA, even in the absence of gRNA modifications that could render it inactive, could result in inefficient gRNA loading and the lack of gene-editing activity in a three assayed cell lines. Furthermore, the lack of gene-editing activity observed using the wild-type BPK4410 plasmid plus a positive control B2M gRNA and the lack of HypaCas9 protein detection on western blots suggests that this endonuclease is generally poorly expressed in the wild-type plasmid.

Future work could attempt to troubleshoot possible reasons why HypaCas9 translation was not detected. This could involve cloning of STU-transcript downstream of an alternate promoter, although sequencing did not reveal any mutations with the CMV promoter in BPK4410. Equally an EGFP protein could be fused to HypaCas9 component to offer fluorometric or visual confirmation of expression. A norther blot, small RNA next generation sequencing, or oxford nanopore RNA modification analysis set of experiments would inform gRNA maturation outcomes.

Chapter 6. Concluding Remarks

Chronic genetic lung disorders such as CF and IPF are life-limiting conditions with significant disease-related morbidities impacting upon patients daily activities. As global life expectancy continues to increase, the debilitating effects of these genetic disorders gets extended for longer periods of time, to the detriment of a patient's quality of life. Current interventions for CF and IPF only serve to tackle disease symptoms long-term, extend life short-term, can be toxic and very costly. Therefore, there is

a need to develop targeted treatment for such conditions to ensure life expectancy and human health span are both increased in tandem.

Initially, to add to the pool of CF-targeting gRNAs already available, multiple gRNAs targeting the Δ F508 locus in the human *CFTR* gene were created. Initial designs of CF gRNAs were successful in targeting the Δ F508 locus across three different cell lines. Although there was variation in editing efficiency across the three different cell lines, the presence of observable editing events does demonstrate, to a degree, the versatility of these gRNA designs. Since the transfection system used in these experiments typically yields comparable delivery across these cell lines, activity levels can be attributed to phenotypic or karyotypic differences. Thus, chromatin condensation and epigenetic profiles, or immortalisation-related genome changes at the chromosome or gene level could account for the observed difference. Attempts at precision genome editing using HDR templates were unsuccessful despite alterations made to the donor design. Therefore, further optimisation steps would need to be taken to optimise donor template design to knock-in the wild-type *CFTR* allele. This finding does not necessarily question the efficacy presented for other HDR templates in the context of Δ F508; rather, their utility needs consideration in the production of a STU pro-drug for genome editing. Thus, plasmid HDR templates and similar solutions were not evaluated in this study. Of note, however, was the confirmation of CF gRNA 1 as the most efficacious gRNA design, serving as a prime candidate for future gene-editing studies are the Δ F508 *CFTR* locus.

The rs35705950 SNP has been previously identified as the greatest genetic risk factor for developing IPF. This SNP is located within a lncRNA AC061979.1 which has been predicted to be expressed *in silico* (Frankish et al., 2019), however there are no publications on validating the *in cellulo* expression of this lncRNA. Therefore, another aspect of this project attempted to validate expression of this lncRNA within A549 and CFBE41o- cell lines. Results indicated that this lncRNA molecule was in fact expressed within these two cell lines, admittedly at a very low copy number. Furthermore, the effects of cellular contact inhibition were found to negatively impact upon AC061979.1 expression. The confirmation of AC061979.1 presence has laid the foundation for future work to study the exact mechanisms of this lncRNA in the aetiology of IPF.

In an attempt to target the *MUC5B* locus using CRISPR-Cas9 gene-editing technology in the context of IPF, a set of three gRNAs targeting the rs35705950 SNP were designed then assayed *in vitro* and *in cellulo*. While gRNAs were proven functional *in vitro*, they were unsuccessful at generating indels at the desired genomic locus when transfected into cells. Inactivity of gRNA could have been attributable to gRNA sequence composition, a non-editable genomic locus, or a decrease in the cell viability of gene-edited cells. Hence future work could aim to optimise gene-editing approaches, which will

permit creation of IPF disease models by knocking-in the rs35705950 allele into cell lines and animals. Alternatively, gene-editing approaches could be used to knock-in the wild-type allele in the rs35705950 locus of IPF patient-derived cells, as a first step to eventually attempting to treat the disease genotype at early disease stages (pre-symptomatic) or mitigate against the risk of future disease development.

The use of CRISPR-Cas9 systems for the purpose of conventional genome-editing possess numerous challenges, including size of endonuclease component, issues with co-ordinated expression, uncharacterised nature of pol III promoters and the ubiquitous expression profile of gRNA pol III promoters. To circumvent these expression challenges, we assayed the potential use of STU technology to perform genome-editing in CFBE41o- cells. Initial attempts to use $tRNA^{Gly}$ to interspace gene-editing components proved problematic during the cloning process, although this strategy was in line with previous attempts that made use of overlap-extension PCR for the same process (Xie, Minkenberg and Yang, 2015). Other studies achieved the successful generation of $tRNA^{Gly}$-interspaced STU by applying type IIS restriction cloning strategy, which is a simpler approach than the numerous PCR reactions and Gibson assembly deployed in this study (Tang *et al.*, 2019; Dong *et al.*, 2017).

In a subsequent simplified cloning attempt, HH and HDV ribozymes were adopted to interspace gene-editing components, which eventually proved successful. Attempts to gene-edit the CFTR Δ F508 locus using this ribozyme-interspaced STU proved unsuccessful. Subsequent qPCR and western blot analysis suggested that gene-editing components were potentially being transcribed but not translated. These results differ from previous studies were gene-editing was achieved using the same technology in yeast, plants, and human cells (Gao and Zhao, 2014; Yoshioka *et al.*, 2015; Tang *et al.*, 2019).

Future work will aim to decipher the potential reasons why ribozyme-interspaced gene-editing STU proved non-functional. Eventually, successful development of this ribozyme-interspaced gene-editing STU will support future studies where this technology could be deployed in the context of prime editing. Briefly, prime-editing is a gene-editing approach where a prime-editing gRNA (pegRNA) which encodes the knock-in allele, is co-delivered along with an endonuclease and reverse-transcriptase to mediate gene-editing (Anzalone *et al.*, 2019). This prime-editing STU approach has numerous advantages including; (i) reduction in the frequency of indels generated characteristic of conventional CRISPR-gRNA-mediated gene-editing, (ii) Reducing the cargo size gene-editing cargo by eliminating the use of a donor template, (iii) fewer off target effects than conventional Cas9-initiated HDR. Previous gene-editing attempts in human cells have been successful using pegRNA delivered as an RNP, as well as using a plasmid encoding a multicistronic pegRNA cassette (Anzalone *et al.*, 2019; Dong

et al., 2017). However, advantages derived from usage of an STU design combined with a prime-editing approach, shall serve to advance the study and correction of chronic genetic lung diseases.

Another future approach to improve the safety of gene-editing approaches includes the shift to more transient gene-editing approaches. Here the entire gene-editing or prime-editing STU could be introduced into cells, in the form of IVT-RNA, using either conventional lipofection or the more novel delivery methods such as clinically-approved lipid nanoparticles. The use of IVT-RNA has numerous advantages including (i) the reduced risk of integration of gene-editing DNA cargo into the host genome, (ii) the highly transient nature of the anticipated gene-editing events, and (iii) the reduced off-target effects as a result of therapy transience. Chemical modification of gRNA molecules to improve gene-editing efficiency, as well as chemical modification of RNA molecules to permit gymnotic delivery have been attempted previously (Yin *et al.*, 2017; Stein *et al.*, 2010; González-Barriga *et al.*, 2017), however no study has combined the two approaches. Therefore, in future, a much-advanced version of the ribozyme-separated STU would be a chemically-modified IVT gene-editing or prime-editing STU, capable of mediating specific genome alterations with or without the aid of a delivery vehicle.

In summary, this work adds to the wider efforts to improve gene-editing approaches, contributing findings supporting safer and more transient gene-editing technologies. Moreover, it presents evidence for the *in cellulo* expression of a pancRNA molecule potentially implicated in the development of IPF, which had only previously been predicted *in silico* to be present in the lung.

References

Aalipour, A., Dudley, J. C., Park, S. M., Murty, S., Chabon, J. J., Boyle, E. A., Diehn, M. and Gambhir, S. S. (2018) 'Deactivated CRISPR Associated Protein 9 for Minor-Allele Enrichment in Cell-Free DNA', *Clin Chem,* 64(2), pp. 307-316.

Accurso, F. J., Rowe, S. M., Clancy, J. P., Boyle, M. P., Dunitz, J. M., Durie, P. R., Sagel, S. D., Hornick, D. B., Konstan, M. W., Donaldson, S. H., Moss, R. B., Pilewski, J. M., Rubenstein, R. C., Uluer, A. Z., Aitken, M. L., Freedman, S. D., Rose, L. M., Mayer-Hamblett, N., Dong, Q., Zha, J., Stone, A. J., Olson, E. R., Ordoñez, C. L., Campbell, P. W., Ashlock, M. A. and Ramsey, B. W. (2010) 'Effect of VX-770 in persons with cystic fibrosis and the G551D-CFTR mutation', *N Engl J Med,* 363(21), pp. 1991-2003.

Adikusuma, F., Pfitzner, C. and Thomas, P. Q. (2017) 'Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression', *PloS one,* 12(12), pp. e0187236-e0187236.

Adli, M. (2018) 'The CRISPR tool kit for genome editing and beyond', *Nature Communications,* 9(1), pp. 1911.

Aitken, M. L., Moss, R. B., Waltz, D. A., Dovey, M. E., Tonelli, M. R., McNamara, S. C., Gibson, R. L., Ramsey, B. W., Carter, B. J. and Reynolds, T. C. (2001) 'A Phase I Study of Aerosolized Administration of tgAAVCF to Cystic Fibrosis Subjects with Mild Lung Disease', *Human Gene Therapy,* 12(15), pp. 1907-1916.

Alapati, D., Zacharias, W. J., Hartman, H. A., Rossidis, A. C., Stratigis, J. D., Ahn, N. J., Coons, B., Zhou, S., Li, H., Singh, K., Katzen, J., Tomer, Y., Chadwick, A. C., Musunuru, K., Beers, M. F., Morrisey, E. E. and Peranteau, W. H. (2019) 'In utero gene editing for monogenic lung disease', *Science Translational Medicine,* 11(488), pp. eaav8375.

Alder, J. K., Chen, J. J., Lancaster, L., Danoff, S., Su, S. C., Cogan, J. D., Vulto, I., Xie, M., Qi, X., Tuder, R. M., Phillips, J. A., 3rd, Lansdorp, P. M., Loyd, J. E. and Armanios, M. Y. (2008) 'Short telomeres are a risk factor for idiopathic pulmonary fibrosis', *Proc Natl Acad Sci U S A,* 105(35), pp. 13051-6.

Allen, F., Crepaldi, L., Alsinet, C., Strong, A. J., Kleshchevnikov, V., De Angeli, P., Páleníková, P., Khodak, A., Kiselev, V., Kosicki, M., Bassett, A. R., Harding, H., Galanty, Y., Muñoz-Martínez, F., Metzakopian, E., Jackson, S. P. and Parts, L. (2019) 'Predicting the mutations generated by repair of Cas9-induced double-strand breaks', *Nature Biotechnology,* 37(1), pp. 64-72.

Amaral MD, Pacheco P, Beck S. (2001). Cystic fibrosis patients with the 3272-26A>G splicing mutation have milder disease than F508del homozygotes: a large European studyJournal of Medical Genetics 2001;38:777-783

Álvarez MM, Biayna J, Supek F (2022). TP53-dependent toxicity of CRISPR/Cas9 cuts is differential across genomic loci and can confound genetic screening. *Nat Commun*.13(1):4520

Anderson, E. M., Haupt, A., Schiel, J. A., Chou, E., Machado, H. B., Strezoska, Ž., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A. and Smith, A. (2015) 'Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity', *J Biotechnol,* 211, pp. 56-65.
Anne M. Akkerman-Nijland, Onno W. Akkerman, Floris Grasmeijer, Paul Hagedoorn, Henderik. W. Frijlink, Bart. L. Rottier, Gerard. H. Koppelman & Daniel. J. Touw (2021) The pharmacokinetics of antibiotics in cystic fibrosis, *Expert Opinion on Drug Metabolism & Toxicology*, 17:1, 53-68

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A. and Liu, D. R. (2019) 'Search-and-replace genome editing without double-strand breaks or donor DNA', *Nature,* 576(7785), pp. 149-157.

