Correction of the ΔF508 mutation in the
CFTR Gene by CRISPR/Cas9 system

Ahmad Mohammad Aldossary

Great Ormond Street Institute of Child Health

University College London

A thesis submitted for the degree of Doctor of Philosophy

July 2018
Declaration

I, Ahmad Aldossary declare that the work presented in this thesis is my own. Any contributions from colleagues in the collaboration or diagrams driven from other sources are explicitly referenced in the text.
Abstract

Cystic Fibrosis (CF) is one of the most common autosomal recessive genetic diseases. It is caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR). CF causes chronic lung disease including thickened mucus, bacterial infection and inflammation, with a progressive loss of pulmonary function and, ultimately, death. Several clinical trials have been performed to date assessing the potential of gene therapy to limit the progression of CF lung disease. However, a clinically relevant treatment has yet to emerge. The major challenges in gene therapy for CF relate to the limited levels of gene transfer achieved in the lung airway epithelium, and the persistence of transgene expression. Here we are investigating the potential of genome editing to develop a genetic therapy for CF using the CRISPR/Cas system that allows for gene-specific, targeted correction of disease-related mutations to be introduced at the chromosomal level. In particular, we aim to investigate the therapeutic potential for ΔF508 mutation cystic fibrosis, which is the most common CF mutation and affects more than 70% of patients.

Initially, multiple guide RNAs were screened for double strand break (DSB), targeting the CFTR gene close to the ΔF508 mutation on the CFBE41o- cell line. The efficient gRNAs were used for the mutation correcting through homology directed repair (HDR) after which the cells were cloned. The correction was confirmed at the molecular level, followed by restoring the electrophysiology function and the mRNA expression. This work was also extended to correct the mutation on primary CFBE cells where the editing was optimized with Cas9 mRNA/gRNA and ribonucleoprotein (RNP) was delivered by Receptor-Targeted Nanocomplex (RTN). To improve the editing efficiency, the homology-independent targeted integration (HITI) strategy as an alternative for HDR, was used to investigate the CFTR exon 10 in CFBE41o- being replaced with wildtype exon. The gene editing in vivo was explored successfully in a mice reporter model to restore tdTomato expression by paired gRNAs excision the stop cassette where targeting nanocomplex was used as a safe and non-immunogenic delivery method.
Impact statement

Cystic fibrosis (CF) is the most common recessively inherited genetic disease. This lethal disease predominantly affects the lungs, and is caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR).

To date, several clinical trials have been performed assessing the potential of gene therapy to limit the progression of CF lung disease; however, a clinically relevant treatment has yet to emerge. The major challenges in gene therapy for CF relate to the limited levels of gene transfer achieved in the lung airway epithelium and the persistence of transgene expression. This work describes advances towards the development of a gene therapy targeting the airway disease in CF patients with ΔF508 mutation, the most common cause of CF using CRISPR/Cas9 gene editing tool.

The data presented in this thesis could be beneficial to multiple fields of research. First of all, the concept of correcting the mutation can extended to other cells harbouring the same mutation or other non-CF mutations. In addition, the in vitro and in vivo reporter models explored here can help for future gene editing optimisation and pre-clinical assessment. Moreover, the targeted nanoparticles show huge potential for a safe and non-immunogenic gene therapy delivery method, not limited to CF.

This study provides fundamental step for developing a new gene therapy for patients suffering from CF, and could help healthcare systems to reduce the cost from the disease complications.

This work has been presented at both national and international conferences. Furthermore, manuscripts of this work are in preparation and are expected to be submitted to journals in the near the future, where we believe it will have impact on the gene editing field.
Acknowledgements

First of all, I thank Allah for giving me strength and ability to complete this project. I would like to express my sincerest and deepest gratitude and thanks for my supervisors Prof. Stephen Hart and Prof. Christopher O'Callaghan for giving me the opportunity to do a PhD at UCL, their guidance and encouragement all the time throughout the PhD. Without their support, infinite wisdom and constant feedback, this PhD would not have been achievable.

I would like to thank for the current and previous postdoc in our group: Aris Tagalakis, Ruhina Maeshima, Ileana Guerrini, Melis Dalbay, Martin Attwood. Special thanks for Aris Tagalakis for his countless suggestions which impacted a lot on my research and helping for in vivo experiments. Also, thanks for Mustafa Munye the previous PhD student in our group for semi-immortalise primary epithelium with BMI-1; proto-oncogene.

Also, I cannot forget to thanks the currant PhD students in our group: Amy Walker, Dania Grant-Serroukh, Preetha Purushothaman, Mahmoud Fassad, William Baird, Afrodit Avgerinou for their friendship and makes my PhD journey is enjoyable. Special thanks for Maximillian Woodall from St George's, University of London for helping in cells electrophysiological studies.

Special thanks to my sponsor King Abdulaziz City for Science and Technology (KACST) and Saudi government for financial support throughout my PhD.

I would like to thank everyone in the Molecular and Cellular Immunology and Genetics and Genomic Medicine sections, both past and present members.

Finally, I would like to thank my parents, my wife and my children: Mais, Mohammad, Moath and Maan for their love and support throughout my life.
Publications


# Contents page

Declaration ........................................................................................................................................... 2  
Abstract ............................................................................................................................................... 3  
Impact statement .............................................................................................................................. 4  
Acknowledgements .......................................................................................................................... 5  
Publications ....................................................................................................................................... 6  
Contents page ..................................................................................................................................... 7  
List of Figures .................................................................................................................................... 14  
List of Tables ...................................................................................................................................... 20  
Abbreviations ..................................................................................................................................... 21  
1. Introduction .................................................................................................................................... 28  
   1.1. Cystic fibrosis ........................................................................................................................... 28  
      1.1.1. History of cystic fibrosis ..................................................................................................... 28  
      1.1.2. CFTR gene and protein ...................................................................................................... 29  
      1.1.3. CFTR mutations .................................................................................................................. 30  
         1.1.3.1. ΔF508 mutation .............................................................................................................. 32  
      1.1.4. Cystic fibrosis pathophysiology ....................................................................................... 33  
   1.2. CFTR and another channels interaction .................................................................................. 35  
      1.2.1. The calcium-activated chloride channels (CaCCs) ............................................................ 36  
      1.2.2. The Epithelial sodium channel (ENaC) ............................................................................ 37  
   1.3. Cystic fibrosis study models ..................................................................................................... 39  
      1.3.1. Cystic fibrosis cells ............................................................................................................. 39  
      1.3.2. Cystic fibrosis animal models ............................................................................................ 42  
   1.4. Cystic fibrosis therapy .............................................................................................................. 43  
      1.4.1. Targeting non-CFTR channels .......................................................................................... 44  
      1.4.2. CF gene therapy ............................................................................................................... 45  
         1.4.2.1. Viral vector .................................................................................................................... 45
1.4.2.2. Non-viral vector .................................................................................................................. 47
1.4.2.3. CF mRNA therapy ............................................................................................................. 49
1.4.3. CF modulator therapies ......................................................................................................... 49
1.4.4. Cystic fibrosis cellular therapy .............................................................................................. 53
1.5. Genetic recombination ............................................................................................................. 55
1.5.1. Non-homologous end joining .............................................................................................. 56
1.5.2. Homologous directed repair ............................................................................................... 58
1.6. Gene editing .............................................................................................................................. 59
1.6.1 Zinc finger endonuclease ...................................................................................................... 60
1.6.2. Transcription activator-like effector nucleases .................................................................. 64
1.6.3. Clustered regularly interspaced palindromic repeats(CRISPR)/ CRISPR associated system (Cas) .......................................................................................................................... 66
1.6.3.1. CRISPR/Cas variants ...................................................................................................... 68
1.6.3.2. CRISPR/Cas9 application .............................................................................................. 71
1.6.3.3. CRISPR/Cas off-targeting ............................................................................................ 74
1.7. Gene editing for cystic fibrosis ............................................................................................... 74
1.8. Barriers for cystic fibrosis gene editing .................................................................................... 76
1.9. Receptor-targeted nanocomplexes (RTN) .............................................................................. 78
1.10. Aims and objectives ............................................................................................................... 81
2. Materials and Methods .............................................................................................................. 84
2.1. Materials ................................................................................................................................. 84
2.1.1. Equipment ........................................................................................................................... 84
2.1.2. Kits and reagents ................................................................................................................ 84
2.1.3. Bacterial Strains .................................................................................................................. 86
2.1.4. Eukaryotic cells .................................................................................................................. 86
2.1.5. Lipids ................................................................................................................................... 86
2.1.6. Composition of Lipid nanoparticle .................................................................................... 87
2.1.7. Peptides .............................................................................................................................. 87
2.1.8. plasmids ................................................................. 87
2.1.9. Taqman Assays (FAM-labelled probes) ........................................... 88
2.1.10. Antibodies for immunoblotting .......................................................... 88
   2.1.10.1. Primary antibodies .................................................................. 88
   2.1.10.2. Secondary antibodies ............................................................... 88
2.1.11. Antibodies for Flow Cytometry .......................................................... 88
2.1.12. Recipe ......................................................................................... 88
2.1.13. primers ......................................................................................... 89
2.1.14. Oligonucleotides for gRNA cloning .................................................. 92
2.1.15. Single strand DNA (ssDNA) ............................................................ 94
2.1.16. Double strand DNA fragment (gBlock) ............................................ 94
2.2. Methods ......................................................................................... 96
   2.2.1. Production of competent bacteria ................................................... 96
   2.2.2. Bacterial transformation ................................................................. 97
   2.2.3. Growth and Maintenance of Bacteria ............................................. 97
   2.2.3. Plasmid DNA extraction ................................................................. 97
      2.2.3.1. Miniprep DNA extraction ......................................................... 97
      2.2.3.2. Maxiprep DNA extraction ......................................................... 98
   2.2.4. DNA extraction ............................................................................ 98
   2.2.5. RNA extraction ............................................................................ 99
   2.2.6. DNA manipulation ....................................................................... 100
      2.2.6.1. Restriction digests .................................................................. 100
      2.2.6.2. Agarose gels electrophoresis ................................................. 100
      2.2.6.3. Gel purification ....................................................................... 100
   2.2.7. Polymerase chain reaction (PCR) .................................................. 101
   2.2.8. PCR purification ........................................................................... 101
   2.2.9. Quantitative real time RT-PCR (qRT-PCR) .................................... 101
   2.2.10. T7 endonuclease I assay .............................................................. 102
2.2.11. PCR-Restriction Fragment Length Polymorphism (RFLP)................................. 102
2.2.12. Transepithelial electrical resistance (TEER) measurements ................................ 102
2.2.13. Ligation of DNA fragments.............................................................................. 102
2.2.14. Ligation of oligonucleotides........................................................................... 103
2.2.15. PCR product cloning......................................................................................... 103
2.2.16. Site directed mutagenesis................................................................................ 103
2.2.17. Sanger sequencing............................................................................................ 104
2.2.18. Gibson assembly cloning.................................................................................. 104
2.2.19. Cell culturing................................................................................................... 104
2.2.20. Collagen coating.............................................................................................. 105
2.2.21. Air-Liquid Interface (ALI) Culture ................................................................... 105
2.2.22. Liquid-liquid interface (LLI) culture................................................................. 105
2.2.23. Flow cytometry for in vitro transfections .......................................................... 105
2.2.24. Biophysical characterisation of RTNs.............................................................. 106
2.2.25. Western blotting............................................................................................. 106
2.2.26. Synthesis of gRNA by in vitro transcription ...................................................... 107
2.2.27. Electrophysiological studies............................................................................. 108
2.2.28. Annealing synthetic gRNA.............................................................................. 109
2.2.29. Receptor targeting nanocomplexes (RTNs) preparation ................................... 109
2.2.30. Complex preparation with lipofectamine 2000 (L2K)...................................... 110
2.2.31. Transfection of submerged cultures................................................................. 110
2.2.32. Transfection of ALI cultures .......................................................................... 110
2.2.33. In vivo transfections......................................................................................... 111
2.2.34. Flow cytometry for in vivo transfections........................................................... 111
2.2.35. Data analysis................................................................................................... 112
2.2.36. Statistical Analysis.......................................................................................... 112

3. CRISPR/Cas9 mediated NHEJ in cell models .......................................................... 114
3.1 Introduction ........................................................................................................................................... 114
3.3. Results .................................................................................................................................................. 116
  3.3.1. Optimising transfection on Neuro-2A............................................................................................ 116
    3.3.1.1. DNA transfection .......................................................................................................................... 116
    3.3.1.2. mRNA transfection ....................................................................................................................... 118
  3.3.2. Optimising the CRISPR/Cas9 transfection in a GFP model ............................................................ 121
    3.3.2.1. Plasmid DNA encoding CRISPR/Cas9 system ............................................................................. 122
    3.3.2.2. SpCas9 as mRNA and protein .................................................................................................... 129
  3.3.2. Ai9 cell line model ........................................................................................................................... 132
    3.3.3.1. Creating the Ai9 model in Neuro-2A cells .................................................................................... 133
    3.3.3.2. CRISPR/Cas9 mediate NHEJ in the Ai9 cell model ................................................................. 136
      3.3.3.2.1. Optimising transfection with a plasmid encoding Cas9 and gRNA .................................. 136
      3.3.3.2.2. Optimising transfection with mRNA and protein Cas9 ................................................. 143
  3.3.3.4. Gene editing in the Ai9 in vivo model .......................................................................................... 147
  3.4. Discussion ........................................................................................................................................... 151

4. CRISPR/Cas9 mediated correction of ΔF508 mutation in CFBE41o- cells .............................................. 158
  4.1. Introduction ......................................................................................................................................... 158
  4.2. Aims .................................................................................................................................................... 159
  4.3. Results ................................................................................................................................................. 160
    4.3.1. Constructing gRNAs targeting near the ΔF508 mutation and screening for DSB ... 160
    4.3.2. CRISPR/Cas9 – mediated correction of ΔF508 in CFBE41o- ...................................................... 168
      4.3.2.1. gRNAs transfection on CFBE41o- .............................................................................................. 168
      4.3.2.2. Correcting the ΔF508 mutation by homologous direct repair (HDR) at CFBE41o- ............. 172
      4.3.2.3. Functional analysis for corrected CFBE41o- clones .............................................................. 180
        4.3.2.3.1. CFTR and SCNN1 mRNA expression after the correction of CFBE41o-..... 180
        4.3.2.3.2. Electrophysiology study using an Ussing chamber ......................................................... 183
    4.3.3. Correcting the ΔF508 mutation on CFBE41o- cells without selection .................................... 186
    4.3.4. CFTR gRNA off-target screening ................................................................................................. 187
  4.4 Discussion .............................................................................................................................................. 190
5. CRISPR/Cas9 mediated correction of ΔF508 mutation in CFBE primary cells ........................................... 199

5.1. Introduction ................................................................................................................................. 199
5.2. Aims............................................................................................................................................... 200
5.3. Results........................................................................................................................................... 201

5.3.1. Optimising transfection in CFBE primary cells ................................................................. 201

5.3.1.1. Transfection of submerged CFBE primary cells ............................................................. 201
5.3.1.2 Transfection on Air-Liquid Interface (ALI) Cultures ....................................................... 204
5.3.2. CRISPR/Cas9 - mediated correction of ΔF508 in primary cells from the bronchial epithelium ........................................................................................................................................... 205

5.3.2.1. Optimising CRISPR/Cas9 transfection in primary CFBE BMI-1 cells ................. 205
5.3.2.2. Modify the selection marker at the donor ................................................................... 210
5.3.2.3. HDR in CFBE BMI-1 primary cells .............................................................................. 211
5.3.2.3. Functional analysis for corrected CFBE BMI-1 cells ................................................ 214
  5.3.2.4.1. CFTR and SCNN1 mRNA expression after the correction ........................................ 214
  5.3.2.4.2. Correction analysis at protein level ........................................................................ 216
  5.3.2.4.3. Electrophysiological study using an Ussing chamber ........................................... 217

5.4 Discussion...................................................................................................................................... 220

6. Homology-independent targeted integration (HITI) as alternative approach for gene editing ........................................................................................................................................... 226

6.1. Introduction ................................................................................................................................... 226
6.2. Aims............................................................................................................................................... 227
6.3. Results........................................................................................................................................... 228

6.3.1. HITI in neuro-2A Ai9 cell model ............................................................................................ 228

6.3.1.1. Integrating GFP at Ai9 with single DSB ........................................................................ 228
  6.3.1.1.1. Constructing the donor plasmid and transfection ..................................................... 229
  6.3.1.1.2. RNP transfection with donor plasmid .................................................................... 232
  6.3.1.1.3. GFP clone analysis by Sanger sequencing .............................................................. 233

6.3.1.2. Integrating GFP at Ai9 with double cuts .................................................................... 234
  6.3.1.2.1 Constructing the donor plasmid and transfection ..................................................... 235
  6.3.1.2.2. GFP clone analysis by Sanger sequencing .............................................................. 238
  6.3.1.2.3. HITI and HDR techniques together ...................................................................... 239
6.3.2. Replacement of *CFTR* exon 10 by HITI technique ..................................................... 243

6.3.2.1. Targeting *CFTR* intron 9 and 10 by CRISPR/Cas9 ............................................. 243

6.3.2.2. Constructing *CFTR* exon 10 donor plasmid ............................................................ 247

6.3.2.3. Transfection of HEK293 cells to replace *CFTR* exon 10 ....................................... 249

6.3.2.4. Transfection on CFBE41o- to replace *CFTR* exon 10 ......................................... 251

6.3.2.5. *CFTR* intron 9-1 and 10-4 off-target screening ....................................................... 253

6.4. Discussion .......................................................................................................................... 257

7. General discussion and conclusion .................................................................................... 266

8. References ............................................................................................................................ 274
List of Figures

Chapter 1.
Fig. 1.1. Schematic of the CFTR protein. ................................................................. 30
Fig. 1.2. Figure 1.2: The six CFTR mutation classes................................................. 32
Fig. 1.3. Western blot for CFTR protein ................................................................. 33
Fig. 1.4. Cross-sectional of the airway epithelium tissue .......................................... 34
Fig. 1.5. CF aetiology hypotheses ............................................................................. 35
Fig. 1.6. Structure of anoctamin-1 (TMEM16A) ....................................................... 36
Fig. 1.7. Structure of ENaC ...................................................................................... 38
Figure 1.8. Modulator approach for CFTR potential therapy according to the mutation class.................................................................................................................. 50
Figure 1.9. Localisation of progenitor cells at different niches of the human airway system ...................................................................................................................... 54
Figure 1.10 c-NHEJ and alt-NHEJ repair pathway ..................................................... 57
Figure 1.11. Schematic of repair of the DNA break by HDR ...................................... 59
Figure 1.12. Genome editing after DSB ................................................................. 60
Figure 1.13. ZFNs................................................................................................. 61
Figure 1.14. Modular assembly technique for ZFNs .................................................. 62
Figure 1.15. OPEN method for assemble ZFNs....................................................... 63
Figure 1.16. CoDA method for assemble ZFNs.......................................................... 63
Figure 1.17. TALENs............................................................................................. 65
Figure 1.18 CRISPR/Cas9 system .......................................................................... 67
Figure 1.19. Receptor-targeted nanocomplexes (RTN).......................................... 79
Figure 1.20. Examples of lipids (DOPE (neutral), DTDTMA, DHDTMA and DOTMA (cationic), which have been used to deliver nucleic acid molecules formulated with a target peptide ............................................................................................................ 80

Chapter 2.
Figure 2.1. Replace the U6 promoter with a T7 promoter in order to perform in vitro transcription (IVT) .......................................................... 108

Chapter 3.
Figure 3.1. GFP plasmid transfection on Neuro-2A cells using RTN and L2K .......................................................... 117
Figure 3.2. mRNA transfection on Neuro-2A cells using RTN and L2K ................. 121
Figure 3.31. Modify GFP gene to BFP as approach for HDR gene editing.............153

Chapter 4.

Figure 4.1. CFTR gene sequence near the ΔF508 mutation with the eight possible SpCas9 gRNAs targeting the region.................................................................160
Figure 4.2. T7 endonuclease I assay for CFTR gRNAs on HEK293 cells...........162
Figure 4.3. Indels analysis of targeting CFTR by CRISPR/SpCas9 on HEK293 cells…165
Figure 4.4. CFTR gene sequence near the ΔF508 mutation with two SaCas9 gRNAs targeting the region.................................................................167
Figure 4.5. CFTR gene sequence near the ΔF508 mutation with two Cpf1 gRNAs targeting the region.................................................................167
Figure 4.6. Sanger sequence after SaCas9 and Cpf1 transfection....................168
Figure 4.7. T7 endonuclease I assay for CFTR gRNA targets on CFBE41o- cells......169
Figure 4.8. Indels analysis of targeting CFTR by CRISPR/SpCas9 on CFBE41o- cells....................................................................172
Figure 4.9. Strategy to correct the ΔF508 mutation in CFBE41o- using CRISPR/SpCas9 system and donor plasmid template...........................................173
Figure 4.10. gRNA T7 and T5+ PAM motif coding amino acid.......................174
Figure 4.11. The strategy to mutate the PAM region at the donor template of gRNA target 5+ without changing the encoding amino acid..............................175
Figure 4.12. Confirming the HDR event at CFBE41o- by junction PCR..............176
Figure 4.13. HDR screening on CFBE41o- clones...........................................177
Figure 4.14. Allele correction screening on CFBE41o- corrected clones...........178
Figure 4.15. Sanger sequencing for corrected CFBE41o- clones.....................180
Figure 4.16. CFTR mRNA expression in CFBE41o-, corrected CFBE41o- and 16HBE14o- cells.................................................................181
Figure 4.17. SCNN1 mRNA expression on CFBE41o-, corrected CFBE41o- and 16HBE.................................................................183
Figure 4.18. Electrophysiology study on epithelial cell lines............................185
Figure 4.19. The single strand oligo template strategy to correct the ΔF508 mutation.187
Figure 4.20. Sanger sequence screening for five off-target gRNA target 5+ and 7 on HEK293 cells.............................................................................189

Chapter 5.

Figure 5.1. GFP plasmid transfection on primary CFBE BMI-1 cells using RTN and L2K.................................................................201
Figure 5.2. mRNA transfection on submerged primary CFBE BMI-1 cells using RTN and L2K........................................................................................................................................204
Figure 5.3. Transfection efficiency on primary CFBE BMI-1 cells (ALI) using RTN and L2K........................................................................................................................................205
Figure 5.4. Indel analysis in CFTR after transfections with by CRISPR/SpCas9 on CFBE BM-1 cells........................................................................................................................................208
Figure 5.5. Indels analysis of multiple transfection targeting CFTR by CRISPR/SpCas9 on CFBE BMI-1 cells........................................................................................................................................209
Figure 5.6. Modify the selection gene in the CF donor plasmid........................................210
Figure 5.7. The strategy to correct the ΔF508 mutation by using the CRISPR/SpCas9 system and donor plasmid on CFBE BMI-1 cells........................................................................................................................................212
Figure 5.8. Confirming the HDR event at CFBE BMI-1 cells by junction PCR..............212
Figure 5.9. Replacement the puromycin resistance gene at the donor plasmid with neomycin resistance gene...............................................................................................................................213
Figure 5.10. CFTR mRNA expression on primary epithelium cells.................................214
Figure 5.11. SCNN1 mRNA expression on CFBE and corrected CFBE..........................215
Figure 5.12 SCNN1 mRNA expression on CFBE transfected with CFTR mRNA........216
Figure 5.13. Immunoblot for CFTR protein on primary epithelial cells..........................217
Figure 5.14. Electrophysiology study on epithelium primary BMI-1cells.......................219
Figure 5.15. Examples of chemical modification at gRNA synthesis..............................222

Chapter 6.

Figure 6.1. Homology-independent targeted integration (HITI) strategy......................227
Figure 6.2. Schematic of GFP knock-in at gRNA T3 site by HITI..................................228
Figure 6.3. Schematic of constructing GFP HITI donor template with Ai9 target 3 gRNA........................................................................................................................................229
Figure 6.4. GFP knock-in efficiency with HITI strategy at gRNA target 3 site with a donor which has one or two cutting sites.................................................................230
Figure 6.5. The GFP knock-in efficiency with HITI strategy at the gRNA target 3 sites with multiple transfections.................................................................231
Figure 6.6. The GFP knock-in efficiency with the HITI strategy at the gRNA target 3 site using Cas9 protein and donor plasmid.................................................................232
Figure 6.7. The Sanger sequence results at the 5’prime end of the HITI T3 clones.......234
Figure 6.8. Schematic of the GFP knock-in at the stop codons site by the HITI technique.................................................................................................................................235
Figure 6.9. Schematic of constructing the GFP HITI donor template with target 1 and 4 gRNAs ................................................................. 236
Figure 6.10. GFP knock-in efficiency with HITI strategy at the stop codons site ...... 237
Figure 6.11. Sanger sequence at the 3 prime end of HITI T1+4 ............................ 238
Figure 6.12. Schematic of constructing GFP HITI HDR donor template with target 1 and 4 gRNAs ................................................................. 239
Figure 6.13. Schematic of GFP knock-in at stop codons site by HITI and HDR technique ........................................................................ 240
Figure 6.14. Schematic of constructing GFP HDR donor template ....................... 240
Figure 6.15. Schematic of the GFP knock-in at the stop codons site by the HDR pathway ......................................................................... 241
Figure 6.16. GFP knock-in efficiency with HITI HDR strategy at stop codons site analysed by flow cytometry ................................................................. 242
Figure 6.17. GFP knock-in efficiency with HDR strategy at stop codons site analysed by flow cytometry ................................................................. 242
Figure 6.18. CFTR introns 9 and 10 with the eight SpCas9 gRNAs targeting the region ........................................................................... 245
Figure 6.19. T7 endonuclease I assay for CFTR introns 9 and 10 gRNAs .......... 245
Figure 6.20. Indels analysis of targeting CFTR introns 9 and 10 by CRISPR/SpCas9.. 247
Figure 6.21. Removing the CFTR exon 10 using two gRNAs .................................. 247
Figure 6.22. DNA fragment for HITI CFTR exon 10 replacement ......................... 248
Figure 6.23. Cloning the HITI CFTR exon 10 donor ............................................. 249
Figure 6.24. Schematic of replacement of the mutated exon 10 by the HITI technique ........................................................................... 250
Figure 6.25. Replacement of the CFTR exon 10 on HEK293 cells using the HITI technique ........................................................................... 251
Figure 6.26. Replacement of the CFTR exon 10 on CFBE41o- cells using the HITI technique ........................................................................... 253
Figure 6.27. Sanger sequence screening for five off-target sequences for intron 9-1 gRNA on HEK293 cells ........................................................................... 255
Figure 6.28. Sanger sequence screening for five off-target sequences for intron 10-4 gRNA on HEK293 cells ........................................................................... 256
Figure 6.29 Replacement of the stop cassette at Neuro-2A Ai9 by GFP using single strand DNA (ssDNA) template ........................................................................... 261
Chapter 7.

Figure 7.1 *CFTR* gene sequence near the ΔF508 mutation with the three possible Cas9 gRNAs (CjCas9 and expanded SpCas9 variant) targeting the region................................270

Figure 7.2 The potential future gene editing therapy for cystic fibrosis using CRISPR system using ex vivo approach.................................................................273
List of Tables

Chapter 1.
Table 1.1. Examples of airway epithelium cell lines that have been used for the study of CF.................................................................40
Table 1.2. Variant of Cas9 protein from different bacteria species. .........................69

Chapter 2.
Table 2.1. List of primers which have been used to amplify gRNA prior to IVT ..........108
Table 2.2. RTN preparation and weight ratio..........................................................110

Chapter 3.
Table 3.1. List of gRNAs targeting GFP gene.........................................................123
Table 3.2. List of gRNAs targeting Ai9 stop cassette.................................................138

Chapter 4.
Table 4.1. List of SpCas9 gRNAs targeting the CFTR gene near the ΔF508 mutation ..161
Table 4.2. List of SaCas9 and Cpf1 targeting gRNAs near to ΔF508 mutation .............166
Table 4.3. Summary of screening selected clones either with gRNA target 5+ or 7 ..179
Table 4.4. Table 4.4 Potential off-target for gRNA target 5+ ......................................188
Table 4.5. Potential off-target for gRNA target 7....................................................188

Chapter 6.
Table 6.1. List of SpCas9 gRNAs targeting the CFTR introns 9 and 10 .................245
Table 6.2. Potential off-target for 9-1 gRNA..........................................................254
Table 6.3. Potential off-target for intron 10-4 gRNA..............................................255
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>2- ΔΔCt</td>
<td>Delta-delta Ct</td>
</tr>
<tr>
<td>μA</td>
<td>microampere</td>
</tr>
<tr>
<td>μm</td>
<td>micrometer</td>
</tr>
<tr>
<td>α</td>
<td>alpha</td>
</tr>
<tr>
<td>β</td>
<td>beta</td>
</tr>
<tr>
<td>γ</td>
<td>gamma</td>
</tr>
<tr>
<td>δ</td>
<td>delta</td>
</tr>
<tr>
<td>λ</td>
<td>lambda</td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
</tr>
<tr>
<td>µl</td>
<td>Microlitre</td>
</tr>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
</tr>
<tr>
<td>ΔF508</td>
<td>CFTR mutation with deletion in phenylalanine at position 508</td>
</tr>
<tr>
<td>A</td>
<td>adenosine base</td>
</tr>
<tr>
<td>A 83-01</td>
<td>3-(6-Methyl-2-pyridinyl)-N-phenyl-4-(4-quinolinyl)-1H-pyrazole-1-carbothioamide</td>
</tr>
<tr>
<td>AAV</td>
<td>adeno-associated virus</td>
</tr>
<tr>
<td>AAVS1</td>
<td>adeno-associated virus integration site 1</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>ALI</td>
<td>Air-liquid interface</td>
</tr>
<tr>
<td>Alt-NHEJ</td>
<td>alternative NHEJ</td>
</tr>
<tr>
<td>ANO1</td>
<td>anoctamin 1</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ASL</td>
<td>Airway Surface Liquid</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>B2H</td>
<td>bacterial two-hybrid</td>
</tr>
<tr>
<td>BEBM</td>
<td>Bronchial epithelial basal media</td>
</tr>
<tr>
<td>BEGM</td>
<td>Bronchial epithelial growth media</td>
</tr>
<tr>
<td>BFP</td>
<td>blue fluorescent protein</td>
</tr>
<tr>
<td>BGHpA</td>
<td>Bovine Growth Hormone Polyadenylation Signal</td>
</tr>
<tr>
<td>BMI1</td>
<td>B-cell-specific Moloney murine leukemia virus integration site 1</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>C</td>
<td>cytosine base</td>
</tr>
<tr>
<td>C14</td>
<td>DTDTMA/DOPE</td>
</tr>
<tr>
<td>C16</td>
<td>DHDTMA /DOPE</td>
</tr>
<tr>
<td>C18</td>
<td>DOTMA/DOPE</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>Calcium ions</td>
</tr>
</tbody>
</table>
CaCCs  calcium-activated chloride channels
CAG  promoter
CAMP  cyclic adenosine monophosphate
Cas  CRISPR associated system
CBF  ciliary beat frequency
CBh  Chicken Beta Actin Short) promoter.
CD  cluster of differentiation
cDNA  Complementary DNA
CF  Cystic Fibrosis
CFTR  Cystic fibrosis transmembrane conductance regulator
CjCas9  Campylobacter jejuni Cas9
Cl−  chloride ion
cm²  Square Centimeter
CMV  cytomegalovirus
CoDA  context-dependent assembly
Cpf1  CRISPR from Prevotella and Francisella
CRISPR  Clustered regularly interspaced short palindromic repeats
CRISPR/Cas9  Clustered Regularly Interspaced Short Palindromic Repeats and the CRISPR-associated protein 9
crRNA  CRISPR-RNA
Ct  cycle threshold
CtIP  C-terminal-binding protein interacting protein
dATP  Deoxyadenosine triphosphate
dCas9  dead Cas9
DC-Chol  (3 beta [N-(N',N'-dimethylaminoethane)-carbamoyl] cholesterol
DDT  Dithiothreitol
DHDTMA  1,2-Di-((Z)-hexadec-11-enyloxy)-N,N,N trimethylammonium propane iodide
DMD  Duchenne muscular dystrophy
DMEM  Dulbecco's Modified Eagle Medium
DMH-1  4-[6-[4-(1-Methylethoxy)phenyl]pyrazolo[1,5-a]pyrimidin-3-yl]-quinoline
DMPE-PEG₅₀₀₀  1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine-N-methoxy(polyethylene glycol)-5000
DMSO  dimethyl sulfoxide
DNA  Deoxyribonucleic acid
DNMT3A  DNA Methyltransferase 3 Alpha
dNTP  Deoxyribonucleotide triphosphate
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOPE</td>
<td>Dioleoyl L-α phosphatidyl ethanolamine</td>
</tr>
<tr>
<td>DOTMA</td>
<td>N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium chloride</td>
</tr>
<tr>
<td>DSB</td>
<td>Double-strand break</td>
</tr>
<tr>
<td>DSBR</td>
<td>Double-strand break repair</td>
</tr>
<tr>
<td>DTDMA</td>
<td>1,2-Di-(Z)-tetradec-11-onyloxy)-N,N,N trimethylammonium propane chloride, C-11-</td>
</tr>
<tr>
<td>E peptide</td>
<td>K16GACSERSMNFCG</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>eGFP</td>
<td>Enhanced green fluorescent protein</td>
</tr>
<tr>
<td>ENaC</td>
<td>epithelial sodium channel</td>
</tr>
<tr>
<td>EpCAM</td>
<td>Epithelial cell adhesion molecule</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAD</td>
<td>endoplasmic reticulum-associated degradation</td>
</tr>
<tr>
<td>eSpCas9</td>
<td>enhanced specificity Streptococcus pyogenes Cas9</td>
</tr>
<tr>
<td>F</td>
<td>Fusion protein</td>
</tr>
<tr>
<td>FBS</td>
<td>Foetal bovine serum</td>
</tr>
<tr>
<td>FEV1</td>
<td>forced expiratory volume in one second</td>
</tr>
<tr>
<td>G</td>
<td>guanine base</td>
</tr>
<tr>
<td>G1</td>
<td>gap 1 phase</td>
</tr>
<tr>
<td>G2</td>
<td>gap 2 phase</td>
</tr>
<tr>
<td>G418</td>
<td>Geneticin</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GL67</td>
<td>Genzyme lipid 67A/1,2-Dioleoyl-sn-Glycero-3-Phosphoethanolamine</td>
</tr>
<tr>
<td>gRNA</td>
<td>Guide RNA</td>
</tr>
<tr>
<td>GSEA</td>
<td>set enrichment analysis</td>
</tr>
<tr>
<td>GSH</td>
<td>glutathione</td>
</tr>
<tr>
<td>GSNO</td>
<td>S-Nitrosothiol reductase</td>
</tr>
<tr>
<td>hCFTR</td>
<td>human CFTR</td>
</tr>
<tr>
<td>HCO⁻³</td>
<td>bicarbonate</td>
</tr>
<tr>
<td>HDR</td>
<td>Homology directed repair</td>
</tr>
<tr>
<td>HITI</td>
<td>homology-independent targeted integration</td>
</tr>
<tr>
<td>hMSCs</td>
<td>human mesenchymal stem cells</td>
</tr>
<tr>
<td>HN</td>
<td>hemagglutinin/neuraminidase protein</td>
</tr>
<tr>
<td>HNH</td>
<td>histidine-asparagine–histidine nuclease</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous recombination</td>
</tr>
<tr>
<td>HSF1</td>
<td>heat shock factor 1</td>
</tr>
<tr>
<td>HypaCas9</td>
<td>hyper-accurate Cas9 variant</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>IBMX</td>
<td>3-isobutyl-1-methylxanthine</td>
</tr>
<tr>
<td>ICAM-1</td>
<td>Intercellular Adhesion Molecule 1</td>
</tr>
<tr>
<td>IL</td>
<td>interleukin</td>
</tr>
<tr>
<td>iPSC</td>
<td>induced pluripotent stem cell</td>
</tr>
<tr>
<td>Isc</td>
<td>short-circuit current</td>
</tr>
<tr>
<td>IVT</td>
<td>in vitro transcribed gRNA</td>
</tr>
<tr>
<td>Kbp</td>
<td>Kilo base pair</td>
</tr>
<tr>
<td>KRAB</td>
<td>Krüppel-associated box</td>
</tr>
<tr>
<td>L</td>
<td>lipid</td>
</tr>
<tr>
<td>L2K</td>
<td>Lipofectamine 2000</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>LIG4</td>
<td>DNA ligase 4</td>
</tr>
<tr>
<td>LLI</td>
<td>Liquid-liquid interface</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>M1</td>
<td>transmembrane domains 1</td>
</tr>
<tr>
<td>M2</td>
<td>transmembrane domains 2</td>
</tr>
<tr>
<td>mAbs</td>
<td>monoclonal antibodies</td>
</tr>
<tr>
<td>MBL-2</td>
<td>mannose-binding lectin 2</td>
</tr>
<tr>
<td>MEM</td>
<td>Minimum Essential Media</td>
</tr>
<tr>
<td>MFI</td>
<td>Mean fluorescence intensity</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>ml</td>
<td>Millilitre</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>MMEJ</td>
<td>microhomology-mediated end joining</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MSCs</td>
<td>mesenchymal stromal cells</td>
</tr>
<tr>
<td>MSD</td>
<td>membrane-spanning domains</td>
</tr>
<tr>
<td>Na+</td>
<td>Sodium ion</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium chloride</td>
</tr>
<tr>
<td>NBD</td>
<td>nucleotide-binding domains</td>
</tr>
<tr>
<td>NBF-1</td>
<td>first-nucleotide binding domain</td>
</tr>
<tr>
<td>neo</td>
<td>Neomycin-resistance gene</td>
</tr>
<tr>
<td>ng</td>
<td>Nanogram</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td>NHBE</td>
<td>Normal human bronchial epithelial</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-Homologous End Joining</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization signals</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>NmCas9</td>
<td><em>Neisseria meningitidis</em> Cas9</td>
</tr>
<tr>
<td>NS</td>
<td>Not significant</td>
</tr>
<tr>
<td>OCT</td>
<td>ornithine-cytosine transferase</td>
</tr>
<tr>
<td>OPEN</td>
<td>oligomerized pool engineering</td>
</tr>
<tr>
<td>P</td>
<td>Probability value</td>
</tr>
<tr>
<td>P2Y2</td>
<td>purinergic receptors</td>
</tr>
<tr>
<td>pac</td>
<td>Puromycin-resistance gene</td>
</tr>
<tr>
<td>PAM</td>
<td>Protospacer adjacent motif</td>
</tr>
<tr>
<td>PARP1</td>
<td>poly(ADP-ribose) polymerase 1</td>
</tr>
<tr>
<td>PAXX</td>
<td>paralog of XRCC4 and XLF</td>
</tr>
<tr>
<td>PBAEs</td>
<td>Poly-beta amino esters polymer</td>
</tr>
<tr>
<td>PCD</td>
<td>primary ciliary dyskinesia</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PHA1</td>
<td>Pseudohypoaldosteronism type I</td>
</tr>
<tr>
<td>PKA</td>
<td>protein kinase A</td>
</tr>
<tr>
<td>pol θ</td>
<td>DNA polymerase theta</td>
</tr>
<tr>
<td>ppFEV1</td>
<td>percent predicted forced expiratory volume in one second</td>
</tr>
<tr>
<td>pre-crRNA</td>
<td>precursor-CRISPR-RNA</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative reverse transcription-PCR</td>
</tr>
<tr>
<td>RES</td>
<td>reticuloendothelial system</td>
</tr>
<tr>
<td>RFLP</td>
<td>restriction fragment length polymorphism</td>
</tr>
<tr>
<td>rhDNase</td>
<td>recombinant human deoxyribonuclease I</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>ROCK</td>
<td>rho-associated protein kinase</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>RTN</td>
<td>receptor-targeted nanocomplexes</td>
</tr>
<tr>
<td>RuvC</td>
<td>crossover junction endodeoxyribonuclease</td>
</tr>
<tr>
<td>RVD</td>
<td>repeat-variable di-residue</td>
</tr>
<tr>
<td>S</td>
<td>synthesis phase</td>
</tr>
<tr>
<td>SaCas9</td>
<td><em>Staphylococcus aureus</em> Cas9</td>
</tr>
<tr>
<td>SAM</td>
<td>synergistic activation mediator protein complex</td>
</tr>
<tr>
<td>SCID</td>
<td>severe combined immunodeficiency</td>
</tr>
<tr>
<td>SCNN1</td>
<td>sodium channel non-neuronal 1</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SDSA</td>
<td>synthesis-dependent strand annealing</td>
</tr>
</tbody>
</table>
SIN  self-inactivating
siRNA  Small-interfering RNA
SLC26A9  Solute Carrier Family 26 Member 9
SOC  Super optimal broth with catabolite repression
SP-C  surfactant protein C
SpCas9  *Streptococcus pyogenes* Cas9
SPLUNC1  short palate lung and nasal epithelial clone 1
SRSRs  short regularly spaced repeats
ssDNA  single strand DNA
ssODN  Single-stranded oligonucleotide
StCas9  *Streptococcus thermophilus* Cas9
T  thymine base
TAE  Tris-acetate-EDTA buffer
TALEN  Transcription activator-like effector nuclease
TdCas9  *Treponema denticola* Cas9
TEER  Transepithelial electrical resistance
TET1  Ten-eleven translocation methylcytosine dioxygenase 1 enzyme
TF  tissue factor
TGF-β1  Transforming growth factor beta 1
TIDE  Tracking of Indels by Decomposition
TMEM16A  transmembrane member 16A
tracrRNA  Trans-activating crRNA
U  uracil base
UTR  untranslated region
V  volt
w/v  weight/volume
wt-CFTR  wild-type Cystic fibrosis transmembrane conductance regulator
XLF  XRCC4-like factor
XRCC4  X-ray cross-complementing protein 4
Y peptide  K16GACYGLPHKFCG
ZFN  Zinc-finger nuclease
CHAPTER 1

Introduction
1. Introduction

1.1. Cystic fibrosis
Cystic fibrosis (CF) is a recessive autosomal genetic disease that has multi-organic effects such as those relating to sweat glands, the pancreas, the liver, the gastrointestinal tract and the respiratory tract system (1). The incidence rate of CF in Caucasian people is approximately 1 in 2,500 births (2). In the UK, there are over 10,000 people who have been diagnosed with CF, with a current survival rate of around 40 years (3).

1.1.1. History of cystic fibrosis
The first description of CF is found in a medical text written in 1595, where there is a link between death at childhood and a damaged pancreas and salty skin. In this text, the infant was described as “hexed” or “ewitched” (4). The first scientific description for the disease came three centuries later (1938) by Dr Dorothy Andersen, who called it “cystic fibrosis of the pancreas” based on a study of autopsies of children (5). Later, she linked this disorder with a vitamin A deficiency; however, this hypothesis was disproven once penicillin was discovered, together with the patients’ response (6, 7). In the 1940s, CF was reported by Dr Sydney Farber as a multiorgan disorder with an excess of mucus, which leads to secondary staphylococcal infection (8). The disease etiology has been investigated by numerous researchers since that time. For example, Dr Howard studied two families in 1944 with a high number of deaths as result CF and he suggested the possibility of a heterozygous inheritance (9). A large study two years later led by Drs Andersen and Hodges on 113 families concludes “the disease, although hereditary, requires more than one factor for the expression” (10). A great achievement for CF diagnosis occurred in 1948 when Dr Paul di Sant’Agnese observed the increase of salt in the sweat of children suffering from CF. Over the next three decades, the research focus was on improving the patient’s life by using enzyme replacement and proper diet (11-13). The CF transmembrane conductance regulator (CFTR) gene was localized in the mid-1980s at chromosome 7 long arm using restriction fragment length polymorphism (RFLP) analysis (14, 15). Cloning and characterization of the complementary
DNA for the \textit{CFTR} gene was performed in 1989 by a collaborative effort between Lap Chee Tsui, Francis Collins and Jack Riordan (16, 17). These researchers also noticed that the majority of CF patients have three missing nucleotides (CTT) at this gene, causing deletion of phenylalanine amino acid at position 508 (ΔF508) (16, 18).

\subsection*{1.1.2. \textit{CFTR} gene and protein}

\textit{CFTR} is located in humans at chromosome 7 (long arm, region q31–q32). It is 189 Kb in length with 27 exons, which encode 1,480 amino acids protein (19). The CFTR protein is located in the apical membrane of epithelium cells and consists of two nucleotide-binding domains (NBD) and one regulatory domain (R domain). In addition, there are two transmembrane domains (TMDs) (TMD-1 and TMD-2) comprising six alpha helix segments, which together form the channel pore (Figure 1.1A).

The CFTR is a unique adenosine triphosphate (ATP)-binding cassette (ABC) protein, which is an ABC transporter-class ion channel protein (ABC sub-family C, member 7 protein). CFTR regulates the chloride flow, which is essential for fluid hemostasis, especially in the airway system (20). In addition, the CFTR protein shows permeation to other molecules such as bicarbonate (HCO\textsuperscript{−}\textsubscript{3}) and glutathione (GSH) (21).

CFTR activation starts with phosphorylation of the R domain by cAMP-dependent protein kinase A (PKA), which is followed by the binding of the ATP to NBD-1. The CFTR channel opens when the two NBDs are dimerized as a result of ATP binding which is then followed by conformational changes of the TMDs. The final step starts by hydrolyzing the ATP at NBD-2 and releasing ADP and Pi, following the same process as in NBD-1 (Figure 1.1B) (22).
The synthesis of mature CFTR protein has multiple steps before reaching the apical side of the cell membrane. It begins in the nucleus where the CFTR gene is transcribed into messenger RNA (mRNA) and then translated into protein once reaching the endoplasmic reticulum (ER). The protein undergoes core post-translation modification (such as ubiquitination, glycosylation and phosphorylation) and folding. After that, CFTR trafficking occurs from the ER to ER-Golgi-intermediate compartment to the Golgi, where the protein becomes mature by complex glycosylation. The CFTR protein completes its journey to the apical plasma membrane, where it works as an ion transport channel (23).

1.1.3. CFTR mutations

CF patients have a large range of clinical phenotypes, with more than 2000 variants having been identified at the CFTR gene. However, only 336 mutations have been confirmed as causing CF according to the Clinical and Functional Translation of CFTR (CFTR2) project database (www.CFTR2.org; August 2018 update).

The CFTR mutation can be classified into six groups according to the molecular mechanism (Figure 1.2.). Class I is the most severe type and affects 10% of CF patients (24). No protein synthesis occurs as a result of the premature stop codon (TAA, TAG, or TGA) seen in the nonsense mutation G542X (a stop codon in place of glycine at position 542). In addition, no protein production can be caused by frameshift mutations with large deletions (CFTRdele2,3(21 kb): 21kb deletion at
introns 1-3 of the CFTR gene) or canonical splice mutations (1717-1G→A: mutation at the 3’ end of intron 10, where guanine changes to adenine) (25, 26). Class II is where there is a defect in protein processing and trafficking, which prevents the protein from reaching the cell apical membrane. More than 80% of CF patients can be classified under this class, including those with the most common CFTR mutation (ΔF508). Another example of this class is the N1303K mutation (c.3909C>G: mutation at nucleotide binding domain 2 (NBD2)), which is more common in CF Mediterranean patients and is the fourth most common CFTR mutation in the world (27). In class III, the CFTR protein reaches the cell membrane. However, the mutation prevents the chloride channel from gating. The most common mutation that can be categorized under this class is the G551D missense mutation (occurring in almost 4% of CF patients) (28). The rest of the CFTR class mutations are less severe. In class IV, the mutation leads to a defect in conductance, as in the case of the R117H missense mutation (where arginine is replaced by histidine at residue at 117), which represents almost 1% of CF patients (29). Class V refers to mutation caused by a decreasing synthesis of the functional protein as result of a splicing defect; for instance, at the alternative acceptor splice site in intron 17a (3272-26A→G mutation) (30). The last class (VI) of the CFTR mutation causes instability and turnover of the protein at the cell membrane. An example of this mutation is 4326delTC, which causes a frame shift mutation in exon 23 (31).

In addition to the classification above, some CF patients might have a compound heterozygous mutation (each allele has different mutation); however, they can be classified according to which mutation has the dominant effect. For example, a patient with CF mutation G542X (class I) and ΔF508 (class II) is considered to fall into class I, as the G542X mutation has the dominate functional effect (32). Class IV and V mutations are functionally dominant when occurring with class I, II or III mutations (33).
Figure 1.2. The six CFTR mutation classes. The classification is according to molecular mechanism: Class I: lack of CFTR synthesis; class II: defective protein processing; class III: defective channel regulation or channel gating; class IV: defective chloride conductance; class V: reduced amount of CFTR protein; and class VI: increased turnover of CFTR channel at the cell surface. Adapted from (34).

1.1.3.1. ΔF508 mutation
The most prevalent CF mutation is the ΔF508 mutation (70% of all mutations), which belongs to the class II CFTR mutation (premature degradation or incomplete maturation). The percentage increases to 90% of all CF patents if we include the compound heterozygous (35). The mutation is a result of the deletion of three nucleotides (CTT) and, as a consequence, the phenylalanine residue in the CFTR protein at position 508 is missing in the first nucleotide-binding domain (NBD1). This mutation causes a misfolding of the protein during processing in the endoplasmic reticulum, which traps it there and is then degraded by the endoplasmic reticulum-associated degradation (ERAD) pathway. The ΔF508 CFTR mutation can be visualized by a western blot assay with protein size around 150KDa (band B, immature core-glycosylated), while wildtype wt-CFTR protein has an immature core-glycosylated band (B) and mature glycosylated (C, around 170KDa) (Figure 1.3.). The trafficking of ΔF508 CFTR protein can be rescued by growing the cells at lower temperatures (27°C), which allows them to reach the cell surface (36, 37).
Figure 1.3. Western blot for CFTR protein. The wild type CFTR has both mature band C for fully glycosylated CFTR and immature band B for the immature CFTR, while ΔF508 CFTR has the immature band B. Adapted from (38).

1.1.4. Cystic fibrosis pathophysiology

CF has multi-organic effects such as those relating to sweat glands, the pancreas, the liver and the gastrointestinal tract. However, lung disease is the major cause of morbidity and mortality (1). The severity of the CF disease relies on numerous factors such as the class of CF mutation. In addition, exogenous factors have an influence on the disease manifestation, which can explain why siblings may have the same mutation but the disease severity is different (39). Transforming growth factor beta 1 (TGF-β1) and mannose-binding lectin 2 (MBL2) expression have been linked as gene modifiers for cystic fibrosis and infection severity (40). The MBL, for example has a role in innate immunity where the deficiency of expression has been linked to bacterial and viral infection (41), while polymorphisms at codon 10 and 25 in TGF-β1 gene have been associated with worse the lung function in CF patients (42). Other CF modifier genes have been reported through genome-wide association studies (GWAS) including Ets homologous factor (EHF) and APAF1 interacting protein (APIP) (43). In addition, non-genetic factor could have an effect, such as outdoor pollution and stress (44).
In the respiratory tract, the epithelium lining of the airways has various barrier mechanisms that act as defences against pathogenic microbes. Anatomically, this tissue is formed from pseudostratified ciliated cells covered by a watery layer (30 µm thick) called the periciliary sol or airway surface liquid (ASL), which is necessary for cilia beating. Furthermore, a mucus layer secreted by goblet and submucosal glands in the epithelium tissue contains ions, glycoproteins, and antimicrobial and antioxidant agents, which play an important role in terms of innate immunity (Figure 1.4) (45, 46).

CF patients suffer from recurrent inflammation due to bacterial infection of the lung. The pathophysiology of this disease is debatable but there are two main hypotheses (Figure 1.5). The first hypothesis proposed by Dr Jeffrey Smith in 1996 relates to high amounts of salt. He claims that the mutation in the chloride channel leads to alteration of the salt concentration in the ASL. The NaCl concentration increases from ~ 50 mM to 100 mM in the apical side of the epithelium tissue and, as a result, bacteria are able to colonize there because the high salt concentration inhibits the antimicrobial substrate at the ASL (48). In contrast, the second hypothesis argues that the ASL layer is isotonic in both normal and CF patients. However, the bacteria colonization and inflammation happen when the cilia do not function properly as a result of a decrease in the
ASL depth from \( \sim 7 \, \mu m \) to 3-4 \( \mu m \). The hydration of this layer is caused by an abnormal over-absorption of water and sodium ions on the apical epithelium side (46, 49).

**Figure 1.5. CF aetiology hypotheses.** A1) The high salt hypothesis argues that the normal ASL has a low salt concentration as a result of the salt absorbed in an excess of water. A2) In CF cases where the ion channel is defective, the salt accumulates in a high concentration at the ASL layer, inactivating the antimicrobial elements. B1) The low-volume hypothesis claims that the salt concentration at the ASL surface does not change. However, in the case of a CF disorder (B2), the mutation in the CFTR causes an over-activation of the epithelial sodium channel (ENaC), with the result that sodium and water are over absorbed, which negatively affects the ASL thickness and the cilia function. Adapted from (50).

1.2. CFTR and another channels interaction
Apart from the secretory function for the CFTR channel, it regulates and interacts with other channels, such as calcium-activated chloride channels (CaCCs) and epithelial sodium channel (ENaC).
1.2.1. The calcium-activated chloride channels (CaCCs)

The CaCCs distributed in the body tissue include excitable (such as cardiac muscles and olfactory sensory neurons) (51, 52) and non-exitable (such as secretory epithelia cells) (52). The CaCC is a class of chloride channels that is activated by intracellular calcium. The gene encoding CaCCs was identified in 2008, initially called transmembrane member 16A (TMEM16A) (53-55), and renamed by Yang et al. 2008 as anoctamin 1 (ANO1) (Figure 1.6.) (55).

Figure 1.6. Structure of anoctamin-1 (TMEM16A). It consists of ten transmembrane segments (TM). The protein channel functions as a chloride channel activated by binding cytosolic Ca$^{2+}$ at two glutamic residues. Adapted from (56).

In mammals, there are 10 anoctamin homologs: ANO1 (TMEM16A) to ANO10 (TMEM16K). However, ANO1 is most abundant anoctamin expressed in the epithelia tissues. The interaction between TMEM16A and CFTR has been studied in vitro and in vivo. Early studies thought that there was no interaction mechanism between CFTR channel and CaCC, whereas CFTR expression causes the down-regulation the CaCC and vice versa (57, 58). However, with the molecular identify of the encoding gene and knockout TMEM16A$^{-/-}$ mice show impaired mucociliary clearance and a CF-like lung phenotype (59, 60). Additionally, in a recent study by Roberta Benedetto and colleagues on tissue-
specific knocked-out \textit{TMEM16A}, mice show that that not only eliminates the Ca\textsuperscript{2+} activated Cl\textsuperscript{−} function, but also inhibits the chloride gating through the CFTR channel in both intestine and airways tissues. \textit{In vitro}, the overexpression of TMEM16A on CFBE41o- (transduced with wt-CFTR) cells enhances the cAMP-activated CFTR currents as well as in CFBE41o- (transduced with ΔF508 CFTR) cells (61). In primary bronchial epithelium cells from healthy and CF patients, Manon Ruffin and colleagues studied the interaction activity between the two channels where they found that the CaCC activity significantly decreased in CF patients in comparison with non-CF individuals (62).

These previous studies prove the functional interaction between the CFTR and CaCCs (TMEM16A). The mechanism of the interaction between them has been proposed by numerous hypothesis. Both channels have different means of activation in order to conductance chloride. The CaCC channel is activated by calcium while CFTR is stimulated by cAMP. However, there are some studies that suggest that the CFTR channel can be a response to intracellular calcium signals. The physical interaction between the two channels has been proposed through PDZ domain (61, 63), whereas ΔPDZ from both channels causes a reduction in their expression. In addition, immunoprecipitation of TMEM16A in CFBE cells shows pulldown CFTR protein (band B and C) (61).

\subsection*{1.2.2. The Epithelial sodium channel (ENaC)}

The ENaC has different names such as sodium channel non-neuronal 1 (SCNN1) and amiloride-sensitive sodium channel (64). The ENaC exists in the apical side of epithelium membrane of numerous tissues such as the sweat glands, the nephron and the airway epithelium and functions in the reabsorption of sodium ions (65). The sodium channel is composed of three homologous subunits: α, β and γ, which are encoded by the \textit{SCNN1A}, \textit{SCNN1B} and \textit{SCNN1G} genes respectively (Figure 1.7) (66). In brain and testis tissues, the δ subunit has been reported as additional subunit, which with β and γ subunits forms a functional ENaC channel (67). Mutation in the \textit{SCNN1} gene has been linked to other
diseases such as Liddle’s syndrome (pseudohyperaldosteronism) and Pseudohypoaldosteronism type I (PHA1) (68).

Despite the formation of ENaC from the three subunits at epithelium tissues, the most essential subunit for the proper channel functional has been debated. An early study in 1990s of SCNN1A (−/−) in mice showed a defect in terms of the clearance of liquid from the lung, which consequently caused early death after 40 hours of birth (69). Despite the fact that no βENaC knockout mouse has been created to evaluate the outcome effect. A transgenic mouse with airway-specific overexpression of βENaC subunit has shown the manifestation of CF-like lung disease, which includes airway mucus obstruction and chronic inflammation. In contrast, a transgenic mouse overexpressed either αENaC or γENaC alone did not increase amiloride-sensitive Na+ transport (70).

**Figure 1.7. Structure of ENaC.** It is composed of three subunits: α, β and γ. Each subunit consists of two transmembrane domains (M1 and M2) and an extracellular loop. In addition, there is a PY motif at the end of the carboxyl side of β, a γ unit, which consists of three amino acids, two of which are proline and tyrosine, while the third can be any amino acid. Adapted from (65).

The direct interaction between the ENaC and CFTR is a matter of debate. The interaction between the two channels proposed by the overexpression of the βENaC subunit in particular leads to a CF-like disease in mice with a decreased level of ASL and an increase in mucus production, which is correlated with the
low-volume CF pathophysiology hypothesis (70, 71). A recent study from our lab showed the knockdown αENaC subunit by siRNA molecule on primary CFBE cells increases ASL levels as well as ciliary beat frequency (CBF) (72). On the other hand, other studies have challenged the direct interaction between CFTR and ENaC. In a study led by Dr Barbara Grubb showed the breeding of a βENaC transgenic mouse with another transgenic mouse that expressed human CFTR (hCFTR), resulting in a new strain of mouse (hCFTR/β-ENaC), failed to reduce sodium channel activity even though there was a remarkable improvement in Cl⁻ conductance (73). Furthermore, on CF pig models, the knockout of CFTR does not demonstrate an alternation of the ENaC expression or liquid absorption (1).

The mechanism interaction between ENaC and CFTR has been proposed by numerous studies. The feedback inhibition hypothesis suggests that the increased level of intracellular chloride ions works as a self-inhibitor for sodium channel activity (74). However, this hypothesis cannot apply to the sweat gland, as the activation of CFTR by cAMP causes an increase in the activity of the sodium channel (75). Other studies suggest that the interaction between the two proteins could be through cellular proteins as interference such as in the cytoskeletal elements, the C-terminal PDZ-binding domain of CFTR and the first-nucleotide binding domain (NBD-1) of CFTR (76-79).

**1.3. Cystic fibrosis study models**

**1.3.1. Cystic fibrosis cells**

An assessment of the functional consequence of variants requires the appropriate cellular context. Immortalizing epithelium cells especially from various airway tissues of healthy and CF patients has been achieved by different immortalization techniques (Table 1.1). These cell lines have different genotype (wt-CFTR or mutant) and phenotype features, such as the formation of a tight junction when they grow on an air-liquid interface (ALI). Despite the useful application of these cells for CF research and their advantages of long lifespans and being cost effective, there are some limitations such as changing phenotype
over time of multiple passaging, with changing gene expression or lack of some phenotypes from the origin tissue.

<table>
<thead>
<tr>
<th>Name</th>
<th>Origin</th>
<th>Genotype</th>
<th>Immortalization</th>
<th>Tight Junctions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A549</td>
<td>alveolar epithelial cells</td>
<td>wt-CFTR</td>
<td>adenocarcinoma</td>
<td>no</td>
<td>(80)</td>
</tr>
<tr>
<td>CFTE-290-</td>
<td>tracheal epithelia cells</td>
<td>ΔF508/ ΔF508</td>
<td>simian virus 40 (SV40)</td>
<td>no</td>
<td>(81)</td>
</tr>
<tr>
<td>IB3-1</td>
<td>bronchial epithelia cells</td>
<td>ΔF508/1282X</td>
<td>hybrid virus, adenovirus type 12 -SV40</td>
<td>no</td>
<td>(82)</td>
</tr>
<tr>
<td>56FHTc8o-</td>
<td>fetal tracheal cells</td>
<td>wt-CFTR</td>
<td>simian virus 40 (SV40)</td>
<td>no</td>
<td>(83)</td>
</tr>
<tr>
<td>Calu-3</td>
<td>submucosal tracheal epithelial cells</td>
<td>wt-CFTR</td>
<td>adenocarcinoma</td>
<td>yes</td>
<td>(84)</td>
</tr>
<tr>
<td>CFBE41o-</td>
<td>bronchial epithelia cells</td>
<td>ΔF508/ ΔF508</td>
<td>simian virus 40 (SV40)</td>
<td>yes</td>
<td>(85)</td>
</tr>
<tr>
<td>16HBE14o</td>
<td>bronchial epithelia cells</td>
<td>wt-CFTR</td>
<td>simian virus 40 (SV40)</td>
<td>yes</td>
<td>(86)</td>
</tr>
<tr>
<td>1HAeo</td>
<td>bronchial epithelia cells</td>
<td>wt-CFTR</td>
<td>simian virus 40 (SV40)</td>
<td>yes</td>
<td>(87)</td>
</tr>
<tr>
<td>9HTc8o-</td>
<td>tracheal epithelia cells</td>
<td>wt-CFTR</td>
<td>simian virus 40 (SV40)</td>
<td>No</td>
<td>(88)</td>
</tr>
</tbody>
</table>

Table 1.1. Examples of airway epithelium cell lines that have been used for the study of CF.
The use of primary cells is a valuable approach to the study of CF. The primary epithelial cells are often used for representing the situation in vivo and for personalised studies. These cells were first used in 1982 when Frank Lechner and colleagues were able to grow human bronchial cells in defined media of growth factors on coated plastic with fibronectin (89). The primary airway epithelial cells can be obtained from nasal or bronchial brushing. These cells can be differentiated to pseudostratified cells that form tight junctions and cilia and produce mucin when they are grown as an ALI culture, which is an in vitro model of the human airway. However, the primary epithelial cells are only passaged two to three times before they lose their proliferation and differentiation, which makes work on these cells limited, especially in cell engineering and drug screening (90).

Different approaches have been investigated to overcome this problem. For example, growing the primary epithelium cells on irradiated fibroblast feeder layer in presence of rho-associated protein kinase (ROCK) inhibitors extends proliferation. The feeder layer is responsible for the growth factors while ROCK inhibitors increase cell proliferation and adhesion (91). Recently, a culture technique free of a feeder layer was developed using bone morphogenetic protein antagonist (DMH-1) and transforming growth factor β antagonist (A-83-01) to block the signal pathway of SMAD proteins. This technique showed that the basal epithelium cells were on average able to double 70 times (21 passages) in comparison to the six passages without both inhibitors (92). Another approach is to extend the proliferation and delay senescence by overexpression of integrating the B-cell-specific Moloney murine leukemia virus integration site 1 (BMI-1) gene. The transduction of the basal epithelial cells with lentivirus encoding the human BMI1 was shown to extend the lifespan up to 25 passages without losing their ability for differentiation (93). The high expression of BMI-1 has been reported in many types of stem cells such as neural and hematopoietic stem cells, as it is necessary for self-renewal (94, 95).

Other advanced models used to conduct studies on epithelial airway cells using organoid (mini-organs) are “three-dimensional (3D) structures derived from stem cells and consist of organ-specific cell types which self-organise through cell sorting and spatially restricted lineage commitment in a manner reminiscent of
the native organ with some degree of organ functionality” (96). The culture technique is based on using culture media containing specific growth factors and an appropriate extracellular matrix such as Matrigel to mimic the \textit{in vivo} status. The organoid model has been successfully created for different organs such as lung (97), intestine (98), kidney (99) and cornea (100) from either adult stem cells or induced pluripotent stem cells (iPSCs). Recently, a research group from University of North Carolina at Chapel Hill described a new way to form an organoid (nasospheroids) from a nose biopsy using a minimally invasive procedure in comparison to obtaining intestinal stem cells by rectal biopsy. The nasospheroids formed spontaneously within five days of nasal epithelial culturing with a single layer of differentiated pseudostratified ciliated epithelial cells and non-ciliated cells, where the apical membranes face the outside and basolateral membranes face the inside, which is opposite to organoids driven from intestinal stem cells (101).

The primary cells have some limitation such as the short lifespan with limited passages. In addition, this procedure requires access to internal organs such as the lung. Furthermore, rare CFTR genotypes might be limited as sources for primary cells. The maintenance of primary cells on general is expensive and requires skill in growing them.

\textbf{1.3.2. Cystic fibrosis animal models}

Despite the recent advances in primary cells culturing and modelling, there is still a need to perform \textit{in vivo} studies. The animal models are needed to represent the physiology and pathology of CF, which consequently helps us to understand the disease and develop a new therapy. Early in the 1990s after the identification of the \textit{CFTR} gene, several groups of researchers began to develop mice models for CF by making null (\textit{−/−}) mice (102-104), then creating murine models with specific CFTR mutations (such as G551D and ΔF508) (105, 106). Despite the high similarity between the human and mice CFTR amino acid sequence (78%), the pathophysiology of the disease is different (107). The murine model failed to develop spontaneous lung obstruction or chronic bacterial infections with a short lifespan. In addition, CF mice have a chloride CFTR independent channel (108).
To develop a CF similar phenotype with increased mucus secretion and poor bacterial clearance on mice is achieved by overexpression of the sodium channel (beta subunit) (71).

For these reasons above, another CF animal model is needed. Ferret and pig CF models have been generated either by disrupting the CFTR or introducing ΔF508 CFTR mutation (109, 110). The longer lifespan of both animals, physiological similarity to human lung and spontaneous CF manifestation make them preferable for CF in vivo models. However, there are some limitations for the two models, such as difficulty of gene manipulation, longer gestation time and the high cost of maintenance.

1.4. Cystic fibrosis therapy

The current conventional treatment for CF is based on the reduction of the symptoms of the disease and improvement of the patient’s quality of life. Antibiotics are prescribed for CF patients with acute and chronic infection, especially due to the bacteria species Pseudomonas aeruginosa, given to the patients orally or intravenously depending on the severity of the case. However, the long-term use of antibiotics has a significant risk in terms of developing toxicity with regard to body organs such as the kidneys (111). The clearance of thick mucus from the airway system could also be achieved by physical exercise, through coughing or by inhaling a hypertonic saline where the high salt solution helps to rehydrate the ASL and consequently aids mucociliary clearance (112). In addition, mucolytic agents such as dornase alfa (Pulmozyme), which contains recombinant human deoxyribonuclease I (rhDNase) enzyme, can improve the lung function (113). Another molecule called OligoG extracted from Laminaria hyperborean shows promising results in reducing the mucus thickness by chelating calcium, which is required for normal mucin unfolding. AlgiPharma finished recently the clinical trial phase IIb for OligoG (ClinicalTrials.gov Identifier: NCT02157922) (114).

Furthermore, patients with exocrine pancreatic insufficiency treated with pancreatic enzyme replacement with proper nutrition is recommended for all CF
patients. In chronic cases with no improvement being apparent, a lung transplant might be the best choice with regard to saving CF patients (115).

1.4.1. Targeting non-CFTR channels
Another potential cystic fibrosis therapy is targeting non-CFTR interacting protein channels such as ENaC or CaCCs by activation or inhibition. Activating CaCCs directly or indirectly is an alternative approach to stimulating Cl⁻ efflux in CF cells. Duramycin (Moli1901, Lancovutide) as an activator for CaCCs through release of intracellular Ca²⁺ has been investigated as a potential therapy for cystic fibrosis. However, the results were not encouraging, with only minimal effects (79, 116). Another drug (Denufosol) has reached clinical trial stage III as an activator for CaCCs through the release of intracellular Ca²⁺ by binding to purinergic receptors (P2Y2). However, no significant improvement for lung function have been shown, which could be related to the short half-life of the compound in vivo (17 minutes) compared to the three hours in differentiated nasal primary epithelium cells (117). New modulators for the CaCCs with specific action without activating the airway smooth muscles cells need more investigation.

On the other hand, dehydration of ASL in CF patients combined with hyper activation of the sodium channel targeted, as discussed above in the CFTR and ENaC interaction (section 1.2.2.). For that reason, inhibiting the ENaC is proposed as a potential therapy for CF. This treatment was seen in the 1990s when amiloride – a transport blocker – was given to CF patients through inhalation (118, 119). However, the outcome of this treatment was limited, probably because of the short half-life of amiloride in the airway surface. Several small molecules as modulators for ENaC such as such as P-552, P-1037 and P-1055 have been investigated by Parion Sciences, Inc. and Vertex Pharmaceuticals as therapeutic agents for CF and primary ciliary dyskinesia (PCD), some of which are undergoing clinical trials.

The other approach to knock down the ENaC expression using a small-interfering RNA (siRNA) has a longer duration effect. siRNA targeting ENaC has been evaluated in our lab, where we show the increase of ASL depth and the return of CBF to normal levels after transfection of siRNA-targeting α-subunit on
differentiated primary CFBEs delivered by receptor-targeted nanocomplexes (RTN), where in vivo we demonstrated a 30% reduction in SCNN1A expression after one dose, which increased to almost 50% silencing after three doses (72). Finally, short palate lung and nasal epithelial clone 1 (SPLUNC1) is a multifunctional protein involved in the innate defence of airways and has a regulatory effect on the epithelium sodium channel through dissociating αβγ-ENaC and generating a new SPLUNC1–β-ENaC complex (120). Spyryx Biosciences and Cystic Fibrosis Foundation are evaluating the effect of SPX-101 (SPLUNC1) on CF patients, and have finished clinical trial stage I, showing it to be safe and well tolerated in treated subjects. However, the drug did not show a significant change in lung function for patients evaluated by the forced expiratory volume in one second (FEV1) test (120). Currently, the SPX-101 is on stage II of the clinal trial (NCT03229252).

1.4.2. CF gene therapy
1.4.2.1. Viral vector
For more than two decades, viral vectors have been successfully used for gene therapy. They have the ability to effectively transduce cells, and some viral vectors integrate to host genomes such as lentivirus, which provides long-term expression (121). The earliest clinical study in this field goes back to 1993 when CFTR cDNA was administrated though the adenovirus, where early-transcribed regions E1 were deleted to prevent the virus replication. The adenovirus has been attractive for CF gene therapy to the lung because of the efficiency of transducing non-dividing cells of the airway without integration in the genome. However, the low transfection efficiency and immune inflammation response against the virus even with low doses make this therapy approach infeasible for CF patients (122).

The adenovirus as a vector for gene therapy in general has some issues pertaining to safety, especially after the death in 1999 of a patient with ornithine-cytosine transferase (OCT) deficiency disease as result of severe immune reaction of administration adenovirus vector on clinical trial (123).

The other vector that has been investigated for CF gene therapy is the adeno-associated virus (AAV), which is attractive as a gene therapy vector because of
the lack of pathogenicity, broad tissue tropism, long expression as an episomal vector and ability to transduce non-dividing cells (124). However, the main disadvantage of this vector is the relatively small genome (4.7kb), which makes it challenge for larger genes without codon optimisation. AAV serotypes 1,2,5 and 6 have been used in pre-clinical and clinical studies as delivery vectors (125).

Despite the safety outcome from the clinical trials with AAV from single and multiple doses, the CFTR mRNA expression from most vectors is undetectable, even though it was detected in animal pre-clinal studies (126). The improvement of the AAV for CFTR cDNA delivery is focused on using an engineered effective infection serotype and improving the expression promoter instead of CFTR as endogenous promoter (127, 128).

Lentivirus is another viral vector that has been investigated for CF gene therapy. It is a genus of retrovirus and can transduce both dividing and non-dividing cells with a long expression, as it is integrated at the host genome, which makes it a preferable vector for targeting differentiated airways cells. The lentivirus has been used effectively for gene therapy for many diseases such as severe combined immunodeficiency (SCID)(129) and Wiskott-Aldrich syndrome (130). Despite the efficiency of the lentivirus vector, the main concern is the mutagenesis when the virus is integrated close to oncogenes and activated, as seen in patients treated from X-SCID who developed leukemia in early clinical studies in this field (131). Currently, the third generation of lentivirus is safer, with a self-inactivating (SIN) feature, where the 3' long terminal repeat (LTR) sequence has been deleted (132). The insertion of the mutagenesis has been investigated with SIN-lentivirus in clinical trials with no evidence of oncogenesis (133, 134).

Despite the efficiency of lentivirus to infect numerous tissue, this virus does not naturally have tropism for airway epithelium. Various modifications have been performed on the virus envelope with pseudotyping. An example of pseudotyping is the simian immunodeficiency virus (SIV) with the Sendai virus (SeV) envelope (135). The pre-clinical studies on SIV pseudotyped with the SeV envelope which has Fusion protein (F) and hemagglutinin/neuraminidase protein (HN) used to develop vector encode CFTR gene with elongation factor 1α promoter and the cytomegalovirus enhancer in mice and human ALI show encouraging results,
which have led the UK Cystic Fibrosis Gene Therapy Consortium to announce that the first clinical trial with lentivirus on CF patient will be in 2017 (136). However, the main drawback of lentivirus in therapeutic approach is the possibility of inducing an acute immune response against virus particles when the virus is re-administered (137).

1.4.2.2. Non-viral vector

Non-viral vectors such as liposome and polymers provide a new and safe approach for the delivery of many molecules such as drugs and well as nucleic acid to cells. Liposome is a spherical lipid bilayer with a hydrophobic and hydrophilic tail. Mixing a cationic lipid such DOTMA (1,2-Di-((Z)-octadec-9- enyloxy)-N,N,N trimethyl ammonium propane chloride) with nucleic acid (anionic molecule) in an appropriate ratio allows for the formation of a self-assembled nanocomplex with a cationic surface charge. The nanocomplex interacts with the cell surface (anionic), which leads to internalization, typically by endocytosis. Inside the cells, the nanocomplex needs the endosomal escape mechanism to avoid degradation by lysosomes (138). The non-viral vector in general is less efficient for gene delivery as it does not have specific components for cell entry until nuclear uptake, while viral vectors naturally have these mechanisms. Nevertheless, the non-viral delivery method has numerous advantages over the viral vector method. The main advantage associated with this delivery method is the low immunogenicity response, which is important when repeated doses need to be administered, as in the case of CF therapy. In addition, it has the capacity to deliver larger molecules. It can also be undertaken on a large scale due to the simplicity of preparation (139). For these reasons, the non-viral vectors have been attractive for CF therapy. The first clinical trial using a non-viral vector in the UK took place in 1994 on 15 CF patients (9 received cDNA CFTR DNA encapsulated with liposome and 6 subjects received only the liposome) (140). Proof of concept for safety and tolerance of repeat administration (3 doses) of non-viral vector was studied in 2000 by Hyde and colleagues, where DC-Chol/DOPE (3 beta [N-(N',N'-dimethylaminoethane)-carbamoyl] cholesterol/dioleoylphosphatidylethanolamine) cationic liposomes were used to deliver
CFTR cDNA to the nasal epithelium of 10 patients (141). In the UK, the Cystic Fibrosis Gene Therapy Consortium has, since it was formed in 2001, aimed to move the clinical trials for CF from the safety stage to advanced stages where the evaluation of benefits can be studied. To do so, preclinical studies have focused on improving the encoding plasmid as well as the non-viral vector. To reduce the chance of inflammation and increase the expression a codon optimized CFTR, CpG-free plasmid (pGM169) has been developed (142). The liposome vector (GL67A) (143) has also been optimized for nebulizer-directed aerosol delivery which has three components: 1) cationic lipid (GL67) (Genzyme lipid 67A/1,2-Dioleoyl-sn-Glycero-3-Phosphoethanolamine) (144); 2) DOPE; and 3) DMPE-PEG5000 (1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine-N-methoxy(polyethylene glycol)-5000), with the following molar ratio: (1 : 2 : 0.05) (GL67:DOPE:DMPE-PEG5000).

The first clinical study (phase I/IIa) (NCT00789867) using a non-viral gene transfer agent (pGM169/GL67A) aimed to evaluate the safety outcome of a single dose with three volumes of nebulization (5 ml, 10 ml and 20 ml). The study performed on 35 patients concluded the 5 ml dose the most suitable for next phase clinical trial (145). In 2012, phase IIb started on 140 patients (116 subjects completed the study) to evaluate the clinical benefit from multiple dose of 5 ml of pGM169 (13.3mg)/GL67A (75mg) in a 12-month study (NCT01621867). The primary efficacy endpoint of this study was the relative change in ppFEV₁ (before dosing to the end of the study). The statistical analysis for data showed a significant increase at ppFEV₁ (3.7%, p=0.046) on the treated group compared to the placebo group at the end of the study. However, the authors conclude the pGM169/GL67A treatment causes stabilization in lung function rather than an improvement (146). Despite this finding, this study demonstrated that non-viral based therapy could be the potential approach in the future for treatments of CF that use increasement in dosage for example, or co-administration with modulator drugs.
1.4.2.3. CF mRNA therapy
The other approach for cystic fibrosis is mRNA molecules as an alternative for DNA-based therapy. The mRNA molecules are translated in the cytoplasm of the transfected cells without the need to translocate into the nucleus, which makes them more effective. In addition, the risk of potential insertion to host gene is less than with DNA transfection. The main limitations of this approach is the short half-life expression after administration, instability and potential cytotoxicity through toll-like receptor activation (147). The chemical modification by various techniques shows great improvement in this field, especially in cancer and infection vaccination (148). The *in vitro* work on differentiated human nasal epithelium primary cells shows that the minimal dose of wt-\textit{CFTR}-mRNA (0.6 \( \mu \)g/cm\(^2\)) delivered by lipofectamine 2000 rescues the electrophysiology functional for the CFTR channel over a period of 72 hours (149). For CF, the first clinical trial (I/II) (NCT03375047) has been given approval to start this year in USA sponsored by Translate Bio, Inc., where the MRT5005 drug (\textit{CFTR} mRNA) will be given to 32 participants regardless of the CF class mutation.

1.4.3. CF modulator therapies
Modulators are small molecules designed to target the \textit{CFTR} mRNA or protein in order to correct the gene malfunctioning. These modulators can be classified into different categories: potentiator, corrector, stabilizer and amplifier (Figure 1.8.).
The potentiator is kind of modulator that binds to the CFTR protein on the cell membrane and increases the open probability for the CFTR channel. The first potentiator ivacaftor (Kalydeco®) was approved in 2012 for a patient with a G551D class III mutation causing a defect in CFTR channel gating (151). Ivacaftor was identified initially in 2009 by screening over 200,000 small-molecule compounds using high throughput screening on a cell-based fluorescence membrane assay. The in vitro works on mutated cells show an increase of CFTR probability of opening, while the apical fluid height also increased (152). In clinical trials, the treated patients showed improvement in the lung function, had fewer exacerbations and indicated a better quality of life. The lung function was assessed by the FEV1 test, where the patients showed an improvement in absolute FEV1 by almost 10% (153). Furthermore, the evaluation on non-G551D mutation with a reduction on CFTR channel gating including G178R, S549N, S549R, G551S, G970R, G1244E, S1251N, S1255P, and G1349D showed the significant improvement of 7.5% in the FEV1 test (154). In addition to ivacaftor, there are other potentiator drug undergoing clinical trials, such as BW251 developed by Novartis (NCT02449018), which has completed phase II, whereas PTI-808 (NCT03251092) developed by Proteostasis Therapeutics is on phase I in combination with other modulators.