Aravind L, Koonin EV (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res*.11(8):1365-74.

Auerbach, C., Robson, J. M. and Carr, J. G. (1947) 'The Chemical Production of Mutations', *Science,* 105(2723), pp. 243-7.

Azzam, T., Eliyahu, H., Shapira, L., Linial, M., Barenholz, Y. and Domb, A. J. (2002) 'Polysaccharide−Oligoamine Based Conjugates for Gene Delivery', *Journal of Medicinal Chemistry,* 45(9), pp. 1817-1824.

Barbezier, N., Canino, G., Rodor, J., Jobet, E., Saez-Vasquez, J., Marchfelder, A. and Echeverría, M. (2009) 'Processing of a dicistronic tRNA-snoRNA precursor: combined analysis in vitro and in vivo reveals alternate pathways and coupling to assembly of snoRNP', *Plant Physiol,* 150(3), pp. 1598-610.

Bednarski, C., Tomczak, K., vom Hövel, B., Weber, W.-M. and Cathomen, T. (2016) 'Targeted Integration of a Super-Exon into the CFTR Locus Leads to Functional Correction of a Cystic Fibrosis Cell Line Model', *PLOS ONE,* 11(8), pp. e0161072.

Benjaminsen, R. V., Mattebjerg, M. A., Henriksen, J. R., Moghimi, S. M. and Andresen, T. L. (2013) 'The Possible "Proton Sponge " Effect of Polyethylenimine (PEI) Does Not Include Change in Lysosomal pH', *Molecular Therapy,* 21(1), pp. 149-157.

Benoit, R. M., Wilhelm, R. N., Scherer-Becker, D., & Ostermeier, C. (2006). An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids. *Protein expression and purification*, 45(1), 66–71.

Bente, H., Mittelsten Scheid, O. and Donà, M. (2020) 'Versatile in vitro assay to recognize Cas9-induced mutations', *Plant Direct,* 4(9), pp. e00269.

Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K. and Carroll, D. (2006) 'Efficient gene targeting in Drosophila with zinc-finger nucleases', *Genetics,* 172(4), pp. 2391-2403.

Beumer, K. J., Trautman, J. K., Bozas, A., Liu, J.-L., Rutter, J., Gall, J. G. and Carroll, D. (2008) 'Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases', *Proceedings of the National Academy of Sciences of the United States of America,* 105(50), pp. 19821-19826.

Bibikova, M., Beumer, K., Trautman, J. K. and Carroll, D. (2003) 'Enhancing gene targeting with designed zinc finger nucleases', *Science,* 300(5620), pp. 764.

Bibikova, M., Golic, M., Golic, K. G. and Carroll, D. (2002) 'Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases', *Genetics,* 161(3), pp. 1169-1175.

Bitinaite, J., Wah, D. A., Aggarwal, A. K. and Schildkraut, I. (1998) 'FokI dimerization is required for DNA cleavage', *Proceedings of the National Academy of Sciences of the United States of America,* 95(18), pp. 10570-10575.

Bjursell, M., Porritt, M. J., Ericson, E., Taheri-Ghahfarokhi, A., Clausen, M., Magnusson, L., Admyre, T., Nitsch, R., Mayr, L., Aasehaug, L., Seeliger, F., Maresca, M., Bohlooly-Y, M. and Wiseman, J. (2018)

'Therapeutic Genome Editing With CRISPR/Cas9 in a Humanized Mouse Model Ameliorates α1-antitrypsin Deficiency Phenotype', *EBioMedicine,* 29, pp. 104-111.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) 'Breaking the code of DNA binding specificity of TAL-type III effectors', *Science,* 326(5959), pp. 1509-12.

Bolhassani, A., Khavari, A. and Oraf, Z. (2014) 'Electroporation – Advantages and Drawbacks for Delivery of Drug, Gene and Vaccine'.

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S. D. (2005) 'Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin', *Microbiology,* 151(Pt 8), pp. 2551-61.

Borie, R., Le Guen, P., Ghanem, M., Taillé, C., Dupin, C., Dieudé, P., Kannengiesser, C. and Crestani, B. (2019) 'The genetics of interstitial lung diseases', *European Respiratory Review,* 28(153), pp. 190053.

Boussif, O., Lezoualc'h, F., Zanta, M. A., Mergny, M. D., Scherman, D., Demeneix, B. and Behr, J. P. (1995) 'A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine', *Proceedings of the National Academy of Sciences of the United States of America,* 92(16), pp. 7297-7301.

Bradford, J. and Perrin, D. (2019) 'Improving CRISPR guide design with consensus approaches', *BMC Genomics,* 20(9), pp. 931.

Brandsma, I. and Gent, D., 2012. Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integrity*, 3(1), p.9.

Briner, A.E. et al. (2014) “Guide RNA functional modules direct cas9 activity and orthogonality,” *Molecular Cell,* 56(2), pp. 333–339.

Brinkman, E. K., Chen, T., Amendola, M. and van Steensel, B. (2014) 'Easy quantitative assessment of genome editing by sequence trace decomposition', *Nucleic acids research,* 42(22), pp. e168-e168.

Bruegmann, T., Deecke, K. and Fladung, M. (2019) 'Evaluating the Efficiency of gRNAs in CRISPR/Cas9 Mediated Genome Editing in Poplars', *International journal of molecular sciences,* 20(15), pp. 3623.

Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V. M., Gregory, P., Urnov, F., Weinstock, D. M. and Jasin, M. (2009) 'Chromosomal translocations induced at specified loci in human stem cells', *Proceedings of the National Academy of Sciences of the United States of America,* 106(26), pp. 10620-10625.

Bryksin, A. V., & Matsumura, I. (2010). Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques,* 48(6), 463–465. https://doi.org/10.2144/000113418

Carroll, D. (2011) 'Genome engineering with zinc-finger nucleases', *Genetics,* 188(4), pp. 773-782.
Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J. and Voytas, D. F. (2011) 'Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting', *Nucleic acids research,* 39(12), pp. e82-e82.

Chen, F., Alphonse, M. and Liu, Q. (2020) 'Strategies for nonviral nanoparticle-based delivery of CRISPR/Cas9 therapeutics', *WIREs Nanomedicine and Nanobiotechnology,* 12(3), pp. e1609.

Chen, G., Ribeiro, C. M. P., Sun, L., Okuda, K., Kato, T., Gilmore, R. C., Martino, M. B., Dang, H., Abzhanova, A., Lin, J. M., Hull-Ryde, E. A., Volmer, A. S., Randell, S. H., Livraghi-Butrico, A., Deng, Y., Scherer, P. E., Stripp, B. R., O'Neal, W. K. and Boucher, R. C. (2019) 'XBP1S Regulates MUC5B in a Promoter Variant-Dependent Pathway in Idiopathic Pulmonary Fibrosis Airway Epithelia', *American journal of respiratory and critical care medicine,* 200(2), pp. 220-234.

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A. and Doudna, J. A. (2017) 'Enhanced proofreading governs CRISPR-Cas9 targeting accuracy', *Nature,* 550(7676), pp. 407-410.

Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., O'Riordan, C. R. and Smith, A. E. (1990) 'Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis', *Cell,* 63(4), pp. 827-34.

Cheng, W.-J., Chen, L.-C., Ho, H.-O., Lin, H.-L. and Sheu, M.-T. (2018) 'Stearyl polyethylenimine complexed with plasmids as the core of human serum albumin nanoparticles noncovalently bound to CRISPR/Cas9 plasmids or siRNA for disrupting or silencing PD-L1 expression for immunotherapy', *International journal of nanomedicine,* 13, pp. 7079-7094.

Cho, S. W., Kim, S., Kim, J. M. and Kim, J. S. (2013) 'Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease', *Nat Biotechnol,* 31(3), pp. 230-2.

Choi, P. S. and Meyerson, M. (2014) 'Targeted genomic rearrangements using CRISPR/Cas technology', *Nat Commun,* 5, pp. 3728.

Choi, W., Choe, S. and Lau, G. W. (2020) 'Inactivation of FOXA2 by Respiratory Bacterial Pathogens and Dysregulation of Pulmonary Mucus Homeostasis', *Frontiers in Immunology,* 11.

Chong, Z. X., Yeap, S. K. and Ho, W. Y. (2021) 'Transfection types, methods and strategies: a technical review', *PeerJ,* 9, pp. e11165-e11165.

Ciobanasu, C., Siebrasse, J. P. and Kubitscheck, U. (2010) 'Cell-Penetrating HIV1 TAT Peptides Can Generate Pores in Model Membranes', *Biophysical Journal,* 99(1), pp. 153-162.

Ciuffi, A. (2008) 'Mechanisms governing lentivirus integration site selection', *Curr Gene Ther,* 8(6), pp. 419-29.

Colosimo, A., Goncz, K. K., Holmes, A. R., Kunzelmann, K., Novelli, G., Malone, R. W., Bennett, M. J. and Gruenert, D. C. (2000) 'Transfer and expression of foreign genes in mammalian cells', *Biotechniques,* 29(2), pp. 314-8, 320-2, 324 passim.

Concordet, J.-P. and Haeussler, M. (2018) 'CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens', *Nucleic Acids Research,* 46(W1), pp. W242-W245.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. and Zhang, F. (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science (New York, N.Y.),* 339(6121), pp. 819-823.

Connett, G. J. (2019) 'Lumacaftor-ivacaftor in the treatment of cystic fibrosis: design, development and place in therapy', *Drug design, development and therapy,* 13, pp. 2405-2412.

Cox, C. M., Mandell, E. K., Stewart, L., Lu, R., Johnson, D. L., McCarter, S. D., Tavares, A., Runyan, R., Ghosh, S. and Wilson, J. M. (2015) 'Endosomal regulation of contact inhibition through the AMOT:YAP pathway', *Molecular Biology of the Cell,* 26(14), pp. 2673-2684.

Cox, D. B. T., Platt, R. J. and Zhang, F. (2015) 'Therapeutic genome editing: prospects and challenges', *Nature medicine,* 21(2), pp. 121-131.

Crane, A. M., Kramer, P., Bui, J. H., Chung, W. J., Li, X. S., Gonzalez-Garay, M. L., Hawkins, F., Liao, W., Mora, D., Choi, S., Wang, J., Sun, H. C., Paschon, D. E., Guschin, D. Y., Gregory, P. D., Kotton, D. N., Holmes, M. C., Sorscher, E. J. and Davis, B. R. (2015) 'Targeted correction and restored function of the CFTR gene in cystic fibrosis induced pluripotent stem cells', *Stem cell reports,* 4(4), pp. 569-577.

Crystal, R. G. (2014) 'Adenovirus: the first effective in vivo gene delivery vector', *Human gene therapy,* 25(1), pp. 3-11.

Cutting, G. R. (2015) 'Cystic fibrosis genetics: from molecular understanding to clinical application', *Nature reviews. Genetics,* 16(1), pp. 45-56.

Damdindorj, L., Karnan, S., Ota, A., Hossain, E., Konishi, Y., Hosokawa, Y. and Konishi, H. (2014) 'A comparative analysis of constitutive promoters located in adeno-associated viral vectors', *PLoS One,* 9(8), pp. e106472.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. and Charpentier, E. (2011) 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III', *Nature,* 471(7340), pp. 602-607.

Dever, D. P., Bak, R. O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C. E., Pavel-Dinu, M., Saxena, N., Wilkens, A. B., Mantri, S., Uchida, N., Hendel, A., Narla, A., Majeti, R., Weinberg, K. I. and Porteus, M. H. (2016) 'CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells', *Nature,* 539(7629), pp. 384-389.

Di Stazio, M. et al. (2021) “Systematic analysis of factors that improve homologous direct repair (HDR) efficiency in CRISPR/cas9 technique,” *PLOS ONE*, 16(3).

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J. and Root, D. E. (2016) 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', *Nature biotechnology,* 34(2), pp. 184-191.

Dong, F., Xie, K., Chen, Y., Yang, Y. and Mao, Y. (2017) 'Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells', *Biochemical and biophysical research communications,* 482(4), pp. 889-895.

Drews, K., Calgi, M. P., Harrison, W. C., Drews, C. M., Costa-Pinheiro, P., Shaw, J. J. P., Jobe, K. A., Nelson, E. A., Han, J. D., Fox, T., White, J. M. and Kester, M. (2019) 'Glucosylceramidase Maintains Influenza Virus Infection by Regulating Endocytosis', *Journal of virology,* 93(12), pp. e00017-19.