Another class of modulator is the stabilizer, which aims to enhance the CFTR protein stability by reducing the degradation in the endoplasmic reticulum and increase the protein residence time on the cell membrane. An example is Cavosonstat (N91115), developed by Nivalis Therapeutics Inc., which acts on S-Nitrosoglutathione reductase (GSNO) as increased in CF patients. However, the outcome from stage II of the clinical study (NCT02589236), where it was tested in combination with ivacaftor on CF homozygous for the ΔF508 mutation, was no significant change in lung function after the 28-day treatment period (155). Currently, the N91115 stabilizer molecule with ivacaftor is in phase II, to evaluate the effect on heterozygous ΔF508 mutation patients with other gating mutations (NCT02724527).
The encouraging outcome results from the potentiator modulator on CF patients with G551D mutation led to more research to find molecules correct the most common CF mutation (ΔF508). By screening more than 64,000 small molecules in a recombinant cell-based assay led to discover VX-809 corrector molecule which is able to rescue the CFTR protein from the endoplasmic reticulum through enhancing interactions between the NBD1, MSD1, and MSD2 domains (156, 157). However, monotherapy by VX-809 alone fail to improve the lung function of ΔF508 CF patients (158). The reason for that linked to impair the rescued CFTR protein for gating (159). To overcome of that the corrector molecule (VX-809) combined with potentiator modulator (Ivacaftor) which called this combination later Orkambi® (from Vertex Pharmaceuticals). This drug got the FDA approval for clinical use on CF patient with homozygous ΔF508 mutation in July 2015. However, the patients' benefits from Orkambi® is moderate in comparison with the benefits received by patients with CFTR class III mutation from ivacaftor. In clinical study in phase III, the improvement in the percentage of predicted FEV1 was on average 3% from the mean baseline FEV1 61% of the predicted value (160). Another issue is some side effects such as liver damage, dyspnea, abnormal respiration and the potential interaction with other drugs, which means that it is not suitable for all ΔF508 homozygous CF patients (161). Earlier in this year, a new drug called Symdeko™ (from Vertex Pharmaceuticals) was approved by US FDA for CF patients with a ΔF508 mutation for patients aged 12 years and older. This drug is a combination of two modulators; tezacaftor as corrector and ivacaftor as a potentiator. The outcome from clinical trial phase III on 475 patients over 24 weeks with a mean FEV1 at baseline 60.0% of the predicted value increased by 4% in comparison with placebo subjects, while the incidence of adverse events was similar between the two groups (162). In addition, a new strategy to treat ΔF508 homozygous CF patients with two correctors has a different mechanism of action that might help to restore the protein function. Two new generation of corrector (VX-445 and VX-659) completed recently phase II study in combination with tezacaftor and ivacaftor (NCT03227471, NCT03224351 respectively) which they show increase the absolute FEV1 between 7 to 12% to the tezacaftor-ivacaftor effect (163). The plan
for Phase III trials expect to start later this year to more evaluation. Other correctors under clinical trials are PTI-801 from Proteostasis Therapeutics (NCT03140527, stage I), GLPG2222 from Galapagos (NCT03045523, finished phase II) and FDL169 from Flatley Discovery Lab (NCT03093714, finished phase I).

Amplifier modulators aim to increase the amount of CFTR mRNA and consequently more protein loaded to ER. However, for therapy they need to be combined with corrector and potentiator modulators. PTI-428 developed by Proteostasis Therapeutics has completed phase I/II of study (NCT02718495) to evaluate safety, tolerability and pharmacokinetics. Currently, the amplifier modulator (PTI-428) is under another clinical study on homozygous ΔF508 patients in combination with PTI-808 (potentiator) and PTI-801 (corrector) (NCT03500263).

The previous modulators can work in most CFTR class mutation except class I, where the mutation causes no protein synthesis as result of frameshift by deletions or insertions or premature stop codon mutation. In premature stop codon mutation, such as G542X mutation, a read-through molecule could be the strategy for rescuing the CFTR protein. In early studies, the aminoglycoside antibiotics such as gentamicin and tobramycin have been investigated as potential therapies for CFTR premature stop codon mutation. However, by nasal administration of the aminoglycoside, 11 patients demonstrated no changes in nasal ion transport in comparison with the placebo group (164). Ataluren (PTC124) is a read-through molecule identified by using high throughput screening. This drug conditionally approved by the European Medicines Agency for the treatment of nonsense mutation DMD in patients aged 5 years and older with ambulatory decline-phase (165). The Ataluren has been investigated for CF patients with nonsense mutation, inclusive results from phase II clinical study (NCT00351078) led to another study on large number of objectives (NCT02456103). However, the study terminated as result of very low change in percentage predicted FEV1 (0.6%) (166).

Another type of small molecule used to restore the channel function by correcting the CFTR mRNA using oligonucleotide molecules. Examples include QR-010,
developed by ProQR Therapeutics for ΔF508 homozygous mutation. The QR-010 is a 33mer antisense oligonucleotide (wildtype template) that binds to the CFTR mRNA sequences adjacent to the ΔF508 mutation in order to correct the mutation. This drug is currently in phase Ib (NCT02532764) of study. In addition, antisense oligonucleotide molecules have been investigated to correct splice CFTR mutation (class V) such as c.2657+5G>A (167).

1.4.4. Cystic fibrosis cellular therapy
Cellular therapy could be alternative therapeutic approach for CF, whereby correcting the patient’s own stem cells ex vivo then administered back or using allogeneic stem cells from a healthy donor. The source of the stem cells as therapy could be endogenous, such as from the airway system, or from other, exogenous tissues. The basal cells in the airway tissue could be the most suitable for this kind of therapy as they are easy to access. In vitro culturing of basal cells on an ALI can differentiate them as pseudostratified epithelium with Club, goblet and ciliated cells (168). In addition, the xenograft model confirms that human basal cells selected by two markers tetraspanin CD151 and tissue factor (TF) can be seeded on rat trachea, engrafted into nude mice for 35 days. The histological and immunohistochemical analyses for the samples show the ability of basal cells to differentiate and form mucociliary epithelium (169).

Stem and progenitor cells were identified in the airway tract of humans and mice, which are important for homeostasis and repair of this system. The stem niches were located either in the upper airway tract (trachea and bronchi) or in the lower tract (bronchioles and alveoli) (Figure 1.9.) (170, 171).
Figure 1.9. Localisation of progenitor cells at different niches of the human airway system. Basal cells are identified as progenitor cells in the trachea and bronchi. Furthermore, variant Club cells are characterized as stem cells in the lower airway with the ability to regenerate the bronchiole and the proximal alveolar tissue. Finally, in the alveoli (bronchoalveolar duct junction), type II alveolar epithelial cells show the ability to differentiate into squamous cells. Adapted from (171).

The manipulation of cells outside the body has more advantages than the in vivo approach. Barriers before reaching the basal cells, such as mucus, which is abundant in CF patients, will not be there with the ex vivo approach. The progenitor cells in the airway system have a slow turnover, a transient injury agent could help the corrected cells to proliferate and differentiate to replace mutated cells. Inhalation of sulphur dioxide or naphthalene in an animal model caused injury to the airway system with damage to differentiated cells but not the basal cells. After 24 hours, the trachea started to re-line with new differentiated cells and reached a peak after two weeks, with fully ciliated pseudostratified
epithelial cells (172, 173). However, this approach was taken in a pre-clinical study, thus needs more work to determine the safety aspect in humans.

Bone marrow has different types of stem cells such as hematopoietic stem cells and mesenchymal stromal cells (MSCs), it has been used successfully for decades as cell therapy for bloods origin diseases. However, the ability of stem cells from bone barrow to home at non-hematopoietic tissues and cure other diseases is debatable. Dr Erica Herzog and colleagues show that by performing different sex bone barrow transplant at murine model lack surfactant protein C (SP-C) were lethally irradiated first. The analysis six months after the transplant showed that 65% of male recipients’ SP-C+ cells contained the Y chromone while the Y chromosome was lost in 72% of female SP-C+ cells, which indicates that their origin came from fusion (174). This phenomenon has been reported in humans as well after bone marrow or hematopoietic stem cells transplant. Similarly, hematopoietic stem cells transplant from male donors to female specimens shows the presence of Y chromosome in the lung epithelial cells (175-177). To study the benefit of the stem cells in bone marrow on CF, a study was performed on CFTR null mice (−/−). The results show improvement in the CFTR mRNA expression by 7% in comparison to the wildtype and restored the three CFTR protein bands (A, B and C) at the Western Blot. In addition, the transplant improved bacterial clearance and survival. However, to show this effect the recipient mice were given a lung injury using busulfan and naphthalene (178). A safety study (phase I) on 15 adult CF patients (NCT02866721) started in August 2016, with estimate for primary completion data in August 2018. The patients receive a single intravenous allogeneic human mesenchymal stem cells (hMSCs) up to $5 \times 10^6$ cells/kg of their body weight. The study will focus on safety first and report any alteration in the disease status (for example, infection and inflammation biomarkers and FEV1 %).

**1.5. Genetic recombination**

Genome correction naturally occurs in the case of damaged DNA where double strands break (DSB) is the result of endogenous factors such as oxidative free radicals or exogenous factors such as chemotherapeutics. This correction
happens in more than ten thousand cells per day to avoid mutation and initiation of tumours. Cells correct this tendency by non-homologous end joining (NHEJ) pathways or homologous directed repair (HDR). HDR involves repairing a DSB by using another undamaged chromatid as a template, while non-homologous recombination relies on direct end joining of the break, which can lead to insertion or deletion (indel) in that region (179, 180).

### 1.5.1. Non-homologous end joining

NHEJ is essential for V(D)J recombination during early B and T cell development (181). In addition, cells use this pathway to join the DSB when the DNA template is not available. Consequently, this process can take place in any phase of the cell cycle (182). The NHEJ can divided into two subclasses: 1) canonical or classical NHEJ (c-NHEJ) and 2) alternative NHEJ (Alt-NHEJ), which is also known as microhomology-mediated end joining (MMEJ).

The c-NHEJ is a predominant repair pathway in mammalian cells and is initiated by binding of heterodimer Ku proteins (Ku80 and Ku70) to the break site, which protects and stabilizes the DNA ends. The end joining of the break strands requires both ends to be blunt in order to join them. The broken ends can be trimmed by the Artemis protein, which in complex DNA-PKcs cleaves 5’ and 3’ overhangs or resynthesizes the missing nucleotides using DNA polymerases (such as DNA polymerase mu (μ) or DNA polymerase lambda (λ)). DNA ligase (ligase IV) then joins the DSB with X-ray cross-complementing protein 4 (XRCC4), XRCC4-like factor (XLF) and a paralog of XRCC4 and XLF (PAXX) (Figure 1.10) (183, 184).

The Alt-NHEJ works more as backup pathway for c-NHEJ when is it not active or fails as result of mutation or chemical inhibition. The result from Alt-NHEJ is intensive insertion or deletion compared with the c-NHEJ, which causes minimal indels at the repair junction. This repair pathway starts by joining poly(ADP-ribose) polymerase 1 (PARP1) at DSB, which works as a signal and recruit other proteins. Then the break end sectioning, which creates complementary 1-10 microhomology nucleotides, is helped by MRN (MRE11–RAD50–NBS1) and C-terminal-binding protein interacting protein (CtIP)(185). The ligation process is
independent of ligase IV and other factors such as XRCC4 and XLF. DNA ligase III plays a major role for the ligation in Alt-NHEJ, helping low-fidelity DNA polymerase θ (Pol θ) to fill the gaps (Figure 1.10.) (186).

**Figure 1.10. c-NHEJ and alt-NHEJ repair pathway.** c-NHEJ is the common mechanism for repairing the DSB with minimal sequence alterations, whereas the alt-NHEJ causes extensive insertions or deletions. The c-NHEJ is initiated by dimerizing Ku (Ku70- Ku80) and DNA-PK catalytic subunits (DNA-PKcs) at the DNA break site, which is trimmed by the Artemis protein. Following that, the DNA ends recruit additional enzymes and finally rejoins the ends using the ligase LIG4/XRCC4 complex. One other hand, Alt-NHEJ is a Ku protein and DNA-PK independent repair mechanism. It starts by joining poly(ADP-ribose) polymerase 1 (PARP1), then the break end resection (5′–3′) with the help of MRN (MRE11–RAD50–NBS1) and CtBP-interacting protein (CtIP). The LIG3 plays major role in the ligation in Alt-NHEJ, helping low-fidelity DNA polymerase θ (Pol θ) to fill the gaps. Adapted from (187).
1.5.2. Homologous directed repair

HDR is a fundamental step during meiosis in germ cells by DSB and in the crossover of genetic material. Cells can correct DSB by HDR during the mitotic DNA replication phase (S and G2 phase) in the presence of the DNA template (187). There are two models for this recombination: DSB repair (DSBR) and synthesis-dependent strand annealing (SDSA). Both pathways are initiated by binding MRE11–RAD50–NBS1 complex (MRN) to the end of a DSB, which makes a 3’ over-hang by re-sectioning (188). Nucleoproteins (protein replication protein A (RPA) and Rad51) then start to form a filament to invade the undamaged homologous template, which generates a D-loop structure (189).

After that, the common pathway is divided into a DSBR or SDSA. In the case of DSBR, the double Holliday junction structure forms four strands with the assistance of a non-invasion strand. A cross-over between strands commonly happens in this model. This is necessary for genetic segregation, otherwise the Holliday junction dissociates without cross-over after DNA synthesis (190). On the other hand, in the SDSA model, the double strand correction takes place without either forming a Holliday junction or cross over by displacing the invasion strand, then re-annealing with the other 3’ over-hang strand (Figure 1.11.) (191).
Figure 1.11. Schematic of repair of the DNA break by HDR. The HDR pathway is a more complicated process and needs a DNA template to occur. The first step of this pathway starts by re-sectioning the site of the break by recruiting the MRN complex (MRE11–RAD50–NBS1) and other proteins. Then, nucleoprotein (RPA), Rad52 and Rad51 paralogs (RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3) mediate the formation of a filament to invade the undamaged template. The common pathway after that is divided into two models: SDSA and DSBR. In the SDSA model, the recombination occurs due to invasion by one 3′-ssDNA strand without crossover between chromatids. On the other hand, the DSBR model starts by forming a Holliday junction. The recombination process ends by a crossover from the donor template which is either with or without crossover. Adapted from (192).

1.6. Gene editing
Gene editing is a genetic technique that aims to modify the genome by the correction, insertion or deletion of nucleotides (Figure 1.12.). The recombinant between donor DNA and target sequence is very low (one per $10^4$ to $10^7$ cells) (193). This incidence rate increases 100-fold when a DSB is applied (194).
Figure 1.12. Genome editing after DSB. The DSB can be introduced by engineering nuclease-targeting specific locus which is repaired either by NHEJ or HR. a) The repair by NHEJ causes insertion (green) or deletion (red), which can cause a gene knockout function. b) Insertion DNA fragment could be happening through NHEJ when a linearized donor DNA is used. Precise genome editing can be achieved but at lower frequency by HR in presence donor DNA with homology arms and causes gene modification (c) or gene insertion (d). Adapted from (195).

The history of targeting a specific DNA sequence and making DSB reaches back to the 1990s when meganuclease I-SceI showed the ability to promote homologous recombination in mammalian chromosomes with a frequency two folds higher than spontaneous HR after genome double break. The meganuclease I-SceI was derived from the mitochondria of *Saccharomyces cerevisiae*, which is able to target a specific 18 bp (5′-TAGGGATAACAGGGTAAT-3′) and achieve a DSB with four-base, 3′ overhangs. However, the rare availability of this sequence in the genome means that this technique has limited application in genome editing (196, 197).

1.6.1 Zinc finger endonuclease

ZFNs overcome this drawback and have become the predominant technique for gene editing. ZFN is composed of two sets of Cys2His2 zinc fingers binding domains separated by a *FokI* nuclease as the DNA-cleavage domain. Typically, designing a target sequence using this technique requires 18 bp recognized by six zinc fingers (three each side) with a 5 to 7 bp spacer that allows for the *FokI*
nuclease to dimerase. Each set of finger-like structures (approximately 30 amino acids sequence and folds into ββα structure) is able to recognize a triple DNA nucleotide (Figure 1.13.) (198, 199).

![Figure 1.13. ZFNs. Two arrays of ZFNs targeting a DNA sequence in opposite strand where the 3’ end of each is fused by FolkI endonuclease as a DNA-cleavage domain. Each zinc finger binds to three nucleotides. Adapted from (200).](image)

The fundamental step for constructing zinc finger domain is the identification of a motif (finger) that recognizes each of the 64 possible DNA triplets. The phage display technique is initially used to select and identify proteins that bind to a desired DNA sequence from a library of zinc finger motifs. However, this methodology is too laborious (201). So, other simpler techniques have been introduced, especially for academia.

The modular assembly is based on assembling the zinc finger one by one to create a full set of constructs (Figure 1.14.). A library of known of zinc finger recognizes most of the triplicates of 3bp DNA (all GNN, most ANN, many CNN, and some TNN) made commercially (Sangamo Biosciences), which need to assembled to target the full desired DNA sequence (202). However, this methodology shows high failure rates in recognizing the DNA target (about 75%) after the assembly, which could be the affinity of each domain for binding independently from the neighbouring domain, and interaction between them needed (203). So, new a methodology proposed to overcome the low success rate of the modular technique and the independent context of DNA binding effects is called oligomerized pool engineering (OPEN) (Figure 1.15.). It based on pool of multi three pre-constructed zinc finger arrays that recognize 9 DNA bp then
select the highest affinity using the bacterial two-hybrid (B2H) screening method. The B2H reporter system measures expression of the lacZ reporter gene as a result of transcriptional activation between the multi-finger domain and the target site upstream of the promoter (204). Despite the success rate, this method is still laborious in terms of screening.

The latest methodology which is available publicly as the platform of reagents and software is called context-dependent assembly (CoDA), which is rapid and requires only standard cloning techniques. This approach using a database of library of three-finger arrays which have been identified previously in other arrays to function and share the possible 18 arrangement of the middle motif (F2) (Figure 1.16.) (205).

**Figure 1.14. Modular assembly technique for ZFNs.** Modular assembly of a three-finger protein targeting ACTCTGTGG sequence in IL2Rγ gene for example (206). Each finger is identified to bind a component of three base pairs of the DNA where they link together. Adapted from (200).
Figure 1.15. OPEN method for assemble ZFNs. Random library combination of three fingers created, followed by isolating the high affinities and high specificities set of multi-finger arrays using B2H selection method. Adapted from (207).

Figure 1.16. CoDA method for assemble ZFNs. Library of three-finger arrays identified previously in other arrays to function and share the middle motif (F2). A new three fingers engineered by joining finger 1 (F1) from first array and finger 3 (F3) from second array followed by B2H reporter assay to evaluate the binding affinity to the target sequence. Adapted from (205).
In general, the zinc finger endonuclease technique is able to target different DNA sequences and make a DSB, which allows for knockout and knock-in genes. However, as we mentioned above, the technique is laborious, has a low success rate and is potentially off target because each finger protein recognizes three DNA bases.

1.6.2. Transcription activator-like effector nucleases

A new protein molecule, TALEN, is expressed by Xanthomonas bacteria and is able to recognize a single nucleotide thus improve precise targeting. TALENs bind to the single nucleotide through a domain consisting of 34-amino acids, which provide more accuracy. From the 34 amino acids residue, the high variable amino acid at positions 12 and 13 specify each monomer (repeat-variable di-residue (RVD)). The most common RVDs are HD (Histidine, Aspartic Acid), NG (Asparagine, Glycine), NI (Asparagine, Isoleucine) and NN (Asparagine, Asparagine) which recognize C, T, A, and G nucleotide respectively. The 3’ end of the two TALENs protein needs to be fused by FokI endonuclease as a DNA-cleavage domain (Figure 1.17.) (208). This technique has some limitations, such as the DNA recognition site for TALENs should start with T base (209). Furthermore, there is difficulty in cloning and synthesizing mostly identical repeat sequences for the TALE array (variable only at RVD). To overcome this technical difficulty, different cloning strategies have been used, such as ligation-independent cloning techniques (210), Golden Gate (211) and the FusX TALEN Assembly System (212). Currently, these cloning assembly techniques for TALEN synthesis are available as kits from Addgene for academic use.
65

**Figure 1.17. TALENs.** A) One transcription activator-like effector consists of tandem repeats of 34 amino acids, where the two amino acids at position 12 and 13 (Repeat-variable di-residue (RVD)) ensure specificity for each monomer. The most common RVDs frequently associated with nucleotide are enclosed within the square. B) Two arrays of transcription activator-like effectors targeting a DNA sequence in the opposite strand where the 3’ end of each fused by FokI endonuclease as a DNA-cleavage domain. Adapted from (200)

In comparison to zinc finger endonuclease, TALENs are more specific, as one motif recognizes single nucleotide while each zinc finger binds onto three nucleotides. In addition, transcription activator-like effector synthesis by modular assembly shows a high success rate, which indicates that interaction between each neighbour module is not required, unlike the zinc finger modules (213). The ZFN has an advantage over TALENs in the size of plasmid encoding the construct (one third smaller), which might affect delivery efficiency.
1.6.3. Clustered regularly interspaced palindromic repeats (CRISPR)/CRISPR associated system (Cas)

The latest gene editing technique is CRISPR/Cas. The history of the CRISPR/Cas system dates back to 1987, when Yoshizumi Ishino and colleagues discovered five identical cluster repeat sequences, each having 29 nucleotides, which were separated by non-repetitive spacer segments in iap gene of E. coli (214). This phenomenon was noticed later in other bacteria and archaea species, which represent around 40% and 90% respectively of the whole genome (215). The specific term for this repeat was established in 2000 when it was referred to as short regularly spaced repeats (SRSRs) (216). However, two years later it was renamed CRISPR. In addition, another set of gene was noted to be linked with CRISPR, which was then called Cas. The Cas gene encodes helicases and nuclease proteins (217). The function of this system was described by Barrangou et al. 2007, who explained how the bacteria use the CRISPR/Cas system for adaptive immunity against viral invasion. The mechanism is made up of three stages: adaptation, expression and interference. The bacteria firstly adapted the dedicated viral genome into the CRISPR locus with mediated help from Cas1 and Cas2 proteins. When reinfection occurs, the CRISPR locus transcribes into precursor-CRISPR-RNA (pre-crRNA), which later becomes mature crRNA. In addition, another non-coding RNA molecule is called trans-activating crRNA (tracrRNA), which hybridizes to crRNA and recruits the Cas protein. The final stage ends by recognizing the protospacer adjacent motif (PAM), which subsequently activates the Cas nuclease, leading to degradation of the invaded genome through a DSB (Figure 1.18.) (218).
Figure 1.18. CRISPR/Cas9 system. The Cas9 protein (light blue) is guided by CRISPR RNA which is two parts: crRNA, in black with 20 nucleotides determining the targeting DNA, and tracrRNA, in red and has role in stabilizing the structure. The Cas9 protein with nuclease domain RuvC and HNH cleave the targeting sequence when it follows a specific protospacer-adjacent motif (PAM, in yellow) (5′-NGG-3, Cas9 from *S. pyogenes*). Adapted from (195)

The CRISPR/Cas system can be classified into two main groups (types I, II). Both types share the same adaptive processing but differ in the expression and interference. In the class II CRISPR system, there is only one signature Cas protein (Cas9) responsible for the interference stage in conjugate with gRNA (crRNA and tracrRNA) while in class I, the effector complex involves several Cas proteins (219). The Cas9 protein has six domains: recognition I (REC I); recognition II (REC II); PAM Interacting (PI), which helps initiating binding to target DNA; bridge helix; histidine–asparagine–histidine (HNH) nuclease and crossover junction endodeoxyribonuclease (RuvC). However, only HNH and RuvC have the nuclease activity. The HNH cleaves the DNA complementary strand while the RuvC domain performs its job on the noncomplementary strand (220).

In 2013, two studies in parallel engineered a plasmid-encoded CRISPR/Cas9 from *S. pyogenes* (class II), which has a hybrid of crRNA and the tracrRNA with
codon-optimized Cas9 driven by a U6 promoter. In this system taking from *S. pyogenes*, the DSB occurred by Cas9 when the gRNA found the 20-nucleotide complementary to the target DNA sequence followed by 5'-NGG-3' PAM motif (where N could be any nucleotide; A, T, C or G). The DSB most likely happened as a blunt cut 3-4bp upstream of the PAM sequence. This was the first evidence of the possibility of using this system as a gene editing tool (221, 222). Since that time, there have been thousands of publications in this field.

The CRISPR/Cas system has numerous advantages in comparison with ZFNs and TALENs. First, it is simple in design, because it is based on a recognition of the DNA target by a ribonucleotide complex. ZFNs and TALENs are also laborious and time-consuming techniques because they are based on protein design and construction. In addition, the cost of synthesizing a CRISPR/Cas targeting complex is far cheaper than other gene-editing techniques. The use of a fused activator or inhibitor to allow genetic modification of the Cas9 domain (D10A and H840A), which makes it catalytically inactive, can be another application of this system as a means of interfering with gene expression (223). Finally, the CRISPR/Cas can operate as a multiplex system to target different genes in one transfection event (224).

1.6.3.1. CRISPR/Cas variants

The NGG PAM sequence is abundant throughout the human genome and found approximately every 8 bp (225). However, targeting an area in which AT nucleotides are rich or the DSB needed at a specific position where the NGG is not located is challenging. As such, numerous studies have investigated the possibility of modifying the CRISPR system from *S. pyogenes* to identify a different PAM sequence or using CRISPR system from other bacteria strains to recognize a different PAM motif. By introducing a specific mutation at wildtype *S. pyogenes* Cas9 (SpCas9) domain, Kleinstiver et al. 2015, were able to alternate the PAM recognition sequence. For example, SpCas9 VQR variant (D1135V, R1335Q and T1337R) is able to recognize NGAN or NGNG PAM sequence while EQR variant (D1135E/R1335Q/T1337R mutation) recognizes NGAG. In addition, VRER variant ((D1135V/G1218R/R1335E/T1337R mutation) shows activity
when the PAM sequence is NGCG (226). Other bacteria species can recognize different PAM sequences with different length of targeting sequences (Table 1.2).

<table>
<thead>
<tr>
<th>Cas9 species</th>
<th>Size (bp)</th>
<th>Target length</th>
<th>PAM sequence (5’ to 3’)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. pyogenes</em></td>
<td>4104</td>
<td>20</td>
<td>NGG</td>
<td>(221)</td>
</tr>
<tr>
<td>(SpCas9)</td>
<td></td>
<td></td>
<td></td>
<td>(222)</td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>3345</td>
<td>21 to 23</td>
<td>NGRRN or NGRRT</td>
<td>(227)</td>
</tr>
<tr>
<td>(SaCas9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>N. meningitidis</em></td>
<td>3246</td>
<td>23, 24</td>
<td>NNNNGATT</td>
<td>(228)</td>
</tr>
<tr>
<td>(NmCas9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. jejuni</em></td>
<td>2952</td>
<td>22</td>
<td>NNNNRYPAC or NNNNACAC</td>
<td>(229)</td>
</tr>
<tr>
<td>(CjCas9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. thermophilus</em></td>
<td>4227</td>
<td>19 to 20</td>
<td>NNAGAAW</td>
<td>(221)</td>
</tr>
<tr>
<td>(StCas9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>T. denticola</em></td>
<td>4185</td>
<td>20</td>
<td>NAAAAC</td>
<td>(230)</td>
</tr>
<tr>
<td>(TdCas9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.2. Variant of Cas9 protein from different bacteria species. Each Ca9 protein recognizes different PAM sequences with a different length of the targeting sequence. In the PAM sequence N could be any nucleotide: A, T, C or G; R could be A or G nucleotide; Y could be T or C nucleotide and W could be A or T nucleotide.

Other variants were also created for spCas9 to enhance the specificity of the CRISPR system. Creation of a specific mutation in one of the nuclease domains (either H840A for RuvC or D10A for HNH) makes the Cas9 nick only one strand (231). Using the two gRNAs targeting opposite DNA strands close to each with the nickase Cas9 will make a DSB and increase the specificity as it is unlikely to have similar two off-target sequences (221). In addition, a study carried on reducing the off-target effect by understanding the interaction energy between
the CRISPR/Cas9 and the targeting DNA sequence conclude that this complex might have more energy than is needed for optimal targeting and reducing this energy might help in reducing the off-targeting binding. Kleinstiver and colleagues studied this hypothesis by creating a different combination of the spCas9 variant at the hydrogen residue (N497, R661, Q695, Q926), which forms bonds with the phosphate backbone of targeting DNA. They found by substitution in all four residues to alanine reduce almost all the off-target effects while maintaining most the sensitivity to the SpCas9 (232). In a similar approach, a study led by Dr. Ian Slaymaker improved the specificity of SpCas9 by attenuating the helicase activity. The new variant of Cas9 (enhanced specificity SpCas9 (eSpCas9)) has silent mutations at HNH and RuvC domains, which reduce the interactions between them and the non-targeted DNA strand. As a result, the strand separation is prevented, which is needed for cutting at the off-target sites (233). In addition, the reduction of target binding affinity between the target site and Cas9 is a reason to enhance specificity in a new study. Janice and colleagues found the non-catalytic domain (REC3) has a role for active confirmation the HNH nuclease domain when matched with targeting site. By introducing silent mutation at the REC3 domain, they created a hyper-accurate Cas9 variant (HypaCas9) with more specificity for targeting the site (234).

The Cas9 variants described above guided by gRNA (crRNA and tracrRNA) create blunt DSB near the PAM site. However, another variant of CRISPR system belonging to class II originally named Cpf1 (CRISPR from Prevotella and Francisella) (235) then reclassified as Cas12a (236) shows different features. It requires only one RNA guide, which is crRNA, and lacks tracrRNA while the Cas12a contains RuvC as the nuclease domain and lacks the HNH domain. In addition, this system prefers T-rich region at 5’ prime as the PAM motif. PAM was initially described 5’TTTN (219), but it was found later that the first nucleotide at PAM could be A or C or G but not T (237). Furthermore, Cas12a creates DSB at the opposite side and leaves 4 or 5 overhanging nucleotides. This system might have advantages over CRISPR/Cas9. Firstly, the cleaves from Cas12a occur at the far end from the PAM and are unlikely to have indels either at seed sequence or PAM site. Thus, the same region could
be a target again. Secondly, the overhang cut by Cas12a might give advantage for HDR. Finally, this system requires only one RNA molecule, and thus might easy for transfection.

Another important variant of Cas9 is inactivating the nuclease domains by silent mutation (D10A and H840A mutation on RuvC and HNH respectively), which causes dead Cas9 (dCas9). This variant has will bind to the target region but not make a DSB (238).

1.6.3.2. CRISPR/Cas9 application

The CRISPR/Cas9 system has numerous applications, which depend on the type of Ca9 variant used as well as the presence or absence of the donor template. Gene editing by knockout genes and eliminated function in the cell or animal is an important technique to study the gene effect or create a disease model. The CRISPR allows for targeting the coding region and introducing a premature stop codon or frameshift at the open reading frame as result of indels after the DSB. Cloning the manipulated cells and identifying the indles that have been introduced in each clone and the effect will be needed to make pure knocked-out cells. In animal models, breeding might be necessary to have homozygous knockout. The knockout strategy could be performed by targeting single genes or more than one gene by multiplexing guides (221). Several knockout animals have been created by CRISPR/Cas9 system such as mice (224), pigs (239), zebrafish (240) and rats (241). Another approach to large knockout screening is using a library of pooled gRNAs delivered by lentivirus. Feng Zhang’s lab created a library targeting 18,080 human genes with 64,751 unique gRNA (3-4 gRNAs per gene) then transduced it into A375 melanoma cells. This work lead to the identification of the essential genes for cell viability in cancer using gene set enrichment analysis (GSEA) (242).

NHEJ could be useful for restoring gene function by deleting the mutated DNA sequence. An example is the exon skipping strategy for Duchenne Muscular Dystrophy (DMD) disease. Ousterout et al. 2015, used CRISPR/Cas9 to target mutated exons 45–55 to restore the dystrophin expression in patient myoblasts cells (243). The same principle proved in vivo by injecting mdx mice zygote
carrying nonsense mutation at \textit{Dmd} exon 23 with CRISPR LbCpf1 restore the expression dystrophin in mice cardiomyocytes (244).

The CRISPR system is a flexible tool for genetic manipulation without creating DSB by using the inactive variant of Cas9 (dCas9). One application is to repress gene expression by preventing transcription factor binding at the promoter region, blocking the transcriptional initiation (245). The other way of repressing the gene by blocking RNA polymerase (RNAP) elongation is binding the dCas9 complex at either the untranslated region (UTR) or the non-template DNA strand (238). In addition, fusing transcriptional repressor to dCas9 is another strategy to suppress the gene. A good example of this is the Krüppel-associated box (KRAB), which induces heterochromatinization (223).

Promoting gene expression is feasible by linking the dCas9 with a transcriptional activator. An example is fusing the dCas9 with the VPR activator (VP64, p65AD, and Rta) (246). The synergistic activation mediator (SAM) protein complex, which consists of VP64, P65 and heat shock factor 1 (HSF1), is incorporated into RNA aptamers for the MS2 engineered for the transcriptional activation of endogenous genes (247). Another molecule has been fused to dCas9 called SunTag, which can recruit antibodies such as Gcn4 bound to transcription factors in order to activate gene expression (248).

Another approach to control gene expression is by alternate epigenetics by activation or suppression without sequence modification. Hilton et al. 2015, show the fusion acetyltransferase p300 to dCas9 and, targeting four gene promoters, increase the acetylation of histone H3 lysine 27, which results in gene activation (249). Induce DNA demethylation is another way of epigenetic modification by Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) enzyme (250). However, epigenetic modification can cause gene repression. For example, Peter Stepper and colleagues were able to target EpCAM, CXCR4 and TFRC gene promoters by methylation using DNA Methyltransferase 3 Alpha (DNMT3A) fused to dCas9 (251).

Another application of dCac9 is the base editing technique. Alexis Komor and colleagues successfully fused cytidine deaminase to dCa9 and converted cytosine (at position 4-8 in the non-targeted strand) to uracil without cutting DNA,
where uracil became thymine through DNA replication (252). Another base editor that has been created is adenine base editing (A>G change) through several rounds of molecular evolution. This technique is able to change a single base with efficiency up to 50% in human cells (253).

Correcting the mutation or inserting new elements to a specific location is another feature that can be introduced at a specific location using HDR after a DSB using CRISPR. The HDR requires a donor template with homology arms (right and left) flanks on the cutting site. Numerous attempts have been made to study the effect of homology arm length on HDR. The homology arm between 500 bp and 3 kb is needed to introduce knock-in at eukaryotic cells (254, 255). The donor template could be double strand DNA or single-stranded DNA (ssDNA).

In general, the NHEJ is more efficient than HDR. As such, incorporating section marker gene in the donor template helps to enrich the knock-in cells. The selection marker could be antibiotic resistance genes such as puromycin or fluorescent genes such as GFP. The knock-in cells are usually after the selection grown as single clones, which need to confirm the correction at allelic level (homozygous or heterozygous). Removing the selection marker might be necessary especially for any in vivo work as the selection marker gene might have interference with gene expression, through integration or with neighbor genes (256). The most common system of excision is the Cre-lox system, where the Cre recombinase protein recognizes the 34 bp loxP site at both ends of the selection marker and removes it. While this system is effective, there are concern about the Cre mediation including proliferation defects and cytotoxicity (257). In addition, after the excision, one loxP site remains, which means that this system can be used only at an intron locus. The piggyBac transposase system is another approach for selection marker removal and is footprint-free after the excision. The main limitation for this system is transposition efficiency compared to the Cre-lox system (258).
1.6.3.3. CRISPR/Cas off-targeting

Off-targeting is the main drawback of the gene editing technique and for future work as therapeutic approach. When the CRISPR system targets a specific sequence in the genome, there is a similar site at the genome, which might differ in one nucleotide or more, which might cause DSB at the off-target region. Optimizing the two components of the CRISPR system (gRNA and Cas9) helps to give more specificity. The crucial step is designing a gRNA that has a low off-targeting possibility, especially at the seeding region, which are the last 12 nucleotides before the PAM sequence (238). In fact, one and double mismatches at that region in general does not eliminate the activity of the CRISPR system to make a DSB (259). Interestingly, work led by Yanfang Fu was able to decrease off-targeting dramatically by designing a shorter gRNA (17 or 18 nucleotides) without significantly affecting the efficiency of the system (260). Potential off-targeting is also reduced by using the CRISPR/Cas9 nickase system, which contains two opposite mutated cas9 (D10A) and each guide RNA cleaves only one DNA strand. This technique increases specificity because it is unlikely to have similar two off-target sequences close to each other. The off-target activity by this system shows a reduction by 50–1,000 fold (221). Similarly, two gRNA fused with dCas9 and FokI nucleases only make a DSB when dimerized (261). Another variant of Cas9 (section 1.6.3.1) has been created to increase the system specificity.

1.7. Gene editing for cystic fibrosis

CF is attractive when it comes to gene editing techniques. Editing has been performed using all three genes editing technique with and without selection markers and on different cells. A study detecting the efficiency of the correction in whole population of cells using ZFNs cutting at CFTR exon 10 without selection has been conducted by Lee and colleagues. The HDR correction percentage for ΔF508 mutation in CFTR29- cells was very low (<1%), which was expected to be a result of the low efficiency of double-strand DNA break (DSB) (7.8%) (262). On other hand, transcription activator-like effector nucleases (TALENs) have been used to correct the ΔF508 mutation using single-strand oligo as a donor template,
which has around 40 base homology arms without selection markers. The transfected cells cloned after that then are screening by polymerase chain reaction (PCR) to identify the corrected clone. From the five clones that were successfully isolated with mono-allelic correction, one clone was characterized for preserving the stemness marker and the ability to differentiate to all three germ layers (263). A similar work has been carried out on isolating iPSCs from ΔF508 CF patients then genetically modifying them using TALENs and a small DNA fragment (double strand) carrying the wt-CFTR as a donor template. The isolated corrected clones were achieved by cycles of allele-specific PCR enrichment strategy. Furthermore, the corrected induced pluripotent stem cell (iPSC) clones differentiated to airway-like epithelial cells, restoring the electrophysiology property once co-cultured with CFBE41o- cells (264, 265). The genetic mutation also can be possibly rescued by using a super-exon as strategy to cover more than one mutation or integrating a full length of cDNA either at the endogenous locus or at safe harbor locus (adeno-associated virus integration site 1 (AAVS1)) (266). The super-exon strategy for CFTR mutation has been studied by making DSB by zinc finger endonucleases (ZFNs) at CFTR exon 11 then using a donor template to carry the super-exon wtCFTR 11 to 27 (267). iPSCs are also attractive to validate the idea for correcting the CFTR mutation and as possible therapy in future. An example of this application has been shown by taking skin fibroblasts of patients diagnosed with CFTR ΔF508 mutation then correcting them using ZFNs, which are differentiated to mature epithelium cells and restore CFTR function. The corrected cells show the ability of gene editing to correct the mutation and as well as to restore the full functioning of the CFTR. In addition, the puromycin gene flanked with loxP sequence has been integrated in the intronic locus of the CFTR to facilitate the selection removed by recombination using Cre plasmid (268).