Drummond, D. A. and Wilke, C. O. (2009) 'The evolutionary consequences of erroneous protein synthesis', *Nature reviews. Genetics,* 10(10), pp. 715-724.

Duan, D., Yue, Y., Yan, Z., McCray, P. B. and Engelhardt, J. F. (1998) 'Polarity Influences the Efficiency of Recombinant Adenoassociated Virus Infection in Differentiated Airway Epithelia', *Human Gene Therapy,* 9(18), pp. 2761-2776.

Duchesneau, P., Besla, R., Derouet, M. F., Guo, L., Karoubi, G., Silberberg, A., Wong, A. P., & Waddell, T. K. (2017). Partial Restoration of CFTR Function in cftr-Null Mice following Targeted Cell Replacement Therapy. *Molecular therapy : the journal of the American Society of Gene Therapy,* 25(3), 654–665.

Duncan, G. A., Kim, N., Colon-Cortes, Y., Rodriguez, J., Mazur, M., Birket, S. E., Rowe, S. M., West, N. E., Livraghi-Butrico, A., Boucher, R. C., Hanes, J., Aslanidi, G. and Suk, J. S. (2018) 'An Adeno-Associated Viral Vector Capable of Penetrating the Mucus Barrier to Inhaled Gene Therapy', *Molecular therapy. Methods & clinical development,* 9, pp. 296-304.

Edgeworth, D. et al. (2015) “Exercise improvements in ivacaftor treated G551D cystic fibrosis patients are not solely related to FEV1 and sweat changes,” 7.3 Cystic Fibrosis [Preprint].

Ehrke-Schulz, E., Schiwon, M., Leitner, T., Dávid, S., Bergmann, T., Liu, J. and Ehrhardt, A. (2017) 'CRISPR/Cas9 delivery with one single adenoviral vector devoid of all viral genes', *Scientific Reports,* 7(1), pp. 17113.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. and Jasin, M. (1998) 'Gene conversion tracts from double-strand break repair in mammalian cells', *Mol Cell Biol,* 18(1), pp. 93-101.

Fan, Z. et al. (2018) "A sheep model of cystic fibrosis generated by CRISPR/Cas9 disruption of the CFTR gene," *JCI Insight,* 3(19).

Fazil, M. H. U. T., Ong, S. T., Chalasani, M. L. S., Low, J. H., Kizhakeyil, A., Mamidi, A., Lim, C. F. H., Wright, G. D., Lakshminarayanan, R., Kelleher, D. and Verma, N. K. (2016) 'GapmeR cellular internalization by macropinocytosis induces sequence-specific gene silencing in human primary T-cells', *Scientific reports,* 6, pp. 37721-37721.

Fernández Fabrellas, E., Peris Sánchez, R., Sabater Abad, C. and Juan Samper, G. (2018) 'Prognosis and Follow-Up of Idiopathic Pulmonary Fibrosis', *Medical sciences (Basel, Switzerland),* 6(2), pp. 51.

Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. and Shendure, J. (2014) 'Saturation editing of genomic regions by multiplex homology-directed repair', *Nature,* 513(7516), pp. 120-123.

Firth, A. L., Menon, T., Parker, G. S., Qualls, S. J., Lewis, B. M., Ke, E., Dargitz, C. T., Wright, R., Khanna, A., Gage, F. H. and Verma, I. M. (2015) 'Functional Gene Correction for Cystic Fibrosis in Lung Epithelial Cells Generated from Patient iPSCs', *Cell Rep,* 12(9), pp. 1385-90.

Fisher, K. J., Choi, H., Burda, J., Chen, S.-J. and Wilson, J. M. (1996) 'Recombinant Adenovirus Deleted of All Viral Genes for Gene Therapy of Cystic Fibrosis', *Virology,* 217(1), pp. 11-22.

Flanagan, W. M., Su, L. L. and Wagner, R. W. (1996) 'Elucidation of gene function using C-5 propyne antisense oligonucleotides', *Nature Biotechnology,* 14(9), pp. 1139-1145.

Fleischer, A., Vallejo-Díez, S., Martín-Fernández, J. M., Sánchez-Gilabert, A., Castresana, M., del Pozo, A., Esquisabel, A., Ávila, S., Castrillo, J. L., Gaínza, E., Pedraz, J. L., Viñas, M. and Bachiller, D. (2020)

'iPSC-Derived Intestinal Organoids from Cystic Fibrosis Patients Acquire CFTR Activity upon TALEN-Mediated Repair of the p.F508del Mutation', *Molecular Therapy - Methods & Clinical Development,* 17, pp. 858-870.

Flotte, T., Carter, B., Conrad, C., Guggino, W., Reynolds, T., Rosenstein, B., Taylor, G., Walden, S. and Wetzel, R. (1996) 'A Phase I Study of an Adeno-Associated Virus-CFTR Gene Vector in Adult CF Patients with Mild Lung Disease. Johns Hopkins Children's Center, Baltimore, Maryland', *Human Gene Therapy,* 7(9), pp. 1145-1159.

Foster, K. A., Oster, C. G., Mayer, M. M., Avery, M. L. and Audus, K. L. (1998) 'Characterization of the A549 Cell Line as a Type II Pulmonary Epithelial Cell Model for Drug Metabolism', *Experimental Cell Research,* 243(2), pp. 359-366.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., Gonzalez, J. M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O. G., Lagarde, J., Martin, F. J., Martínez, L., Mohanan, S., Muir, P., Navarro, F. C. P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B. M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J. S., Gerstein, M., Guigó, R., Hubbard, T. J. P., Kellis, M., Paten, B., Reymond, A., Tress, M. L. and Flicek, P. (2019) 'GENCODE reference annotation for the human and mouse genomes', *Nucleic acids research,* 47(D1), pp. D766-D773.

Frit, P., Barboule, N., Yuan, Y., Gomez, D. and Calsou, P. (2014) 'Alternative end-joining pathway(s): bricolage at DNA breaks', *DNA Repair (Amst),* 17, pp. 81-97.

Fujimoto, H., Kobayashi, T. and Azuma, A. (2016) 'Idiopathic Pulmonary Fibrosis: Treatment and Prognosis', *Clinical medicine insights. Circulatory, respiratory and pulmonary medicine,* 9(Suppl 1), pp. 179-185.

Gao, Y., Chuai, G., Yu, W., Qu, S. and Liu, Q. (2020) 'Data imbalance in CRISPR off-target prediction', *Brief Bioinform,* 21(4), pp. 1448-1454.

Gao, Y. and Zhao, Y. (2014) 'Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing', *Journal of Integrative Plant Biology,* 56(4), pp. 343-349.

Gao, Z., Herrera-Carrillo, E. and Berkhout, B. (2018) 'Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III', *Molecular therapy. Nucleic acids,* 10, pp. 36-44.

Gaspar, H. B., Cooray, S., Gilmour, K. C., Parsley, K. L., Adams, S., Howe, S. J., Al Ghonaium, A., Bayford, J., Brown, L., Davies, E. G., Kinnon, C. and Thrasher, A. J. (2011) 'Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency', *Sci Transl Med,* 3(97), pp. 97ra79.

Gaspar, H. B., Parsley, K. L., Howe, S., King, D., Gilmour, K. C., Sinclair, J., Brouns, G., Schmidt, M., Von Kalle, C., Barington, T., Jakobsen, M. A., Christensen, H. O., Al Ghonaium, A., White, H. N., Smith, J. L., Levinsky, R. J., Ali, R. R., Kinnon, C. and Thrasher, A. J. (2004) 'Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector', *Lancet,* 364(9452), pp. 2181-7.

Geurts, M.H. et al. (2021) "Evaluating CRISPR-based prime editing for cancer modeling and CFTR repair in organoids," *Life Science Alliance,* 4(10).

Ghosh, R., Gilda, J. E. and Gomes, A. V. (2014) 'The necessity of and strategies for improving confidence in the accuracy of western blots', *Expert review of proteomics,* 11(5), pp. 549-560.

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M. and Weissman, J. S. (2014) 'Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation', *Cell,* 159(3), pp. 647-661.

Ginzinger, D. G. (2002) 'Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream', *Exp Hematol,* 30(6), pp. 503-12.

Gong, H., Liu, M., Klomp, J., Merrill, B. J., Rehman, J. and Malik, A. B. (2017) 'Method for Dual Viral Vector Mediated CRISPR-Cas9 Gene Disruption in Primary Human Endothelial Cells', *Scientific reports,* 7, pp. 42127-42127.

González-Barriga, A., Nillessen, B., Kranzen, J., van Kessel, I. D. G., Croes, H. J. E., Aguilera, B., de Visser, P. C., Datson, N. A., Mulders, S. A. M., van Deutekom, J. C. T., Wieringa, B. and Wansink, D. G. (2017)

'Intracellular Distribution and Nuclear Activity of Antisense Oligonucleotides After Unassisted Uptake in Myoblasts and Differentiated Myotubes In Vitro', *Nucleic acid therapeutics,* 27(3), pp. 144-158.

Gray, Steven & Foti, Stacey & Schwartz, Joel & Bachaboina, Lavanya & Taylor-Blake, Bonnie & Coleman, Jennifer & Ehlers, Michael & Zylka, Mark & Mccown, Thomas & Samulski, Richard. (2011). Optimizing Promoters for Recombinant Adeno-Associated Virus-Mediated Gene Expression in the Peripheral and Central Nervous System Using Self-Complementary Vectors. *Human gene therapy*. 22. 1143-53.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G. L., White, F. F. and Yin, Z. (2005) 'R gene expression induced by a type-III effector triggers disease resistance in rice', *Nature,* 435(7045), pp. 1122-5.

Guan, Y., Ma, Y., Li, Q., Sun, Z., Ma, L., Wu, L., Wang, L., Zeng, L., Shao, Y., Chen, Y., Ma, N., Lu, W., Hu, K., Han, H., Yu, Y., Huang, Y., Liu, M. and Li, D. (2016) 'CRISPR/Cas9-mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse', *EMBO molecular medicine,* 8(5), pp. 477-488.

Hacein-Bey-Abina, S., Hauer, J., Lim, A., Picard, C., Wang, G. P., Berry, C. C., Martinache, C., Rieux-Laucat, F., Latour, S., Belohradsky, B. H., Leiva, L., Sorensen, R., Debré, M., Casanova, J. L., Blanche, S., Durandy, A., Bushman, F. D., Fischer, A. and Cavazzana-Calvo, M. (2010) 'Efficacy of gene therapy for X-linked severe combined immunodeficiency', *The New England journal of medicine,* 363(4), pp. 355-364.

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J. P., Thrasher, A. J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., Fischer, A., Davies, E. G., Kuis, W., Leiva, L. and Cavazzana-Calvo, M. (2002) 'Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy', *N Engl J Med,* 346(16), pp. 1185-93.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J.-S. and Concordet, J.-P. (2016) 'Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR', *Genome biology,* 17(1), pp. 148-148.

Hammann, C., Luptak, A., Perreault, J. and de la Peña, M. (2012) 'The ubiquitous hammerhead ribozyme', *RNA (New York, N.Y.),* 18(5), pp. 871-885.

Hancock, L. A., Hennessy, C. E., Solomon, G. M., Dobrinskikh, E., Estrella, A., Hara, N., Hill, D. B., Kissner, W. J., Markovetz, M. R., Grove Villalon, D. E., Voss, M. E., Tearney, G. J., Carroll, K. S., Shi, Y., Schwarz, M. I., Thelin, W. R., Rowe, S. M., Yang, I. V., Evans, C. M. and Schwartz, D. A. (2018) 'Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice', *Nature Communications,* 9(1), pp. 5363.

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science (New York, N.Y.),* 329(5997), 1355–1358.

Haurwitz, R. E., Sternberg, S. H., & Doudna, J. A. (2012). Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *The EMBO journal,* 31(12), 2824–2832.

Hawkins, F., Kramer, P., Jacob, A., Driver, I., Thomas, D. C., McCauley, K. B., Skvir, N., Crane, A. M., Kurmann, A. A., Hollenberg, A. N., Nguyen, S., Wong, B. G., Khalil, A. S., Huang, S. X. L., Guttentag, S., Rock, J. R., Shannon, J. M., Davis, B. R. and Kotton, D. N. (2017) 'Prospective isolation of NKX2-1–expressing human lung progenitors derived from pluripotent stem cells', *The Journal of Clinical Investigation,* 127(6), pp. 2277-2294.

Helling, B. A., Gerber, A. N., Kadiyala, V., Sasse, S. K., Pedersen, B. S., Sparks, L., Nakano, Y., Okamoto, T., Evans, C. M., Yang, I. V. and Schwartz, D. A. (2017) 'Regulation of MUC5B Expression in Idiopathic Pulmonary Fibrosis', *American journal of respiratory cell and molecular biology,* 57(1), pp. 91-99.

Henke, M.O. et al. (2004) "MUC5AC and MUC5B mucins are decreased in cystic fibrosis airway secretions," *American Journal of Respiratory Cell and Molecular Biology*, 31(1), pp. 86–91.

Herbers, K., Conradsstrauch, J. and Bonas, U. (1992) 'RACE-SPECIFICITY OF PLANT-RESISTANCE TO BACTERIAL SPOT DISEASE DETERMINED BY REPETITIVE MOTIFS IN A BACTERIAL AVIRULENCE PROTEIN', *Nature,* 356(6365), pp. 172-174.

Hobbs, B. D., Putman, R. K., Araki, T., Nishino, M., Gudmundsson, G., Gudnason, V., Eiriksdottir, G., Zilhao Nogueira, N. R., Dupuis, J., Xu, H., O'Connor, G. T., Manichaikul, A., Nguyen, J., Podolanczuk, A. J., Madahar, P., Rotter, J. I., Lederer, D. J., Barr, R. G., Rich, S. S., Ampleford, E. J., Ortega, V. E., Peters,

S. P., O'Neal, W. K., Newell, J. D., Jr., Bleecker, E. R., Meyers, D. A., Allen, R. J., Oldham, J. M., Ma, S.-F., Noth, I., Jenkins, R. G., Maher, T. M., Hubbard, R. B., Wain, L. V., Fingerlin, T. E., Schwartz, D. A., Washko, G. R., Rosas, I. O., Silverman, E. K., Hatabu, H., Cho, M. H. and Hunninghake, G. M. (2019) 'Overlap of Genetic Risk between Interstitial Lung Abnormalities and Idiopathic Pulmonary Fibrosis', *American journal of respiratory and critical care medicine,* 200(11), pp. 1402-1413.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R. C., Katibah, G. E., Amora, R., Boydston, E. A., Zeitler, B., Meng, X., Miller, J. C., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D. and Jaenisch, R. (2009) 'Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases', *Nature biotechnology,* 27(9), pp. 851-857.

Hollywood, J. A., Lee, C. M., Scallan, M. F. and Harrison, P. T. (2016) 'Analysis of gene repair tracts from Cas9/gRNA double-stranded breaks in the human CFTR gene', *Scientific Reports,* 6(1), pp. 32230.

Hommelsheim, C. M., Frantzeskakis, L., Huang, M. and Ülker, B. (2014) 'PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications', *Scientific Reports,* 4(1), pp. 5052.

Horibe, T., Torisawa, A., Akiyoshi, R., Hatta-Ohashi, Y., Suzuki, H. and Kawakami, K. (2014) 'Transfection efficiency of normal and cancer cell lines and monitoring of promoter activity by single-cell bioluminescence imaging', *Luminescence,* 29(1), pp. 96-100.

Horii, T., Arai, Y., Yamazaki, M., Morita, S., Kimura, M., Itoh, M., Abe, Y. and Hatada, I. (2014) 'Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering', *Scientific reports,* 4, pp. 4513-4513.

Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., Fields, A. P., Park, C. Y., Corn, J. E., Kampmann, M. and Weissman, J. S. (2016) 'Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation', *eLife,* 5, pp. e19760.

Hryhorowicz, M., Grzeskowiak, B., Mazurkiewicz, N., Sledzinski, P., Lipinski, D. and Slomski, R. (2019) 'Improved Delivery of CRISPR/Cas9 System Using Magnetic Nanoparticles into Porcine Fibroblast', *Mol Biotechnol,* 61(3), pp. 173-180.

Hryhorowicz, M., Lipiński, D., Zeyland, J. and Słomski, R. (2017) 'CRISPR/Cas9 Immune System as a Tool for Genome Engineering', *Archivum Immunologiae et Therapiae Experimentalis,* 65(3), pp. 233-240.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G. and Zhang, F. (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', *Nature Biotechnology,* 31(9), pp. 827-832.

Hunninghake, G. M., Hatabu, H., Okajima, Y., Gao, W., Dupuis, J., Latourelle, J. C., Nishino, M., Araki, T., Zazueta, O. E., Kurugol, S., Ross, J. C., San José Estépar, R., Murphy, E., Steele, M. P., Loyd, J. E., Schwarz, M. I., Fingerlin, T. E., Rosas, I. O., Washko, G. R., O'Connor, G. T. and Schwartz, D. A. (2013) 'MUC5B promoter polymorphism and interstitial lung abnormalities', *N Engl J Med,* 368(23), pp. 2192-200.

Idiopathic Pulmonary Fibrosis Clinical Research, N., Raghu, G., Anstrom, K. J., King, T. E., Jr., Lasky, J. A. and Martinez, F. J. (2012) 'Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis', *The New England journal of medicine,* 366(21), pp. 1968-1977.

'Idiopathic Pulmonary Fibrosis: Diagnosis and Treatment', (2000) *American Journal of Respiratory and Critical Care Medicine,* 161(2), pp. 646-664.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) 'Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product', *Journal of bacteriology,* 169(12), pp. 5429-5433.

Jacob, A., Morley, M., Hawkins, F., McCauley, K. B., Jean, J. C., Heins, H., Na, C.-L., Weaver, T. E., Vedaie, M., Hurley, K., Hinds, A., Russo, S. J., Kook, S., Zacharias, W., Ochs, M., Traber, K., Quinton, L. J., Crane, A., Davis, B. R., White, F. V., Wambach, J., Whitsett, J. A., Cole, F. S., Morrisey, E. E., Guttentag, S. H., Beers, M. F. and Kotton, D. N. (2017) 'Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells', *Cell stem cell,* 21(4), pp. 472-488.e10.

Janas, M. M., Zlatev, I., Liu, J., Jiang, Y., Barros, S. A., Sutherland, J. E., Davis, W. P., Liu, J., Brown, C. R., Liu, X., Schlegel, M. K., Blair, L., Zhang, X., Das, B., Tran, C., Aluri, K., Li, J., Agarwal, S., Indrakanti, R., Charisse, K., Nair, J., Matsuda, S., Rajeev, K. G., Zimmermann, T., Sepp-Lorenzino, L., Xu, Y., Akinc, A., Fitzgerald, K., Vaishnaw, A. K., Smith, P. F., Manoharan, M., Jadhav, V., Wu, J.-T. and Maier, M. A.

(2019) 'Safety evaluation of 2'-deoxy-2'-fluoro nucleotides in GalNAc-siRNA conjugates', *Nucleic acids research,* 47(7), pp. 3306-3320.

Jia, J., Bai, F., Jin, Y., Santostefano, K. E., Ha, U.-H., Wu, D., Wu, W., Terada, N. and Jin, S. (2015) 'Efficient Gene Editing in Pluripotent Stem Cells by Bacterial Injection of Transcription Activator-Like Effector Nuclease Proteins', *STEM CELLS Translational Medicine,* 4(8), pp. 913-926.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science (New York, N.Y.),* 337(6096), pp. 816-821.

Joung, J. K. and Sander, J. D. (2013) 'TALENs: a widely applicable technology for targeted genome editing', *Nature reviews. Molecular cell biology,* 14(1), pp. 49-55.

Ju, A., Lee, S. W., Lee, Y. E., Han, K.-C., Kim, J.-C., Shin, S. C., Park, H. J., EunKyeong Kim, E., Hong, S. and Jang, M. (2019) 'A carrier-free multiplexed gene editing system applicable for suspension cells', *Biomaterials,* 217, pp. 119298.

Kalva, S., Boeke, J. D. and Mita, P. (2018) 'Gibson Deletion: a novel application of isothermal in vitro recombination', *Biological procedures online,* 20, pp. 2-2.

Kapur, M. and Ackerman, S. L. (2018) 'mRNA Translation Gone Awry: Translation Fidelity and Neurological Disease', *Trends in genetics : TIG,* 34(3), pp. 218-231.

Karikó, K., Buckstein, M., Ni, H. and Weissman, D. (2005) 'Suppression of RNA Recognition by Toll-like Receptors: The Impact of Nucleoside Modification and the Evolutionary Origin of RNA', *Immunity,* 23(2), pp. 165-175.

Kay, M. A., Glorioso, J. C. and Naldini, L. (2001) 'Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics', *Nature Medicine,* 7(1), pp. 33-40.

Kay, S. and Bonas, U. (2009) 'How Xanthomonas type III effectors manipulate the host plant', *Curr Opin Microbiol,* 12(1), pp. 37-43.

Kay, S., Hahn, S., Marois, E., Hause, G. and Bonas, U. (2007) 'A Bacterial Effector Acts as a Plant Transcription Factor and Induces a Cell Size Regulator', *Science,* 318(5850), pp. 648.

Kikuchi, Y., Sasaki-Tozawa, N. and Suzuki, K. (1993) 'Artificial self-cleaving molecules consisting of a tRNA precursor and the catalytic RNA of RNase P', *Nucleic acids research,* 21(20), pp. 4685-4689.

Kedmi R, Ben-Arie N, Peer D. (2010) 'The systemic toxicity of positively charged lipid nanoparticles and the role of Toll-like receptor 4 in immune activation'. *Biomaterials,* 31(26):6867-75.

Kim, N., Duncan, G. A., Hanes, J. and Suk, J. S. (2016) 'Barriers to inhaled gene therapy of obstructive lung diseases: A review', *Journal of controlled release : official journal of the Controlled Release Society,* 240, pp. 465-488.

Kim, Y. G., Cha, J. and Chandrasegaran, S. (1996) 'Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain', *Proceedings of the National Academy of Sciences of the United States of America,* 93(3), pp. 1156-1160.

King, T. E. and Lee, J. S. (2022) 'Cryptogenic Organizing Pneumonia', *New England Journal of Medicine,* 386(11), pp. 1058-1069.

King, J., Nichols, A., Bentley, S., Carr, S. and Davies, J., 2022. An Update on CFTR Modulators as New Therapies for Cystic Fibrosis. *Pediatric Drugs,* 24(4), pp.321-333.

Konstantakos, V., Nentidis, A., Krithara, A. and Paliouras, G. (2022) 'CRISPR-Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning', *Nucleic acids research,* 50(7), pp. 3616-3637.

Kouranova, E., Forbes, K., Zhao, G., Warren, J., Bartels, A., Wu, Y. and Cui, X. (2016) 'CRISPRs for optimal targeting: Delivery of CRISPR components as DNA, RNA and protein into cultured cells and single-cell embryos', *Human Gene Therapy,* 27.

Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S., Marchfelder, A. and Echeverría, M. (2003) 'Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z', *The EMBO journal,* 22(3), pp. 621-632.

Kuchipudi SV, Tellabati M, Nelli RK, White GA, Perez BB, Sebastian S, Slomka MJ, Brookes SM, Brown IH, Dunham SP, Chang KC., (2012) '18S rRNA is a reliable normalisation gene for real time PCR based on influenza virus infected cells', *Virol J*; 9:230.

Kuhn, A., Ackermann, M., Mussolino, C., Cathomen, T., Lachmann, N. and Moritz, T. (2017) 'TALEN-mediated functional correction of human iPSC-derived macrophages in context of hereditary pulmonary alveolar proteinosis', *Scientific Reports,* 7(1), pp. 15195.

Kumar, M. and Moschos, S. A. (2017) 'Oligonucleotide Therapies for the Lung: Ready to Return to the Clinic?', *Molecular Therapy,* 25(12), pp. 2604-2606.

Küppers, M., Ittrich, C., Faust, D. and Dietrich, C. (2010) 'The transcriptional programme of contact-inhibition', *Journal of Cellular Biochemistry,* 110(5), pp. 1234-1243.

Kurkulos, M., Weinberg, J. M., Roy, D. and Mount, S. M. (1994) 'P element-mediated in vivo deletion analysis of white-apricot: deletions between direct repeats are strongly favored', *Genetics,* 136(3), pp. 1001-1011.

Kwon, J. S., Park, S. H., Baek, J. H., Dung, T. M., Kim, S. W., Min, B. H., Kim, J. H. and Kim, M. S. (2016) 'Gene Transfection of Human Turbinate Mesenchymal Stromal Cells Derived from Human Inferior Turbinate Tissues', *Stem Cells International,* 2016, pp. 4735264.