The genetic engineering technique has also been used to knockout the CFTR to study the effect on animal models. Zebrafish is an excellent research model for in vivo; however, the role of CFTR in fluid regulation during the development is not well studied. TALENs were injected in the yolk of the zebrafish targeting the
exon 6 of CFTR. The knockout showed damage to Kupffer's vesicle (KV) lumen expansion and function, which is a ciliated organ in zebrafish (269). The CRISPR/Cas9 system has also been tested to correct the ΔF508 mutation in the presence of a donor template. Selected clones of transfected intestinal stem cells taken from CF patients were able to respond to a forskolin assay, which led to the swelling of the organoid (98). Similar work performed on iPSCs with a ΔF508 homozygous mutation proved the ability of this system to perform gene editing. Interestingly, the piggyBac transposase system used to excise the selection cassette was footprint-free compare to other systems (270).

1.8. Barriers for cystic fibrosis gene editing
For proper gene editing therapy for CF, there are multiple concerns and questions, such as; How much of the cells need to correct in order to rescue the CFTR function? Which cells on the airways tissue need to be targeted? How can we overcome the thick mucus layer and cell immunity on CF patients?

Different studies have looked at the amount of CFTR expression necessary to improve lung function. One early study in the 1990s indicated that a correction of 6 to 10% of cells is enough to restore proper CFTR function based on cell mixing experiments (271). However, a later study suggests that a high percentage of cells (>25%) express the CFTR needed to restore mucus clearance in an in-vitro experiment (272). An in vivo study on CF knockout pigs infected by adenovirus-mediated CFTR gene showed increased CFTR expression to 50% of non-CF levels when 17 to 28% of sinus epithelial cells has been infected (273). These percentages need further investigation especially in vivo as genetic and non-genetic modifiers might have an effect. In addition, the assessment might need to be extended and involve other channels interacting with CFTR such ENaC to evaluate the restoring function needed for clinical benefit.

The other question for efficient gene therapy or gene editing is, Which types of cells on the airways tissue are the most appropriate for targeting? Studies on human proximal bronchus using in situ hybridisation and immunochromehemistry to characterize the distribution of CFTR in cells showed the CFTR expression (protein and mRNA) primarily localized in submucosal glands while the surface
airway epithelium has less CFTR expression (274, 275). These findings were challenged by a later study using high-affinity CFTR mAbs with ion transport assays, where it was found that the CFTR expression is more abundant in the apical membrane of ciliated cells within the superficial epithelium airway tissue (276). Recently, a new cell type has been identified within human bronchial epithelial tissue and mouse tracheal epithelial tissue called pulmonary ionocyte. These cells are a major source of CFTR expression (60%) compared to the 4% expression in ciliated cells (277, 278). Targeting the superficial epithelial cells is more feasible for nebulization non-viral gene editing therapy in comparison with submucosal glands, which lie beneath the surface epithelium of the airway tissue. In addition, targeting stem cells in the airway epithelium tissue might appropriate for persistence effects; however, these cells in general are located deeper in the tissues (279).

The other consideration for any editing therapy is the lifespan of the targeting cells. The average half-life of ciliated epithelial cells is almost 6 months in mice trachea tissue whereas in the lung it is about 17 months (280). The lifespan of human ciliated epithelial cells is unknown but might be similar to the mice, and might be reduced in disease conditions (281).

In addition to the previous concern about appropriate cells for CFTR gene editing, other barriers need to overcome before reaching the target cells, such as airway cell immunity and thick mucus specially in CF patients. In the airway tract, the immune cells such as macrophages can destroy the delivery carrier, as seen on the adenovirus vector (282, 283). In a non-viral therapy approach, Tammy and colleagues showed the modification of the nanoparticles with PEGylation delaying macrophage clearance as well as increasing the residency time in the lung (284). Mucus in CF patient is also a barrier for any kind of gene editing or gene therapy. To increase the transfection efficiency, two techniques can be taken: 1) reducing the mucus thickness and improving the clearance as described previously in CF therapy (section 1.4) and 2) minimising the mucus adherent effect by coating the cationic nanocomplex with polyethylene glycol (PEG). Benjamin and his colleagues show an improvement of penetration of the nanoparticles through human mucus by around 25-fold, in comparison with
uncoated particles (285). In addition, poly (β-amino esters) (PBAEs) polymer demonstrates the ability to penetrate through mucus freshly expectorated by CF patients (286).

1.9. Receptor-targeted nanocomplexes (RTN)
The liposome on general has lower transfection in comparison with viral vectors, especially in vivo due to lack of specificity in terms of targeting, and also is easily cleared by the reticuloendothelial system (RES) (287). To improve the specificity of liposome targeting, many models have been proposed, for example using mannose ligand, which is expressed by many bacteria and viruses to target mannose receptors on the cells (288). In addition, pharmaceutical molecules encapsulated by liposome coated with specific monoclonal antibodies have been successfully introduced for cancer therapy to decrease the side effects of toxicity on normal cells (289). The third type is conjugate liposome with a targeting peptide sequence to deliver nucleic acid into specific cells (Figure 1.19.). This formulation is called receptor-targeted nanocomplexes (RTN), formed from three components: 1) cationic lipid with DOPE to enhance the endosomal escaping, 2) targeting peptide and 3) nucleic acid. The aim of peptide is to bind with nucleic acid from one side through a chain of cationic amino acids such as lysine, arginine or ornithine, and on other end the specific peptide ligand can target definite receptors on the cell membrane. In addition, a spacer in the middle can give flexibility to the receptor binding. Previous work on lung epithelium cells has led to the development of a novel targeting peptide (SERSMNF) through a phage display technique. This peptide (E) sequence shows the ability to bind to intercellular adhesion molecule-1 (ICAM-1) receptor on the airway epithelium, which interestingly is targeted by rhinovirus through a similar protein (SDRSMN) (290). This peptide in nanocomplex (RTN) mediates targeting to airway cells in vitro and in vivo (291, 292). In addition, another peptide (Y) with the peptide ligand (YGLPHKF) shows a similar targeting protein expressed by the intracellular pathogen Legionella pneumophila (293). Although the receptor on the cells is unknown, this peptide shows the ability to mediate targeting delivery
of plasmid DNA and siRNA in nanocomplexes to airway cells as well as cell of neuronal origin (294, 295)

In addition, cationic lipid with different alkyl tail such as DTDTMA (1,2-Di-((Z)-tetradec-11-enyloxy)-N,N,N trimethylammonium propane chloride, C-11-), DHDTMA (1,2-Di-((Z)-hexadec-11-enyloxy)-N,N,N trimethylammonium propane iodide), DOTMA (1,2-Di-((Z)-octadec-9-enyloxy)-N,N,N trimethylammonium propane chloride) has effect on the transfection efficiency (Figure 1.20.). Tagalakis et al. 2011, showed that the DHDTMA is more efficient than DOTMA for DNA transfection on B104, 1HAEo- and Neuro-2A cells. On the other hand, DHDTMA was less efficient when it was used to deliver siRNA in comparison with DOTMA lipid (294).

Figure 1.19. Receptor-targeted nanocomplexes (RTN). A) The RTN scheme, which consists of DNA encapsulated with lipid and peptide. B) Example of the peptide chain structure: nucleic binding side (K16) connected to the peptide receptor, which here is SERSMNF target ICAM-1 on the cells. The spacer amino acid (glycine–alanine) in the middle provides flexibility. Adapted from (296).
Figure 1.20. Examples of lipids (DOPE (neutral), DTDTMA, DHDTMA and DOTMA (cationic), which have been used to deliver nucleic acid molecules formulated with a target peptide. Adapted from (296, 297).
1.10. Aims and objectives

Cystic fibrosis (CF) is a common genetic disease with multi-organic effects and is lethal when it involves the lungs. The disease is linked to mutations in the *CFTR* gene. More than 300 mutations have been confirmed as causing CF. Since the *CFTR* gene was identified in 1989 (16, 17), several clinical trials have attempted to develop a gene therapy approach for CF lung disease; however, a clinically relevant treatment has yet to emerge (298). The major challenges in gene therapy for CF relate to the limited levels of gene transfer achieved in the lung airway epithelium, and the persistence of transgene expression. A great breakthrough in terms of CF treatment is the modulator therapy where small molecules target the malfunctioning CFTR protein. Two CFTR modulators have been approved for the ∆F508 CFTR mutation which represents almost 70% of all cases: Orkambi® (299) and Symdeko™ (162). However, the high cost of the modulator therapy for ∆F508 patients and the moderate benefit, means that it is limited to a few patients for whom the healthcare system is able to cover the cost (161). In the UK this therapy has yet to be approved.

In this study we aim to investigate the potential of genome editing to develop a genetic therapy for CF using the CRISPR/Cas system. This allows the gene-specific, targeted correction of disease-related mutations to be introduced at the chromosomal level, with a therapeutic potential for the ∆F508 CFTR mutation.

In this study we aim to investigate the potential of genome editing to develop a genetic therapy for CF using the CRISPR/Cas system. This allows the gene-specific, targeted correction of disease-related mutations to be introduced at the chromosomal level, with a therapeutic potential for the ∆F508 CFTR mutation.

Moving towards gene editing for ∆F508 cystic fibrosis the objectives of this thesis are:

1. To optimise CRISPR/Cas9 transfection in cell models *in vitro* before it is used later for *in vivo* transfection;
2. To correct the ΔF508 mutation in a cystic fibrosis human bronchial epithelial cell line (CFBE41o-) mediated by CRISPR/Cas9 through Homology directed repair (HDR) pathway and assessing the functional outcome following the correction;

3. To optimise transfection of the CRISPR system in cystic fibrosis human bronchial epithelial primary BMI-1 cells using Cas9 as mRNA or protein delivered by Receptor Targeted Nanocomplexes (RTN);

4. To attempt ΔF508 mutation correction in CF primary BMI-1 cells mediated by CRISPR/Cas9 through the HDR pathway and assessing functional outcome following the correction;

5. To evaluate the homology-independent targeted integration (HITI) strategy as an alternative approach for HDR to knock-in GFP in the Ai9 cell model and in the replacement of the CFTR exon 10.
CHAPTER 2

Materials and Methods
2. Materials and Methods

2.1. Materials

2.1.1. Equipment

<table>
<thead>
<tr>
<th>Name</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ussing chamber</td>
<td>World Precision Instruments</td>
</tr>
<tr>
<td>Bio-Rad 96 CFX</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Canon Rebel XS DS126191 Digital Camera</td>
<td>Canon</td>
</tr>
<tr>
<td>BD LSRII</td>
<td>BD Biosciences</td>
</tr>
<tr>
<td>FACS Calibur</td>
<td>BD Biosciences</td>
</tr>
<tr>
<td>MoFlo XDP Cell Sorter</td>
<td>Beckman Coulter</td>
</tr>
<tr>
<td>Zeiss Axioplan Scopes</td>
<td>Zeiss</td>
</tr>
<tr>
<td>Trans-Blot Turbo Transfer System</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>NanoDrop ND-1000 Spectrophotometer</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>Olympus IX70 microscope</td>
<td>Olympus</td>
</tr>
<tr>
<td>Thermocycler (Mastercycler ep)</td>
<td>Eppendorf</td>
</tr>
<tr>
<td>Gel Doc XR System</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Ohmmeter (EVOM)</td>
<td>World Precision Instruments</td>
</tr>
<tr>
<td>Zetasizer Nano ZS</td>
<td>Malvern</td>
</tr>
</tbody>
</table>

2.1.2. Kits and reagents

<table>
<thead>
<tr>
<th>Name</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMEM+GlutaMAX (Dulbecco's modified eagle media)</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>GeneRuler 1 kb Plus DNA Ladder</td>
<td>Thermo Scientific Scientific</td>
</tr>
<tr>
<td>6x DNA Loading Dye</td>
<td>Thermo Scientific Scientific</td>
</tr>
<tr>
<td>Q5 Hot Start High-Fidelity DNA Polymerase</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>L-glutamine (100x)</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>NuPAGE 4x loading dye buffer</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>NP40 Cell Extraction Buffer</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>dNTPs</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>Eagle's Minimal Essential Medium (MEM)</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Bolt Sample Reducing Agent 10X</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>0.05% Trypsin-EDTA</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>Item</td>
<td>Supplier</td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Foetal Bovine Serum (FBS)</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>B-Mercaptoethanol</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Lysogeny broth (LB) Agar</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>LB Broth</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>GeneJET Gel Extraction</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>GeneJET PCR Purification</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>GeneJET Plasmid Maxiprep</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>GeneJET Plasmid Miniprep</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>GeneJET Genomic DNA Purification</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>PureCol, bovine collagen</td>
<td>Nutacon</td>
</tr>
<tr>
<td>Pierce ECL Western Blotting Substrate</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>SuperSignal West Femto Maximum Sensitivity Substrate</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>Penicillin/Streptomycin (10000U/mL)(100x)</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>OptiMEM</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>BEBM Basal Medium</td>
<td>Lonza</td>
</tr>
<tr>
<td>SingleQuot Kit Suppl. &amp; Growth Factors</td>
<td>Lonza</td>
</tr>
<tr>
<td>SOC Media</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>T4 DNA Ligase and Buffer</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>4-20% pre-cast polyacrylamide gels</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>10x Tris Buffered Saline (TBS)</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Trans-Blot Turbo Mini PVDF Transfer Packs</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>CloneJET PCR Cloning Kit</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>Non-Essential Amino Acids (NEAA) (100x)</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>Precision Plus Protein Dual Color standards</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>NEBuilder HiFi DNA Assembly</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>HiScribe T7 High Yield RNA Synthesis Kit</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>T7 Endonuclease I</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>CleanCap Cas9 mRNA (modified)</td>
<td>TriLink BioTechnologies</td>
</tr>
<tr>
<td>CleanCap Cre mRNA (5moU)</td>
<td>TriLink BioTechnologies</td>
</tr>
<tr>
<td>Alt-R S.p. Cas9 Nuclease 3NLS</td>
<td>Integrated DNA Technologies</td>
</tr>
<tr>
<td>SensiFAST Probe Hi-ROX one-step kit</td>
<td>Bioline</td>
</tr>
<tr>
<td>SYBR Safe DNA Gel Stain</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>Qiagen RNeasy mini kit</td>
<td>Qiagen</td>
</tr>
<tr>
<td>Pierce Protein assay BCA kit</td>
<td>Thermo Fisher Scientific</td>
</tr>
</tbody>
</table>
### 2.1.3. Bacterial Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Supplier</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP10</td>
<td>Invitrogen, Life Technologies</td>
<td>F– mcrA Δ (mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ (ara leu) 7697 galU galK rpsL (StrR) endA1 nupG</td>
</tr>
<tr>
<td>DH5α</td>
<td>New England Biolabs</td>
<td>F– Φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17 (rK-, mK+) phoA supE44 λ– thi-1 gyrA96 relA1</td>
</tr>
</tbody>
</table>

### 2.1.4. Eukaryotic cells

<table>
<thead>
<tr>
<th>Name</th>
<th>Supplier</th>
<th>Growth Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>16HBE14o-1</td>
<td>Prof. Dieter Gruenert</td>
<td>MEM with 10% FBS, 2mM L-glutamine and NEAA</td>
</tr>
<tr>
<td>CFBE41o-1</td>
<td>Prof. Dieter Gruenert</td>
<td>MEM with 10% FBS, 2mM L-glutamine and NEAA</td>
</tr>
<tr>
<td>HEK293T</td>
<td>ATCC</td>
<td>DMEM+GlutaMAX with 10% FBS</td>
</tr>
<tr>
<td>Neuro-2A</td>
<td>ATCC</td>
<td>DMEM+GlutaMAX with 10% FBS</td>
</tr>
<tr>
<td>NHBE BMI-1</td>
<td>Dr. Mustafa Munye (93)</td>
<td>BEGM or ALI media</td>
</tr>
<tr>
<td>CFBE BMI-1</td>
<td>Dr. Mustafa Munye (93)</td>
<td>BEGM or ALI media</td>
</tr>
</tbody>
</table>

### 2.1.5. Lipids

<table>
<thead>
<tr>
<th>Lipid</th>
<th>Structure</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE)</td>
<td><img src="image1" alt="DOPE Structure" /></td>
<td>Avanti Polar Lipids</td>
</tr>
<tr>
<td>1,2-Di-(Z)-tetradec-11-enoxy-N,N,N-trimethylammonium propane I(II)Cl, C-11-DTDTMA</td>
<td><img src="image2" alt="DTDTMA Structure" /></td>
<td>Avanti Polar Lipids</td>
</tr>
<tr>
<td>1,2-Di-(Z)-hexadec-11-enoxy-N,N,N-trimethylammonium propane iodide (DHDTMA)</td>
<td><img src="image3" alt="DHDTMA Structure" /></td>
<td>Avanti Polar Lipids</td>
</tr>
<tr>
<td>1,2-Di-(Z)-octadec-9-enoxy-N,N,N-trimethylammonium propane chloride (DOTMA)</td>
<td><img src="image4" alt="DOTMA Structure" /></td>
<td>Avanti Polar Lipids</td>
</tr>
</tbody>
</table>
### 2.1.6. Composition of Lipid nanoparticle

<table>
<thead>
<tr>
<th>Liposome name</th>
<th>Lipid1</th>
<th>Lipid 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOTMA:DOPE (C18)</td>
<td>DOTMA (50.0%)</td>
<td>DOPE (50.0%)</td>
</tr>
<tr>
<td>DHDTMA:DOPE (C16)</td>
<td>DHDTMA (50.0%)</td>
<td>DOPE (50.0%)</td>
</tr>
<tr>
<td>DTDTMA:DOPE (C14)</td>
<td>DTDTMA (50.0%)</td>
<td>DOPE (50.0%)</td>
</tr>
</tbody>
</table>

### 2.1.7. Peptides

<table>
<thead>
<tr>
<th>Name</th>
<th>Composition</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>K16GACYGLPHKFCG</td>
<td>China Peptides Co.</td>
</tr>
<tr>
<td>E</td>
<td>K16GACSERSMNFCG</td>
<td>China Peptides Co.</td>
</tr>
</tbody>
</table>

### 2.1.8. plasmids

<table>
<thead>
<tr>
<th>Name</th>
<th>Transgene and resistance gene</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>pX-330</td>
<td>SpCas9 and chimeric guide RNA expression</td>
<td>Addgene #42230 (221)</td>
</tr>
<tr>
<td>pX-601</td>
<td>SaCas9 and chimeric guide RNA expression</td>
<td>Addgene #61591 (227)</td>
</tr>
<tr>
<td>pY094</td>
<td>huAsCpf1-T2A-GFP and crRNA guide</td>
<td>Addgene #84743 (300)</td>
</tr>
<tr>
<td>Ai9</td>
<td>tdTomato with stop codons (donor template at Rosa26) Neomycin-resistance (neo) gene</td>
<td>Addgene #22799 (301)</td>
</tr>
<tr>
<td>pEGFP-N1</td>
<td>EGFP</td>
<td>Clontech</td>
</tr>
<tr>
<td>pY004</td>
<td>FnCpf1</td>
<td>Addgene #69976 (219)</td>
</tr>
<tr>
<td>Donor CFTR-G</td>
<td>Donor template to correct ΔF508 mutation puromycin-resistance (pac) gene</td>
<td>Dr. Brian Davis (268)</td>
</tr>
<tr>
<td>pMC.BESPX.MCS2</td>
<td>Non, parental plasmid for minicircle</td>
<td>System Biosciences</td>
</tr>
<tr>
<td>pCAG-Cre:GFP</td>
<td>Mammalian expression of Cre recombinase fused to GFP</td>
<td>Addgene #13776 (302)</td>
</tr>
</tbody>
</table>
2.1.9. Taqman Assays (FAM-labelled probes)

All probes were purchased from Thermo Fisher Scientific.

<table>
<thead>
<tr>
<th>Name</th>
<th>Catalogue #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human ACTB</td>
<td>Hs01060665_g1</td>
</tr>
<tr>
<td>Human CFTR</td>
<td>Hs00357011_m1</td>
</tr>
<tr>
<td>Human SCNN1A</td>
<td>Hs00168906_m1</td>
</tr>
<tr>
<td>Human SCNN1B</td>
<td>Hs01548617_m1</td>
</tr>
<tr>
<td>Human SCNN1G</td>
<td>Hs00168918_m1</td>
</tr>
</tbody>
</table>

2.1.10. Antibodies for immunoblotting

2.1.10.1. Primary antibodies

<table>
<thead>
<tr>
<th>Name</th>
<th>Supplier</th>
<th>Catalogue #</th>
<th>Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFTR 596</td>
<td>UNC-Chapel Hill</td>
<td>A4 (antibody code)</td>
<td>1:2000</td>
</tr>
<tr>
<td>β-Actin (AC-74)</td>
<td>Sigma-Aldrich</td>
<td>A2228</td>
<td>1:10000</td>
</tr>
<tr>
<td>Na+/K+-ATPase α Antibody (H-3)</td>
<td>Santa Cruz</td>
<td>sc-48345</td>
<td>1:20000</td>
</tr>
</tbody>
</table>

2.1.10.2. Secondary antibodies

The Secondary antibodies for immunoblotting were horseradish peroxidase conjugated (HRP-conjugated) anti-IgG antibodies (Dako, Agilent Technologies) at dilution (1:10000).

2.1.11. Antibodies for Flow Cytometry

<table>
<thead>
<tr>
<th>Name</th>
<th>Supplier</th>
<th>Catalogue #</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC Rat Anti-Mouse CD45</td>
<td>BD Biosciences</td>
<td>561018</td>
</tr>
<tr>
<td>BV421 Rat Anti-Mouse CD326</td>
<td>BD Biosciences</td>
<td>563214</td>
</tr>
</tbody>
</table>

2.1.12. Recipe

<table>
<thead>
<tr>
<th>Solution</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBST</td>
<td>50 mM Tris/HCl, pH 7.5, 150 mM NaCl, 0.1% Tween 20™</td>
</tr>
<tr>
<td>Innoue Transformation Buffer</td>
<td>10mM PIPES, 15mM CaCl₂.2H₂O and 55mM MnCl₂.4H₂O; adjusted to pH6.7</td>
</tr>
</tbody>
</table>
### 2.1.13. primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-&gt;3’)</th>
<th>Amplified product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP_primer_Forward</td>
<td>ACGTAACGGCCACAGTTC</td>
<td>546</td>
</tr>
<tr>
<td>GFP_primer_Reverse</td>
<td>TGTCAGGGTAGGGTGTGGTCG</td>
<td></td>
</tr>
<tr>
<td>ROSA26_primer_Forward</td>
<td>ACGTTTCCGACTGAGTGTCG</td>
<td>660</td>
</tr>
<tr>
<td>ROSA26_primer_Reverse</td>
<td>ACTCGGGTGAGCATGTCTTT</td>
<td></td>
</tr>
<tr>
<td>CFTR_primer_Forward</td>
<td>AGAACACGACTCGACACAGA</td>
<td>629</td>
</tr>
<tr>
<td>CFTR_primer_Reverse</td>
<td>ACCGATTGAATATGGAGCCAA</td>
<td></td>
</tr>
<tr>
<td>CFTR_CFBE41o-- 5’end HDR primer_Forward</td>
<td>AAATCAGTGCTTTTTTCAGGGTTAGGAG</td>
<td>1785</td>
</tr>
<tr>
<td>CFTR_CFBE41o-- 5’end HDR primer_Reverse</td>
<td>CGACTGTGCCCTTCTAGGGTC</td>
<td></td>
</tr>
<tr>
<td>CFTR_CFBE41o-- 3’end HDR primer_Forward</td>
<td>CCTGCAGGCCGCGGGGATCTAT</td>
<td>923</td>
</tr>
<tr>
<td>CFTR_CFBE41o-- 3’end HDR primer_Reverse</td>
<td>GCCATTTCATGTAAGCATGCAACCC</td>
<td></td>
</tr>
<tr>
<td>CFTR_ outside donor primer_ Forward</td>
<td>AAATCAGTGCTTTTTTCAGGGTTAGGAG</td>
<td>1835</td>
</tr>
<tr>
<td>CFTR_ outside donor primer_ Reverse</td>
<td>TTCTCTGCTGGCAGTCAATGCTCA</td>
<td></td>
</tr>
<tr>
<td>CFTR_ 5’end outside donor primer_Forward</td>
<td>AAATCAGTGCTTTTTTCAGGGTTAGGAG</td>
<td>1450</td>
</tr>
<tr>
<td>CFTR_ 5’end outside donor primer_Reverse</td>
<td>ACCGATTGAATATGGAGCCAA</td>
<td></td>
</tr>
<tr>
<td>CFTR_CFBE BMI-1 3’end HDR primer_Forward</td>
<td>TCGAGCCTGGAAATGTTGTCAGTTAG</td>
<td>2620</td>
</tr>
<tr>
<td>CFTR_CFBE BMI-1 3’end HDR primer_Reverse</td>
<td>GCCATTTCATGTAAGCATGCAACCC</td>
<td></td>
</tr>
<tr>
<td>Primer Set</td>
<td>5' End Primer</td>
<td>3' End Primer</td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>HITI_Ai9_5'end HDR primer</td>
<td>GCAACGTGCTGGTTATTGTG</td>
<td>TGCCTAGGTAAGTGTTGCG</td>
</tr>
<tr>
<td>HITI_Ai9_3'end HDR primer</td>
<td>ACGTAACCCGACACAAGGC</td>
<td>TCTTTGATGACGGCCATGT</td>
</tr>
<tr>
<td>intron 9-1_off-target-1_prime</td>
<td>TACAGGGGAGGGAACACAC</td>
<td>GCCATCCAATTAGGATTCTTC</td>
</tr>
<tr>
<td>intron 9-1_off-target-2_prime</td>
<td>TGGATGTCGAGAAAAACCTGTG</td>
<td>ACTTCTGAGCACCTGGCATTT</td>
</tr>
<tr>
<td>intron 9-1_off-target-3_prime</td>
<td>AAGTGCGCTGGATGATGAAG</td>
<td>CCAGGCTCCATTTCAAGA</td>
</tr>
<tr>
<td>intron 9-1_off-target-4_prime</td>
<td>CGTGAAGCAGCACAGAAAC</td>
<td>CGCAAGGTATCAGCAGACAA</td>
</tr>
<tr>
<td>intron 9-1_off-target-5_prime</td>
<td>TGGATTGCGCATATTGTG</td>
<td>CTGAAAAGGCAGCATGGTAA</td>
</tr>
<tr>
<td>intron 10-4_off-target-1_prime</td>
<td>TGCAAAATGCTGGACCAAGTG</td>
<td>CTAACTGCTTGCCCTTGAC</td>
</tr>
<tr>
<td>intron 10-4_off-target-2_prime</td>
<td>TCTGCTCTGGTCTGCTTCT</td>
<td>TGGCATGCAAGTCAACAAT</td>
</tr>
<tr>
<td>intron 10-4_off-target-3_prime</td>
<td>ACTCCTTACCACCACCACAG</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Primer Name</td>
<td>Forward Primer Sequence</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>intron 10</td>
<td>10-4_off-target-3 primer Reverse</td>
<td>TGAGAGGCTCGCTATTTCTCT</td>
</tr>
<tr>
<td>intron 10</td>
<td>10-4_off-target-4 primer Forward</td>
<td>CCATGGGCAAGGAATATAACA</td>
</tr>
<tr>
<td></td>
<td>10-4_off-target-4 primer Reverse</td>
<td>AGGTGCTACCTGGCTTTCA</td>
</tr>
<tr>
<td>intron 10</td>
<td>10-4_off-target-5 primer Forward</td>
<td>GCAATTGGAAGGCGTGTAECT</td>
</tr>
<tr>
<td></td>
<td>10-4_off-target-5 primer Reverse</td>
<td>AAACATTTTTAAAGACGTTTTCG</td>
</tr>
<tr>
<td>CFTR-T5+</td>
<td>T5+_off-target-1 primer Forward</td>
<td>ATGTATGGCTGTGTCGCCAG</td>
</tr>
<tr>
<td></td>
<td>T5+_off-target-1 primer Reverse</td>
<td>CAGGGGACAGTGACCAAGG</td>
</tr>
<tr>
<td>CFTR-T5+</td>
<td>T5+_off-target-2 primer Forward</td>
<td>CCCTGTGGATCTGTGTCTTCT</td>
</tr>
<tr>
<td></td>
<td>T5+_off-target-2 primer Reverse</td>
<td>CGGGGTGCTGAGTGTCATA</td>
</tr>
<tr>
<td>CFTR-T5+</td>
<td>T5+_off-target-3 primer Forward</td>
<td>CTGCTCTTCTGCTGCTACCT</td>
</tr>
<tr>
<td></td>
<td>T5+_off-target-3 primer Reverse</td>
<td>CCCACCTAAACCAACAGC</td>
</tr>
<tr>
<td>CFTR-T5+</td>
<td>T5+_off-target-4 primer Forward</td>
<td>TGTTGAGTGGTTTCTGGCGG</td>
</tr>
<tr>
<td></td>
<td>T5+_off-target-4 primer Reverse</td>
<td>TTACAGATGATGCTGAACCAC</td>
</tr>
<tr>
<td>CFTR-T5+</td>
<td>T5+_off-target-5 primer Forward</td>
<td>CCACGTTAGTCCCATGCTTT</td>
</tr>
<tr>
<td></td>
<td>T5+_off-target-5 primer Reverse</td>
<td>TCTCAGCACTTTGGGAGGTC</td>
</tr>
<tr>
<td>CFTR-T7</td>
<td>T7_off-target-1 primer Forward</td>
<td>ATAACTACCCCGCGCTGTTC</td>
</tr>
<tr>
<td></td>
<td>T7_off-target-1 primer Reverse</td>
<td>AACCTGGACCTTTCTGCCCT</td>
</tr>
<tr>
<td>CFTR-T7</td>
<td>T7_off-target-2 primer Forward</td>
<td>GGCTCAGCAGCTGAATCCCA</td>
</tr>
<tr>
<td></td>
<td>T7_off-target-2 primer Reverse</td>
<td>ATCAAGACCTTCCCCAACC</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>oligonucleotides</th>
<th>Sequence (5'-&gt;3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP_gRNA_Forward_SpCas9 T1</td>
<td>CACCGTGAAACCGCATCGAGCTGAA</td>
</tr>
<tr>
<td>GFP_gRNA_Reverse_SpCas9 T1</td>
<td>AAACCTTCAGCTCGATCAGGCCCTAC</td>
</tr>
<tr>
<td>GFP_gRNA_Forward_SpCas9 T2</td>
<td>CACCGAGCGGACCATCTTCTTCA</td>
</tr>
<tr>
<td>GFP_gRNA_Reverse_SpCas9 T2</td>
<td>AAACGGAAGAGATGGTGCGTC</td>
</tr>
<tr>
<td>GFP_gRNA_Forward_SaCas9 T1</td>
<td>CACCAGGAGGAGAGCTGACAC</td>
</tr>
<tr>
<td>GFP_gRNA_Reverse_SaCas9 T1</td>
<td>AAACGGAAGACGCTC</td>
</tr>
<tr>
<td>ROSA26_gRNA_Forward</td>
<td>CACCGCAGATCACGAGGGAGAGG</td>
</tr>
<tr>
<td>ROSA26_gRNA_Reverse</td>
<td>AAACCCCTCTCTCGTCTGATGCG</td>
</tr>
<tr>
<td>CFTR_gRNA_Forward_T1</td>
<td>CACCGCAGTTTTCTGATTATG</td>
</tr>
<tr>
<td>CFTR_gRNA_Reverse_T1</td>
<td>AAACGGAATATCCAGGAAAAAAC</td>
</tr>
<tr>
<td>CFTR_gRNA_Forward_T2</td>
<td>CACCGATATTTCTTTAATG</td>
</tr>
<tr>
<td>CFTR_gRNA_Reverse_T2</td>
<td>AAACCCGCATTAAGAAATA</td>
</tr>
<tr>
<td>CFTR_gRNA_Forward_T3</td>
<td>CACCGACATTTCTTTAATG</td>
</tr>
<tr>
<td>CFTR_gRNA_Reverse_T3</td>
<td>AAACATGATATTTCTTTAATG</td>
</tr>
<tr>
<td>CFTR_gRNA_Forward_T4</td>
<td>CACCGCACCATTAAAGAAATA</td>
</tr>
<tr>
<td>CFTR_gRNA_Reverse_T4</td>
<td>AAACCTTAAGAAATCATC</td>
</tr>
<tr>
<td>CFTR_gRNA_Forward_T5</td>
<td>CACCGGAAACTGAGCCTTCAG</td>
</tr>
<tr>
<td>CFTR_gRNA_Reverse_T5</td>
<td>AAACCTGAAAGCTCAGTTC</td>
</tr>
<tr>
<td>CFTR_gRNA_Forward_T5+</td>
<td>CACCGGAAACTGAGCCTTCAG</td>
</tr>
<tr>
<td>CFTR_gRNA_Reverse_T5+</td>
<td>AAACCTGAAAGCTCAGTTC</td>
</tr>
<tr>
<td>CFTR_gRNA_Forward_T6</td>
<td>CACCGCAGAGCATGCCACTGAA</td>
</tr>
</tbody>
</table>

All primers were ordered from Sigma Aldrich.

### 2.1.14. Oligonucleotides for gRNA cloning
| CFTR_gRNA_Reverse_T6           | AAACCTTCTAGGGCATGCTTTTGAC |
| CFTR_gRNA_Forward_T7          | CACCGAGGGTAAATTAGACAG     |
| CFTR_gRNA_Reverse_T7          | AAACCTGTGCTTAATTTACCTTC  |
| CFTR_gRNA_Forward_SaCas9 T1   | CACCGATATCATTTGGGTTCCTATG |
| CFTR_gRNA_Reverse_SaCas9 T1   | AAACCATAGGAACACAAATGATATC |
| CFTR_gRNA_Forward_SaCas9 T2   | CACCGTTCTGTATCTATATTCATCAT |
| CFTR_gRNA_Reverse_SaCas9 T2   | AAACATGATGAATATGACGAGAAG |
| CFTR_gRNA_Forward_cpf1 T1     | AGATGCTGGATTATGCTGCGACATTAA |
| CFTR_gRNA_Reverse_cpf1 T1     | AAATTAATGTTGGCCAGGATAATCCAGC |
| CFTR_gRNA_Forward_cpf1 T2     | AGATGATGACGCTTTGCTATATATTCAG |
| CFTR_gRNA_Reverse_cpf1 T2     | AAATTAATGATTACAGAAGTCAGAG |
| Ai9_gRNA_Forward_loxp         | CACCCGCGTATAATGCTGATACAG  |
| Ai9_gRNA_Reverse_loxp         | AAACGTATAGCATACATTATACG   |
| Ai9_gRNA_Forward_T1           | CACCGAAAGATTTGATGCTACCAG  |
| Ai9_gRNA_Reverse_T1           | AAACGTGCTAATAGATTTACCTTC |
| Ai9_gRNA_Forward_T2           | CACCGTAGAATCTTTAGGGCCCG   |
| Ai9_gRNA_Reverse_T2           | AAACCGGGCCCTAAGAGGTTCCTAC |
| Ai9_gRNA_Forward_T3           | CACCGTATGCTGATACGAGAAG   |
| Ai9_gRNA_Reverse_T3           | AAAACATAACTTTGCATTAGCATAC |
| Ai9_gRNA_Forward_T4           | CACCGCTAGCTTTGGGCTGCGAGG |
| Ai9_gRNA_Reverse_T4           | AAACCGACCTGCGCCAAAGCTAGC |
| Ai9_gRNA_Forward_T5           | CACCGAAACCTCTACAAATGTGTA |
| Ai9_gRNA_Reverse_T5           | AAACCTACACATTGTTAGGAGTTTC |
| Ai9_gRNA_Forward_T6           | CACCGAACTAAACCTTCAAAATG   |
| Ai9_gRNA_Reverse_T6           | AAACCAGTTTGGAGGTTCCTACTC |
| Ai9 HiT1 Forward_T3           | AGCTTAGGAACCTTTAGGGGCCCGG |
| Ai9 HiT1 Reverse_T3           | GATCCTCGCGGCGCCCTAAGAGGTTCCTA |
| Ai9 HiT1 Back Forward_T3      | AGCTTAGGAACCTTTAGGGGCCCGG |
| Ai9 HiT1 Back Reverse_T3      | GGCCTCGCGGCGCGCTAAGAGGTTCCTA |
| Ai9 HiT1 Forward_T1           | AGCTTAGGCGCCATAGGGTAAAGAAG |
| Ai9 HiT1 Reverse_T1           | GATCTTTTCTAACTAAACTATAGCCGC |
| Ai9 HiT1 Forward_T4           | AGCTCTAGCCTGGCTGCGAGGTAGG |
| Ai9 HiT1 Reverse_T4           | GGCCTGCGCGCCCGCTACTAGAGCCAGT |
| CFTR_gRNA_Forward_intron 9-1  | CACCGCTTGCTTAGGATGATAATTGG |
| CFTR_gRNA_Reverse_intron 9-1  | AAACCAATATCATCCCTAAGCAGC |
| CFTR_gRNA_Forward_intron 9-2  | CACCGTAGAGGTTCACAGAGAAG |
| CFTR_gRNA_Reverse_intron 9-2  | AAACCTTCTCTGTGAACCTCTATC |
**CFTR_gRNA_Forward_intron 9-3**  
CACC GTC ATT ATCAAATCACGCTC

**CFTR_gRNA_Reverse_intron 9-3**  
AAAC GAG GGT GAT TTG GATAATGAC

**CFTR_gRNA_Forward_intron 10-1**  
CACC GA AAATGC ATTT ATGAAA

**CFTR_gRNA_Reverse_intron 10-1**  
AAAC TCTAA ATCT GCT GG CAGAC

**CFTR_gRNA_Forward_intron 10-2**  
CACC GTC CTG CAC GAG AAT TATGAG

**CFTR_gRNA_Reverse_intron 10-2**  
AAAC TCT AAAT CTG CTG G ACAC

**CFTR_gRNA_Forward_intron 10-3**  
CACC GAG A AAATGC ATTT ATGAAA

**CFTR_gRNA_Reverse_intron 10-3**  
AAAC TCT AAAT CTG CTG G ACAC

**CFTR_gRNA_Forward_intron 10-4**  
CACC GTC ATG CAT TATG AT TATGAG

**CFTR_gRNA_Reverse_intron 10-4**  
AAAC CAT ATC ACT TATG CAT GAC

**CFTR_gRNA_Forward_intron 10-5**  
CACC GTC CAT TAA AT A ACC AT TGC

**CFTR_gRNA_Reverse_intron 10-5**  
AAAC CAT GTT ATT ATAT TG GAC

**G** nucleotide has been added to the gRNA sequence at the beginning if the gRNA did not start with a **G**. The underlined sequences indicate the extra bases needed for overhang ligation. All oligonucleotides were ordered from Sigma Aldrich.

### 2.1.15. Single strand DNA (ssDNA)

<table>
<thead>
<tr>
<th>CFTR_ssDNA (199 base)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGGGTTTTATTTCCAGACTTCACTTTCAATGTTGATTATGGGAGAACTCGAGCCTTCAGAGGGTAAAATTAAGCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTGGTGTTTTCTCTATATGATATC GATACAGAAGCGTCATCAAAGCATGCCA AAGAGGTAAGAAACTATGTAATTTTGATTATGCATATGA ACCCTTCACACTACCCAAATTATTTGGCTCCATATTCAATCGGTTAGTCTACATATTTTTATGTTTCCTCTATGGGTAAGCTACTGTGAATGGATCAATTAATAAACACATGACCTATG</td>
</tr>
</tbody>
</table>

The **Bold** sequences were used to make the correction to the gene while the silent mutation creates a ClaI restriction site highlighted in grey. The ssDNA was ordered from Integrated DNA Technologies (IDT).

### 2.1.16. Double strand DNA fragment (gBlock)

<table>
<thead>
<tr>
<th>CFTR_HITI_gBlocks Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTGACGTCCACCTGGAGGTTAATAGTAGATGGATTCGCTCAAGTCTGAGCGTGATTTGA TAATGACCTAATAATGATGGGTTTTATTTTCTCCAGACTTCACTTTCAATGTTGATTATGGGAGAACTCGAGCCTTCAGAGGGTAAAATTAAGCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTGGTGTTTTCTCTATATGATATC GATACAGAAGCGTCATCAAAGCATGCCAAGAGGTAAGAAACTATGTAATTTTGATTATGCATATGAACCCTTCACACTACCCAAATTATTTGGCTCCATATTCAATCGGTTAGTCTACATATTTTTATGTTTCCTCTATGGGTAAGCTACTGTGAATGGATCAATTAATAAACACATGACCTATG</td>
</tr>
</tbody>
</table>

94
The **Bold** sequences were used to make the correction to the gene while the silent mutation creates a ClaI restriction site highlighted in grey. The DNA fragment was ordered from Integrated DNA Technologies (IDT).

### Ai9_HDR_Left arm_gBlocks Fragment

| ATTAATGGGGTCATTTCATCACAAAATGCAATTTATGAAATGGTGAATTTTTGTTTATCAGCCTCAATGTTATATTATATGGAATGTATAGTGAATAATACGTAGCCAGG |
| GTCGGGGCCAGGCAATAGCCACCCCAAGCTAGGGGCGAGGGCACCTAGCCAGCCGGCCAGCCGGCCAGGCGGCCCAGTCG |

### Ai9_HDR_Right arm_gBlocks Fragment

| GGGCCGTCAGCTCCACTCTGACGGCGGCTGCTGCTGCTATGCTGAGGGGTCCTAGGCTGGGAGGCGCGGGGACG |
| GGGCCGTCGAGCTGAGGCTCAGGAGCAGGGCACTCCAGGCCAGGGCGAGGGGACGGGAGGTCAGCCCAGGAGG |

### Ai9_HITI_HDR_Left arm_gBlocks Fragment

| ATTAATGGGGTCATTTCATCACAAAATGCAATTTATGAAATGGTGAATTTTTGTTTATCAGCCTCAATGTTATATTATGGAATGTATAGTGAATAATACGTAGCCAGG |
| GTCGGGGCCAGGCAATAGCCACCCCAAGCTAGGGGCGAGGGCACCTAGCCAGCCGGCCAGCCGGCCAGGCGGCCCAGTCG |

### Ai9_HITI_HDR_Right arm_gBlocks Fragment

| GGGCCGTCAGCTCCACTCTGACGGCGGCTGCTGCTGCTATGCTGAGGGGTCCTAGGCTGGGAGGCGCGGGGACG |
| GGGCCGTCGAGCTGAGGCTCAGGAGCAGGGCACTCCAGGCCAGGGCGAGGGGACGGGAGGTCAGCCCAGGAGG |
The underlined sequence was used to digest with a restriction enzyme to facilitate the ligation. The DNA fragment were ordered from Integrated DNA Technologies (IDT).

### 2.2. Methods

#### 2.2.1. Production of competent bacteria

The chemical competent TOP10 strain of *E. coli* was prepared following the inoue method (303). The bacteria strain was grown first on an LB plate without antibiotic, then a single colony was inoculated into 25ml of LB media in a 250ml flask. The incubation time was 8 hours at 37°C with shaking at 250rpm. After that, 1 ml of the inoculate was placed into 250ml of fresh LB which was shaken at 150rpm at room temperature (18-22°C) until it reached OD₆₀₀ of 0.5-0.6. The culture was then transferred to an ice-water bath for 10 minutes, followed by centrifugation at 2500 x g for 10 minutes at 4°C. 80ml of ice-cold Inoue transformation buffer was used to re-suspend the bacteria cells, followed by centrifugation at 2500 x g for 10 minutes at 4°C. After that, the cells were again suspended in 20ml of ice-cold Inoue transformation buffer, then 1.5ml of dimethyl sulfoxide (DMSO) was added and gently mixed. Finally, the competent cells were snap frozen in liquid nitrogen at 50µl aliquots. For long term storage, the competent cells were stored at -80°C.
2.2.2. Bacterial transformation

50µl aliquots of competent *E. coli* were thawed on ice. 1ng to 50ng of plasmid DNA was added to the competent cells followed by 30 minutes incubation on ice. After that, the cells were then heat shocked by incubation at 42°C for 45 seconds, then transferred to the ice for 5 minutes. 200µl of SOC media was added to the cells which were then transferred to an orbital shaker (250rpm) at 37°C for 1 hour. 100µl of the cell suspension was then spread on an LB agar plate with an appropriate antibiotic, and incubated for 16-20 hours at 37°C.

2.2.3. Growth and Maintenance of Bacteria

Individual colonies from LB plates were inoculated in 5ml of LB broth (miniprep) or 150 ml of LB broth (maxiprep) with an appropriate antibiotic where necessary (100µg/mL or 50µg/mL of ampicillin or kanamycin, respectively) which was incubated overnight at 37°C with shaking at 250rpm.

2.2.3. Plasmid DNA extraction

2.2.3.1. Miniprep DNA extraction

The Plasmid DNA was extracted using GeneJET Plasmid Miniprep Kit following the manufacturer’s guidelines which is based on using the alkaline lysis method. Bacteria from 5ml LB media were harvested by centrifugation (13000 rpm) for 1 minute at room temperature. The supernatant was decanted, then the pellet was resuspended in 250µl of resuspension solution containing RNase. After that, 250µl of lysis solution were added for 3 minutes incubation time. The lysis step was terminated by adding 350µl of neutralization solution followed by centrifugation at 13000 rpm for 5 minutes. The supernatant was then transferred into a GeneJET Spin Column then centrifuged for 1 minute for binding. The column was washed twice with 500µl wash solution each time. Finally, the plasmid DNA was eluted into a new tube with 50µl of nuclease-free water. The Plasmid DNA quantity was measured by OD260 using NanoDrop ND-1000. Plasmid DNA samples with 260/280 of 1.8-2.0 were deemed to be sufficiently
pure for use in further experiments. The plasmid DNA obtained was stored at 4°C or -20°C for long term storage.

### 2.2.3.2. Maxiprep DNA extraction
The Plasmid DNA was extracted using a GeneJET Plasmid Maxiprep Kit following the manufacturer’s guidelines which were based on using the alkaline lysis method. Bacteria from 150 ml LB media were harvested by centrifugation (5,000×g) for 10 minutes at 4°C. The supernatant was decanted and then the pellet was resuspended in 6 ml of resuspension solution containing RNase. After that, 6 ml of lysis solution were added with incubation for a maximum of 3 minutes to avoid denaturation of the supercoiled plasmid DNA. The lysis step was terminated by adding 6 ml of neutralization solution followed by 0.8 ml of the Endotoxin binding reagent and 6 ml of 96% ethanol. The mixture was centrifuged at 4,000×g for 40 minutes to pellet both cell debris and chromosomal DNA. The supernatant was then transferred to a new 50 ml tube, and mixed with 6 ml of 96% ethanol. After that, the sample was transferred to the GeneJET Spin Column, then centrifuged for 3 minutes at 2,000×g for binding. The column was washed firstly with 8 ml of Wash Solution I (containing isopropanol) then washed again with 8 ml of Wash Solution II (containing 96% ethanol). Finally, the plasmid DNA was eluted into a new tube with 1 ml of nuclease-free water. The Plasmid DNA quantity was measured by OD$_{260}$ using NanoDrop ND-1000. Plasmid DNA samples with 260/280 of 1.8-2.0 were deemed to be sufficiently pure for use in further experiments. The plasmid DNA obtained was stored at 4°C or -20°C for long term storage.

### 2.2.4. DNA extraction
Total DNA was extracted using a GeneJET Genomic DNA Purification Kit following the manufacturer’s guidelines which are based on reversibly binding of DNA to a silica-gel membrane followed by elution in low-salt conditions. The cells needed for DNA extraction were harvested then pelleted at 13000 rpm for 5 minutes. The cells were resuspended in 200μl of Lysis Solution and 20μl of Proteinase K Solution, followed by 10 minutes incubation at 56°C. After that,
20µl of RNase A Solution was added and the cell lysates were incubated for 10 minutes at room temperature, followed by the addition of 400µl of 50% ethanol, followed by being transferred to the binding column. Two washes were required for the column, first with 500µl of washing buffer I then with 500µl of washing buffer II, followed by 2 minutes spinning at 13000 rpm to ensure that no ethanol was carried over during the DNA elution step. Finally, the genomic DNA was eluted into a new tube with 50µl of nuclease-free water. The DNA quantity was measured by OD\textsubscript{260} using NanoDrop ND-1000. DNA samples with 260/280 of 1.8-2.0 were sufficiently pure for use in further experiments. The genomic DNA obtained was stored at 4°C or -20°C for long term storage.

2.2.5. RNA extraction

Total RNA was extracted using an RNeasy mini kit following the manufacturer's guidelines which are based on a silica-based membrane column selective for RNA binding of samples followed by elution in low-salt conditions. The cells needed for RNA extraction are harvested then pelleted at 13000 rpm for 5 minutes. The cells were resuspended in 350µl RLT buffer supplemented with 1% β-mercaptoethanol. 350µl of 70% Ethanol was then added into cell lysates and were transferred to a binding column. Three washes were required for the column - first with 700µl RW1 washing buffer, then twice with 500µl of RPE washing buffer, followed by 2 minutes spinning at 13000 rpm to ensure that no ethanol was carried over during the RNA elution step. Finally, total RNA was eluted by adding 30 µl RNase-free water to the column membrane, then centrifuging at 13000 rpm for 1 minute. The RNA quantity was measured by OD\textsubscript{260} using NanoDrop ND-1000. RNA samples with 260/280 of 1.9-2.1 were sufficiently pure for use in further experiments. The RNA obtained was stored at -80°C for long term storage.
2.2.6. DNA manipulation

2.2.6.1. Restriction digests
DNA was digested with an appropriate restriction enzyme following the manufacturer's instructions in an appropriate 1x reaction buffer, and at the appropriate reaction temperature.

2.2.6.2. Agarose gels electrophoresis
The 1% agarose gel for DNA electrophoresis was prepared in 1X TAE (2% agarose gels were used for T7 endonuclease I assays) with 1µl of SYBR Safe DNA Gel Stain per 10ml of 1X TAE. The agarose was dissolved by heating in a microwave oven, then poured into an appropriately-sized casting tray. To run samples on the agarose gel, the DNA was mixed with 6X DNA loading dye at a final concentration of 1X, then separated by running on a TAE buffer at 120V for 60 minutes. The DNA bands were visualised and documented using the UVIdoc system.

2.2.6.3. Gel purification
The DNA fragment was purified from agarose gel after electrophoresis when needed using a GeneJET Gel Extraction Kit following the manufacturer's guidelines. The agarose gel was visualized on an ultraviolet transilluminator, and the DNA fragment excised using a scalpel. Then 1:1 volume of binding buffer was added to the gel slice (w/v) and incubated at 60°C until the gel slice was completely dissolved. The solubilized gel was transferred to a binding column, followed by washing with 700µl of wash buffer. Finally, the DNA was eluted into a new tube with 50 µl of nuclease-free water. The DNA quantity was measured by OD$_{260}$ using NanoDrop ND-1000. DNA samples with 260/280 of 1.8-2.0 were sufficiently pure for use in further experiments. The DNA obtained was stored at 4°C or -20°C for long term storage.
2.2.7. Polymerase chain reaction (PCR)
The PCR reaction was performed in 50μl reaction volume using 100ng of DNA template, 10μl of 5X Q5 Reaction Buffer, 2.5μl of 10μM primers (forward and reverse), 1μl of 10mM dNTPs, 0.5μl Q5 Hot Start High-Fidelity DNA Polymerase and made up to 50μl with nuclease-free water. The cycling conditions were as follows: an initial denaturation and a polymerase activation step at 98°C for 30 secs, followed by 35 cycles of denaturation (10 seconds, 98°C), annealing [30 seconds, 3°C above the melting temperature (Tm) of the lower Tm prime] and extension (30 seconds/kb, 72°C). A final extension cycle (5 mins, 72°C) completed the PCR reaction. The PCR reaction was then analysed on a 1% agarose gel and visualised using SYBR Safe staining.

2.2.8. PCR purification
The PCR fragment was purified when it was necessary using a GeneJET PCR Purification Kit following the manufacturer’s instructions. The PCR reaction was mixed with same volume of binding buffer, then transferred into a binding column, followed by washing with 700µl of wash buffer. Finally, the PCR product was eluted into a new tube with 50µl of nuclease-free water. The DNA quantity was measured by OD\textsubscript{260} using NanoDrop ND-1000. DNA samples with 260/280 of 1.8-2.0 were sufficiently pure for use in further experiments. The DNA obtained was stored at 4°C or -20°C for long term storage.

2.2.9. Quantitative real time RT-PCR (qRT-PCR)
The qRT-PCR was performed in 20μl reaction volume using 50ng total RNA template, appropriate 1μl of TaqMan probe, 10μl of 2x SensiFAST Probe Hi-ROX One-Step Mix, 0.2 μl reverse transcriptase, 0.4μl RiboSafe RNase Inhibitor and made up to 20μl with nuclease-free water. The qRT-PCR reactions were preheated at 45°C for 20 minutes to perform reverse transcription, followed by 95°C for 2 minutes to activate the polymerase. After that there were 40 cycles of 95°C for 10 seconds (denaturing) and at 60°C for 30 seconds (annealing and extension). The qRT-PCR was performed in the Bio-Rad 96 CFX and Ct values
were obtained using the Bio-Rad CFX manager software. The relative expression levels were calculated using the delta-delta Ct \( (2^{-\Delta\Delta Ct}) \) method (304).

2.2.10. T7 endonuclease I assay
200ng of purified PCR product was mixed with 2μl of NEB buffer 2, and nuclease-free water up to 19μl. This mixture was heated at 95°C for 5 minutes to denature it, then it was allowed to re-anneal at room temperature for 20 minutes. After that 1μl of T7 endonuclease I enzyme was added, then it was incubated at 37°C for 20 minutes. Immediately, the reaction was loaded on 2% agarose gel which was visualised using SYBR Safe staining.

2.2.11. PCR-Restriction Fragment Length Polymorphism (RFLP)
200ng of purified PCR product was digested with an appropriate restriction enzyme following the manufacturer’s instructions, in an appropriate 1x reaction buffer and at the appropriate reaction temperature. Then the reaction was loaded on 2% agarose gel which was visualised using SYBR Safe staining.

2.2.12. Transepithelial electrical resistance (TEER) measurements
TEER measurements are used to assess the integrity of the epithelial monolayer on snapwells. The measurement is performed with an Ohmmeter (EVOM) after replacing the apical media with 250μl of PBS. The EVOM has two electrodes, one of which was placed in the upper compartment while the other was placed in the lower compartment. The PBS was removed immediately after measuring the TEER, and replaced by complete media for cell lines, or exposure to air in the case of primary cells.

2.2.13. Ligation of DNA fragments
Linearised plasmid backbone and inserted DNA was ligated using 1μl of T4 DNA ligase and 2μl of 10X T4 DNA ligase buffer in a final volume of 20μl. The ligation reaction was performed at room temperature for 2 hours for sticky ends fragments at a ratio of 1 backbone: 4 insert, while the blunt end ligation reaction was kept
for 16 hours where the ratio was 6 backbone: 1 insert. 10μl of the ligation reaction was transformed into competent cells.

### 2.2.14. Ligation of oligonucleotides

The two pairs of oligonucleotides (forward and reverse) were resuspended in nuclease-free water at a concentration of 100μM. 1μl of each was mixed with 43μl nuclease-free water and 5μl of 10X T4 DNA Ligase Buffer followed by heating at 95°C for 5 minutes and then cool down at room temperature for 20 minutes to allow base pair annealing. The ligation reaction was performed in 20μl between 3μl of the annealed oligonucleotides pairs, with 200ng of linearised plasmid backbone in the presence of 2μl of 10X T4 DNA Ligase Buffer and 1μl of T4 DNA Ligase at room temperature for 2 hours. 10μl of the ligation reaction was used for transformation into competent cells.

### 2.2.15. PCR product cloning

The PCR product cloning was performed by using the CloneJET PCR Cloning Kit following the manufacturer’s guidelines which were based on a lethal gene disrupted by ligation of a DNA insert into the cloning site. 1μl of PCR product was mixed with 10μl of 2X Reaction Buffer, 1μl of pJET1.2/blunt Cloning Vector, 1μl of T4 DNA Ligase and 17μl of nuclease-free water, followed by incubation at room temperature for 5 minutes. The ligated PCR product was then transformed into competent cells (section 2.2.2).

### 2.2.16. Site directed mutagenesis

For single base pair mutagenesis (substitution), mutagenic primers were designed using the NEBaseChanger website: (https://nebasechanger.neb.com). 10ng of the desired plasmid for mutagenesis was amplified by a PCR reaction for 25 cycles, using Q5 Hot Start High-Fidelity DNA Polymerase and mutagenic primers, followed by digesting the parental template using a methylation-dependent endonuclease (DpnI). Both ends of the PCR product were ligated using T4 DNA Ligase, then transformed into competent cells. To confirm the mutagenesis, purified plasmid from a miniprep was sent for sequencing.
2.2.17. Sanger sequencing
The DNA sample at a minimum concentration 50ng with 5 µM of appropriate primer was sent for sequencing to the GENEWIZ company. The sequence results were examined using the SnapGene Viewer version 4.1 software.

2.2.18. Gibson assembly cloning
Gibson cloning was used to assemble multiple linear DNA fragments with the assistance of three enzymes: T5 Exonuclease, Phusion DNA Polymerase and Taq DNA Ligase. The DNA needed for cloning was amplified by a PCR reaction with 15-20bp sequence homologous to the next DNA fragment using NEBuilder tools (https://nebuilder.neb.com) to design the PCR primer. The PCR products were assembled using an NEBuilder HiFi DNA Assembly Cloning Kit following the manufacturer's instructions. The construct was confirmed by restriction of the enzyme digest first, followed by Sanger sequencing.

2.2.19. Cell culturing
16HBE14o- and CFBE41o- cell lines were maintained in complete media based on Eagle’s Minimum Essential Medium (MEM) complete media. Neuro-2A (mouse neuroblastoma cell line) and HEK293 were cultured in complete media based on Dulbecco’s Modified Eagle Medium (DMEM). Normal Human Bronchial Epithelial Cells (NHBE) and Cystic Fibrosis Bronchial Epithelial Cells (CFBE) have been semi-immortalised with BMI-1 (93). These cells were cultured in a Bronchial Epithelial Cell Growth Medium (BEGM) on coated collagen plastic. All cell lines and primary cells were grown at 37°C in incubators with 5% CO₂ and passaged when they were 70-80% confluent. To passage the cells, they were first washed with PBS, then 0.05% trypsin-EDTA added where the cells returned to the incubator to detach. After that the cells were neutralised with complete media, then centrifuged for 5 minutes at 1500rpm. One tenth of the cells were transferred into a new flask with complete media.
2.2.20. Collagen coating

To grow primary cells, tissue culture plates, flasks and snapwells were first coated with 1% (v/v) of a 3mg/ml bovine collagen solution in phosphate buffered saline (PBS) for 1 hour, then washed once with sterilized distilled water and dried in a laminar hood.

2.2.21. Air-Liquid Interface (ALI) Culture

Primary cells (NHBE or CFBE) were seeded onto Snapwell collagen-coated permeable supports (0.4µm pore size, 1.12 cm² cell growth area) at a seeding density of 1,000,000 cells/well. The cells were fed with BEGM on both sides (apical and basolateral) and incubated at 37°C and 5% CO₂. After 48 hours, the apical and basolateral media were removed and replaced in the basolateral compartment only with ALI media (Recipe in Section 2.1.12). The cells were left in the incubator with changes of the ALI media in the basolateral compartments every 2-3 days.

2.2.22. Liquid-liquid interface (LLI) culture

Cell lines (CFBE41o- or 16HBE14o-) were seeded onto Snapwell collagen-coated permeable supports (0.4µm pore size, 1.12 cm² cell growth area) at a seeding density of 500,000 cells/well. The cells were fed with MEM on both sides (apical and basolateral) and incubated at 37°C and 5% CO₂. The cells were left in the incubator with changes in the media on both sides every 2-3 days.

2.2.23. Flow cytometry for in vitro transfections

Cells grown on 24 well plates were detached using 200µl Trypsin-EDTA in each well. The cells were re-suspended with 1000µl per well of the respective culture media and transferred to FACS tubes. The cell suspensions were then centrifuged at 1500 rpm for 5 minutes at room temperature. The supernatants were discarded by decantation then the cells were re-suspended in 200µl PBS. The acquisition is performed with a FACScalibur flow cytometer where the cells are plotted first in the forward scatter area (FSC-A) versus the side scatter area (SSC-A), while excluding the debris. The Doublet cells in the next step were
excluded by using side scatter height (SSC-H) versus a FSC-A plot. The GFP fluorescence was detected using a FL-1 detector (using a blue laser for excitation) while tdTomato fluorescence was detected using a FL-2 detector (using a blue laser for excitation). The data analysis was conducted using the FlowJo software v.10 software.

2.2.24. Biophysical characterisation of RTNs

The size and charge (ζ potential) of the nanocomplexes were measured using a Zetasizer nano ZS. The nanocomplex was first prepared on water with a minimum of 1-2 µg of nucleic acid (DNA or mRNA) or RNP, then incubated for 30 minutes at room temperature. The sample was then diluted to a final volume of 1 ml, then transferred to a cuvette where the size was measured first and then the ζ potential.

2.2.25. Western blotting

The culture medium was removed from each well of the plate then washed twice with cold PBS. For lysing, the cells were kept cold using an NP40 Cell Extraction Buffer supplemented with 1 mM Phenylmethanesulfonyl fluoride (PMSF) and 1x Protease Inhibitor cocktail for 10 minutes on ice. The cells were scraped then transferred into 1.5 ml Eppendorf tubes, and were then incubated on ice for 20 minutes, followed by centrifugation at 15,000 rpm for 20 minutes in a 4°C pre-cooled centrifuge. The supernatant was then transferred to a new tube. The total protein concentration was measured using a Peirce BCA Protein Assay Kit. 10µg of each sample was mixed with a 4x loading dye buffer supplemented with 10x Dithiothreitol (DDT) as a reducing agent, then incubated for 30 minutes at 37°C. The samples were then loaded into wells of 4-20% pre-cast polyacrylamide gels (BioRad) and run for 90 minutes at 125 V. They were then transferred using a Trans-Blot Turbo Transfer System at 25 V for 30 minutes. The membrane was then blocked in TBST with 5% milk for 1 hour, followed by incubation with the anti-CFTR UNC-596 Primary Antibody in a blocking solution overnight at 4°C. After that, the membrane was washed three times in TBST before incubation with
anti-mouse IgG-HRP labelled secondary antibody in a blocking buffer at room temperature for 1 hour. The membrane was incubated with SuperSignal West Femto Substrate and the image was acquired using a Gel Doc XR System. The membrane was washed twice before incubation with anti-Na/K-ATPase primary antibody or anti-β-Actin in blocking buffer at room temperature, followed by washing and incubation with anti-mouse IgG-HRP labelled secondary antibody in a blocking buffer at room temperature for 1 hour. The membrane was washed before signal development, using Clarity Western ECL Substrate, and the image was acquired using a Gel Doc XR System.

2.2.26. Synthesis of gRNA by in vitro transcription
The pX330 plasmid with a target gRNA was used as a template for the in vitro transcription of the gRNA. The U6 promoter in the plasmid was replaced by a T7 promoter which was necessary for the RNA in vitro transcription process (Figure 2.1). First, the PCR reaction was performed using a forward and reverse primer. The cycling conditions were as follows: an initial denaturation step at 98°C for 30 seconds, followed by 15 cycles of denaturation (10 seconds, 98°C), annealing (30 seconds, temperature as shown in Table 3) and extension (10 seconds, 72°C). Then this was followed by another 20 cycles of denaturation (10 seconds, 98°C), annealing (30 seconds, temperature as shown in Table 2.1) and extension (20 seconds, 72°C). A final extension cycle (5 minutes, 72°C) completed the PCR reaction. The PCR reaction was then analysed on a 1% agarose gel and visualised using SYBR® Safe staining. The product was purified using a GeneJET PCR Purification kit, following the manufacturer’s protocols. Next the RNA was synthesised using a HiScribe T7 Quick High Yield RNA Synthesis Kit following the manufacturer’s protocols by using 1μg of DNA template from the previous step. The gRNA product was purified using a RNeasy Mini kit. The gRNA quantity was measured by OD260 using NanoDrop ND-1000. gRNA samples with 260/280 of 1.9-2.1 were sufficiently pure for use in further experiments. The gRNA obtained was stored at -80°C for long term storage.
Figure 2.1. Replace the U6 promoter with a T7 promoter in order to perform in vitro transcription (IVT).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-&gt;3’)</th>
<th>Amplified product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP_gRNA IVT_Forward_SpCas9 T1</td>
<td>TAATACGACTCTATAGGTTGA ACCGCATCGAGCTGAA A AAAGCACCGACTCGGTGCC</td>
<td>119</td>
</tr>
<tr>
<td>IVT_primer_Reverse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ai9_gRNA IVT_Forward_T1</td>
<td>TAATACGACTCTATAGGAAAG AATTGATTTGATACCG AAAAGCACCGACTCGGTGCC</td>
<td>119</td>
</tr>
<tr>
<td>IVT_primer_Reverse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ai9_gRNA IVT_Forward_T4</td>
<td>TAATACGACTCTATAGGCTAGCTTG GCTGCAGGTCG AAAAGCACCGACTCGGTGCC</td>
<td>119</td>
</tr>
<tr>
<td>IVT_primer_Reverse</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1. List of primers which have been used to amplify gRNA prior to IVT. The underlined sequences indicate T7 promoter.

2.2.27. Electrophysiological studies

Primary epithelium cells grown as Air-liquid interface (ALI) or epithelium cell lines grown as Liquid-liquid interface (LLI) on snapwells were then bathed on Ussing chambers in The Krebs-Henseleit (K-H) solution consisting of NaCl (117 mM), CaCl2 (2.5 mM), KCl (4.7 mM), MgSO4 (1.2 mM), NaHCO3 (25 mM), KH2PO4 (1.2 mM) and D-glucose (11 mM) where the pH was 7.4 and temperature was 37°C. The buffer bubbled at 21% O2 and 5% CO2 gas. The cell monolayers were maintained first at open-circuit conditions until the transepithelial potential difference (Vt) and resistance (Rt) were stable. Following that, the monolayer cells were then short-circuited applied through the DVC-4000 voltage/current clamp to
bring $V_t$ to 0 mV. The short circuit current $I_{sc}$ was measured and recorded using a PowerLab computer interface as $\mu$A/cm$^2$. During the experiment every 30 seconds the settings returned to open circuit for 3 seconds so that the spontaneous $V_t$ could be measured and resistance could be calculated using the Ohm’s law ($R_t = \Delta V$ [the open circuit transepithelial potential difference]/$\Delta I_{sc}$ [the short-circuit current difference]) (305). Drugs were circulated in bathing buffer where amiloride (10 $\mu$M) was added first to the monolayer at the apical side to block the sodium channel activity. The CFTR activator drugs were added to both sides of the monolayers (forskolin 25 $\mu$M and 100 $\mu$M of 3-isobutyl-1-methylxanthine (IBMX)). To block CFTR-dependent $I_{sc}$, 10 $\mu$M of CFTRinh-172 was added to the apical side. The electrophysiological studies were performed in collaboration with Maximillian Woodall from St George's, University of London.

2.2.28. Annealing synthetic gRNA

The chemically-modified targeting gRNA was purchased from Integrated DNA Technologies (IDT) as two systems (crRNA and tracrRNA). The Lyophilized vials were first re-suspended in Nuclease-Free Duplex Buffer to make 100 $\mu$M stock solutions, then the final equimolar duplex concentration (5 $\mu$M) was prepared followed by denaturion at 95ºC for 5 min and resting at room temperature for 20 min to anneal.

2.2.29. Receptor targeting nanocomplexes (RTNs) preparation

RTN formulations consisted of lipids, peptide and payload which could be DNA or mRNA or Ribonucleoprotein (RNP). The lipids that were used are listed in section 2.1.6, while peptides are in section 2.1.7. The RTN was prepared at different weight ratios and way of mixing in Opti-MEM depending on the nature of the molecule that was delivered to the cells (table 2.2) unless mentioned differently in the relevant result section. After mixing all components the formulation was incubated for 30 minutes at room temperature to allow self-assembly of nanocomplexes. The RNP of Cas9 and gRNA assemble first in Opti-MEM for 5 minutes at room temperature prior to mixing with other components.
### Table 2.2 RTN preparation and weight ratio.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Mixing order</th>
<th>Weight ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>lipid/peptide/DNA</td>
<td>1:4:1</td>
</tr>
<tr>
<td>mRNA</td>
<td>lipid/mRNA/peptide</td>
<td>3:1:4</td>
</tr>
<tr>
<td>Cas9 mRNA + gRNA</td>
<td>lipid/mRNA+gRNA/peptide</td>
<td>3:1:4</td>
</tr>
<tr>
<td>RNP (Cas9 protein + gRNA)</td>
<td>lipid/RNP/peptide</td>
<td>3:1:4</td>
</tr>
</tbody>
</table>

#### 2.2.30. Complex preparation with lipofectamine 2000 (L2K)

The molecule to be delivered (DNA or mRNA or Ribonucleoprotein (RNP)) was formulated with lipofectamine 2000 (L2K) at weight ratio 1:3 (each component was prepared in Opti-MEM at a separate tube, then mixed together). The mixture was then incubated for 10 minutes at room temperature to allow self-assembly of nanocomplexes. The RNP of Cas9 and gRNA assemble first in Opti-MEM for 5 minutes at room temperature prior to mixing with L2K.

#### 2.2.31. Transfection of submerged cultures

Cells were seeded on suitable plastic well plate 24 hours before the transfection. The transfection was performed when the cell confluency reached 70-80%. The next day, the complexes were prepared in OptiMEM. Following the incubation time, media from the wells was removed and the complex added to the cells. The plate was centrifuged at 1500 rpm for 5 minutes then incubated at 37 °C and 5% CO₂ for 4 hours. After the incubation time the transfectants were removed and replaced by complete media.

#### 2.2.32. Transfection of ALI cultures

The primary CFBE BMI-1 cells were transfected after 4 weeks of differentiation at ALI culture. The complexes were prepared in OptiMEM. Following the incubation time, ALI media from the basolateral side was removed and the complex added to both the apical and basolateral sides. The plate was incubated at 37 °C and 5% CO₂ for 4 hours, after which the transfectants were removed and replaced by ALI media at the basolateral side only.
2.2.33. In vivo transfections

Complexes were prepared in water and 55μl of the complexes were delivered via the oropharyngeal route to Ai9 mice (B6.Cg-Gt(ROSA)26Sor<sup>tm9(CAG-tdTomato)Hze/J</sup>) that were purchased from The Jackson Laboratory. The mice were initially anaesthetised with 2% isoflurane, then the mice’s noses were closed using tweezers to ensure that breathing took place through the mouth. The nanocomplex formulations were pipetted to the rear of the oral cavity while continuing closing the nose, until all the solution had been inhaled. All procedures were approved by UCL animal care policies and were carried out under Home Office Licenses issued in accordance with the UK Animals (Scientific Procedures) Act 1986.

2.2.34. Flow cytometry for in vivo transfections

The transfected mice were euthanized using pentobarbital administered intraperitoneally. The chest cavity was exposed by midline sternotomy, then the heart was perfused with 5ml of PBS. The lungs were kept in PBS then transferred to a falcon tube for digestion in 2 mg/ml of Collagenase D, then incubated for 30 minutes at 37°C with automated rotation. The cells were passed through a strainer (40μm) to ensure single cell suspension. The cells were washed twice with PBS then counted, and 2x10<sup>6</sup> cells per sample were stained with DAPI for viability, CD45 (APC) and CD326 (BV421) 30 minutes at room temperature then acquired with a BD LSRII flow cytometer where the cells were first plotted in the forward scatter area (FSC-A) versus the side scatter area (SSC-A) while excluding the debris. The Doublet cells in the next step were excluded by using side scatter height (SSC-H) versus FSC-A plot. The GFP fluorescence was detected using a FL-1 detector (using a blue laser for excitation) while tdTomato fluorescence was detected using FL-2 detector (using a blue laser for excitation). The FL-4 detector was used to detect allophycocyanin (APC) (CD45) (red laser for excitation) while the violet laser was used to excite the BV421 (CD326). The data analysis was conducted using the FlowJo software v.10 software.
2.2.35. Data analysis

The efficiency of the T7 endonuclease I assay or HDR = 1 – [(1–fraction cleaved) \( \frac{1}{2} \)]. The fraction cleaved was calculated by dividing the sum of the cleaved band intensities by the sum of the cleaved and uncleaved band. The band’s intensity was determined using ImageJ software.

To determine the indel % using the Tracking of Indels provided by Decomposition software (TIDE), a purified PCR product from the transfected and untransfected cell pool were sent first for Sanger sequencing. A chromatogram sequence was then uploaded to http://tide.nki.nl/.

2.2.36. Statistical Analysis

Data were analysed using GraphPad version 7.0, and Microsoft Excel 2016, and were then expressed as Mean ± Standard Deviation (SD). The significant differences between the 2 groups were calculated using two-tailed, unpaired Student’s t-test. On the other hand, when more than 2 groups were needed for comparison, the one-way analysis of variance (ANOVA) test was used, followed by Bonferroni’s post hoc. The p values of less than 0.05 were marked with *, p values of less than 0.01 were marked with **, p values of less than 0.001 were marked with ***, and p values of less than 0.0001 were marked with ****. The independent experiment is indicated as (N) where (n) indicates the number of biological repeats.
CHAPTER 3

Results

CRISPR/Cas9 mediated NHEJ in cell models
3. CRISPR/Cas9 mediated NHEJ in cell models

3.1 Introduction

The CRISPR/Cas9 system has emerged as a powerful gene editing tool that can knock-out genes, correct mutations, or interfere with gene expression by activation or repression. In the CRISPR system, the Cas9 endonuclease complex with the aid of the guideRNA (crRNA and tracrRNA) recognises and cleaves the site-specific DNA locus creating a double strand break (DSB). The DSB in the cells is repaired either by an error-prone pathway (NHEJ pathway) which is associated with insertions or deletions (indels) at the break site or in a precise manner with HDR pathways in the presence of donor templates (306).

Detection of the indels' efficiency is a crucial step for optimising gene editing. Several techniques have been used to determine the indels after CRISPR/Cas9 transfection, such as T7 endonuclease I or TIDE assay (Tracking of Indels by Decomposition), which tend to be both expensive and labour-intensive. A high-throughput reporter cell model which does not require laborious work with high sensitivity is needed to optimise CRISPR/Cas9 transfections. An example is reporter cells with fluorescent genes such as GFP where the efficiency of gene editing could be determined by quenching the fluorescent gene as a result of gene editing by knock-out (307). In addition, other cell models could be used to determine the NHEJ by activating a reporter gene. An example of that is the Ai9 model which has the tdTomato gene (red fluorescent protein variant) in an inactive state where the transcription is prevented by multiple stop codons located at the front of the gene. Targeting the stop cassette using the CRISPR/Cas9 system will enable tdTomato expression as an indication of NHEJ effect. This reporter model is available for in vivo application where the Ai9 construct is integrated into mice in the ROSA26 locus (301).

To deliver CRISPR/Cas, we use a self-assembling receptor-targeted nanocomplex (RTN) formulation comprising a mixture of liposomes, a receptor-targeting peptide, and the CRISPR system (plasmid, RNA, or protein). This formulation displays receptor-targeted transfection mediated by the peptide, with endosomal release of delivery molecules to the cytoplasm enhanced by the liposome component, and demonstrates high efficiency of transfection (292, 294, 308).
3.2 Aims

In this chapter, we aim to:

1. Develop a GFP cell model for optimising CRISPR/Cas9 transfection;
2. Find an optimal lipid-peptide nanocomplex formulation to package and deliver the CRISPR/Cas9 components;
3. Develop the Ai9 model in Neuro-2A cells;
4. Optimise the CRISPR/Cas9 transfection on Ai9 model by removing the stop cassette in order to activate the tdTomato gene;
5. Assess the gene editing in vivo mediated by CRISPR/Cas9 delivery by the receptor-targeted nanocomplex.
3.3. Results
3.3.1. Optimising transfection on Neuro-2A
Neuro-2A is a neuroblastoma cell line of mouse origin, where we aimed to optimise the CRISPR/Cas9 system on two models (GFP and Ai9). We first created the two model Neuro-2A cell lines, and then aimed to optimise the transfection on these cells before moving to the CRISPR system.

3.3.1.1. DNA transfection
In the beginning we evaluated the DNA transfection on Neuro-2A cells using plasmid pEGFP-N1 encoding the green fluorescent protein (GFP) reporter gene. The transfection was performed on the cells using 500ng of the plasmid and 48 hours after the transfection fluorescent imaging (Figure 3.1. A) and flow cytometry analysis (Figure 3.1. B) were performed. The optimisation of transfection is performed using Receptor-Targeted Nanocomplexes (RTN) as well as lipofectamine 2000 (L2K). The GFP plasmid is formulated with DTDTMA/DOPE (C14), DHDTMA/DOPE (C16), and DOTMA/DOPE (C18) with targeting peptide (Y) at specific weight ratios: 1 Lipid:4 peptide:1 DNA for the RTNs, while for L2K it was 3 Lipid:1 DNA plasmid. The GFP plasmid was firstly transfected with individual components of the RTN, either peptide (Y) alone at weight ratio of 1 DNA: 4 peptide or lipid alone [DTDTMA/DOPE (C14) or DHDTMA/DOPE (C16) or DOTMA/DOPE (C18)] at a weight ratio of 1:1 with plasmid DNA. The results showed that the transfection with peptide alone was the least efficient with almost 4% of the cells transfected. On the other hand, the transfection efficiency with C14, C16, and C18 were 18.3%, 32%, and 10%, respectively. Formulating the three components of RTN together at a specific weight ratio (1 Lipid: 4 peptide: 1 DNA), where the lipid is mixed first with the peptide then adding the plasmid DNA showed significant increase in the transfection efficiency: the transfection efficiency with C14Y was 60.8% (P=0.0007; unpaired Student’s t-test) and for C18Y it was 52.3% (P=0.023; unpaired Student’s t-test). The highest transfection efficiency was achieved by L2K which was on average 88.2%. In summary, pDNA can be efficiently delivered through RTN, especially with C14 and C18 lipids.
Figure 3.1. GFP plasmid transfection of Neuro-2A cells using RTN and L2K. A) Fluorescent images following GFP plasmid transfection. B) The %GFP expression encoded by plasmid DNA was assessed by flow cytometry. *P<0.05; ***P<0.001; unpaired Student’s t-test was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from three independent experiments (N=3) is shown. (B.F: bright field, U.V: ultraviolet light).
3.3.1.2 mRNA transfection
The transfection efficiency with pDNA in the previous experiment was almost 50% delivered as a RTN formulation. Here we want to investigate the possibility of enhancing the transfection efficiency using mRNA instead of pDNA. In general, mRNA delivery has advantages over pDNA delivery. For example, mRNA does not require transfer into the nucleus and there is no risk of integration into the genome, which is possible in pDNA transfection (309). To investigate the effect of the amount of lipid on the transfection efficiency, the cells were transfected with GFP mRNA (100ng per well of a 24 well plate) first at the same weight ratio (1 mRNA: 1 lipid) as was used in pDNA transfections or at a higher ratio (1 mRNA: 3 lipid). The flow cytometry analysis showed no statistical significant difference between the two weight ratios for C14 and C18 (unpaired Student’s t-test) (Figure 3.2 A and D). Then we formulated the GFP mRNA as RTN and we focused on two lipids (C14 and C18) as they show higher transfection efficiency when compared to C16. The other parameter we attempted to investigate was the effect of the order mixing of the RTN to establish the most efficient nanocomplex for mRNA delivery. The mixing order of RTN was: a) lipid mixed with peptide then the mRNA was added last, b) mRNA mixed with lipid first and finally the peptide was added and c) mixing the mRNA with peptide followed by the addition of the lipid. The analysis showed that when the peptide was mixed first with mRNA then the lipid was added this caused a reduction in the transfection efficiency (P<0.05, one-way ANOVA test with Bonferroni post-test analysis) in comparison with the other two mixing orders for the C14 lipid only (Figure 3.2 B,C, and E).