Lahdaoui, F., Messager, M., Vincent, A., Hec, F., Gandon, A., Warlaumont, M., Renaud, F., Leteurtre, E., Piessen, G., Jonckheere, N., Mariette, C. and Van Seuningen, I. (2017) 'Depletion of MUC5B mucin in gastrointestinal cancer cells alters their tumorigenic properties: implication of the Wnt/β-catenin pathway', *Biochemical Journal,* 474(22), pp. 3733-3746.

Lazrak, A., Fu, L., Bali, V., Bartoszewski, R., Rab, A., Havasi, V., Keiles, S., Kappes, J., Kumar, R., Lefkowitz, E., Sorscher, E. J., Matalon, S., Collawn, J. F. and Bebok, Z. (2013) 'The silent codon change I507-ATC->ATT contributes to the severity of the ΔF508 CFTR channel dysfunction', *FASEB journal : official publication of the Federation of American Societies for Experimental Biology,* 27(11), pp. 4630-4645.

Leclercq, S., Gilbert, C. and Cordaux, R. (2012) 'Cargo capacity of phages and plasmids and other factors influencing horizontal transfers of prokaryote transposable elements', *Mobile genetic elements,* 2(2), pp. 115-118.

Lee, C. M., Flynn, R., Hollywood, J. A., Scallan, M. F. and Harrison, P. T. (2012a) 'Correction of the ΔF508 Mutation in the Cystic Fibrosis Transmembrane Conductance Regulator Gene by Zinc-Finger Nuclease Homology-Directed Repair', *BioResearch open access,* 1(3), pp. 99-108.

Lee, C. S., Bishop, E. S., Zhang, R., Yu, X., Farina, E. M., Yan, S., Zhao, C., Zheng, Z., Shu, Y., Wu, X., Lei, J., Li, Y., Zhang, W., Yang, C., Wu, K., Wu, Y., Ho, S., Athiviraham, A., Lee, M. J., Wolf, J. M., Reid, R. R. and He, T.-C. (2017a) 'Adenovirus-Mediated Gene Delivery: Potential Applications for Gene and Cell-Based Therapies in the New Era of Personalized Medicine', *Genes & diseases,* 4(2), pp. 43-63.

Lee, H. J., Kim, E. and Kim, J.-S. (2010) 'Targeted chromosomal deletions in human cells using zinc finger nucleases', *Genome research,* 20(1), pp. 81-89.

Lee, H. J., Kweon, J., Kim, E., Kim, S. and Kim, J.-S. (2012b) 'Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases', *Genome research,* 22(3), pp. 539-548.

Lee, K., Conboy, M., Park, H. M., Jiang, F., Kim, H. J., Dewitt, M. A., Mackley, V. A., Chang, K., Rao, A., Skinner, C., Shobha, T., Mehdipour, M., Liu, H., Huang, W.-C., Lan, F., Bray, N. L., Li, S., Corn, J. E., Kataoka, K., Doudna, J. A., Conboy, I. and Murthy, N. (2017b) 'Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair', *Nature biomedical engineering,* 1, pp. 889-901.

Ley, B., Collard, H. R. and King, T. E. (2011) 'Clinical Course and Prediction of Survival in Idiopathic Pulmonary Fibrosis', *American Journal of Respiratory and Critical Care Medicine,* 183(4), pp. 431-440.
Li, J., Yao, Q. and Liu, D. (2011) 'Hydrodynamic cell delivery for simultaneous establishment of tumor growth in mouse lung, liver and kidney', *Cancer Biology & Therapy,* 12(8), pp. 737-741.

Li, L., Wu, L. P. and Chandrasegaran, S. (1992) 'Functional domains in Fok I restriction endonuclease', *Proceedings of the National Academy of Sciences of the United States of America,* 89(10), pp. 4275-4279.

Li, B. et al. (2019) “A QPCR method for genome editing efficiency determination and single-cell clone screening in human cells,” *Scientific Reports*, 9(1).

Li, N., Cooney, A. L., Zhang, W., Ehrhardt, A. and Sinn, P. L. (2019) 'Enhanced Tropism of Species B1 Adenoviral-Based Vectors for Primary Human Airway Epithelial Cells', *Molecular Therapy - Methods & Clinical Development,* 14, pp. 228-236.

Lin, S., Staahl, B. T., Alla, R. K. and Doudna, J. A. (2014) 'Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery', *Elife,* 3, pp. e04766.

Lino, C. A., Harper, J. C., Carney, J. P. and Timlin, J. A. (2018) 'Delivering CRISPR: a review of the challenges and approaches', *Drug delivery,* 25(1), pp. 1234-1257.

Liou, T. G, Adler, FR, Cox, DR., Cahill BC. (2007) Lung transplantation and survival in children with cystic fibrosis. *N Engl J Med*;357:2143–2152.

Liu, G., Zhang, Y. and Zhang, T. (2020) 'Computational approaches for effective CRISPR guide RNA design and evaluation', *Computational and Structural Biotechnology Journal,* 18, pp. 35-44.

Liu, J., Chang, J., Jiang, Y., Meng, X., Sun, T., Mao, L., Xu, Q. and Wang, M. (2019) 'Fast and Efficient CRISPR/Cas9 Genome Editing In Vivo Enabled by Bioreducible Lipid and Messenger RNA Nanoparticles', *Advanced materials (Deerfield Beach, Fla.),* 31(33), pp. e1902575-e1902575.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods,* 25(4), pp. 402-8.

Loenarz, C., Sekirnik, R., Thalhammer, A., Ge, W., Spivakovsky, E., Mackeen, M. M., McDonough, M. A., Cockman, M. E., Kessler, B. M., Ratcliffe, P. J., Wolf, A. and Schofield, C. J. (2014) 'Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy', *Proceedings of the National Academy of Sciences of the United States of America,* 111(11), pp. 4019-4024.

Lombardo, A., Genovese, P., Beausejour, C., Colleoni, S., Lee, Y.-L., Kim, K., Ando, D., Urnov, F., Galli, C., Gregory, P., Holmes, M. and Naldini, L. (2007) 'Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery', *Nature biotechnology,* 25, pp. 1298-306.

Lopes-Pacheco, M. (2016a) 'CFTR Modulators: Shedding Light on Precision Medicine for Cystic Fibrosis', *Front Pharmacol,* 7, pp. 275.

Lopes-Pacheco, M. (2016b) 'CFTR Modulators: Shedding Light on Precision Medicine for Cystic Fibrosis', *Frontiers in Pharmacology,* 7(275).

Lu, B., Javidi-Parsijani, P., Makani, V., Mehraein-Ghomi, F., Sarhan, W. M., Sun, D., Yoo, K. W., Atala, Z. P., Lyu, P. and Atala, A. (2019) 'Delivering SaCas9 mRNA by lentivirus-like bionanoparticles for transient expression and efficient genome editing', *Nucleic acids research,* 47(8), pp. e44-e44.

Lu, Z. H., Dmitriev, I. P., Brough, D. E., Kashentseva, E. A., Li, J. and Curiel, D. T. (2020) 'A New Gorilla Adenoviral Vector with Natural Lung Tropism Avoids Liver Toxicity and Is Amenable to Capsid Engineering and Vector Retargeting', *Journal of Virology,* 94(10), pp. e00265-20.

Lundstrom, K. (2004) 'Gene Therapy Applications of Viral Vectors', *Technology in Cancer Research & Treatment,* 3(5), pp. 467-477.

Lyu, P., Javidi-Parsijani, P., Atala, A. and Lu, B. (2019) 'Delivering Cas9/sgRNA ribonucleoprotein (RNP) by lentiviral capsid-based bionanoparticles for efficient 'hit-and-run' genome editing', *Nucleic Acids Research,* 47(17), pp. e99-e99.

Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., Townsend, J. A., Unger-Wallace, E., Sander, J. D., Müller-Lerch, F., Fu, F., Pearlberg, J., Göbel, C., Dassie, Justin P., Pruett-Miller, S. M., Porteus, M. H., Sgroi, D. C., Iafrate, A. J., Dobbs, D., McCray, P. B., Cathomen, T., Voytas, D. F. and Joung, J. K. (2008) 'Rapid "Open-Source" Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification', *Molecular Cell,* 31(2), pp. 294-301.

Maeder, M. L., Thibodeau-Beganny, S., Sander, J. D., Voytas, D. F. and Joung, J. K. (2009) 'Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays', *Nature protocols,* 4(10), pp. 1471-1501.

Maes, T., Tournoy, K. G. and Joos, G. F. (2011) 'Gene therapy for allergic airway diseases', *Curr Allergy Asthma Rep,* 11(2), pp. 163-72.

Maitra, M., Wang, Y., Gerard, R. D., Mendelson, C. R. and Garcia, C. K. (2010) 'Surfactant protein A2 mutations associated with pulmonary fibrosis lead to protein instability and endoplasmic reticulum stress', *The Journal of biological chemistry,* 285(29), pp. 22103-22113.

Makarova, K. S. and Koonin, E. V. (2015) 'Annotation and Classification of CRISPR-Cas Systems', *Methods in molecular biology (Clifton, N.J.),* 1311, pp. 47-75.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L. and Church, G. M. (2013a) 'CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering', *Nature biotechnology,* 31(9), pp. 833-838.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013b) 'RNA-guided human genome engineering via Cas9', *Science (New York, N.Y.),* 339(6121), pp. 823-826.

Mangera, Z., Panesar, G. and Makker, H. (2012) 'Practical approach to management of respiratory complications in neurological disorders', *Int J Gen Med,* 5, pp. 255-63.

Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988) 'Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes', *Nature,* 336(6197), pp. 348-52.

Marconi, P., Argnani, R., Berto, E., Epstein, A. and Manservigi, R. (2009) 'HSV as a Vector in Vaccine Development and Gene Therapy', *Human vaccines,* 4, pp. 91-105.

Marshall, E. (1999) 'Gene therapy death prompts review of adenovirus vector', *Science,* 286(5448), pp. 2244-5.

Martin, C., Hamard, C., Kanaan, R., Boussaud, V., Grenet, D., Abély, M., Hubert, D., Munck, A., Lemonnier, L. and Burgel, P.-R. (2016) 'Causes of death in French cystic fibrosis patients: The need for improvement in transplantation referral strategies!', *Journal of Cystic Fibrosis,* 15(2), pp. 204-212.

Martínez-Gil, L., Goff, P. H., Hai, R., García-Sastre, A., Shaw, M. L. and Palese, P. (2013) 'A Sendai virus-derived RNA agonist of RIG-I as a virus vaccine adjuvant', *Journal of virology,* 87(3), pp. 1290-1300.

Martinez-Pitre, P. J., Sabbula, B. R. and Cascella, M. (2022) 'Restrictive Lung Disease', *StatPearls*. Treasure Island (FL): StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC.

Martufi, M., Good, R. B., Rapiteanu, R., Schmidt, T., Patili, E., Tvermosegaard, K., New, M., Nanthakumar, C. B., Betts, J., Blanchard, A. D. and Maratou, K. (2019) 'Single-Step, High-Efficiency CRISPR-Cas9 Genome Editing in Primary Human Disease-Derived Fibroblasts', *The CRISPR Journal,* 2(1), pp. 31-40.

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R. and Ploegh, H. L. (2015) 'Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining', *Nature biotechnology,* 33(5), pp. 538-542.

Maurisse, R., De Semir, D., Emamekhoo, H., Bedayat, B., Abdolmohammadi, A., Parsi, H. and Gruenert, D. C. (2010) 'Comparative transfection of DNA into primary and transformed mammalian cells from different lineages', *BMC Biotechnology,* 10(1), pp. 9.

Mehta, A. and Merkel, O. M. (2020) 'Immunogenicity of Cas9 Protein', *Journal of pharmaceutical sciences,* 109(1), pp. 62-67.

Miah, K., Hyde, S. and Gill, D., 2019. Emerging gene therapies for cystic fibrosis. *Expert Review of Respiratory Medicine,* 13(8), pp.709-725.

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O. and Rebar, E. J. (2007) 'An improved zinc-finger nuclease architecture for highly specific genome editing', *Nat Biotechnol,* 25(7), pp. 778-85.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D. and Rebar, E. J. (2011) 'A TALE nuclease architecture for efficient genome editing', *Nat Biotechnol,* 29(2), pp. 143-8.

Milone, M. C. and O'Doherty, U. (2018) 'Clinical use of lentiviral vectors', *Leukemia,* 32(7), pp. 1529-1541.