In conclusion, the mRNA transfection efficiency was robust with the targeting nanocomplex, which showed almost the same efficiency as that of commercial transfection reagents. In addition, the mRNA transfection was achieved with less amount of nucleic acid compared to the pDNA transfection (100ng mRNA vs 500ng pDNA).
Figure 3.2. mRNA transfection of Neuro-2A cells using RTN and L2K. A), B), C) Fluorescent images following GFP mRNA transfection. A) mRNA transfection with three lipids at weight ratio (1:1) or (1:3). B) mRNA transfection with C14 lipid and peptide Y at different weight ratios (1 mRNA:1 lipid:4 peptide) or (1 mRNA:3 lipid:4 peptide) using different mixing order. C) mRNA transfection with C18 lipid and peptide Y at weight ratio (1 mRNA:1 lipid:4 peptide) or (1 mRNA:3 lipid:4 peptide) using different mixing orders. D) Summary of GFP mRNA transfections with lipid only and assessed by flow cytometry. *P<0.05; unpaired Student's t-test was used to assess the significance. E) Summary of GFP mRNA transfections with lipid and peptide Y with different order of mixing during the preparation. The GFP expression assessed by flow cytometry. *P<0.05; one-way ANOVA with Bonferroni post-test analysis was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from three independent experiments (N=3) is shown. (B.F: bright field, U.V: ultraviolet light).

3.3.2 Optimising the CRISPR/Cas9 transfection in a GFP model
The first part of this project focuses on optimising the CRISPR/Cas9 system in a GFP cell line model, as it is easy to monitor the gene editing by NHEJ, which causes a loss of GFP expression.

The idea is to test the efficiency of the CRISPR/Cas 9 by designing gRNAs targeting the GFP gene in Neuro-2A cells (these cells show stable GFP expression as a result of the integration of the gene in the genome by lentivirus vector transduction in house) in order to create indels (NHEJ) and cause knock-out of the gene. The CRISPR system can be delivered to the cells as plasmid
DNA (needing transcription and translation for both Cas9 and gRNA) or Cas9 mRNA (needing translation only) with gRNA (in vitro transcribed or synthetic), or Cas9 protein with gRNA (in vitro transcribed or synthetic) (Figure 3.3).

![Figure 3.3 schematic diagram of knock-out strategy of GFP cell model using CRISPR/Cas9 editing technique.](image)

### 3.3.2.1 Plasmid DNA encoding CRISPR/Cas9 system

Initially two gRNAs (Figure 3.4. and Table 3.1.) were designed and inserted into the pX-330 plasmid (Figure 3.5. A) (221), which has a codon-optimised SpCas9 that recognises the 5′ NGG 3′ PAM region (N can be any nucleotide: A, T, C or G) and a gRNA backbone where we need to change the first 20 bases in order to target a specific locus (Figure 3.5. A). The targeting sequence can be inserted by digesting the plasmid with the BbsI enzyme (Figure 3.5. B) followed by ligating the guide sequence (without PAM), which is ordered as oligonucleotides including extra bases at the beginning and end in order to facilitate the ligation (Figure 3.5. C). It is important to have a G base at the beginning of the gRNA sequence in order to improve the transcription by the U6 promoter which can be added to the guide sequence without any effect on the efficiency.
Figure 3.4. Shows part of a sequence of GFP in the Neuro-2A cell line with two SpCas9 gRNAs (Sp T1 and Sp T2) and one SaCas9 gRNA (Sa T1), which have been used to knock-out the gene.

Table 3.1 List of gRNAs targeting GFP gene. The table shows each gRNA sequence, targeting the sense DNA strand (1), the PAM motif sequence, the predicted efficiency score where a score near 100 is more efficient and the specificity score (in mouse genome) where a high score indicates potentially more off-targeting. The on-target and off-target scores were calculated by using the Benchling program (https://benchling.com/crispr).
Figure 3.5. The pX330 plasmid and how to insert the gRNA. A) The overview of the plasmid. The gRNA is driven by a RNA polymerase III U6 promoter, while the hSpCas9 (human codon-optimised Streptococcus pyogenes Cas9) is driven by a CBh (Chicken Beta Actin) promoter. The nuclear localisation signals (NLS) flanked the hSpCas9 sequence. In addition, the plasmid has a Bovine Growth Hormone Polyadenylation Signal (BGHpA). B) Sequence of the chimeric gRNA region. The red arrow indicates the removable sequence of the BbsI restriction site which allows the insertion of the desired gRNA. 89 nucleotides (shown in red) show the trans-activating crRNA (tracrRNA) sequence. C) Guide sequence insert. The blue colour indicates the sequence of 20 nucleotides target sequence without the PAM region (NGG). If the target sequence does not start with a G base, the addition of a G nucleotide is necessary to initiate the U6 promoter. The black letters indicate an additional sequence needed to facilitate the overhang cloning. Adapted from (221).

GFP gRNAs T1 and T2 were cloned into pX-330 and this was confirmed by colony PCR where we use the forward oligonucleotides of each gRNA from the ligation step as forward PCR primer, while a common reverse primer at the gRNA backbone was used (the positive integration gave around 120bp PCR product) (Figure 3.6. A). In order to confirm the integration, the miniprep plasmid preparations were sent Sanger sequenced (Figure 3.6. B).
After the confirmation step we transfected gRNA SpT1 and SpT2 separately into the Neuro-2A GFP cells, formulated with peptide Y and C18 lipid. In addition, another transfection was performed where both gRNA SpT1 and SpT2 were used in the same well to detect any synergism effect. Cells were harvested after 48 hours and then analysed by flow cytometry. However, we noticed that the GFP protein still was expressed at that time after the knockdown of the gene. So, the transfected cells were passaged twice to determine more accurate results. The results showed that the efficiency of SpT1 resulted in the highest knock-out with an average of 42.3%, while one third of the cells lost the GFP expression when
they were transfected with SpT2. In addition, the combination of both gRNAs in one transfection gave a knock-out average of 39.6%, which is the mean between the results of SpT1 and SpT2 when they were transfected separately (Figure 3.7).

Figure 3.7. GFP knock-out following SpT1 and SpT2 gRNA transfection. The flow cytometry analysis for Neuro-2A GFP cells after transfection with gRNAs targeting the GFP gene (SpT1, SpT2 separately or in a combination of both gRNAs in one transfection). The cells were harvested after 48 hours of transfection, then passaged twice. The results represent single well transfection analysed by flow cytometry.

This result is also confirmed by another assay (T7 endonuclease I), where a PCR product covering the region that is being targeted is denatured by heating, followed by re-annealing at room temperature. The next step is treatment with T7 endonuclease I enzyme, which recognises and cuts mismatches between the
DNA strands. The treated products were then separated on a 2% agarose gel where it showed the main band (546bp) and two extra bands (313bp and 233bp in SpT1 and 309bp and 237bp in SpT2) that were present, while the control sample (transfect only with the original pX-330 plasmid) did not show extra bands as the GFP does not have indels (Figure 3.8.). By measuring the band's intensity, the indels' efficiency was calculated (section 2.2.35), showing that the indels' percentage for SpT1 was 22% while for SpT2 was 24%.

Figure 3.8. T7 endonuclease I assay for two SpCas9 gRNAs targeting GFP in Neuro-2A cells. The cells were transfected with two gRNAs (SpT1 and SpT2) then PCR was performed (546bp product). The PCR products were denatured then annealed followed by T7 endonuclease I enzyme treatment then separated on a 2% agarose gel. The T7 endonuclease I recognises and cuts the mismatching as a result of indels which can be detected as two extra fragments indicated by red and green arrows in the figure.

In the previous experiments the GFP knock-out was performed by the CRISPR/SpCas9 system. Here we assess another variant of the Cas9 system which is the SaCas9 (from *Staphylococcus aureus*) that recognises the 5’ NGRRN 3’ PAM motif. The SaCas9 has an advantage in size, as it is almost 3.4kb compared to 4.1kb coding sequence in SpCas9 (227), making it easier to clone, and the resulting plasmid is smaller in size which in turn can result in higher transfection efficiencies. Here we used the pX-601 plasmid containing the
SaCas9 driven by the cytomegalovirus (CMV) promoter while the gRNA is driven by the RNA polymerase III U6 promoter. The guide sequence (21bp) was cloned after linearising the plasmid with BsaI restriction enzyme upstream of the gRNA backbone (Figure 3.4. and Table 3.1.). The GFP Neuro-2A cells were transfected with the pX-601 plasmid using L2K after the ligation was confirmed. The flow cytometry analysis showed that the SaCas9 was able to knock-down 55% of the GFP cells. The efficiency of SaCas9 was comparable to SpCas9 in this work as well as with the efficiencies reported in the literature (310).

Figure 3.9. The pX-601 plasmid and how to insert the gRNA. A) The overview of the plasmid. The gRNA was driven by the RNA polymerase III U6 promoter, while the SaCas9 (human codon-optimised S. aureus Cas9) was driven by a cytomegalovirus (CMV) promoter. The nuclear localisation signals (NLS) flanked the hSaCas9 sequence. In addition, the plasmid has a Bovine Growth Hormone Polyadenylation Signal (BGHpA). B) Sequence of the chimeric gRNA region. The red arrow indicates the removable sequence of the BsaI restriction site which allows the insertion of the desired gRNA. 89 nucleotides (shown in red) show the trans-activating crRNA (tracrRNA) sequence. C) Guide sequence insert. The blue colour indicates the sequence of 21 nucleotides target sequence without the PAM region (NNGRR). If the target sequence does not start with a G-base, the addition of a G-nucleotide is needed to initiate the U6 promoter. The black letters indicate an additional sequence needed to facilitate the overhang cloning. Adapted from (227).
Figure.3.10. Flow cytometry analysis of Neuro-2A GFP cells after transfection with pX-601 plasmid encoding SaCas9 and gRNA. The cells were harvested after 48 hours of transfection, then passaged twice. The results represent single well transfection analysed by flow cytometry.

3.3.2.2. SpCas9 as mRNA and protein
The next step in the optimisation was to use another form of Cas9, either mRNA or protein, as we saw higher transfection efficiency with the GFP mRNA (Figure 3.2.). The other component of CRISPR system is gRNA which is synthesised by \textit{in vitro} transcription using the pX-330 plasmid as template and where two primers were used (Figure 3.11.). The reverse primer located at the end of the backbone of gRNA in the plasmid while the forward primer has two parts: the T7 promoter sequence and 20 bases matching the targeting sequence (SpT1). By performing PCR with two annealing temperatures, the T7 promoter was incorporated into it followed by \textit{in vitro} transcription using a commercial kit (see section 2.2.26.).
In order to optimise the transfection, different ratios and amounts between Cas9 mRNA and gRNA were used. We started with a small amount of nucleic acid (300ng) with a weight ratio of mRNA 2: gRNA 1 for transfections with L2K. The GFP knock-out was almost 38%. By using a 1:1 weight ratio (200ng RNA:200ng gRNA), the efficiency increased to reach almost 51%. Using 100ng of gRNA with 500ng Cas9 almost (1:1 in molar ratio) increased the knock-out efficiency to 61%. Interestingly, increasing further the amount of Cas9 (above 500ng) or gRNA (above 100ng) did not show any effect on the efficiency (Figure 3.12.).
Figure 3.12. Optimising the Cas9 mRNA and gRNA ratio and amount. The Neuro-2A GFP cells were transfected with \textit{in vitro} transcribed gRNA and Cas9 mRNA. The analysis was performed by flow cytometry to determine the GFP knocked-out cells. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.

The other form of Cas9 that can be used for transfecting the cells is Cas9 protein. This form of Cas9 is available immediately for gene editing inside the cells without the translation step needed with Cas9 mRNA. Additionally, the short half-life and turn-over of the protein might help to reduce any undesirable off-target effect (311). Formulating gRNA with the Cas9 protein forms a molecule called ribonucleoprotein (RNP). To perform the transfection we use the same weight ratio (5 Cas9: 1 gRNA) which was optimised in the previous experiment for Cas9 mRNA, where it showed the optimal ratio for the RNP transfection in the literature (312). The Neuro-2A cells were transfected with RNP formulations that were first assembled at room temperature for five minutes before mixing with L2K. The results showed that the knock-out efficiency of RNP was almost half of the efficiency shown with Cas9 mRNA (Figure 3.13.).
Figure 3.13. Cas9 mRNA and Cas9 protein transfection with gRNA targeting the GFP gene of the Neuro-2A GFP cells. 100ng of gRNA were formulated with either 500ng of Cas9 mRNA or with 500 ng of Cas9 protein and transfected with L2K. The analysis performed by flow cytometry determined the GFP knocked-out cells. The results represent single well transfection analysed by flow cytometry.

3.3.2. Ai9 cell line model
The previous model (GFP cells) was useful to study the NHEJ (GFP knock-out) mediated by CRISPR/Cas9. However, this model has limitations such as waiting to detect the NHEJ event after the transfection as the expressed GFP protein stays in the cells for a few days before it vanishes. Consequently, the application for this approach for in vivo work might be more challenging.

Here, we examined and optimised another in vitro model in order to apply it later on as an in vivo gene editing approach. The Ai9 model has the tdTomato gene (red fluorescent protein variant) in an inactive state where the transcription is prevented by multiple stop codons located upstream of the gene, which can be removed by Cre recombinase as it is flanked by loxp sequence (301) (Figure 3.14.). We aimed to remove the stop codons cassette using the CRISPR/Cas9 technique to enable tdTomato expression as an indication of the NHEJ effect.
Figure 3.14 The Ai9 construct model. The construct has the tdTomato gene at an inactive status where multiple stop codon (orange pentagon) prevent the CAG promoter from driving the transcription. The stop cassette flanked by loxP site is a target site for Cre recombinase. Targeting the stop codons themselves or the flanked sequence by CRISPR system allows for the tdTomato to be expressed as an indication for NHEJ efficiency.

3.3.3.1 Creating the Ai9 model in Neuro-2A cells
The first step was creating the Ai9 tdTomato model in cell lines in order to do the optimisation. The Ai9 was constructed by Linda Madisen and colleagues (301) in a donor plasmid that contains homology arms of the Rosa 26 locus in the mouse genome (Figure 3.15. A). In order to integrate this construct by homologous recombination, a DSB by CRISPR/Cas9 at Rosa26 locus is needed to facilitate the recombination (Figure 3.15. B). A gRNA (Rosa26 gRNA) was designed to target the sequence between the homology arms, which is then cloned into pX-330 plasmid. The gRNA Cas9 plasmid was transfected first to determine the cutting efficiency. The T7 endonuclease assay shows that the indels' efficiency was approximately 24% (Figure 3.16. A). The DSB was confirmed by sequencing with 32% indels being produced as calculated by Tracking of Indels by Decomposition (TIDE) (https://tide.deskgen.com), (313) where Sanger traces made from PCR products of targeted regions from the control sequence of an un-transfected sample and the sequence of the targeted sample were compared (Figure 3.16 B).
Figure 3.15 The strategy to knock-in the Ai9 construct by using CRISPR/SpCas9 system and donor plasmid template. A) Shows the donor plasmid map which contains the right (green) and left (red) homology arm with the Ai9 construct and the Neomycin resistance gene for selection purpose. B) Shows the site of the Ai9 integration by HDR after DSB by CRISPR/Cas9. The donor template has homology at the Rosa26 locus of the mouse genome between exons 1 and 2.
Figure 3.16 CRISPR/Cas9 targeting the Rosa26 locus. The Rosa26 gRNA encoded by pX-330 plasmid was transfected in Neuro-2A cells. A) T7 endonuclease I assay shows the indels' efficiency on a 2% agarose gel. The red and green arrows indicate the fragments after cutting by T7EI while the blue arrow shows the main PCR product. B) Indels analysis assessed by TIDE (Tracking of Indels by Decomposition) where the transfected sample sequence is compared to the control sample sequence.

As we had confirmed the Rosa26 gRNA was able to make DSB, we transfected the Neuro-2A in 24 well plates with two plasmids: the first plasmid encodes both Cas9 and gRNA while the other plasmid has a donor template (Ai9). The transfected cells were passaged twice before the knock-in cells were selected under G418 as the donor plasmid has the Neomycin resistance gene. The G418 concentration for the selection was optimised by a kill curve assay where we found that 1mg/ml of the drug was the minimum concentration to kill all untransfected cells within one week. The resistant cells for G418 were expanded and were then cloned in 96 well plates by serial dilution where one cell was seeded per well. After two weeks the 96 well plates were examined to determine the number of clones per well. We expanded the wells that had only one clone first in 48 well plates, then in 12 well plates. In order to confirm the integration of the Ai9 construct we transfected the clones with Cre plasmid (pCAG-Cre) as the stop cassette is flanked by loxp sites. In addition, the pCAG-Cre has the GFP fluorescent gene which can be used to determine the transfection efficiency.
From 12 clones that were examined three positive clones for the Ai9 knock-in were found. In figure (3.17.), for example, the negative and positive clones were examined under a fluorescent microscope after the transfection and it is shown that the negative clone express GFP only while the knocked-in clone (clone 4) expressed both fluorescent genes (GFP and tdTomato).

![Figure 3.17 Clonal screening for Neuro-2A knock-in with the Ai9 construct.](image.png)

**Figure 3.17 Clonal screening for Neuro-2A knock-in with the Ai9 construct.** Neuro-2A cells were transfected with Rosa26 gRNA encoded by the pX-330 plasmid and the Ai9 donor plasmid. The screening for knock-in was performed by Cre recombinase plasmid which has a GFP gene as indication for transfection efficiency. The positive clone showed tdTomato expression as the stop cassette has been removed. The negative clone does not show tdTomato as the Ai9 is not integrated.

### 3.3.3.2 CRISPR/Cas9 mediate NHEJ in the Ai9 cell model

#### 3.3.3.2.1 Optimising transfection with a plasmid encoding Cas9 and gRNA

The aim here is to target the stop codon cassette by CRISPR/Cas9 in order to remove it and allow the exogenous promoter to drive tdTomato expression. There are two options to perform that, either targeting both ends of the stop cassette and removing it or within the stop cassette in order to disrupt it by the formation of indels (Figure 3.18., Table 3.2.).

In the first option, where the stop cassette is targeted by two gRNAs, we searched for a gRNA targeting both ends. However, the targeting sequence has a 5’NAG’3 PAM motif which has been reported as a non-canonical PAM for SpCas9 that has
lower efficiency (314). This gRNA is located in the loxP site and was named loxP gRNA. In order to determine the gRNA efficiency, we cloned this gRNA in pX-330 plasmid. The efficiency of the loxP gRNA was determined by flow cytometry after 48 hours from the transfection using 500ng of the plasmid delivered to the cells by L2K, where the analysis showed that only 0.69% of the cells expressed the red fluorescent gene (tdTomato) (Figure 3.19.).

Figure 3.18 Stop cassette at Ai9 construct with eight SpCas9 gRNAs targeting the region. The direction of the arrow indicates the strand targeting (sense or antisense) while the PAM motif of each gRNA is indicated by a similar- coloured rectangle.
<table>
<thead>
<tr>
<th>Target</th>
<th>Strand</th>
<th>Sequence</th>
<th>PAM</th>
<th>Efficiency Score</th>
<th>Specificity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>loxP</td>
<td>1 and -1</td>
<td>CGTATAATGTATGCTATACG</td>
<td>AAG</td>
<td>86.79</td>
<td>56.68</td>
</tr>
<tr>
<td>T1</td>
<td>1</td>
<td>AAAGAATTGATTTGATACCG</td>
<td>CGG</td>
<td>78.84</td>
<td>18.37</td>
</tr>
<tr>
<td>T2</td>
<td>1</td>
<td>GTATGCTATACGAAGTTATT</td>
<td>AGG</td>
<td>76.82</td>
<td>19.51</td>
</tr>
<tr>
<td>T3</td>
<td>-1</td>
<td>TAGGAACTTCTTAGGCCGCG</td>
<td>CGG</td>
<td>82.74</td>
<td>17.52</td>
</tr>
<tr>
<td>T4</td>
<td>-1</td>
<td>CTAGCTTGGGCTGCAGGTCA</td>
<td>AGG</td>
<td>75.26</td>
<td>1.43</td>
</tr>
<tr>
<td>T5</td>
<td>-1</td>
<td>AAACCTCTACAAATGGTTGA</td>
<td>TGG</td>
<td>62.18</td>
<td>9.09</td>
</tr>
<tr>
<td>T6</td>
<td>-1</td>
<td>AAGTAAAAACCTCTACAAATG</td>
<td>TGG</td>
<td>55.25</td>
<td>7.12</td>
</tr>
</tbody>
</table>

Table 3.2. List of gRNAs targeting the Ai9 stop cassette. The table shows each gRNA sequence (targeting the DNA strand sense as 1 or antisense as -1), PAM motif sequences, the predicting efficiency score (where a score near 100 is more efficient) and the specificity score (where high score indicates potential more off-targeting). The on-target and off-target scores were calculated by [https://benchling.com/crispr](https://benchling.com/crispr) for the mouse genome.
The efficiency of loxP gRNA was very low despite the high on-target score given by benchling analysis (Table 3.2.), which could be related to the fact that the PAM motif is not the classical NGG. Therefore, we decided to target the stop cassette using two gRNAs, one upstream and one downstream. In order to do that, we cloned four gRNAs into pX-330 plasmid where gRNAs target 1 and 3 are located upstream of the stop codons while the gRNAs target 2 and 4 are downstream (Figure 3.18., table 3.2.). To remove the stop cassette, we transfected the cells with all four possible combinations, which were: T1+T2, T1+T4, T3+T2, and T3+T4. The transfection was performed in 24 well plates with 500ng of each plasmid using L2K.

The flow cytometry analysis showed that the combination of gRNA T1+4 was the most efficient to remove the stop cassette by NHEJ as a result of double strand break (DSB) with 23.7% tdTomato expression. The second highest efficient was T3+T4 with an average tdTomato expression of 17.6% while the T1+T2 gave 13% and the lowest combination was T2+T3 which was almost 9% (Figure 3.20). This results also indicate the gRNA T4 was the most efficient gRNA as both combinations T1+T4 and T3+T4 gave the highest tdTomato expression.
Figure 3.20 Two gRNA approach to remove the stop cassette in Neuro-2A Ai9 cells. Four gRNAs targeting the stop cassette where gRNA target 1 and 3 are upstream while T2 and T4 are downstream the stop cassette. Four gRNA combinations were transfected (T1+T2, T1+T4, T3+T2 and T3+T4) in Neuro-2A Ai9 cells. A) Fluorescent images for tdTomato expression. B) Flow cytometry plot for cells after the transfection. C) The average of tdTomato expression as represented by the NHEJ effect for the four different gRNA transfections. The transfection was performed with L2K and 500ng of each plasmid. *P<0.05; **P<0.01; one-way ANOVA with Bonferroni post-test analysis used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from three independent experiments (N=3) is shown.

As we discussed above the other approach was to target the stop cassette by a gRNA that has multiple targets within this region in order to disrupt it. To confirm the efficiency of this approach we designed two gRNAs (Target 5 and T6) where both have two targeting sites within the stop cassette (Figure 3.18., table 3.2.). The two gRNAs were cloned into pX-330 plasmid. The efficiency for T5 was on average 14% while that of T6 was 15.7% (Figure 3.21). The efficiency of both T5 and T6 was lower than the best two gRNA combinations (T1+T4). In addition, the tdTomato expression after transfection with either T5 and T6 gave low fluorescent
intensity (dim tdTomato) with mean fluorescent intensity (MFI) (around 50) while the MFI for T1+T4 was around 350.

Figure.3.21. CRISPR/Cas9 targeting sequence within Ai9 stop cassette. The target 5 and 6 gRNAs have two loci within the stop cassette. The efficiency determined by the tdTomato expression was analysed by flow cytometry. The results represent single well transfection analysed by flow cytometry.

We conclude from all the above comparisons that T1+T4 was the most efficient approach for removing the stop cassette.

The accumulated effect from multiple dosages might be necessary to reach certain levels where the function of the gene is restored or repressed by knock-out or knock-in. Here we compare the L2K and RTN (C18, peptide Y) to deliver both plasmids encoding the T1 and T4 gRNAs. The results showed that nanoformulation was as effective as the L2K for delivering the Cas9 and gRNA with no statistically significant difference between them. The accumulated expression of tdTomato increased after the second transfection where it reached on average 16% with L2K and 14.6% for RTN after the third dose (Figure 3.22.).
Figure 3.22. Cumulative effect by targeting the stop cassette in Neuro-2A Ai9 cells. Multiple dosage transfection using pX-330 plasmid encoding T1 and T4 gRNAs delivered by L2K or RTN (C18, peptide Y). The tdTomato expression was measured by flow cytometry. One-way ANOVA with Bonferroni post-test analysis was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.

3.3.3.2.2 Optimising transfection with mRNA and protein Cas9
The transfections performed above with CRISPR/Cas9 were done with the plasmid DNA approach. We next compared it with other forms of Cas9 as mRNA or protein transfected in parallel with a synthetic chemically modified gRNA (crRNA+transRNA) or in vitro transcribed gRNA (IVT). The Cas9 mRNA transfection with the synthetic gRNAs showed on average 14.5% tdTomato expression, while Cas9 protein was 12.3% when transfected with L2K (Figure 3.23. B). On the other hand the efficiency of both types of Cas9 were very low when the gRNA was delivered as IVT with an efficiency approximately 3% (figure 3.23. A). In this transfection we used 500ng of the Cas9 (either mRNA or protein) and 62.5ng of each gRNA (T1 and T4).
Figure 3.23 Cas9 mRNA/gRNAs and RNP transfection in Neuro-2A Ai9 cells. The transfection was performed using L2K and 500ng of Cas9 and 62.5ng of each gRNA. A) T1 and T4 gRNAs synthesised by the in vitro transcription technique (IVT). B) T1 and T4 gRNAs synthesised and chemically modified by a manufacturer as two parts (crRNA+transRNA). The results represent single well transfection analysed by flow cytometry.

In the previous GFP model we optimised the transfection of Cas9 mRNA with one gRNA where we found that 500ng of Cas9 and 100ng of gRNA was the optimal amount for the transfection (Figure 3.12.). However, here in Ai9 we transfected the cells with two gRNAs simultaneously. To optimise that we transfected Neuro-2A Ai9 cells with both Cas9 (mRNA and protein) with either 125ng of both gRNAs (62.5ng of each) or 125ng of each gRNA to asses the optimal amount of gRNA needed. The analysis showed that increasing the amount of each gRNA did not show a significant statistical difference (unpaired Student’s t-test) (Figure 3.24.).
In order to optimise the Cre recombinase transfection as postive control for *in vitro* and *in vivo* applications, we transfected the Neuro-2A Ai9 cells with 100ng of Cre mRNA in 24 well plates either with L2K or with nanocomplex (RTN). The ratio for L2K was 3:1, while we performed three mixing orders for RTN (1:3:4, 1:4:3, 3:4:1). The Cre mRNA was 100 ng (1), lipid was 300ng (3), and peptide Y was 400ng (4). Here we used C18 and C14 lipids as well as C18 with 30% cholesterol as the latter is proposed that it stabilises the liposome (315). The analysis by flow cytometry showed the average efficiency for all nanocomplexes and L2K that removed the stop casette by recombination was almost 60% with no statistical significant difference between them (one-way ANOVA test with Bonferroni post-test analysis). In addition, adding 30% of cholesterol to the nanocomplex did not show a significant effect in terms of efficiency (Figure 3.25.).
Figure 3.25 Cre recombinase mRNA transfection in Neuro-2A Ai9 cells. The cells were transfected with 100ng of cre recombinase mRNA delivered as RTN at a set weight ratio ratio (1 mRNA : 3 lipid: 4 peptide Y) with three different orders of mixing and also delivered with L2K. The tdTomato expression was assessed by flow cytometry. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.

To perform the transfection in the Neuro-2A Ai9 cells using RNP we used the same ratio as for mRNA transfection when comparing the mixing order to investigate whether it had any effect. We used 500ng of the Cas9 protein and 62.5ng of each gRNA, the lipid was 1875ng and 2500ng was used for peptide Y. The results showed that nanocomplexes produced with any order of mixing have a similar delivery efficiency for both the Cas9 protein and synthetic gRNAs with no statistically significant difference between them (Figure 3.26.).
Figure 3.26 RNP transfection in Neuro-2A Ai9 cells using RTN. The transfection was performed by two gRNAs T1 and T4 (each 62.5ng) and Cas9 protein (500ng) delivered as RTN at a set weight ratio (1 RNP : 3 lipid: 4 peptide Y) with three different orders of mixing. The tdTomato expression was assessed by flow cytometry. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.

3.3.4. Gene editing in the Ai9 in vivo model
We aimed by creating the Neuro-2A Ai9 cells to establish a model for gene editing and to optimise the delivery techniques in a high-throughput screening manner. Following the in vitro optimisation we decided to assess the gene editing in vivo using the Ai9 mouse model (Gt(ROSA)26Sor<sup>hml9(CAG-tdTomato)Hze</sup>) (301). The in vivo experiment included three groups of mice: 1) control; where no nanocomplexes were administered (n=3). 2) Cre recombinase; 16 µg of Cre recombinase mRNA was formulated with C18 lipid and peptide Y (n=4). 3) RNP; 15 µg of RNP (12 µg of Cas9 protein, 1.5 µg of Ai9 gRNA T1 and 1.5 µg of Ai9 gRNA T4 formulated with C18 lipid and peptide Y (n=4). The targeting nanocomplexes were prepared by mixing the Cre mRNA or RNP with C18 lipid, then adding peptide Y, at a weight ratio of 1 (mRNA or RNP) : 3 (C18) : 4 (peptide Y). Each mouse was given three sequential doses (every other day) of the respective formulation. The lung tissues were harvested 4 days after the last administration and were treated with
DNase/collagenase in order to obtain single-cell suspensions which were then analysed with flow cytometry. To examine the edited (tdTomato+) cells, we performed two types of analyses; either gating on the whole lung cell population (Figure 3.27. A, B and C), or gating on epithelial cells that were negative for CD45 (leukocyte marker) and positive for CD326 (Ep-CAM) (316) (Figure 3.27. A, B, D and E).

Figure 3.27 Flow cytometry gating analysis in Ai9 lung. A) The debris cells were excluded by using side scatter area (y-axis) and forward scatter area (x-axis). B) Single cells were gated using side scatter width (y-axis) and forward scatter height (x-axis). C) Gating on the cells that express tdTomato following gating in B. D) The epithelium cells in mouse lung (negative for CD45 and positive for CD326 markers), followed by E) gating of the cells that express tdTomato. The analysis was performed in lung cells of Ai9 mice transfected with three doses of Cre mRNA delivered with targeting complex (mouse number 2 is depicted here).

The analyses of the whole lung cell population showed on average 0.282% ±0.18 of cells have been edited (expressed tdTomato) with cre mRNA treatment, while the RNP showed 0.07% ±0.03 editing efficiency (Figure 3.28.) with no statistically significant difference between them (p=0.066; one-way ANOVA test followed by
Bonferroni’s post hoc test). The epithelial cells (CD45-, CD326+) in mice lung we found that almost represent only 1% of the whole lung cell population which also is reported in the literature at the same percentage (317). On epithelium cells, the edited cells after Cre mRNA transfection were 0.498% ± 0.32 of the cells which was higher on average by 1.7-fold compared to the editing in the whole lung. On the other hand delivery of RNP resulted in tdTomato expression in 0.458% ±0.12 of lung epithelium cells with no statistically significant between RNP and the positive control Cre mRNA mouse group (p=0.98; one-way ANOVA test followed by Bonferroni’s post hoc test) (Figure 3.29.). When comparing the editing by RNP at the epithelium cell population to that of the whole lung cell population, the editing efficiency increased by 6.5-fold.

**Figure 3.28 tdTomato expression in whole lung of Ai9 mice.** The tdTomato expression was measured following three doses of RTN either with Cre mRNA (16µg) or with RNP (15µg), with the control Ai9 mice not receiving any dose. The tdTomato expression was measured by flow cytometry in the whole lung cell population as explained in Figure 3.27. *P<0.05; One-way ANOVA with Bonferroni post-test analysis was used to assess the significance.
Figure 3.29 tdTomato expression in epithelial Ai9 mouse lung cells. The tdTomato expression was measured after three doses of RTN either with Cre mRNA (16µg) or with RNP (15µg), with the control Ai9 mice not receiving any dose. The tdTomato expression was measured by flow cytometry in epithelial cells by gating on CD45-,CD326+ cells as explained in Figure 3.27. *P<0.05; One-way ANOVA with Bonferroni post-test analysis was used to assess the significance.
3.4. Discussion
In this chapter we aimed to optimise the gene editing in two different reporter cell models (Neuro-2A GFP and Neuro-2A Ai9) mediated by CRISPR/Cas9. The optimisation included transfection with different forms of the CRISPR system [plasmid DNA, mRNA, and ribonucleoprotein (RNP)]. In addition, the optimisation also involved using Receptor-Targeted Nanocomplexes (RTN) as an approach for in vivo gene editing. We showed that the reporter cell model is an inexpensive, high-throughput bioassay for optimising CRISPR delivery that does not involve specialised techniques with high sensitivity.

In the beginning we started transfecting different RTN formulations on unmodified Neuro-2A cells using plasmid DNA (pDNA) and mRNA molecules encoding GFP as a reporter marker for the transfection efficiency, because the plasmid encoding the Cas9 and targeted gRNA (pX-330) does not have a reporter gene. The previous optimisation from our group on RTN showed that the C16 peptide Y was the most efficient formulation for transfecting Neuro-2A using luciferase reporter vector (pCILuc) in comparison to lipid C18 and C14 with peptide Y (294). Here the three formulations with GFP reporter gene encoded by plasmid showed a similar transfection efficiency with no statistical difference.

On the other hand, optimising mRNA transfection with RTN was not well studied in our group, as the only RNA molecule used previously with RTN was siRNA (72, 294, 318-320). This showed the formulation with C18 lipids was the most efficient in terms of silencing. However, mRNA and siRNA are different in size and hence molecular weight which might affect the formulation. Here, we attempted to establish the optimal formulation for mRNA delivery by testing different molecular ratios between the three components (lipid, mRNA, and targeting peptide), while investigating the effect of the order of mixing on the transfection efficiency. We concluded that C14 and C18 are the most suitable lipids for mRNA transfection where mixing the lipid with mRNA or peptide first gave the most efficient formulation for delivery compared to mixing mRNA with targeting peptide followed by adding the lipid (Figure 3.2.).

Using a fluorescent cell model for gene editing is a useful assay, as is does not involve specialised techniques to determine the NHEJ such as Sanger
sequencing or using advanced methods for HDR like next generation sequencing (NGS). In addition, other techniques might lack the sensitivity; e.g. the T7 endonuclease I assay which usually underestimates the NHEJ efficiency (321).

In this chapter we optimised the NHEJ mediated by CRISPR/Cas9 system on two models (GFP and Ai9) which have been used by others to construct different reporter models. An example of that is the traffic light reporter (TLR) system which is based on using a non-functional GFP sequence (as +1 frame caused premature termination) fused to mCherry through a self-cleaving 2A peptide. Targeting the termination codon by a gRNA will cause indels and as a result of that a frameshift to the reading frame which in turn results in mCherry gene translation. On the other hand, when a donor template is available the non-functional GFP will be restored as a result of HDR (Figure 3.30.) (322). Despite this, the reporter system is able to measure both HDR and NHEJ concurrently. However, there are some limitations such as the fact that expression of mCherry represents only one-third of all NHEJ events when the frameshift to the +3 reading frame happens, while any indels that cause +1 or +2 frameshift do not cause mCherry expression.

Figure 3.30. Diagram of the Traffic Light Reporter (TLR) System. Targeting the stop codon of GFP by CRISPR will cause frame shifting and allow mCherry to be expressed (NHEJ). On the other hand, the correction of GFP can be achieved when a donor template is used to correct the mutation there. Adapted from (322).
In this chapter we used the GFP model to assess the NHEJ mediated by CRISPR/Cas9 system, however this model can be used for monitoring the HDR as well. The similarity between the fluorescent protein in the coding sequence, while changing for example one amino acid, can alter the protein’s fluorescent colour (323). A good example of that is changing the green fluorescent protein (GFP) to blue fluorescent protein (BFP) by changing the encoding sequence for Tyrosine amino acid at position 67 from (TAC) to (CAT) which is coding for Histidine (324) (Figure 3.31.). To do that a gRNA targeting the Tyrosine encoding sequence to make a DSB is needed in parallel with a donor template which has the histidine encoding sequence instead of tyrosine.

![Figure 3.31 Modifying the GFP gene to BFP as an approach for HDR gene editing.](image)

Targeting the GFP gene by CRISPR/Cas9 system using a gRNA (grey arrow colour followed by the PAM site) induces a DSB near the Tyrosine coding sequence at position 67. Using a donor template which could be either ssDNA or dsDNA with 2 bases mutation (from TAC to CAT) will change the amino acid at position 67 from Tyrosine to Histidine which will consequently alter the protein fluorescent colour to BFP. In addition, the mutation creates a restriction site (NcoI) which could be used to confirm the editing. Adapted from (324).

The Ai9 model is a useful tool to investigate the efficiency of gene editing both in vitro and in vivo. Mohammadsharif et al. 2016 studied the ability of the CRISPR system as an exon skipping therapeutic approach for Duchenne muscular...
dystrophy (DMD) with a nonsense mutation at exon 23 where two gRNAs are used for exon excision. To repurpose that, they transfected primary satellite cells (in vitro) from the Ai9 mouse with a plasmid encoding SpCas9 and two gRNAs. The flow cytometry analysis showed that 10% of the cells expressed tdTomato. Interestingly, the gRNAs used here were the gRNA Ai9 T1 and T2. We found these to be less efficient than other combinations such as T1 and T4 or T3 and T4 (Figure 3.20.). For in vivo work, SaCas9 and two gRNAs packaged within an AAV vector were injected intraperitoneally or intramuscularly and showed 4% and 3.8% TdTomato+ gene–edited satellite cells, respectively (325). Furthermore, gold nanoparticles were used to inject RNP that contained gRNA Ai9 T1 and T2 in the tibialis anterior muscle of Ai9 mice. The dose for each mouse was 30 µg of Cas9 and 30 µg sgRNAs (15 µg each gRNA). The results showed the ability of the CRISPR system as RNP formulated with nanoparticles to introduce NHEJ in vivo. However, the main limitation in this study was that the editing quantification was not performed (326). The Ai9 mouse models are also useful to study the possibility of gene editing on other organs such as the brain. In this context, Brett Staahl and colleagues demonstrated that the injection of RNP to the Ai9 mice with dose dependance caused activation of the tdTomato expression (327). The approach taken here to use a single gRNA caused multiple cuts within the stop cassette rather than using two gRNAs which we also investigated using a similar strategy (Figure 3.21.). Despite our finding of similar efficiency of this strategy with the two gRNAs approach, the mean fluorescence intensity of tdTomato after transfecting with the one gRNA that caused multiple cuts was less than six-fold compared to the paired gRNAs approach, which makes it less preferable for in vivo application.