Miyagawa, Y., Marino, P., Verlengia, G., Uchida, H., Goins, W. F., Yokota, S., Geller, D. A., Yoshida, O., Mester, J., Cohen, J. B. and Glorioso, J. C. (2015) 'Herpes simplex viral-vector design for efficient transduction of nonneuronal cells without cytotoxicity', *Proceedings of the National Academy of Sciences,* 112(13), pp. E1632.

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. and Soria, E. (2005) 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements', *J Mol Evol,* 60(2), pp. 174-82.

Mojica, F. J., Juez, G. and Rodriguez-Valera, F. (1993) 'Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites', *Mol Microbiol,* 9(3), pp. 613-21.

Molinski, S. et al. (2012) "Functional rescue of F508del-CFTR using small molecule correctors," *Frontiers in Pharmacology*, 3.

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. and Valen, E. (2014) 'CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing', *Nucleic acids research,* 42(Web Server issue), pp. W401-W407.

Moore, L. D., Le, T. and Fan, G. (2013) 'DNA Methylation and Its Basic Function', *Neuropsychopharmacology,* 38(1), pp. 23-38.

Moore, M., Klug, A. and Choo, Y. (2001) 'Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units', *Proc Natl Acad Sci U S A,* 98(4), pp. 1437-41.

Moschos, S. A., Frick, M., Taylor, B., Turnpenny, P., Graves, H., Spink, K. G., Brady, K., Lamb, D., Collins, D., Rockel, T. D., Weber, M., Lazari, O., Perez-Tosar, L., Fancy, S. A., Lapthorn, C., Green, M. X., Evans, S., Selby, M., Jones, G., Jones, L., Kearney, S., Mechiche, H., Gikunju, D., Subramanian, R., Uhlmann, E., Jurk, M., Vollmer, J., Ciaramella, G. and Yeadon, M. (2011) 'Uptake, Efficacy, and Systemic Distribution of Naked, Inhaled Short Interfering RNA (siRNA) and Locked Nucleic Acid (LNA) Antisense', *Molecular Therapy,* 19(12), pp. 2163-2168.

Moscou, M. J. and Bogdanove, A. J. (2009) 'A simple cipher governs DNA recognition by TAL effectors', *Science,* 326(5959), pp. 1501.

Mou, H., Zhao, R., Sherwood, R., Ahfeldt, T., Lapey, A., Wain, J., Sicilian, L., Izvolsky, K., Musunuru, K., Cowan, C. and Rajagopal, J. (2012) 'Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs', *Cell Stem Cell,* 10(4), pp. 385-97.

Mout, R., Ray, M., Yesilbag Tonga, G., Lee, Y.-W., Tay, T., Sasaki, K. and Rotello, V. M. (2017) 'Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing', *ACS Nano,* 11(3), pp. 2452-2458.

Mrad, A. and Huda, N. (2022) 'Acute Interstitial Pneumonia', *StatPearls*. Treasure Island (FL): StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC.

Muller, H. J. (1927) 'ARTIFICIAL TRANSMUTATION OF THE GENE', *Science,* 66(1699), pp. 84-7.

Muniz, L., Nicolas, E. and Trouche, D. (2021) 'RNA polymerase II speed: a key player in controlling and adapting transcriptome composition', *The EMBO Journal,* 40(15), pp. e105740.

Murphy, C. J., Gole, A. M., Stone, J. W., Sisco, P. N., Alkilany, A. M., Goldsmith, E. C. and Baxter, S. C. (2008) 'Gold Nanoparticles in Biology: Beyond Toxicity to Cellular Imaging', *Accounts of Chemical Research,* 41(12), pp. 1721-1730.

Nakano, Y., Yang, I. V., Walts, A. D., Watson, A. M., Helling, B. A., Fletcher, A. A., Lara, A. R., Schwarz, M. I., Evans, C. M. and Schwartz, D. A. (2016) 'MUC5B Promoter Variant rs35705950 Affects MUC5B Expression in the Distal Airways in Idiopathic Pulmonary Fibrosis', *American journal of respiratory and critical care medicine,* 193(4), pp. 464-466.

Nakayama, T., Wu, J., Galvin-Parton, P., Weiss, J., Andriola, M. R., Hill, R. S., Vaughan, D. J., El-Quessny, M., Barry, B. J., Partlow, J. N., Barkovich, A. J., Ling, J. and Mochida, G. H. (2017) 'Deficient activity of alanyl-tRNA synthetase underlies an autosomal recessive syndrome of progressive microcephaly, hypomyelination, and epileptic encephalopathy', *Human mutation,* 38(10), pp. 1348-1354.

Nassif, N., Penney, J., Pal, S., Engels, W. R. and Gloor, G. B. (1994) 'Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair', *Molecular and cellular biology,* 14(3), pp. 1613-1625.

Nayerossadat, N., Maedeh, T. and Ali, P. A. (2012) 'Viral and nonviral delivery systems for gene delivery', *Advanced biomedical research,* 1, pp. 27-27.

Nayfeh, A. S., Chippa, V. and Moore, D. R. (2022) 'Nonspecific Interstitial Pneumonitis', *StatPearls*. Treasure Island (FL): StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC.

Neatu, R., Thompson, D., Enekwa, I., Schwalbe, E., Fois, G., Frick, M., Braubach, P. and Moschos, S. (2022) 'The Idiopathic Pulmonary Fibrosis-associated single nucleotide polymorphism rs35705950 is transcribed in a MUC5B Promoter Associated Long Non-Coding RNA (AC061979.1)'.

Nissim-Rafinia, M., Aviram, M., Randell, S. H., Shushi, L., Ozeri, E., Chiba-Falek, O., Eidelman, O., Pollard, H. B., Yankaskas, J. R. and Kerem, B. (2004) 'Restoration of the cystic fibrosis transmembrane conductance regulator function by splicing modulation', *EMBO reports,* 5(11), pp. 1071-1077.

Nishimasu, H., Ishitani, R. and Nureki, O. (2014) "Crystal structure of streptococcus pyogenes cas9 in complex with guide RNA and Target DNA."

Noth, I., Zhang, Y., Ma, S.-F., Flores, C., Barber, M., Huang, Y., Broderick, S. M., Wade, M. S., Hysi, P., Scuirba, J., Richards, T. J., Juan-Guardela, B. M., Vij, R., Han, M. K., Martinez, F. J., Kossen, K., Seiwert, S. D., Christie, J. D., Nicolae, D., Kaminski, N. and Garcia, J. G. N. (2013) 'Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study', *The Lancet. Respiratory medicine,* 1(4), pp. 309-317.

Nowak, C. M., Lawson, S., Zerez, M., & Bleris, L. (2016). Guide RNA engineering for versatile Cas9 functionality. *Nucleic acids research,* 44(20), 9555–9564.

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W. and Doudna, J. A. (2014) 'Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity', *Nature Structural & Molecular Biology,* 21(6), pp. 528-534.

Nurmemmedov, E. et al. (2007) “Biophysics of viral infectivity: Matching genome length with capsid size,” *Quarterly Reviews of Biophysics*, 40(4), pp. 327–356.

Okamoto, S., Amaishi, Y., Maki, I., Enoki, T. and Mineno, J. (2019) 'Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs', *Scientific Reports,* 9(1), pp. 4811.

Ordovás, L., Boon, R., Pistoni, M., Chen, Y., Wolfs, E., Guo, W., Sambathkumar, R., Bobis-Wozowicz, S., Helsen, N., Vanhove, J., Berckmans, P., Cai, Q., Vanuytsel, K., Eggermont, K., Vanslembrouck, V., Schmidt, B. Z., Raitano, S., Van Den Bosch, L., Nahmias, Y., Cathomen, T., Struys, T. and Verfaillie, C. M. (2015) 'Efficient Recombinase-Mediated Cassette Exchange in hPSCs to Study the Hepatocyte Lineage Reveals AAVS1 Locus-Mediated Transgene Inhibition', *Stem cell reports,* 5(5), pp. 918-931.

Parker, A. L., Newman, C., Briggs, S., Seymour, L. and Sheridan, P. J. (2003) 'Nonviral gene delivery: techniques and implications for molecular medicine', *Expert Rev Mol Med,* 5(22), pp. 1-15.

Pavletich, N. P. and Pabo, C. O. (1991) 'Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A', *Science,* 252(5007), pp. 809-17.

Phase 3 study of VX-445 combination therapy in subjects with cystic fibrosis heterozygous for the F508del mutation and a minimal function mutation (F/MF) - full text view (2020) Full Text View - ClinicalTrials.gov. Available at: https://clinicaltrials.gov/ct2/show/study/NCT03525444 (Accessed: October 27, 2022).

Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) 'CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies', *Microbiology,* 151(Pt 3), pp. 653-63.

Qin, W., Dion, S. L., Kutny, P. M., Zhang, Y., Cheng, A. W., Jillette, N. L., Malhotra, A., Geurts, A. M., Chen, Y.-G. and Wang, H. (2015) 'Efficient CRISPR/Cas9-Mediated Genome Editing in Mice by Zygote Electroporation of Nuclease', *Genetics,* 200(2), pp. 423-430.

Qiu, X. Y., Zhu, L. Y., Zhu, C. S., Ma, J. X., Hou, T., Wu, X. M., Xie, S. S., Min, L., Tan, D. A., Zhang, D. Y. and Zhu, L. (2018) 'Highly Effective and Low-Cost MicroRNA Detection with CRISPR-Cas9', *ACS Synth Biol,* 7(3), pp. 807-813.

Raikwar, S., Raikwar, A., Chaurasia, S. and Mohan, R., 2016. Gene editing for corneal disease management. *World Journal of Translational Medicine*, 5(1), p.1.

Rauluseviciute, I., Drabløs, F. and Rye, M. B. (2020) 'DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation', *BMC Medical Genomics,* 13(1), pp. 6.

Redel, B. K., Beaton, B. P., Spate, L. D., Benne, J. A., Murphy, S. L., O'Gorman, C. W., Spate, A. M., Prather, R. S. and Wells, K. D. (2018) 'Single step production of Cas9 mRNA for zygote injection', *BioTechniques,* 64(3), pp. 118-124.

Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D. and Joung, J. K. (2012) 'FLASH assembly of TALENs for high-throughput genome editing', *Nature biotechnology,* 30(5), pp. 460-465.

Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. and Corn, J. E. (2016) 'Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA', *Nat Biotechnol,* 34(3), pp. 339-44.

Roche, P. J. R., Gytz, H., Hussain, F., Cameron, C. J. F., Paquette, D., Blanchette, M., Dostie, J., Nagar, B. and Akavia, U. D. (2018a) 'Efficient Homology Directed Repair by Cas9:Donor Localization and Cationic Polymeric Transfection in Mammalian Cells', *bioRxiv*, pp. 248179.
Roche, P. J. R., Gytz, H., Hussain, F., Cameron, C. J. F., Paquette, D., Blanchette, M., Dostie, J., Nagar, B. and Akavia, U. D. (2018b) 'Homology Directed Repair by Cas9:Donor Co-localization in Mammalian Cells', *bioRxiv*, pp. 248179.

Rosenfeld, M. A., Yoshimura, K., Trapnell, B. C., Yoneyama, K., Rosenthal, E. R., Dalemans, W., Fukayama, M., Bargon, J., Stier, L. E., Stratford-Perricaudet, L., Perricaudet, M., Guggino, W. B., Pavirani, A., Lecocq, J.-P. and Crystal, R. G. (1992) 'In vivo transfer of the human cystic fibrosis transmembrane conductance regulator gene to the airway epithelium', *Cell,* 68(1), pp. 143-155.

Rothstein, R. J. (1983) 'One-step gene disruption in yeast', *Methods Enzymol,* 101, pp. 202-11.

Rouet P, Smih F, Jasin M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, 14(12):8096-106.

Ruan, J. et al. (2019) “Efficient gene editing at major CFTR mutation loci,” *Molecular Therapy - Nucleic Acids,* 16, pp. 73–81.

Ryu, N., Kim, M. A., Park, D., Lee, B., Kim, Y. R., Kim, K. H., Baek, J. I., Kim, W. J., Lee, K. Y. and Kim, U. K. (2018) 'Effective PEI-mediated delivery of CRISPR-Cas9 complex for targeted gene therapy', *Nanomedicine,* 14(7), pp. 2095-2102.

Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J. and Gale, M., Jr. (2008) 'Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA', *Nature,* 454(7203), pp. 523-527.

Saketkoo, L. A., Russell, A.-M., Jensen, K., Mandizha, J., Tavee, J., Newton, J., Rivera, F., Howie, M., Reese, R., Goodman, M., Hart, P., Strookappe, B., De Vries, J., Rosenbach, M., Scholand, M. B., Lammi, M. R., Elfferich, M., Lower, E., Baughman, R. P., Sweiss, N., Judson, M. A. and Drent, M. (2021) 'Health-Related Quality of Life (HRQoL) in Sarcoidosis: Diagnosis, Management, and Health Outcomes', *Diagnostics (Basel, Switzerland),* 11(6), pp. 1089.

Scherer, S. and Davis, R. W. (1979) 'Replacement of chromosome segments with altered DNA sequences constructed in vitro', *Proceedings of the National Academy of Sciences of the United States of America,* 76(10), pp. 4951-4955.

Schwank, G., Koo, B. K., Sasselli, V., Dekkers, J. F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C. K., Nieuwenhuis, E. E., Beekman, J. M. and Clevers, H. (2013) 'Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients', *Cell Stem Cell,* 13(6), pp. 653-8.

Seibold, M. A., Wise, A. L., Speer, M. C., Steele, M. P., Brown, K. K., Loyd, J. E., Fingerlin, T. E., Zhang, W., Gudmundsson, G., Groshong, S. D., Evans, C. M., Garantziotis, S., Adler, K. B., Dickey, B. F., du Bois, R. M., Yang, I. V., Herron, A., Kervitsky, D., Talbert, J. L., Markin, C., Park, J., Crews, A. L., Slifer, S. H., Auerbach, S., Roy, M. G., Lin, J., Hennessy, C. E., Schwarz, M. I. and Schwartz, D. A. (2011) 'A common MUC5B promoter polymorphism and pulmonary fibrosis', *The New England journal of medicine,* 364(16), pp. 1503-1512.

Seidner M (2020). Modulator Treatments for Cystic Fibrosis: Effectiveness and Value Final Evidence Report and Meeting Summary, *Institute for clinical and economic review.*

Seiler, J., Breinig, M., Caudron-Herger, M., Polycarpou-Schwarz, M., Boutros, M. and Diederichs, S. (2017) 'The lncRNA VELUCT strongly regulates viability of lung cancer cells despite its extremely low abundance', *Nucleic Acids Res,* 45(9), pp. 5458-5469.

Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P. and Pruett-Miller, S. M. (2018) 'A Survey of Validation Strategies for CRISPR-Cas9 Editing', *Scientific reports,* 8(1), pp. 888-888.

Shah, K., Bentley, E., Tyler, A., Richards, K. S. R., Wright, E., Easterbrook, L., Lee, D., Cleaver, C., Usher, L., Burton, J. E., Pitman, J. K., Bruce, C. B., Edge, D., Lee, M., Nazareth, N., Norwood, D. A. and Moschos, S. A. (2017) 'Field-deployable, quantitative, rapid identification of active Ebola virus infection in unprocessed blood', *Chemical science,* 8(11), pp. 7780-7797.

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelson, T., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G. and Zhang, F. (2014) 'Genome-scale CRISPR-Cas9 knockout screening in human cells', *Science (New York, N.Y.),* 343(6166), pp. 84-87.

Sharifi Tabar, M., Hesaraki, M., Esfandiari, F., Sahraneshin Samani, F., Vakilian, H. and Baharvand, H. (2015) 'Evaluating Electroporation and Lipofectamine Approaches for Transient and Stable Transgene Expressions in Human Fibroblasts and Embryonic Stem Cells', *Cell journal,* 17(3), pp. 438-450.

Shei, R.-J., Peabody, J. E., Kaza, N. and Rowe, S. M. (2018) 'The epithelial sodium channel (ENaC) as a therapeutic target for cystic fibrosis', *Current opinion in pharmacology,* 43, pp. 152-165.

Shen, X., Beasley, S., Putman, J. N., Li, Y., Prakash, T. P., Rigo, F., Napierala, M. and Corey, D. R. (2019) 'Efficient electroporation of neuronal cells using synthetic oligonucleotides: identifying duplex RNA and antisense oligonucleotide activators of human frataxin expression', *Rna,* 25(9), pp. 1118-1129.

Shen, Y., Cohen, J. L., Nicoloro, S. M., Kelly, M., Yenilmez, B., Henriques, F., Tsagkaraki, E., Edwards, Y. J. K., Hu, X., Friedline, R. H., Kim, J. K. and Czech, M. P. (2018) 'CRISPR-delivery particles targeting nuclear receptor-interacting protein 1 (Nrip1) in adipose cells to enhance energy expenditure', *The Journal of biological chemistry,* 293(44), pp. 17291-17305.

Shrock, E. and Güell, M. (2017) 'Chapter Six - CRISPR in Animals and Animal Models', in Torres-Ruiz, R. and Rodriguez-Perales, S. (eds.) *Progress in Molecular Biology and Translational Science*: Academic Press, pp. 95-114.

Simsek, D., Brunet, E., Wong, S. Y.-W., Katyal, S., Gao, Y., McKinnon, P. J., Lou, J., Zhang, L., Li, J., Rebar, E. J., Gregory, P. D., Holmes, M. C. and Jasin, M. (2011) 'DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation', *PLoS genetics,* 7(6), pp. e1002080-e1002080.

Soifer, H. S., Koch, T., Lai, J., Hansen, B., Hoeg, A., Oerum, H. and Stein, C. A. (2012) 'Silencing of gene expression by gymnotic delivery of antisense oligonucleotides', *Methods Mol Biol,* 815, pp. 333-46.

Song, J., Cano-Rodriquez, D., Winkle, M., Gjaltema, R. A. F., Goubert, D., Jurkowski, T. P., Heijink, I. H., Rots, M. G. and Hylkema, M. N. (2016) 'Targeted epigenetic editing of SPDEF reduces mucus production in lung epithelial cells', *American Journal of Physiology-Lung Cellular and Molecular Physiology,* 312(3), pp. L334-L347.

Song, Y. K., Liu, F. and Liu, D. (1998) 'Enhanced gene expression in mouse lung by prolonging the retention time of intravenously injected plasmid DNA', *Gene Ther,* 5(11), pp. 1531-7.

Southern, K.W. (1997) “Delta F508 in cystic fibrosis: Willing but not able,” *Archives of Disease in Childhood*, 76(3), pp. 278–282

Speizer, F. E., Horton, S., Batt, J. and Slutsky, A. S. (2006) 'Respiratory Diseases of Adults', in nd, Jamison, D.T., Breman, J.G., Measham, A.R., Alleyne, G., Claeson, M., Evans, D.B., Jha, P., Mills, A. and Musgrove, P. (eds.) *Disease Control Priorities in Developing Countries*. Washington (DC).

Stein, C. A., Hansen, J. B., Lai, J., Wu, S., Voskresenskiy, A., Høg, A., Worm, J., Hedtjärn, M., Souleimanian, N., Miller, P., Soifer, H. S., Castanotto, D., Benimetskaya, L., Ørum, H. and Koch, T. (2010) 'Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents', *Nucleic acids research,* 38(1), pp. e3-e3.

Stephens, C. J., Kashentseva, E., Everett, W., Kaliberova, L. and Curiel, D. T. (2018) 'Targeted in vivo knock-in of human alpha-1-antitrypsin cDNA using adenoviral delivery of CRISPR/Cas9', *Gene Therapy,* 25(2), pp. 139-156.

Sternberg, S. H., LaFrance, B., Kaplan, M. and Doudna, J. A. (2015) 'Conformational control of DNA target cleavage by CRISPR–Cas9', *Nature,* 527(7576), pp. 110-113.

Strachan JB, Dyett BP, Nasa Z, Valery C, Conn CE. (2020) Toxicity and cellular uptake of lipid nanoparticles of different structure and composition. *J Colloid Interface Sci*, 576:241-251

Strutt, S. C., Torrez, R. M., Kaya, E., Negrete, O. A. and Doudna, J. A. (2018) 'RNA-dependent RNA targeting by CRISPR-Cas9', *eLife,* 7, pp. e32724.

Suzuki, S., Sargent, R. G., Illek, B., Fischer, H., Esmaeili-Shandiz, A., Yezzi, M. J., Lee, A., Yang, Y., Kim, S., Renz, P., Qi, Z., Yu, J., Muench, M. O., Beyer, A. I., Guimarães, A. O., Ye, L., Chang, J., Fine, E. J., Cradick, T. J., Bao, G., Rahdar, M., Porteus, M. H., Shuto, T., Kai, H., Kan, Y. W. and Gruenert, D. C. (2016) 'TALENs Facilitate Single-step Seamless SDF Correction of F508del CFTR in Airway Epithelial Submucosal Gland Cell-derived CF-iPSCs', *Molecular Therapy - Nucleic Acids,* 5, pp. e273.

Szwed, M., Torgersen, M. L., Kumari, R. V., Yadava, S. K., Pust, S., Iversen, T. G., Skotland, T., Giri, J. and Sandvig, K. (2020) 'Biological response and cytotoxicity induced by lipid nanocapsules', *Journal of nanobiotechnology,* 18(1), pp. 5-5.

Tang, X., Ren, Q., Yang, L., Bao, Y., Zhong, Z., He, Y., Liu, S., Qi, C., Liu, B., Wang, Y., Sretenovic, S., Zhang, Y., Zheng, X., Zhang, T., Qi, Y. and Zhang, Y. (2019) 'Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing', *Plant biotechnology journal,* 17(7), pp. 1431-1445.

Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., Voytas, Daniel F. and Zhang, Y. (2016) 'A Single Transcript CRISPR-Cas9 System for Efficient Genome Editing in Plants', *Molecular Plant,* 9(7), pp. 1088-1091.

Thabut, G., Christie, J. D., Mal, H., Fournier, M., Brugière, O., Leseche, G., Castier, Y. and Rizopoulos, D. (2013) 'Survival Benefit of Lung Transplant for Cystic Fibrosis since Lung Allocation Score Implementation', *American Journal of Respiratory and Critical Care Medicine,* 187(12), pp. 1335-1340.

Than, B. L. N., Linnekamp, J. F., Starr, T. K., Largaespada, D. A., Rod, A., Zhang, Y., Bruner, V., Abrahante, J., Schumann, A., Luczak, T., Walter, J., Niemczyk, A., O'Sullivan, M. G., Medema, J. P., Fijneman, R. J. A., Meijer, G. A., Van den Broek, E., Hodges, C. A., Scott, P. M., Vermeulen, L. and Cormier, R. T. (2016) 'CFTR is a tumor suppressor gene in murine and human intestinal cancer', *Oncogene,* 35(32), pp. 4179-4187.

Thomas, K. R., Folger, K. R. and Capecchi, M. R. (1986) 'High frequency targeting of genes to specific sites in the mammalian genome', *Cell,* 44(3), pp. 419-28.

Tsai SQ, et al. (2015) 'GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases', *Nature Biotechnology,* 33(2): 187-97

Turowski, T. W. and Tollervey, D. (2016) 'Transcription by RNA polymerase III: insights into mechanism and regulation', *Biochemical Society Transactions,* 44(5), pp. 1367-1375.

UK cystic fibrosis registry (2021). Available at: https://www.cysticfibrosis.org.uk/sites/default/files/2021-12/CF_Annual%20Report%202020.pdf (Accessed: October 22, 2022).

Van der Meijden, C. M., Lapointe, D. S., Luong, M. X., Peric-Hupkes, D., Cho, B., Stein, J. L., van Wijnen, A. J. and Stein, G. S. (2002) 'Gene profiling of cell cycle progression through S-phase reveals sequential expression of genes required for DNA replication and nucleosome assembly', *Cancer Res,* 62(11), pp. 3233-43.

Vistica DT, Skehan P, Scudiero D, Monks A, Pittman A, Boyd MR. (1991) Tetrazolium-based assays for cellular viability: a critical examination of selected parameters affecting formazan production. *Cancer Res,* 51(10):2515-20

Voets, O., Tielen, F., Elstak, E., Benschop, J., Grimbergen, M., Stallen, J., Janssen, R., van Marle, A. and Essrich, C. (2017) 'Highly efficient gene inactivation by adenoviral CRISPR/Cas9 in human primary cells', *PloS one,* 12(8), pp. e0182974-e0182974.

Voltz, J. W., Card, J. W., Carey, M. A., Degraff, L. M., Ferguson, C. D., Flake, G. P., Bonner, J. C., Korach, K. S. and Zeldin, D. C. (2008) 'Male sex hormones exacerbate lung function impairment after bleomycin-induced pulmonary fibrosis', *American journal of respiratory cell and molecular biology,* 39(1), pp. 45-52.