Cas9 from Streptococcus pyogenes is the most popular and well established Cas9 in the CRISPR system for genome editing. However, there are other Cas9s from other bacteria species which might be suitable for certain gene editing conditions. For example, Cas9 from Staphylococcus aureus (SaCas9) has an advantage over the SpCas9 in terms of size (1kb smaller) which allows it to be packaged in the AAV vector. In addition, the SaCas9 recognises the 5'-NGRRN-3' PAM sequence which gives other options when an SpCas9 PAM site is not
available (227). In this chapter we investigated only one SaCas9 gRNA on a GFP model which showed almost 50% knock-out (Figure 3.10.). For a direct comparison between the two systems (SpCas9 and SaCas9), the use of the same targeting gRNA is needed. In this context Xiya, et al. 2016 compared the cleaving efficiency in mouse embryos at the Slx2 and Zp1 loci where both gRNAs overlapped. The analysis showed that the cleavage efficiency at the Slx2 locus was 88.8% for SaCas9 and 53.8% for SpCas9, while in Zp1 both Cas9s demonstrated almost the same activity (328).

Using different forms of Cas9 such as 'ready status' as protein or encoded by plasmid DNA or mRNA has advantages and disadvantages. Plasmid DNA delivery is a simple and low-cost approach which is good for initial screening for an efficient gRNA. Validating the editing strategy on an essay of transfected cells such as HEK293 cells (a human cell line) or Neuro-2A (mouse cell line) is helpful before moving to other cells, as these cells are well-established for plasmid DNA transfection. However, the main drawback for plasmid transfection is the low efficiency on slow or non-dividing cells. On the other hand, long expression from pDNA can increase the chance of undesirable off-targets. The mRNA and RNP approaches showed higher sensitivity and specificity but the cost of them is much higher than plasmid cloning (329).

The efficiency of RNP to induce NHEJ was not always higher than gRNA/Cas9 mRNA transfection as we reported here. The editing by RNP could be cell dependant while the delivery method might had an effect. Xiquan, et al. 2015 compared the efficiency of CRISPR/Cas9 as RNP or gRNA/Cas9 mRNA targeting the HPRT locus on different cells. They found some cells such as HEK293FT and Human iPSCs achieved higher level of indels with gRNA/Cas9 mRNA compared to RNP when it was delivered by lipofection while the opposite happens when the electroporation was used where the RNP showed higher editing in the same cells (312).

In the in vivo experiment in the Ai9 mouse model, we showed almost 0.5% of epithelial lung cells edited after three sequential transfections of RNP or Cre recombinase mRNA (Figure 3.29). Interestingly, the editing in the epithelial cells population was higher than the editing in the whole lung cell population which
might indicate the effect of the targeting peptide. Despite the low number of tdTomato cells after the RNP transfection, the efficiency of NHEJ might be higher as this model only shows the cells edited by excision of the stop cassette itself, whilst the cells that had NHEJ occurring at a single locus either upstream the stop cassette or downstream will be tdTomato negative. Further optimisation in the delivery vehicle and CRISPR/Cas9 system is needed in the future to enhance the gene editing in vivo.

In this chapter we used Neuro-2A cells as a well established mouse cell line in our laboratory for RTN transfection (319, 330-332). In addition, the final goal from the CRISPR/Cas9 optimisation is the in vivo work using an Ai9 mouse model.

In terms of future work for the GFP and Ai9 models, these reporter models might be introduced into airway epithelial cells to perform the optimisation where primary epithelium cells might have different responses to CRISPR systems than Neuro-2A cells. The integration of GFP reporter in the cells can be performed by HDR as has been created by Dirk Hockemeyer and colleagues where a donor plasmid with homology arms at the AAVS1 locus was used with a puromycin resistance gene (333). On the other hand, the Ai9 construct needs a modification in terms of homology arms from Rosa26 sequence to a human sequence.

In conclusion, we showed the reporter cell model (Neuro-2A GFP and Neuro-2A Ai9) is an inexpensive, high-throughput bioassay for optimising CRISPR delivery that does not involve specialised techniques with high sensitivity. In addition, the targeting nanocomplex was able to deliver all forms of the CRISPR system with the added advantage of repeat dosing for both in vitro and in vivo models.
CHAPTER 4

Results

CRISPR/Cas9 mediated correction of ΔF508 mutation in CFBE41o- cells
4. CRISPR/Cas9 mediated correction of ΔF508 mutation in CFBE41o- cells

4.1. Introduction

Cystic fibrosis (CF) is a common genetic disease that is lethal when it involves the lungs. The disease is linked to a mutation in the CFTR gene, the most common mutation of which is the ΔF508 mutation, which represents almost 70% of all cases. Several clinical trials have been performed to date assessing the potential of gene therapy to limit the progression of CF lung disease; however, a clinically relevant treatment has yet to emerge (298). The major challenges in gene therapy for CF relate to the limited levels of gene transfer achieved in the lung airway epithelium and the persistence of transgene expression. The gene editing strategy for genetic diseases has been proposed as a potential therapy for diseases such as CF with permanent effects. CRISPR/Cas systems allow for gene-specific, targeted correction of disease-related mutations to be introduced at the chromosomal level with therapeutic potential for CF.

In theory, gene therapy for CF has advantages over other genetic diseases: first, it is caused by a mutation in a single gene; second, CF carrier patients are phenotypically normal; third, the airway system is easily accessible for the delivery of products through nebulisation; and finally, relatively low levels of correction are potentially therapeutic: 5-10% for electrical properties and 25% for mucociliary correction (271).

The CFBE41o- cell is a well-established bronchial epithelial cell line with which to study CF with homozygosity for the ΔF508 mutation; it is easily maintained and grown on permeable supports, forming polarised monolayers on which electrophysiology study can be performed (334, 335). Correcting the ΔF508 mutation by homology directed repair (HDR) in CFBE41o, where CRISPR/Cas systems are used to create DSB, will be assessed at molecular and functional levels.
4.2. Aims

In this chapter, we aim to:

1. Screen for efficient gRNAs and mediate DSB close to the ΔF508 mutation;
2. Correct the ΔF508 mutation of the CFTR gene in CFBE41o- cells mediated by the CRISPR system;
3. Assess the CFTR and ENaC function after mutation correction;
4. Screen the possible off-target effects of CRISPR/Cas9 systems.
4.3. Results

4.3.1. Constructing gRNAs targeting near the ΔF508 mutation and screening for DSB

The main step to performing HDR is creating a double strand break near the site where the correction is needed. So, using the Benchling algorithm (https://benchling.com/crispr), we analysed CFTR exon 10 with the ΔF508 mutation for potential guide RNAs. The analysis showed 15 possible gRNAs for Cas9 targeting. However, we decided to focus on the closest gRNAs to the mutation (T1, T2, T3, T4, T5, T5+, T6 and T7). All gRNAs locations with respect to the PAM site are shown in Figure 4.1., which are also summarised in Table 4.1., with expected cutting distance from the ΔF508 mutation, specificity (off-target) and efficiency score (on-target).

Figure 4.1. CFTR gene sequence near the ΔF508 mutation with the eight possible SpCas9 gRNAs targeting the region. The direction of the arrow indicates the strand targeting gRNA (sense or antisense) while the PAM motif of each gRNA is indicated by the same colour rectangle.
<table>
<thead>
<tr>
<th>Target</th>
<th>Strand</th>
<th>Sequence</th>
<th>PAM</th>
<th>Efficiency Score</th>
<th>Specificity Score</th>
<th>Cutting distance (bases) from CFTR ΔF508 mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>1</td>
<td>CAGTTTTCCTG GATTATGCC</td>
<td>TGG</td>
<td>21.34</td>
<td>41.44</td>
<td>27</td>
</tr>
<tr>
<td>T2</td>
<td>-1</td>
<td>ATATTTTCCTTA ATGGTGCC</td>
<td>AGG</td>
<td>21.62</td>
<td>38.33</td>
<td>20</td>
</tr>
<tr>
<td>T3</td>
<td>1</td>
<td>ACCATTAAGAA AATATCAT</td>
<td>TGG</td>
<td>17.84</td>
<td>40.11</td>
<td>3</td>
</tr>
<tr>
<td>T4</td>
<td>-1</td>
<td>ACCAATGATATT TTCTTTAA</td>
<td>TGG</td>
<td>2.13</td>
<td>37.60</td>
<td>13</td>
</tr>
<tr>
<td>T5</td>
<td>1</td>
<td>GGAGAACTGGA GCCTTCAGA</td>
<td>GGG</td>
<td>84.24</td>
<td>35.14</td>
<td>88</td>
</tr>
<tr>
<td>T5+</td>
<td>1</td>
<td>GGGAGAACTGGA AGCCTTCAG</td>
<td>AGG</td>
<td>16.59</td>
<td>36.30</td>
<td>87</td>
</tr>
<tr>
<td>T6</td>
<td>1</td>
<td>CAAAGCATGCC AACTAGAAG</td>
<td>AGG</td>
<td>26.60</td>
<td>41.44</td>
<td>65</td>
</tr>
<tr>
<td>T7</td>
<td>1</td>
<td>GAGGTTAAAAT TAAGCACAG</td>
<td>TGG</td>
<td>48.91</td>
<td>36.78</td>
<td>69</td>
</tr>
</tbody>
</table>

Table 4.1. List of SpCas9 gRNAs targeting the CFTR gene near the ΔF508 mutation. The table shows each gRNA sequence, targeting the DNA strand sense as 1 or antisense -1; the PAM motif sequence; the predicted efficiency score where the score near 100 is more efficient and the specificity score where a high score indicates potentially more off-targeting and cutting distance calculated by expecting the DSB to occur at position 17 of the guide. The on-target and off-target scores were calculated using the Benchling design tool (https://benchling.com/crispr).

The gRNAs ligated into the pX-330 plasmid as described in Chapter three. After confirming the cloning using Sanger sequencing. We compared gRNAs through screening for NHEJ efficiency by transfecting the targeting gRNAs in HEK293 cells as a well-established cell line for transfection and CRISPR/Cas9 manipulation. To investigate that, HEK293 cells were seeded on 24 well plates then transfected with 500 ng of each plasmid, thus encoding the gRNAs separately and delivered by L2K at ratio 3:1. The DNA was extracted from each transfected well after 48 hours, and then the region of interest was amplified by PCR (CFTR_primer). Treatment the PCR products with T7 endonuclease I showed that four of the gRNAs were able to introduce indels, which were represented by an extra band on the gel (Figure 4.2.). The efficiency of each gRNA was calculated by fraction intensity in comparison with the sum of the cleaved band intensities (section 2.2.34). The most efficient gRNA was target 1 (38% indels), followed by gRNA T5 (34%), and then target 5+, which was 30%.
T7 had 22% indels efficiency. The rest of the gRNAs did not show detectable indels by this assay (Figure 4.2.). Here, we transfected the HEK293 cells with gRNA target 3 and target 4 to detect any off-target possibility as they should be working only in the mutated cell line. The target 3 gRNA had the PAM motif located directly after the deletion (CTT) while target 4 gRNA had the mutation in its sequence (Figure 4.1.).

![Figure 4.2. T7 endonuclease I assay for CFTR gRNAs on HEK293 cells.](image)

The transfection was performed by L2K using pX-330 plasmid. The red and green arrows indicate the fragments after cutting by T7EI, while the blue arrow shows the main PCR product.

To confirm the T7 endonuclease I assay, we sent the samples for Sanger sequencing and then TIDE analysis. The indels were undetectable for gRNA targets 2 and 3 and we could not analyse target 4 because it did not match the wildtype CFTR sequence (specific to the ΔF508 mutation). The rest of the gRNAs showed indels where target 5+ was the most efficient with indels around 50%. Target 1 gRNA and target 7 were in the second rank with efficiency of almost 34%, while target 5 gRNA showed low efficiency in comparison with other gRNAs at approximately 15% indels. Interestingly, the indels of gRNA target 6 were undetectable by T7 endonuclease I assay, while TIDE analysis showed 12.2% indels (Figure 4.3.).
A)

**Figure 4.3.** Indels analysis of targeting *CFTR* by CRISPR/SpCas9 on HEK293 cells. The analysis shows the efficiency of eight CFTR targeting gRNAs on HEK293 cells assessed by TIDE, whereby the transfected sample sequence was compared to the control sample. They represent one transfection by each gRNA. B) Summary of indels efficiency of all eight gRNAs (n=2) were represented as mean ± SD.

In the above experiments, we showed that we were able to make DSBs in wildtype *CFTR* exon 10, whereby the best gRNA cutting was target 1, which approximately cuts at 27 bases from the mutation, which would be expected to produce low efficiency repair of mutated exon 10. Therefore, we explored other CRISPR systems, SaCas9 and Cpf1, to try to identify PAM targets closer to the ΔF508 mutation that might enhance the HDR efficiency later. We designed closer two targets gRNAs recognised with SaCas9 gRNAs and other two target gRNAs.
for Cpf1 CRISPR system (Table 4.2, Figure 4.4, and Figure 4.5). In order to transfec
t the cells with SaCas9 with designated gRNAs, we used pX-601 plasmid,
which has a BsaI cloning site, while for the Cpf1 CRISPR system we used the
pY-094 plasmid, which expresses Cpf1 from Acidaminococcus and crRNA guide
with BsmBI cutting site for cloning the gRNA targeting sequence. For the SaCas9
system, we designed the gRNA to be 21 bases long, instead of the typical 20
bases, as it had been described to be more efficient (227). On the other hand,
the Cpf1 guide sequence was 24 bases, as recommended by Yamano et al. 2016
(336). The four cloned gRNAs were then transfected on HEK293 cells, except for
SaCas9 T1, which was transfected into CFBE41o- cells (specific for the ΔF508
mutation). However, none showed indels during Sanger sequencing (Figure 4.6).
Figure 4.4. **CFTR** gene sequence near the ΔF508 mutation with two SaCas9 gRNAs targeting the region. The direction of the arrow indicates the strand targeting (sense or antisense) while the PAM motif of each gRNA is indicated by the same colour rectangle.

Figure 4.5. **CFTR** gene sequence near the ΔF508 mutation with two Cpf1 gRNAs targeting the region. The direction of the arrow indicates the strand targeting (sense or antisense) while the PAM motif of each gRNA is indicated by the same colour rectangle. The expected cutting for each guide is indicated with a red triangle.
4.3.2. CRISPR/Cas9 – mediated correction of ΔF508 in CFBE41o-

4.3.2.1. gRNAs transfection on CFBE41o-

In the previous experiment, most of the gRNAs worked and introduced indels after DSB; however, two of the gRNAs were more specific to the ΔF508 mutation (Target 3 and target 4) and needed to transfect on mutated cells. CFBE41o- cells were transfected with all eight gRNAs using L2K transfection reagent; they were then treated and processed as described before in the HEK293 cells. The T7 endonuclease I was performed for all gRNAs, which showed that the gRNA target 5+ was the most efficient with 40% cutting efficiency. Target 7 was the second highest gRNA in terms of efficiency with 34.4%, while target 5 gRNA showed around 20% indels after the transfection. The closest gRNAs to the mutation (targets 2, 3 and 4) did not show DSB by this assay. In addition, regarding target 1 and target 6 gRNAs, which were working on HEK293 cells, the activity for them was undetectable by T7 endonuclease I here (Figure 4.7.).

Figure 4.6. Sanger sequence after SaCas9 and Cpf1 transfection. The highlighted sequence indicates the gRNA targeting site either by SaCas9 or Cpf1 while the red rectangle shows the PAM sequence.
Figure 4.7. T7 endonuclease I assay for CFTR gRNA targets on CFBE41o- cells. The transfection was performed via L2K using pX-330 plasmid. The red and green arrows indicate the fragments after cutting by T7EI while the blue arrow shows the main PCR product.

In order to confirm these results, the purified PCR products (CFTR_primer) were sequenced and then analysed using TIDE. The analysis confirmed that Target 5+ was the most efficient at 40.8% indels in comparison with controlled un-transfected samples. The gRNA T7 was the second most efficient, whereby almost a quarter of the cells showed indels. In addition, the sequencing confirmed that the specific gRNAs (T2, 3 and 4) did not introduce indels or under sensitivity of this technique (Figure 4.8.).
4.3.2.2. Correcting the ΔF508 mutation by homologous direct repair (HDR) at CFBE41o-

The main aim of gRNA screening was to determine the most efficient gRNAs, which might increase the chance for HDR events. We concluded from the previous gRNA screening that target 5+ and target 7 were the most efficient at producing indels, so we decided to focus on them for the correction of CFTR in the CFBE41o- cells.

To correct the ΔF508 mutation, we used a donor plasmid that has homology arms for the CFTR, as well as a puromycin N-acetyl-transferase gene (Pac) in the intron to help select corrected cells. Pac was flanked with loxP sites to allow for the removal of Pac by Cre recombinase at a later time, if required (Figure 4.9.) (268). The left homology arm was 1182 bp in length, while the right arm was 382 bp.
Figure 4.9. Strategy to correct the ΔF508 mutation in CFBE41o- using CRISPR/SpCas9 system and donor plasmid template. The donor template has a wild-type exon 10 sequence with two homology arms to facilitate the correction through the HDR pathway. In addition, the donor plasmid has a puromycin resistance gene as a selective marker driven by a phosphoglycerate kinase (PGK) promoter. To detect the event of homologous recombination, PCR reaction with one primer occurred inside the donor while the other primer only occurred at the CFTR gene and vice versa.

In order to use this donor plasmid, we then mutated the PAM 5'‑NGG‑3' to prevent re‑cutting by Cas9 after the repair. The T7 gRNA sequence in the genome locus was followed by TGGA nucleotides, whereby the GGA encodes glycine amino acid. However, all possible silent mutations did not change the first and second nucleotides (GG) (Figure 4.10. A). On the other hand, the T5+ guide was followed by AGG as PAM motif where the last nucleotide (G) of the guide with the first two bases of the PAM (GG) encode glutamic acid. The glutamic acid could be encoded by either GAG or GAA. Changing the coding of glutamic acid to GAA causes silent mutation of the PAM to AAG without affecting the protein sequence (Figure 4.10. B and 4.11.).
Figure 4.10. gRNA T7 and T5+ PAM motif coding amino acid. The sequences show the possibility of changing the PAM from NGG to another sequence not recognised by SpCas9 without affecting the coding amino acid. The silent mutation in the PAM sequence of donor plasmid is important in preventing re-cutting by Cas9 after the repair.

In order to do that, we used site directed mutagenesis to amplify the whole donor plasmid where the forward primer had silent mutation at the PAM locus (forward primer 5’ GGAGCCTCAGAAGGAAAAATTAAG’3 and reverse primer 5’AGTTCGCCATTATCACC ’3). We did this by performing PCR and then digesting the parental plasmid by DpnI followed by annealing both ends. The new construct was confirmed by Sanger sequencing (Figure 4.11 C).
Figure 4.11. The strategy to mutate the PAM region at the donor template of gRNA target 5+ without changing the encoding amino acid. A) The original sequence of the donor template with the gRNA 5+ targeting the red rectangle at the PAM sequence; B) The modified donor is needed where the protospacer’s sequence has been silently changed from AGG to AAG to prevent the DSB re-cut after repair; C) Modified plasmid on agarose gel after PCR where the one primer has the mutation interest. The blue arrow indicates the PCR product (donor template); and D), Sanger sequence of the donor template after site-directed mutagenesis confirms the modification.

In order to start the correction experiment, the kill curve with various concentrations of puromycin (ranged from 0.5 μg/ml to 10 μg/ml) was tested on CFBE41o- to determine the lowest concentration of the antibiotic able to kill all cells within a week. The results showed that the 2 μg/ml is the optimal concentration for the selection on this cell line. We then transfected cells with the two plasmids, one encoding the SpCas9 and the gRNA (either T5+ or T7) and the other plasmid was the donor template (mutated at target 5+ PAM sequence). To confirm the HDR events, 5′ prime (CFTR_CFBE41o-_5′end HDR primer) and 3′ prime (CFTR_CFBE41o-_3′end HDR primer) junction PCR were performed (Figure 4.9.). A positive result, after transfection by PCR, indicated that homologous recombination had happened in both transfections (T5+ and T7) (Figure 4.12). After that, cells were passaged at least five times to dilute any
remaining episomal plasmid inside the cells follow by selecting the cells using 2 μg/ml puromycin. The puromycin resistant cells were later grown as single clones to screen them for correction. Twenty clones for T5+ and 38 for T7 were expanded, which were then screened by 3’ junction PCR (CFTR_CFBE41o- _3’end HDR primer). The results showed only 8 and 19 clones were positive for HDR from the transfection of target 5+ and target 7, respectively (Figure 4.13.). Then, the corrected clones were screened using the ClaI restriction enzyme after the PCR performed (CFTR_primer) to determine the correction level (either heterozygous or homozygous). The screening results showed that all clones with target 5+ were corrected at one allele, while the screening for corrected clones with gRNA T7 showed two clones corrected at both alleles and the rest corrected at one allele (17 clones). (Figure 4.14, Table 4.3).

Figure 4.12. Confirming the HDR event at CFBE41o- by junction PCR. A) PCR gel for 0.9 kb PCR product (3’ prime) for cells corrected either with T5+ or T7 guide RNAs; B) showing the PCR gel for 1.8 kb PCR product (5’ prime) for cells corrected either with T5+ or T7 guide RNAs.
Figure 4.13. HDR screening on CFBE41o- clones. A) 20 puromycin-resistant clones (T5+) were screened: eight clones were positive for HDR (+); B) 38 puromycin-resistant clones (T7) were screened with 19 clones positive for HDR (+). The screening was performed by 3' junction PCR where the forward primer was inside the donor template, while the reverse primer was outside.
Figure 4.14. Allele correction screening on CFBE41o-corrected clones. The corrected clones were screened for homozygosity or heterozygosity correction level using the ClaI restriction enzyme. Clones corrected at one allele (+/-) are partially digested by ClaI, while clones corrected at both alleles (+/+) are totally digested. A) Eight clones where gRNA T5+ was used in parallel with the donor template and confirmed previously by junction PCR. All eight clones displayed monoallelic correction; B) 19 clones with CFTR targeted by gRNA T7 were analysed after positive junction PCR for template insertion. Two of the 19 clones were corrected at both alleles while the rest were corrected at one allele.
Table 4.3. Summary of screening selected clones either with gRNA target 5+ or 7.

<table>
<thead>
<tr>
<th>Target</th>
<th>Number of clones</th>
<th>Corrected clones</th>
<th>Heterozygous/Homozygous</th>
</tr>
</thead>
<tbody>
<tr>
<td>T5+ (+donor plasmid)</td>
<td>20</td>
<td>8</td>
<td>8/0</td>
</tr>
<tr>
<td>T7 (+donor plasmid)</td>
<td>38</td>
<td>19</td>
<td>17/2</td>
</tr>
</tbody>
</table>

In order to confirm the previous screening for hetero/homozygosity for correction by Clal restriction enzymes, genomic DNA amplified by PCR (CFTR_primer) was analysed by Sanger sequencing. The sequencing results confirmed the Clal digestion analysis. In addition, all clones at the targeting region (T5+ or T7) were intact (no insertion or deletion) (for example, clone 5 with T5+ gRNA) (Figure 4.15. A) excepting two clones with T5+, which showed overlapping peaks at the expected DSB site which indicate for indels at non corrected allele (Figure 4.15. B). By performing another PCR where both primers were outside the donor template (CFTR_outside donor primer,1.8 kb), but using a short extension time of 1.5 minutes to enable amplification of the uncorrected allele, we noticed insertion of two nucleotides (TT) (T5+ clone 1), while clone 6 had three bases inserted (TTT) (Figure 4.15 C). In addition, the corrected clones at both alleles were confirmed by sequencing (Figure 4.15 A).
Figure. 4.15. Sanger sequencing for corrected CFBE41o- clones. A) sequence for two clones (T5+ clone 5 and T7 clone 3) as well as CFBE41o- original cells at the ΔF508 mutation site and PAM locus at the end of T5+; B) sequences for two clones (T5+ clone 1 and 6) have indels at the uncorrected allele as well as clone 5 from the same targeting; C) sequence result when the uncorrected allele was amplified by PCR (1.8 kb) where both primers were located outside the donor template and extension time was 1.5 minutes for amplification.

4.3.2.3. Functional analysis for corrected CFBE41o- clones

4.3.2.3.1. CFTR and SCNN1 mRNA expression after the correction of CFBE41o-

The effect of correcting CFTR at the mRNA level was performed by qRT-PCR on RNA samples extracted from uncorrected CFBE41o-, 16HBE cells and two corrected clones: clone 5, where T5+ gRNA was used for DSB (one allele corrected), and clone 3, where T7 gRNA was used to make the DSBs (both alleles corrected). In addition, we used RNA extracted from CFBE41o- targeted with gRNA T5+ or T7 and donor plasmid, which was then selected with puromycin but not cloned (mixed population). The analysis showed the CFTR mRNA expression increased in all corrected cells (mixed population and clones) (Figure 4.16.). In CFBE41o- clone 5, the CFTR mRNA expression increased by 630-fold in comparison with uncorrected cells; while in CFBE41o- T7 clone 3, the expression was almost 234-fold higher than uncorrected cells. Both values were highly statistically significant, with p values less than 0.0001 using the one-way ANOVA test followed by Bonferroni’s post hoc test. In addition, the mixed
population of corrected cells, which were selected by puromycin but uncloned, demonstrated an increase at *CFTR* mRNA expression in comparison with uncorrected CFBE41o-. The cells where T5+ gRNA was used showed an 8.5-fold increase at *CFTR* expression (not statistically significant; one-way ANOVA test followed by Bonferroni’s post hoc test), while the cells targeted by gRNA 7 then selected showed an 145-fold increase (p value = 0.0003; one-way ANOVA test followed by Bonferroni’s post hoc test). The *CFTR* mRNA expression in 16HBE14o- cells was very high in comparison with uncorrected CFBE41o- cells (142,707-fold difference), while the difference for *CFTR* mRNA in comparison with clone 5 was 226-fold (Figure 4.16.).

![Image](image_url)

**Figure 4.16. CFTR mRNA expression in CFBE41o-, corrected CFBE41o- and 16HBE14o-cells.** *CFTR* gene expression measured in CFBE41o- cells, two corrected CFBE41o- clones (one allele corrected (5) or both alleles corrected (3)), a mix population of CFBE41o- cells with gRNA T5+ or T7 and finally in 16HBE14o- (16HBE) cells. The expression was calculated using the 2−ΔΔCt method. The *CFTR* mRNA expression value of CFBE41o- cells was defined as 1 and expression of all other cells was normalised to this value. **P<0.01; ****P<0.0001; one-way ANOVA with Bonferroni post-test analysis was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.
We also investigated the effect of correcting the CFTR ΔF508 mutation on expression of the epithelial sodium channel (SCNN1). In the airways, it has been reported that the activation of CFTR is associated with reduction of ENaC activity; in the sweat duct, the interaction is different, the activation of CFTR is associated with ENaC over-activity (337). We hypothesised that ENaC activity may decrease after correction of the ΔF508 mutation. By performing qRT-PCR to assess the expression of SCNN1 (SCNN1A, SCNN1B, SCNN1G) and comparing the expression to uncorrected cells, we detected reduction at SCNN1B only with almost 50% for both corrected clones (5 and 3). The reduction was statistically different for T5+ clone 5 (p= 0.0001, one-way ANOVA followed by Bonferroni’s post hoc test) and T7 clone 3 (p= 0.0001, one-way ANOVA followed by Bonferroni’s post hoc test) (Figure 4.17. B). The expression at SCNN1A and SCNN1G after the correction either slightly increased or did not have an effect in comparison with uncorrected CFBE41o- cells (Figure 4.17 A and C).
Figure 4.17. *SCNN1* mRNA expression in CFBE41o-, corrected CFBE41o- and 16HBE cells. *SCNN1* gene expression measured in CFBE41o- cells, two corrected CFBE41o- clones (one allele corrected (5) or both alleles corrected (3)) and finally in 16HBE cells. A) *SCNN1A* expression; B) *SCNN1B* expression; C) *SCNN1G* expression. Expression was calculated using the 2−ΔΔCt method. The *SCNN1* mRNA expression value of CFBE41o- cells was defined as 1 and expression of all other cells was normalised to this value. ****P<0.0001; one-way ANOVA with Bonferroni post-test analysis was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.

4.3.2.3.2. Electrophysiology study using an Ussing chamber

We next examined whether CFBE41o- cells corrected by CRISPR/Cas gene editing in one allele (Figure 3.15. B, T5+ clone 5) or both alleles (Figure 3.15. C, T7 clone 3) restored the activity of the CFTR chloride channel. CFBE41o- cells, corrected and uncorrected (negative control), and 16HBE14o- cells (positive control) were maintained on snapwells in liquid-liquid interface (LLI) cultures. Cells were maintained for two weeks to allow monolayer polarisation and then
the transepithelial electric resistance (TEER) was measured. The presence of tight junctions was shown by TEER values above 350 Ω.cm²; such cultures were then taken forward to ion transport studies. CFBE41o- parental cells and the corrected clone (T5+ clone 5) were first studied on an Ussing chamber where the Cl⁻ concentration in the buffer (Ringer’s solution) at both apical and basolateral sides are symmetric. The first treatment was with amiloride (10 μM), which is an inhibitor for the epithelial Na⁺ channel ENaC; however, both cells did not show any response that could be related to the absence of ENaC at the apical surface in these cells(337). The cells were then treated by forskolin (25 μM), a 3’, 5’-cyclic AMP (cAMP) agonist and 3-isobutyl-1-methylxanthine (IBMX) (100 μM) to activate the CFTR channel. The corrected clone showed an increase in the short circuit current values (I_sc) of 0.026 μA/cm² and the uncorrected cells showed a very small increase of 0.006 μA/cm². Following that, the activity of the CFTR channel was blocked by 10 μM (CFTRinh-172) when the I_sc dropped by 0.46 μA/cm² in the single allele corrected clone while the uncorrected CFBE41o- cell fell by 0.04 μA/cm² (Figure 4.18 A and B).

In an attempt to focus on chloride currents through the apical membrane, we used an asymmetric chloride buffer where the apical bath solution has a lower chloride concentration than the basolateral bath. The gradient chloride technique showed improvement regarding the response for the treatment in the corrected clones. The forskolin and IBMX increased the I_sc by 0.32 μA/cm² and 0.006 μA/cm² at T5+ clone 5 and T7 clone 3, respectively. Following the treatment, the cells with the CFTR inhibitor drug reduced the current by 1.33 μA/cm² for T5+ clone 5 while the same treatment only dropped I_sc by 0.12 μA/cm² with the clone corrected in both alleles (T7 clone 3) (Figure 4.18 B and C).

Electrophysiology studies were also performed on 16HBE14o- cells as a positive control where the CFTR channel conducts normally. The 16HBE14o- showed higher transepithelial electric resistance (800 Ω.cm²) in comparison with either corrected or uncorrected CFBE41o- cells. The CFTR channel activated first with forskolin and IBMX increased the I_sc from 72 to 77.3 (equal 5.3 μA/cm²) while the CFTRinh-172 drug caused a reduction in the current of approximately 15.5-fold (Figure 4.18 E and F).
Figure 4.18. Electrophysiology study on epithelial cell lines. The short-circuit current (Isc) traces of CFBE41o- cells in: A) uncorrected CFBE41o- cells; B) CFBE41o- cells corrected in one allele (T5+ clone 5); C) CFBE41o- cells corrected in both alleles (T7 clone 3); E) 16HBE (wildtype CFTR) and their response for the treatment with forskolin and IBMX (apical and basolateral) followed by CFTR inhibitor-172 (apical) was recorded using the Ussing chamber; and D) aggregated data of short-circuit current measurements from A, B and C experiments. The Cl\(^{-}\) gradient experiment was used where indicated (the buffer at the apical side has a low chloride concentration in comparison with the basolateral side).

In conclusion, from correcting the ΔF508 mutation at CFBE41o-, we showed that we were able to target the CFTR gene by CRISPR and inducing DSB where it
enhances the HDR with donor plasmid has puromycin selection marker. The corrected cell clones showed an increase in the \textit{CFTR} mRNA expression while the CFTR channel showed some response in the electrophysiological studies.

4.3.3. Correcting the ΔF508 mutation on CFBE41o- cells without selection

Correcting the ΔF508 mutation under a selection marker and then cloning the cells is an approach for proving the principle for restoring the gene function and might be feasible for ex vivo therapy. On the other hand, we want to investigate the possibility of correcting the ΔF508 mutation without performing selection, which is more suitable for a therapeutic approach where the selection cannot be used. In order to perform this, a single DNA strand (ssDNA) without a selection marker was used as a donor template. The 199 bases oligonucleotide had 52 bases homology in the right arm and contains a novel ClaI restriction site while the left arm had 144 bases with a Xhol restriction site (Figure 4.19 A). CFBE41o-cells were co-transfected with pX-330 encoding target 7 gRNA and Cas9 nuclease, and the ssDNA. Cells were harvested after 48 hours then PCR was performed (CFTR_primer); they were then digested with the restriction enzymes (either ClaI or Xhol). The restriction fragment length polymorphism (RFLP) analysis does not show any extra bands as an indication of a HDR event with this approach (Figure 4.19 B).
Figure 4.19. The single strand oligo template strategy to correct the ΔF508 mutation. A) 199 ssDNA covers the mutation with two silent mutations that were introduced to make the restriction enzyme cutting site (XhoI and ClaI); B) the gel electrophoresis shows that the RFLP assay was unable to detect the HDR event after transfection of the cells with the donor template (ssDNA) and gRNA (target 7) either with XhoI or ClaI restriction enzyme.

4.3.4. CFTR gRNA off-target screening

Earlier we screened for efficiency of on-target activity for gRNAs target 5+ and 7 in different cells, but here we attempted to screen the specificity for both gRNAs by assessing the potential for off-target DSBs in other similar sequences.

Five possible off-target sites for each gRNA were selected based on off-target sites located in exons and three further sites were selected with the highest off-target score and any possible off-target followed by 5′ NGG 3′ PAM sequence (Table 4.4. and Table 4.5.). HEK293 cells were transfected with pX330 plasmid encode either T5+ or T7 gRNA; we amplified the region of interest by PCR and then analysed the amplified products using Sanger sequencing. The results did not show any alternation on the chromatogram trace at the end of each off-target site where the cutting was expected, which indicates that both gRNA T5+ and T7 did not have off-target effects at the loci we screened or had very low off-target effects, which cannot be detected by Sanger sequencing technique. The Sanger sequence for the second off-target locus of gRNA T7 missed because of the technical difficulty in amplifying that region (Figure 4.20.).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>PAM</th>
<th>Score</th>
<th>Gene</th>
<th>Chromosome</th>
<th>Strand</th>
<th>Mismatches</th>
<th>off-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGGAGAACTGGAGCTTCAG</td>
<td>AGG</td>
<td>100</td>
<td>ENSG00000001626</td>
<td>chr7</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GAGAGAAGCTTGGACTTCAG</td>
<td>CAG</td>
<td>4.249</td>
<td>chr16</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CGAGAGAAGCTTGGCTTCAG</td>
<td>TGG</td>
<td>2.7</td>
<td>chr2</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GGGAGAGCTGGGCGTTCAG</td>
<td>AAG</td>
<td>2.128</td>
<td>chr16</td>
<td>-1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GGCTGACTGGAGCTTCAG</td>
<td>AGG</td>
<td>1.686</td>
<td>chr10</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GGCAGAATTGGCCTTCAG</td>
<td>CAG</td>
<td>1.568</td>
<td>chr12</td>
<td>-1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGGAGAATGGAGCTTCAG</td>
<td>GGG</td>
<td>1.553</td>
<td>chr8</td>
<td>-1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTGGGAAGTGGCTTCAG</td>
<td>CGG</td>
<td>1.52</td>
<td>chr22</td>
<td>-1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGTTAAGATTGGCTTCAG</td>
<td>GGG</td>
<td>1.464</td>
<td>chr19</td>
<td>-1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAAAAGATTGGAGCTTCAG</td>
<td>GGG</td>
<td>1.458</td>
<td>chr17</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GACCTAACTGGAGCTTCAG</td>
<td>TGG</td>
<td>1.326</td>
<td>chr11</td>
<td>-1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGGAGAATGGAGCTTCAG</td>
<td>TGG</td>
<td>1.08</td>
<td>chr17</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGGACAAAGTGGAGCTTCAT</td>
<td>CAG</td>
<td>0.85</td>
<td>chr14</td>
<td>-1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGGAGAAATGGAGCTTCAG</td>
<td>CAG</td>
<td>0.82</td>
<td>chr5</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTGTGAAATGGGCTTCAG</td>
<td>GAG</td>
<td>0.808</td>
<td>chr7</td>
<td>-1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAGTGAACCTGGAGCTTCAG</td>
<td>GAG</td>
<td>0.741</td>
<td>ENSG000000054523</td>
<td>chr1</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 4.4. Potential off-target for gRNA target 5+. The table shows the top 15 off-target and PAM sequences on the human genome with chromosomal location and strand direction (sense (1) or anti-sense (-1)) where the red letter indicates the nucleotide difference with on-target sequence. The blue highlighted sequence indicates targeting CFTR sequence and the grey highlighted sequences are chosen for screening. The Benchling design tool (https://benchling.com/crispr) was used for the off-target screening.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>PAM</th>
<th>Score</th>
<th>Gene</th>
<th>Chromosome</th>
<th>Strand</th>
<th>Mismatches</th>
<th>off-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAGGTTAAATAGCACAG</td>
<td>TGG</td>
<td>100.00</td>
<td>ENSG00000001626</td>
<td>chr7</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GGGAATTAAAGCACAG</td>
<td>AAG</td>
<td>1.77</td>
<td>ENSG000000272163</td>
<td>chr8</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>GAGATTAAATAGCACAG</td>
<td>AAG</td>
<td>1.66</td>
<td>chr8</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>TATGGTAATAGCACAG</td>
<td>AAG</td>
<td>1.56</td>
<td>chr1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GAGTTAAAAATAGCACAG</td>
<td>CAG</td>
<td>1.41</td>
<td>chr6</td>
<td>-1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAATTAAATAGCACAG</td>
<td>GAG</td>
<td>1.39</td>
<td>chr6</td>
<td>-1</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GGGACCTAAATAGCACAG</td>
<td>TGG</td>
<td>1.33</td>
<td>chr3</td>
<td>-1</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GAGGATAATAGCACACG</td>
<td>GAG</td>
<td>1.15</td>
<td>chr5</td>
<td>-1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAGGATAATAGCACACG</td>
<td>GAG</td>
<td>1.15</td>
<td>chr5</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATGGTTAAATAGCACAG</td>
<td>GAG</td>
<td>1.07</td>
<td>ENSG00000141052</td>
<td>chr17</td>
<td>-1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>GTGGGAAAAATAAGCACAG</td>
<td>AAG</td>
<td>1.00</td>
<td>chr5</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAGGCTGATTTAGCACAG</td>
<td>AGG</td>
<td>0.99</td>
<td>chr20</td>
<td>-1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAGTCTCAATTTAGCACAG</td>
<td>CAG</td>
<td>0.95</td>
<td>chr2</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAGGTAATTTAAGCACAG</td>
<td>CAG</td>
<td>0.89</td>
<td>chr1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5. Potential off-target for gRNA target 7. The table shows the top 15 off-target and PAM sequences on the human genome with chromosomal location and strand direction (sense (1) or anti-sense (-1)) where the red letter indicates the nucleotide difference with on target sequence. The blue highlighted sequence indicates targeting CFTR sequence and the grey highlighted sequences are chosen for screening. The Benchling design tool (https://benchling.com/crispr) was used for the off-target screening.
Figure 4.20. Sanger sequence screening for five off-target gRNA target 5+ and 7 on HEK293 cells. The highlighted sequence indicates the potential off-targeting sequence at human genome while the red rectangle shows the PAM sequence for each potential off-target. The Sanger sequence for the second off-target locus of gRNA T7 missed because of the technical difficulty in amplifying that region.