Wagner, J. A., Messner, A. H., Moran, M. L., Daifuku, R., Kouyama, K., Desch, J. K., Manley, S., Norbash, A. M., Conrad, C. K., Friborg, S., Reynolds, T., Guggino, W. B., Moss, R. B., Carter, B. J., Wine, J. J., Flotte, T. R. and Gardner, P. (1999) 'Safety and Biological Efficacy of an Adeno-Associated Virus Vector–Cystic Fibrosis Transmembrane Regulator (AAV-CFTR) in the Cystic Fibrosis Maxillary Sinus', *The Laryngoscope,* 109(2), pp. 266-274.

Wagner, J. A., Moran, M. L., Messner, A. H., Daifuku, R., Conrad, C. K., Reynolds, T., Guggino, W. B., Moss, R. B., Carter, B. J., Wine, J. J., Flotte, T. R. and Gardner, P. (1998a) 'A phase I/II study of tgAAV-CF for the treatment of chronic sinusitis in patients with cystic fibrosis', *Hum Gene Ther,* 9(6), pp. 889-909.

Wagner, J. A., Reynolds, T., Moran, M. L., Moss, R. B., Wine, J. J., Flotte, T. R. and Gardner, P. (1998b) 'Efficient and persistent gene transfer of AAV-CFTR in maxillary sinus', *The Lancet,* 351(9117), pp. 1702-1703.

Wagner, R. W., Matteucci, M. D., Lewis, J. G., Gutierrez, A. J., Moulds, C. and Froehler, B. C. (1993) 'Antisense Gene Inhibition by Oligonucleotides Containing C-5 Propyne Pyrimidines', *Science,* 260(5113), pp. 1510-1513.

Wang, C., Zhuang, Y., Guo, W., Cao, L., Zhang, H., Xu, L., Fan, Y., Zhang, D. and Wang, Y. (2014a) 'Mucin 5B Promoter Polymorphism Is Associated with Susceptibility to Interstitial Lung Diseases in Chinese Males', *PLOS ONE,* 9(8), pp. e104919.

Wang, H.-X., Song, Z., Lao, Y.-H., Xu, X., Gong, J., Cheng, D., Chakraborty, S., Park, J. S., Li, M., Huang, D., Yin, L., Cheng, J. and Leong, K. W. (2018a) 'Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide', *Proceedings of the National Academy of Sciences of the United States of America,* 115(19), pp. 4903-4908.

Wang, L., Li, F., Dang, L., Liang, C., Wang, C., He, B., Liu, J., Li, D., Wu, X., Xu, X., Lu, A. and Zhang, G. (2016) 'In Vivo Delivery Systems for Therapeutic Genome Editing', *International Journal of Molecular Sciences,* 17(5), pp. 626.

Wang, P., Zhang, L., Zheng, W., Cong, L., Guo, Z., Xie, Y., Wang, L., Tang, R., Feng, Q., Hamada, Y., Gonda, K., Hu, Z., Wu, X. and Jiang, X. (2018b) 'Thermo-triggered Release of CRISPR-Cas9 System by Lipid-Encapsulated Gold Nanoparticles for Tumor Therapy', *Angew Chem Int Ed Engl,* 57(6), pp. 1491-1496.

Wang, T., Wei, J. J., Sabatini, D. M. and Lander, E. S. (2014b) 'Genetic screens in human cells using the CRISPR-Cas9 system', *Science (New York, N.Y.),* 343(6166), pp. 80-84.

Wauthoz, N., Bastiat, G., Moysan, E., Cieślak, A., Kondo, K., Zandecki, M., Moal, V., Rousselet, M.-C., Hureaux, J. and Benoit, J.-P. (2015) 'Safe lipid nanocapsule-based gel technology to target lymph nodes and combat mediastinal metastases from an orthotopic non-small-cell lung cancer model in SCID-CB17 mice', *Nanomedicine: Nanotechnology, Biology and Medicine,* 11(5), pp. 1237-1245.

Webb, C.-H. T. and Lupták, A. (2011) 'HDV-like self-cleaving ribozymes', *RNA biology,* 8(5), pp. 719-727.

Wei, Z., Huang, Y., Zhao, D., Hu, Z., Li, Z. and Liang, Z. (2015) 'A Pliable Electroporation Patch (ep-Patch) for Efficient Delivery of Nucleic Acid Molecules into Animal Tissues with Irregular Surface Shapes', *Scientific Reports,* 5(1), pp. 7618.

Wienert, B., Shin, J., Zelin, E., Pestal, K. and Corn, J. E. (2018) 'In vitro-transcribed guide RNAs trigger an innate immune response via the RIG-I pathway', *PLoS biology,* 16(7), pp. e2005840-e2005840.

Wright, A. V., Nunez, J. K. and Doudna, J. A. (2016) 'Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering', *Cell,* 164(1-2), pp. 29-44.

Wilschanski, M. et al. (2003) “Gentamicin-induced correction of CFTR function in patients with cystic fibrosis andcftrstop mutations,” *New England Journal of Medicine,* 349(15), pp. 1433–1441.

Wu, X., Kriz, A. J. and Sharp, P. A. (2014) 'Target specificity of the CRISPR-Cas9 system', *Quantitative biology (Beijing, China),* 2(2), pp. 59-70.

Xia, E., Duan, R., Shi, F., Seigel, K. E., Grasemann, H. and Hu, J. (2018) 'Overcoming the Undesirable CRISPR-Cas9 Expression in Gene Correction', *Molecular therapy. Nucleic acids,* 13, pp. 699-709.
Xia, E., Zhang, Y., Cao, H., Li, J., Duan, R. and Hu, J. (2019) 'TALEN-Mediated Gene Targeting for Cystic Fibrosis-Gene Therapy', *Genes,* 10(1), pp. 39.

Xia, W., Bringmann, P., McClary, J., Jones, P. P., Manzana, W., Zhu, Y., Wang, S., Liu, Y., Harvey, S., Madlansacay, M. R., McLean, K., Rosser, M. P., MacRobbie, J., Olsen, C. L. and Cobb, R. R. (2006) 'High levels of protein expression using different mammalian CMV promoters in several cell lines', *Protein Expr Purif,* 45(1), pp. 115-24.

Xie, K., Minkenberg, B. and Yang, Y. (2015) 'Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system', *Proceedings of the National Academy of Sciences of the United States of America,* 112(11), pp. 3570-3575.

Xie, X., Wang, G., Yang, L., Cheng, T., Gao, J., Wu, Y., & Xia, Q. (2015). Cloning and characterization of a novel Nicotiana tabacum ABC transporter involved in shoot branching. *Physiologia plantarum,* 153(2), 299–306.

Xiong, C., Tang, D. Q., Xie, C. Q., Zhang, L., Xu, K. F., Thompson, W. E., Chou, W., Gibbons, G. H., Chang, L. J., Yang, L. J. and Chen, Y. E. (2005) 'Genetic engineering of human embryonic stem cells with lentiviral vectors', *Stem Cells Dev,* 14(4), pp. 367-77.

Xu, C.L. et al. (2019) "Viral Delivery Systems for CRISPR," *Viruses,* 11(1), p. 28.

Xu, L., Zhao, L., Gao, Y., Xu, J. and Han, R. (2017) 'Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA', *Nucleic acids research,* 45(5), pp. e28-e28.

Xu Tang, Z. Z., Xuelian Zheng, (2017) 'Construction of a Single Transcriptional Unit for Expression of Cas9 and Single-guide RNAs for Genome Editing in Plants ', *bio-protocol,* 7(17).

Xu, X., Gao, D., Wang, P., Chen, J., Ruan, J., Xu, J. and Xia, X. (2018) 'Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation', *Scientific Reports,* 8(1), pp. 11649.

Xu, X. et al. (2018) “Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation,” *Scientific Reports,* 8(1).

Yang, Y., Nunes, F. A., Berencsi, K., Gönczöl, E., Engelhardt, J. F. and Wilson, J. M. (1994) 'Inactivation of E2a in recombinant adenoviruses improves the prospect for gene therapy in cystic fibrosis', *Nature Genetics,* 7(3), pp. 362-369.

Yao, B., Li, J., Huang, H., Sun, C., Wang, Z., Fan, Y., Chang, Q., Li, S. and Xi, J. (2009) 'Quantitative analysis of zeptomole microRNAs based on isothermal ramification amplification', *RNA (New York, N.Y.),* 15(9), pp. 1787-1794.

Yasar, H., Ho, D.-K., De Rossi, C., Herrmann, J., Gordon, S., Loretz, B. and Lehr, C.-M. (2018) 'Starch-Chitosan Polyplexes: A Versatile Carrier System for Anti-Infectives and Gene Delivery', *Polymers,* 10(3), pp. 252.

Ye, L., Wang, J., Beyer, A. I., Teque, F., Cradick, T. J., Qi, Z., Chang, J. C., Bao, G., Muench, M. O., Yu, J., Levy, J. A. and Kan, Y. W. (2014) 'Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection', *Proceedings of the National Academy of Sciences of the United States of America,* 111(26), pp. 9591-9596.

Yin, H., Song, C.-Q., Suresh, S., Wu, Q., Walsh, S., Rhym, L. H., Mintzer, E., Bolukbasi, M. F., Zhu, L. J., Kauffman, K., Mou, H., Oberholzer, A., Ding, J., Kwan, S.-Y., Bogorad, R. L., Zatsepin, T., Koteliansky, V., Wolfe, S. A., Xue, W., Langer, R. and Anderson, D. G. (2017) 'Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing', *Nature Biotechnology,* 35(12), pp. 1179-1187.

Yin, H., Xue, W., Chen, S., Bogorad, R. L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P. A., Jacks, T. and Anderson, D. G. (2014) 'Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype', *Nature biotechnology,* 32(6), pp. 551-553.

Yin, W., Xiang, P. and Li, Q. (2005) “Investigations of the effect of DNA size in transient transfection assay using dual luciferase system,” *Analytical Biochemistry,* 346(2), pp. 289–294

Yoshioka, S., Fujii, W., Ogawa, T., Sugiura, K. and Naito, K. (2015) 'Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells', *Sci Rep,* 5, pp. 18341.

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S. and Qi, L. S. (2015) 'Small molecules enhance CRISPR genome editing in pluripotent stem cells', *Cell Stem Cell,* 16(2), pp. 142-7.

Yu, X., Liang, X., Xie, H., Kumar, S., Ravinder, N., Potter, J., de Mollerat du Jeu, X. and Chesnut, J. D. (2016) 'Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX', *Biotechnology Letters,* 38(6), pp. 919-929.

Zaman, T. and Lee, J. S. (2018) 'Risk factors for the development of idiopathic pulmonary fibrosis: A review', *Current pulmonology reports,* 7(4), pp. 118-125.

Zhang, H., Bahamondez-Canas, T. F., Zhang, Y., Leal, J. and Smyth, H. D. C. (2018) 'PEGylated Chitosan for Nonviral Aerosol and Mucosal Delivery of the CRISPR/Cas9 System in Vitro', *Mol Pharm,* 15(11), pp. 4814-4826.

Zhang, S., Shen, J., Li, D. and Cheng, Y. (2021) 'Strategies in the delivery of Cas9 ribonucleoprotein for CRISPR/Cas9 genome editing', *Theranostics,* 11(2), pp. 614-648.

Zhang, X., Oulad-Abdelghani, M., Zelkin, A. N., Wang, Y., Haikel, Y., Mainard, D., Voegel, J. C., Caruso, F. and Benkirane-Jessel, N. (2010) 'Poly(L-lysine) nanostructured particles for gene delivery and hormone stimulation', *Biomaterials,* 31(7), pp. 1699-706.

Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S. and Yang, S. H. (2015) 'Off-target Effects in CRISPR/Cas9-mediated Genome Engineering', *Mol Ther Nucleic Acids,* 4(11), pp. e264.

Zhuo, C., Zhang, J., Lee, J.-H., Jiao, J., Cheng, D., Liu, L., Kim, H.-W., Tao, Y. and Li, M. (2021) 'Spatiotemporal control of CRISPR/Cas9 gene editing', *Signal Transduction and Targeted Therapy,* 6(1), pp. 238.

Zufferey, R., Dull, T., Mandel, R. J., Bukovsky, A., Quiroz, D., Naldini, L. and Trono, D. (1998) 'Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery', *Journal of virology,* 72(12), pp. 9873-9880.

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., Maeder, M. L., Joung, J. K., Chen, Z. Y. and Liu, D. R. (2015) 'Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo', *Nat Biotechnol,* 33(1), pp. 73-80.