<table>
<thead>
<tr>
<th>Off Target</th>
<th>T5+</th>
<th>T7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Sequence1" /></td>
<td><img src="image2" alt="Sequence2" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image3" alt="Sequence3" /></td>
<td><img src="image4" alt="Sequence4" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image5" alt="Sequence5" /></td>
<td><img src="image6" alt="Sequence6" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image7" alt="Sequence7" /></td>
<td><img src="image8" alt="Sequence8" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image9" alt="Sequence9" /></td>
<td><img src="image10" alt="Sequence10" /></td>
</tr>
</tbody>
</table>
4.4 Discussion

The aim in this chapter was to correct the ΔF508 mutation of the *CFTR* gene using the CRISPR system in CFBE41o- cells and then evaluate the correction at the mRNA level as well as performing an Ussing chamber assay to study the effect on the CFTR function. In addition, we studied the ability to detect the correction without using a selection marker and the off-target effect.

To evaluate the correction of ΔF508 of the *CFTR* gene, we used CFBE41o- as the cell line model because it has numerous advantages. These cells are homozygous for the ΔF508 mutation and can be easily maintained and grown on permeable supports in LLI cultures, forming polarised monolayers enabling both molecular studies of correction as well as analysis of ion transport functionality in the Ussing chamber. However, there are limitations for this cell line, for example, the lack of cilia formation and mucus production, meaning the effect of the correction on these features cannot be assessed (338). In addition, these cells produce a low level of CFTR mRNA and protein (339, 340). Consequently, most studies of these cells involve transduction with wildtype or ΔF508 *CFTR* to generate a higher expression.

The prediction of the efficiency of gRNAs before transfecting cells is challenging. Xu et al. 2015 have attempted to understand the criteria of the influence of *Streptococcus pyogenes* gRNA on target efficiency. First, they found that the cytosine base is preferred at position 18 where there is DNA cleavage by the Cas9. In addition, adenines are favoured from position 9 to 16. Finally, guanine bases are preferred at positions four to seven and at positions 19 and 20. In contrast, some nucleotide bases are not favoured in the gRNA sequence. For example, thymine at position 17 (341). In another observation from the David Root lab, found the greatest influence on the gRNA activity was shown to be position 20 (next to the PAM motif), whereby the (G) nucleotide was strongly preferred instead of the cytosine base, while at position 16 the cytosine was preferred, which is almost the same observation as that from the Zhang lab. In addition, Root showed that adenine nucleotides are preferred at the middle of the gRNA. Additionally, the thymidine (T) base is preferred at the first nucleotide of the PAM while cytosine is disfavoured (342). However, these criteria are still
algorithmic findings and might not apply to all gRNAs. A good example is that of target T5+ gRNA for CFTR, which has a low on-target score (Table 4.1); however, when tested it, it showed the highest efficiency regarding indels in both the T7 endonuclease and TIDE assays (Figure 4.7. and 4.8.). On the other hand, the CFTR gRNA T6 meets most criteria for an efficient gRNA; however, the indels, as a result of the transfection, were low on HEK293 cells (Figure 4.3.) and were almost undetectable at CFBE410- (Figure 4.8.). Actually, there are other factors that might have an effect on this system, such as cell types and the local chromatin structure where there is gRNA targeting (342, 343). In another study from the Zhang lab, the activity of Cas9 to induce DSB was shown not to be affected by the methylation status at the targeting region neither \textit{in vitro} or \textit{in vivo} (344).

Here, we attempted to target the mutation locus (ΔF508) by various gRNAs in the beginning with SpCas9. However, the closest gRNA targets (2, 3 and 4) did not induce detectable indels, neither by T7 endonuclease I nor TIDE assays. For that reason, we investigated targeting with other CRISPR systems (SaCas9 and Cpf1 system) but neither induced indels at that locus. The gRNA T3 has been used previously for correcting intestinal stem cell organoids from patients with an ΔF508 mutation; however, the efficiency has not been determined while corrected cells are screened based on puromycin selection (98). For future work, further studies might be needed to investigate the chromatin structure at ΔF508 mutation locus, which might have an effect on accessibility for the CRISPR system.

To determine the indels as a consequence of NHEJ, we performed T7 endonuclease I and TIDE assay. The TIDE assay in general is more sensitive; for example, T6 on HEK293 cells where the efficiency of indels was 13.3% (Figure 4.3.) while the indels were not recognised by T7 endonuclease I assay (Figure 4.2.). In a recent study by Sentmanat et al. 2018, where they compared the two above techniques as well as next-generation sequencing (NGS) as the gold standard for detecting the indels events, they found that the T7 endonuclease I assay gave an underestimate for the indels because it depended on the diversity of the heteroduplex formation. For example, a gRNA named M1
caused very high indels (96%) when it was measured by NGS; however, it gave 10% indels by T7 endonuclease I because the majority of editing was 1 base deletion (more than 70%). In addition, they suggested that TIDE assay is a cost effective technique and gave almost similar results to deep sequencing by NGS when the indels were within 10 bases from the DSB (321).

In addition, for the three techniques we mention above to determine NHEJ efficiency, there are other assays that have been reported in the literature such as high-resolution melting analysis (HRMA), which is based on analysing the difference in denaturing a double-stranded fluorescently-labelled DNA fragment before and after editing by the CRISPR system (345). In addition, indel detection by amplicon analysis (IDAA) was used to evaluate the editing by NHEJ where it was able to detect indels even at a very low event (1%). The IDAA based on capillary electrophoresis of amplicon labelled with 6-FAM (346).

On the other hand, we used Sanger sequencing as a method for the off-target screening as an alternative method for T7 endonuclease I, whereby the sensitivity of the TIDE method to detect indels produces a resulting DSB of 1-2% (313). However, so as to have more sensitivity and cover more off-targeting locus screening, the Sanger sequencing might not be the most suitable technique, particularly for in vivo gene editing. More sensitive techniques have been proposed as a standard method for off-target and most of them are based on NGS, which can be divided into in vitro genome-wide assays such as Digenome-seq and SITE-Seq or cell-based genome-wide assays such as GUIDE-seq and BLISS.

The Digenome-seq (in vitro Cas9-digested whole genome sequencing) technique, based on digestion of genomic DNA with CRISPR/Cas9 as RNP in vitro, was followed by fragmentation and whole genome NGS. The sensitivity of this technique has been reported to detect 0.1% indels as off-target (347). On the other hand, selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-Seq) using a tagging technique in order to enrich the cleaving area, followed by sequencing then mapping to reference the genome (348). On cell-based genome-wide assays, the screening performed on genomic DNA taken from the cells was treated by the CRISPR system rather than
incubation of the DNA with RNP complex. The GUIDE-seq (genome-wide, unbiased identification of DSBs enabled by sequencing) technique, based on insertion of double-stranded oligodeoxynucleotide (dsODN) at the DSB on living cells, followed amplification then NGS. Despite the sensitivity of this method (0.03%), it was based on the possibility of the integration of dsODN at the break site (349). On other hand, Crosetto et al, 2013 developed a BLESS (direct in situ breaks labelling, enrichment on streptavidin, and NGS) technique for an off-target screening method, which uses biotinylated linking for labelling, then capturing the fragments with streptavidin and, finally, performing NGS. This technique was performed after fixing the cells, which means the ability for detection was limited by the time of permeabilization, and any repair for DSB occurring before that moment could not be detected (350).

In the correcting approach for ΔF508 mutation, we modified the donor plasmids at the PAM sequence following the gRNA T5+ from 5′ NGG 3′ to 5′ NAG 3′ to prevent re-cut after the correction. Upon the clones, we screened where T5+ or T7 was used for DSB; we did not note indels on the corrected allele(s). Paquet et al. 2016 studied the benefit of mutating the PAM on the donor template to prevent the re-cut after HDR occurred; they used deep sequencing at the corrected region and they found the indels occurred only at almost 15% of all corrected alleles, while this percentage increased five-fold when no modification had been incorporated in the donor template (351).

In addition, the other technique has been used to eliminate CRISPR interference after correction when changing the PAM changing not possible in creating multiple silent mutations at gRNA targeting sites, particularly the seeding region (the last 10 nucleotides close to PAM) (352).

In this work, we managed to correct the ΔF508 mutation in both alleles on only two CFBE41o- clones when T7 gRNA was used (2 of 19 clones); we could not find bi-allelic correction with T5+ (Figure 4.14.), which indicates that homozygosity of correction is a rare event. In other similar work, Schwank et al. 2013, in their study to correct ΔF508 in intestinal stem cells by CRISPR/Cas9, managed to correct 16 clones; however, all of them were corrected at one allele (98), while the correction by the TALENs system at CFBE41o- using super-exon
strategy only reported the correction at a monoallelic level (267). The possibility of homozygous correction by HDR increased when the intended incorporated sequence was close to the DSB site while the chance of monoallelic and biallelic events were almost the same when the distance from the DSB was 5 bp (267). This finding correlated to our results as the two clones were corrected by T7 gRNA, which was located closer to the ΔF508 mutation than T5+ (69 bp vs 87 bp, respectively).

To facilitate the selection of homozygous corrected clones, a mix of two donor templates, each with a different selection marker such as puromycin and neomycin, could be used. Despite the importance of having a selection marker in the donor plasmid, however, the antibiotic resistance gene is undesirable, especially for any in vivo work. So, the resistance gene (pac) was flanked by loxp, which can be removed by the Cre recombinase for future work. Another approach that could be used in the future is PiggyBac (PB) transposon, which is a footprint-free system in comparison with the Cre-Lox system (270).

The efficiency of HDR mostly depends on the distance between the DSB and site of modification and the efficiency of DSB. In the correcting strategy for ΔF508, the gRNA T5+ was very efficient (almost 40% indels); however, the HDR was undetectable without antibiotic selection as the distance from the expected locus of cutting and ΔF508 mutation is 87 bp. In general, it is recommended having 10 bp or less between the DSB and locus of the HDR (353). Liang et al. 2017, examined several gRNAs to insert 6 bp using CRISPR/Cas9 and donor template (dsDNA and ssDNA); they found the closest gRNAs, which cut at position −3, +3, and +5, from the insertion site gave the highest HDR (almost 30%); in another gRNA, although there were indels above 70%, HDR was less than 2% as a result of the long distance between the DSB and the modification locus (354). In a similar work by Paquet et al. 2016, where they studied the effect of distance between the DSB and the desire editing locus on HDR efficiency. In HEK293 cells, they tested one gRNA with multiple ssDNA where they made a base change in different distances from the DSB site; they found that the efficiency of HDR decreased dramatically by 50% when the distance was 5 bp or greater (351).
Despite the ΔF508 mutation affecting the CFTR at a protein level by improper folding and, consequently, causing retention in the endoplasmic reticulum and degradation (355), here we showed that the correction of ΔF508 had an unexpected effect on CFTR mRNA expression in CFBE41o-, whereby clone 5 corrected with gRNA T5+ (one allele corrected) showed almost 600-fold increasement in CFTR expression, while clone 3 from T7 gRNA transfection (both allele corrected) showed 234-fold increase compared to parental CFBE41o- cells (Figure 4.16.). This phenomenon has not been described previously, to our knowledge; when the correction of ΔF508 has been performed, it focuses mainly on evaluating the CFTR protein expression and function (268, 270, 356). The increase of the expression of CFTR mRNA could be related to alterations in mRNA structure affecting the stability. Bartoszewski et al. 2010, studied the secondary structure of the CFTR mRNA with and without the ΔF508 mutation by molecular modelling prediction and RNase assays; they found in the secondary structure of ΔF508 mRNA two enlarged single stranded loops near to the mutation site (357). On the other hand, the regulatory elements such as microRNA could have an important influence on the mRNA suppression when the mutation is present. The overexpressing microRNA-145, -223, and -494 have been reported in CF cell lines as well as from patients with ΔF508 mutation in comparison with non-CF cells. In addition, transfecting bronchial epithelium cells (16HBE14o-) with the previous microRNAs (145, 223, and 494) showed a reduction of CFTR mRNA expression, whereby CFTR mRNA expression increased on CFBE41o- when the anti-miRNA oligonucleotides were used (358). In addition, Ramachandran et al. 2012 showed that microRNA-138 regulate the CFTR mRNA and protein expression (359).

In electrophysiology studies on CFBE41o- cells, we assessed the effect of correction of CFTR on ENaC activity as the amiloride-inhibited short-circuit current (I_{sc}) was absent in the Ussing chamber study and mRNA expression of the SCNN1B was inhibited in both corrected clones analysed (one allele and both alleles corrected) (Figure 4.17). The expression of ENaC protein has been reported in this cell but the protein does not reach the cells even when the cells grown as ALI culture. Rubenstein et al. 2011, reported that the treatment of
CFBE41o- cells with dexamethasone can enhance ENaC cell surface expression (337). For future work, experiments might be needed to confirm the downregulation of ENaC expression either by treating the cells with modulator molecules or CFTR overexpression by transfection.

In the Ussing chamber study (Figure 4.18), the response of the corrected cells for both the CFTR activator and the inhibitor was not high. This could be due to the fact that these cells (CFBE41o- cells) do not have a high expression of CFTR protein, as the immunoblotting for the immature protein (band B) by western blot has not been reported in the literature (61, 360). Here we showed how the corrected cells responded to CFTRinh-172 as a channel blocker by reducing the $I_{sc}$ by 1.33 µA/cm². This is comparable to other work on the same cells edited by ZFNs (almost 2 µA/cm² reduction) (267). In future work, the use of modulators as positive control on CFBE41o- cells might be necessary.

Detection of HDR events was a challenge because of the low efficiency on general. Introducing a silent mutation in the corrected sequence to make a restriction enzyme site can facilitate HDR identification although the RFLP technique lacks sensitivity at low correction levels. Advanced techniques such as NGS or droplet digital PCR (ddPCR) are more sensitive, however, they are also more complex and expensive.

The CRISPR/Cas9 system has some limitations and potential disadvantages. The availability of the 5'-NGG-3' PAM sequence for Cas9 from *Streptococcus pyogenes* occurs on average every 8 to 12 nucleotides in the human genome (361, 362). The drawback of this system and other gene editing techniques is the potential off-targeting, especially when this work is going towards *in vivo* therapy.

The crucial step is designing a gRNA that has a low off-targeting possibility, especially at the seeding region, which are the last 12 nucleotides before the PAM sequence (238). A strategy to reduce potential off-target events is using the CRISPR/Cas9 nickase system, which contains two opposite mutated Cas9 (D10A) and each guide RNA cleaves only one DNA strand. The off-target activity produced by this system shows a reduction of 50–1,000 fold (221). A similar idea used where *FokI* nucleases fused to dCas9 that only make a double strand break when dimerized (261). Interestingly, work by Fu et al. 2014 were able to decrease
the off-targeting dramatically by designing a shorter gRNA (17 or 18 nucleotides) without significantly affecting the efficiency of the system (260).

On the other hand, the application of CRISPR Cas9-mediated HDR for future use in vivo is more difficult in comparison with the ex vivo approach. Potential problems are a relatively low percentage of HDR events without the use of a selective marker and the mucus barrier in CF patients.

In summary, we have shown correction of ΔF508 of the CF gene at a molecular level in CFBE41o- cells. In addition, the corrected, cloned cells demonstrated increase CFTR expression at the mRNA level as well as the chloride transport.
CHAPTER 5

Results

CRISPR/Cas9 mediated correction of ΔF508 mutation in CFBE primary cells
5. CRISPR/Cas9 mediated correction of ΔF508 mutation in CFBE primary cells

5.1. Introduction

In the previous chapter we assessed correction of the ΔF508 mutation mediated with the CRISPR/Cas9 system on CFBE41o- cells. However, this cell model lacks the complete phenotype from the tissue of origin, such as mucus production when the cells are polarised as a monolayer (338) which affected the assessment of the mucociliary clearance function despite the fact that the CFBE41o- cells develop microvilli and small cilia (363). In addition, the electrophysiological study of the sodium channel (ENaC) activity is challenging as these cells do not respond to amiloride treatment as inhibitor drugs as the protein did not respond to treatment with the ENaC inhibitor amiloride as the ENaC protein does not reach the apical membrane (337).

Primary bronchial epithelial cells from CF and non-CF donors can be grown in defined media with growth factors on a coated surface (89) where these cells can differentiate to pseudostratified cells that form tight junctions and cilia. These produce mucus when they are grown as an ALI culture, which is an in vitro model of the human airway. However, the primary epithelium cells can only be passaged two to three times before they lose their proliferation and differentiation potential, which makes work on these cells limited, especially in cell engineering and editing (90). Overexpression of the BMI1 gene in the bronchial epithelial cells extends the proliferation potential and delays senescence up to 25 passages without losing their ability for differentiation, and preserving their ion transport function (93).

Transfecting primary cells in vitro and in vivo are challenges for pDNA as the cells are slow or non-dividing which is a barrier for the pDNA transcription (364). In addition, it has been thought that plasmid DNA is eliminated through cell divisions (365). The mRNA molecules have an advantage over pDNA in that they do not require transfer into the cell nucleus for translation, which is important in these cells (366). In addition, transfecting the primary cells with recombinant protein is proposed as an alternate approach for pDNA transfection for a fast onset of action (367).
5.2. Aims
In this chapter, we aim to:
1. Optimise transfection in primary CFBE BMI-1 cells using pGFP and GFP mRNA delivered by RTN;
2. Optimise CRISPR/Cas9 transfection in primary CFBE BMI-1 cells as Cas9 mRNA/gRNA and RNP delivered by RTN;
3. Correct the ΔF508 mutation of the CFTR gene in CFBE BMI-1 cells mediated by the CRISPR system;
4. Access the CFTR and ENaC function after mutation correction.
5.3. Results

5.3.1. Optimising transfection in CFBE primary cells

5.3.1.1. Transfection of submerged CFBE primary cells

The CFBE BMI-1 cells as primary cells are more difficult to transfect. In order to investigate this, these cells were transfected with 500ng DNA plasmid (pEGFP-N1) per well of a 24 well plate using Receptor-Targeted Nanocomplexes (RTN) as well as L2K. The GFP plasmid was formulated with DTDTMA/DOPE (C14), DHDTMA/DOPE (C16) and DOTMA/DOPE (C18) with targeting peptide (E) at a 1:4:1 weight ratio (1 Lipid:4 peptide:1 DNA) while L2K was used at a 3:1 weight ratio with plasmid DNA. The flow cytometry analysis after 48 hours showed ~20% of the cells were transfection with each formulation with no significant difference between them (one-way ANOVA followed by Bonferroni’s post hoc test, Figure 5.1.).

![Figure 5.1. GFP plasmid transfection on primary CFBE BMI-1 cells using RTN and L2K. The GFP expression following transfection were assessed by flow cytometry. One representative experiment (n=3) displayed as mean ± SD from three independent experiments (N=3) is shown.](image)

To improve the transfection efficiency, we attempted to optimise GFP mRNA transfections instead of DNA plasmid. Here we use less nucleic acid (100ng of mRNA) for the transfection while the ratio of components and order of mixing is somewhat different from those in DNA plasmid transfections. The weight ratio and order of mixing for was 1 mRNA: 3 lipid then 4 peptide. Here we used two targeting peptides: peptide E and peptide Y, that target epithelial cells (293). The
GFP mRNA was formulated with three cationic lipids C14, C16, C18 with and without 30% cholesterol in order to investigate the efficiency of cholesterol on the transfection. The transfection was analysed by flow cytometry 24 hours after transfection, and fluorescent images were taken (Figure 5.2.). The results showed that the C14 and C18 with peptide E were the best formulations for mRNA delivery, achieving almost 82% transfection efficiency, with no significant difference between them. Interestingly, C14E and C18E showed a significant difference in comparison with L2K (p<0.05, one-way ANOVA followed by Bonferroni’s post hoc test). On the other hand, incorporating 30% cholesterol into the lipid components of the nanocomplex formulation did not enhance the transfection. Furthermore, peptide (E) showed significant statistical difference in comparison with peptide Y (p<0.0001, one-way ANOVA followed by Bonferroni’s post hoc test) when it was used with C18 lipid. In addition, the mixing order was important for an efficient complex for the transfection as seen when the mRNA was mixed with peptide (E) first then C18 lipid (with and without cholesterol) was added, which caused reduction twofold in comparison with the other order where mRNA mixed first with lipid then peptide added later (P value less than 0.0001, one-way ANOVA followed by Bonferroni’s post hoc test). The same pattern was also seen when peptide Y was used. The transfection with lipid without targeting peptide was inefficient (less fourfold compared to RTN formulation) while the peptide alone was the lowest with 2.8% of the cells transfected (Figure 5.2.).
Figure 5.2. mRNA transfection on submerged primary CFBE BMI-1 cells using RTN and L2K. A) Fluorescent images showing GFP mRNA transfection. B) The GFP expression encoded by mRNA transfection was assessed by flowcytometry. *P<0.05; ***P<0.001; ****P<0.0001; one-way ANOVA with Bonferroni post-test analysis was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from three independent experiments (N=3) is shown.

5.3.1.2 Transfection on Air-Liquid Interface (ALI) Cultures

In the previous experiment, the transfection of submerged CFBE BMI-1 primary cells was significantly increased by using mRNA instead of DNA plasmid. To mimic the situation in vivo, the primary cells were cultured at ALI. By using a higher amount of GFP mRNA (500ng) we transfected the cells either with RTN (C14 peptide E or C18 peptide E) (1 mRNA: 3 lipid then 4 peptide) or with L2K (1 mRNA:3 L2K). L2K had the highest transfection efficiency with 13.7% on average followed by C18 peptide E which had nearly 4% and finally C14 peptide E with 2.8% (Figure 5.3).
5.3.2. CRISPR/Cas9 - mediated correction of ΔF508 in primary cells from the bronchial epithelium

5.3.2.1. Optimising CRISPR/Cas9 transfection in primary CFBE BMI-1 cells

In previous sections we assessed the transfection efficiency in CFBE BMI-1 cells using GFP as plasmid or mRNA, where we showed that the transfection increased dramatically with GFP mRNA. Here, we now evaluate the three options of delivery of SpCas9, i.e. in a plasmid, as mRNA, and as protein.

On submerged cultures, CFBE BMI-1 cells were transfected with pX-330 plasmid (500ng) encoded SpCas9 and CFTR gRNA T7, either with L2K or C18/peptide Y nanoparticles on 24-well plates. The TIDE assay analysis showed a very low indel frequency, of less than 1% (Figure 5.4. A and B). To improve the editing efficiency, the primary cells were transfected first with Cas9 mRNA (500ng) and T7 gRNA (125ng) (in vitro transcribed, IVT) formulated either with L2K (molecular ratio 1:3 L2K) or C18 lipid/peptide Y (at ratio 1:3 C18: 4Y). Despite seeing enormous improvements previously with GFP mRNA compared with DNA, the transfection with Cas9 mRNA (500ng) and IVT gRNA (125ng) did not improve the indel efficiency (Figure 5.4. C and D). To investigate that we used a chemically-modified, synthetic double guide RNA (crRNA and transRNA) instead of IVT gRNA which might not be suitable for primary cells transfection. The indels improved dramatically when L2K was used as the delivery method, showing an average indel frequency of 25.75%, while chemically modified gRNA and Cas9 mRNA caused 27.5% indels when formulated with C18Y (Figure 5.4. E and F).

On the other hand, transfecting the cells with ribonucleoprotein (RNP) (500ng Cas9 + 125ng gRNA) showed only a slight improvement compared with Cas9 mRNA either with L2K (31.4% indels, Figure 5.4. G), C14Y (28.9% indels, Figure 5.4. J) and C18Y (29.6% indels, Figure 5.4. N) in comparison with Cas9 mRNA. Formulating the RNP with another targeting peptide (peptide E) showed almost the same results seen with peptide Y (Figure 5.4. I and M). To study the effect of the order of mixing on RNP delivery, we mixed the RNP with peptide E first then
added the lipid (weight ratio 1 RNP: 4 peptide: 3 lipid). The TIDE assay analysis showed a slight improvement with C14E and C18E (34.9% and 38.9% respectively) but was not statistically significant (one-way ANOVA followed by Bonferroni’s post hoc test) (Figure 5.4. K and O). Finally, we attempted to evaluate the editing using Cas9 protein and synthetic gRNA on the CFBE BMI-1 cells in air-liquid interface (ALI) cultures. The sequence analysis, however, did not show any detectable alternation at the cutting site, with L2K (Figure 5.4. H), or with C14E (Figure 5.4. L) or C18E (Figure 5.4. P).

To evaluate the possibility of improving the editing efficiency by multiple transfections using the targeting nanocomplex (RTN), we transfected the CFBE BMI-1 cells with RNP either with C14Y or C18Y. The first transfection with C18Y showed on average 46.3% indels (figure 5.5. C) which improved by almost double after the second transfection (80.2%) (Figure 5.5 D). On the other hand, the transfection with C14Y caused on average 41.6% indels (Figure 5.5. A) which improved less dramatically after the second transfection to reach 59.7% (Figure 5.5 B).

In summary, the gene editing by CRISPR/Cas9 system worked on primary CFBE BMI-1 cells when the system delivered as gRNA/Cas9 mRNA or RNP, and we showed the RTN formulation was as efficient as the commercial reagents.
Figure 5.4. Indel analysis of CFTR following transfections with CRISPR/SpCas9 in CFBE BM-1 cells. A) to P) show the indel efficiencies of the CRISPR system (CFTR gRNA target 7) as assessed by TIDE (Tracking of Indels by Decomposition), where the sequence from the transfected sample was compared to the control sample. A) and B) transfections were performed with pX-330 plasmid encoding T7 gRNA and SpCas9. C) and D) transfections were performed by Cas9 mRNA and in vitro transcribed (IVT) gRNA. E) to F) transfections were performed by Cas9 mRNA and chemically-modified gRNA (crRNA+transRNA). G) to P) transfections were performed by Cas9 protein and chemically-modified gRNA (crRNA+transRNA). Each condition was represented by one analysis (n=3). Q) summary of the efficiency of indels of all transfection conditions (n=3). The results are presented as mean ± SD. The pX-330 plasmid was transfected with L2K at a weight ratio of 3:1 or as RTN at a weight ratio of 1:4:1 (1 lipid: 4 peptide: 1 plasmid). Cas9 as mRNA or protein was formulated with gRNA (IVT or synthetic) and transfected with L2K or RTN at the same weight ratio and order of mixing (1 Cas9+gRNA:3 lipid: 4 peptide) unless otherwise indicated. The transfections were performed on submerged cells unless indicated as grown at air-liquid interface (ALI). One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.
Figure 5.5. Indel analysis of repeat transfection targeting \textit{CFTR} by CRISPR/SpCas9 in CFBE BMI-1 cells. The analysis shows the indel efficiency of SpCas9 and CFTR target 7 gRNA assessed by TIDE (Tracking of Indels by Decomposition) where the transfected sample sequence was compared to the control sample. A) the first transfection with the CRISPR system using C14 lipid and peptide Y. B) the second transfection with the CRISPR system using C14 lipid/peptide Y nanoparticle. C) first transfection with the CRISPR system using C18 lipid/peptide Y nanoparticle. D) second transfection with the CRISPR system using C18 lipid and peptide Y. Each condition is represented by one transfection \((n=3)\) from two independent experiments \((N=2)\). E) Summary of the efficiency of indels following repeat transfection using RTN. Cas9 protein was formulated with synthetic gRNA and transfected as RTN at the following weight ratio and order of mixing: 1 Cas9+gRNA:3 lipid: 4 peptide. One representative experiment \((n=3)\) displayed as mean \(\pm\) SD from two independent experiments \((N=2)\) is shown.
5.3.2.2. Modify the selection marker at the donor

In the previous chapter, CFBE41o cells were corrected with a donor template containing a puromycin resistance gene as a selective marker, however, here on the CF primary cells we cannot use the same donor template as these BMI-1 transduced cells have puromycin resistance gene (Pac) integrated into their genome as a result of the lentivirus transduction. In order to modify the donor template, we amplified the Neomycin resistance gene (neo) from the pY-004 plasmid (Addgene plasmid # 69976) (Forward primer ccaacctggtttctccgcctAATGTGTCAGTTAGGG, Reverse primer catgatctatatagatcccccgTACAGACATGATAAGATACATTG) as well as the original donor plasmid without the Pac gene (Forward primer CGGGGGATCTATAGATCGAG, Reverse primer AGGCGGAAGAACCCAGCTG) by PCR. By adding homology sequences during the PCR amplification, we assembled them using Gibson cloning (Figure 5.6 A). The growing colonies were screened by restriction digest then confirmed by sequencing (Figure 5.6 B).

![Figure 5.6](image)

Figure 5.6. Replacement of the puromycin resistance gene at the donor plasmid with the neomycin resistance gene. A) Agarose gel for amplified PCR fragment in order to perform Gibson cloning where lane (1) showed around 1.5kb PCR fragment of neomycin gene taken from pY-004 plasmid and in lane (2) the original donor template without puromycin gene (around 6kb in size). B) showed the screening for a positive clone using restriction enzymes. AflIII cuts once
at the new donor template and yields an approximately 7.5kb product, while KasI enzyme digested the plasmid in two sites which resulting in 2.8kb and 4.7kb products. HindIII was able to cut into three locations and results in 0.35kb, 1.6kb, and 5.55kb fragments.

5.3.2.3. HDR in CFBE BMI-1 primary cells
The fundamental step before selecting the corrected cells after the transfection with the donor plasmid and CRISPR/Cas9, was to determine the lowest concentration of the selection drug, G418 (Geneticin), that killed the cells within a week. By using various concentrations, we found that 375 μg/ml was the optimal concentration. Therefore, CFBE primary cells with transfected with two plasmids, one encoding Cas9 and gRNA (T5+), while the other was the donor plasmid. The transfections were performed with C18 lipid and peptide E on a 12-well plate (Figure 5.7.). The corrected cells were selected with G418 for one week, however the antibiotic killed almost all cells. In order to improve the number of drug-resistant cells, the transfection was performed in a T75 flask which had 1x10^6 cells seeded the day before. The cells were maintained in primary cell media supplemented with G418. The resistant cells took almost a month to achieve confluence again in the same flask. To confirm the integration and HDR, junction PCR at 3’prime end (CFTR_CFBE BMI-1_3’end HDR primer) which gave the expected product (2.3kb) followed by Sanger sequencing to confirm the HDR event (Figure 5.8.). On the other hand, the attempt to perform junction PCR at the other side (5’prime end) was failed to produce a PCR product.
Figure 5.7. The strategy to correct the ΔF508 mutation by using the CRISPR/SpCas9 system and donor plasmid on CFBE BMI-1 cells. The donor template has a wild-type exon 10 sequence with two homology arm flanks that facilitate the correction. In addition, the donor plasmid has a Neomycin resistance gene as a selective marker driven by the SV40 promoter. To detect the event of homologous recombination, a PCR was performed with one primer designed inside the donor template while the other primer was designed to within the CFTR gene and vice versa.

Figure 5.8. Confirming the HDR event in CFBE BMI-1 cells by junction PCR. The agarose gel showed the junction PCR product at 3’prime end (2.3kb) where the cells were transfected with a gRNA (target 5+) and donor plasmid.
We aimed then to isolate single clones of corrected cells to expand then perform functional evaluation. The cloned cells started to grow in 96 well plates, however, once they were transferred to a larger well plate size, the cells stopped growing and dividing which could be result of reaching the senescence stage. Accordingly, we decided to continue the work on a mixed population of corrected cells. In order to determine the correction efficiency, we performed PCR for the cells where the forward primer was outside the donor template to avoid amplification of the donor template (CFTR_ 5’end outside donor primer) (Figure 5.9. A), then cloned the PCR product which was followed by a Clal digest to confirm the correct identity of the product. Upon the digest of 10 colonies we found only one colony with an extra Clal site, which indicates the HDR efficiency after the selection was approximately 10% of the whole cell population (Figure 5.9. B).

Figure 5.9. Strategy to screen for the correction efficiency in CFBE BMI-1. A) A) Sanger sequencing was performed for the mixed population of CFBE BMI-1 cells after selection with the G418 drug. The highlighted sequence shows the gRNA 5+ targeting sequence, the PCR product cloned into plasmid for screening. B) The miniprep of PCR clones digested by Clal restriction enzyme to confirm the HDR event.
5.3.2.3. Functional analysis for corrected CFBE BMI-1 cells

5.3.2.4.1. CFTR and SCNN1 mRNA expression after the correction

The next step after confirming CFTR correction by HDR at the molecular level, was to investigate the effects on the mRNA and protein level, and then to evaluate their electrophysiological function in ALI cultured cells.

We first assessed the effect of correcting the ΔF508 mutation directly on CFTR mRNA, or indirectly by measuring the SCNN1 expression (SCNN1A, SCNN1B, SCNN1G). CFTR mRNA analysis by qRT-PCR showed that the expression of the gene after correcting the ΔF508 mutation had increased by ten-fold (Figure 5.10; not significant compared to CFBE un-corrected cells; one-way ANOVA followed by Bonferroni’s post hoc test). The expression of CFTR in the NHBE BMI-1 cells was almost 2,000-fold higher in comparison with the corrected CFBE and 20,000-fold higher than un-corrected CFBE BMI-1 (Figure 5.10; p<0.0001; one-way ANOVA followed by Bonferroni’s post hoc test).

![Figure 5.10. CFTR mRNA expression of primary epithelium cells.](image)

**Figure 5.10. CFTR mRNA expression of primary epithelium cells.** CFTR expression was measured on primary epithelial BMI-1 cells (CFBE, mix corrected CFBE, NHBE). The expression was calculated by the 2−ΔΔCt method. The CFTR mRNA expression value of CFBE BMI-1 cells was defined as 1 and expression of all other cells was normalised to this value. ****P<0.0001; one-way ANOVA with Bonferroni post-test analysis was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.
The other indirect assessment for the CFTR mutation correction was measuring the SCNN1 mRNA expression, CFBE cells express a high level of sodium channel, where we expected the level to decrease after the correction. The qRT-PCR results confirmed SCNN1 mRNA levels were downregulated (SCNN1A, SCNN1B, SCNN1G) (Figure 5.11.). The expression of SCNN1A after the correction in CFBE BMI-1 cells decreased by almost 43% (p=0.006; unpaired Student’s t-test) (Figure 5.11. A). Interestingly, the SCNN1B mRNA expression after the correction showed the highest reduction by 60% (p=0.002; unpaired Student’s t-test) (Figure 5.11B). Finally, in SCNN1G we noticed the same pattern where the expression declined by nearly 48% after the correction (p=0.0062; unpaired Student’s t-test) (Figure 5.11. C).

SCNN1 mRNA expression of CFBE and corrected CFBE cells. A) SCNN1A expression. B) SCNN1B expression. C) SCNN1G expression. The expression was calculated by the 2−ΔΔCt method. The SCNN1 mRNA expression value of CFBE BMI-1 cells was defined as 1 and expression of all other cells was normalised to this value. **P<0.01; unpaired Student’s t-test was used to assess the significance. One representative experiment (n=3) displayed as mean ± SD from two independent experiments (N=2) is shown.

To confirm the effect of restoring the CFTR expression of SCNN1 we transfected CFBE BMI-1 cells with CFTR mRNA (300ng in 6 well plate) where the RNA was extracted after 48 hours after the transfection. By performing qRT-PCR assay, we confirmed that the CFTR was overexpressed by more than 200,000 fold
compared to untransfected CFBE (Figure 5.12. A). On the other hand, the expression of both SCNN1B and SCNN1G was downregulated by 44% (Figure 5.12. B) and 23% (Figure 5.12. C), respectively, while the SCNN1A expression was not downregulated following the CFTR mRNA transfection (Figure 5.12. B).

Figure 5.12. SCNN1 mRNA expression of CFBE cells transfected with CFTR mRNA. A) CFTR mRNA expression B) SCNN1A expression. C) SCNN1B expression. D) SCNN1G expression. The expression was calculated by the 2−ΔΔCt method. The mRNA expression value of CFBE BMI-1 untransfected cells were defined as 1 and the expression of all other cells was normalised to this value. One representative experiment (n=2) displayed as mean ± SD from two independent experiments (N=2) is shown.

5.3.2.4.2. Correction analysis at the protein level

The aim of this experiment was to determine whether the mixed population of corrected cells was able to produce the mature form of CFTR protein (fully glycosylated band C). This was done by performing the Western blot from three primary cells including CFBE (uncorrected), CFBE (mixed population, corrected), and NHBE (wt.CFTR) and positive control (HEK293 transfected with CFTR expressed plasmid). However, the immunoblot did not show either the CFTR C or B in mutated cells or mix corrected CFBE BMI-1 cells. Surprisingly, the CFTR
protein also was not detectable in wild-type primary cells (NHBE) (Figure 5.13.). This indicates that the CFTR protein was expressed in low levels in these cells, or might be as result of BMI-1 overexpression or the from the effects of multiple passaging. Further investigation is needed in future work to evaluate the outcome of correcting ΔF508 mutation on protein level.

![Figure 5.13. Immunoblot for CFTR protein on primary epithelial cells.](image)

**Figure 5.13. Immunoblot for CFTR protein on primary epithelial cells.** Three protein lysates from primary epithelial BMI-1 cells (CFBE, mixed corrected CFBE, NHBE) and a positive sample of HEK293T cells were transfected with the plasmid encoding CFTR. The positive control shows both CFTR bands, immature (B) and mature (C), which we also expected in wildtype epithelial cells (NHBE) and corrected CFBE, while the ΔF508 CFBE only expresses the immature CFTR form (band B).

### 5.3.2.4.3. Electrophysiological study using an Ussing chamber

To examine whether the correction of the CFBE BMI-1 cells as a mixed population will restore the activity of the CFTR chloride channel, we grew these cells, as well as CFBE uncorrected (negative control) and NHBE cells (positive control), as ALI cultures on snapwells. Four weeks after the differentiation, the transepithelial electric resistance (TEER) was measured and showed well developed tight junctions with TEER above 500 Ω.cm², then taken forward to an Ussing chamber, ion transport study. CFBE BMI-1 corrected and uncorrected cells were treated first with amiloride (10 μM), which is an inhibitor for the epithelial Na⁺ channel ENaC. The short-circuit current dropped by 6 μA/cm² in the corrected CFBE epithelium while the amiloride caused a larger reduction of 8
µA/cm² in CFBE uncorrected cells. The sodium channel is over-expressed on epithelium cells with the ΔF508 CFTR mutation and so, in accordance with the results, we expected a reduced amiloride response (Figure 5.14. A and B). In order to activate the CFTR channel by increasing the intracellular cAMP, the cells were treated with forskolin and IBMX (apical and basolateral sides). The corrected cells showed an increase in the I sc by 0.45 µA/cm² while the uncorrected cells did not respond positively. Finally, the inhibitor drug added (CFTRinh-172) apically, showed a reduction in the I sc by 0.15 µA/cm² in both the corrected and uncorrected CFBE cells (Figure 5.14. A and B).

In addition, the same treatment was performed on NHBE BMI-1 cells, primary epithelial cells with wildtype CFTR, as a positive control. The NHBEs showed good resistance with their response to amiloride causing a drop in the short-circuit current from 32 µA/cm² to 4 µA/cm² while the forskolin and IBMX increased it by almost 0.4 µA/cm². The inhibitor drug for the CFTR channel caused a small reduction in the current by -0.3 µA/cm² (Figure 5.14 C and D).
Figure 5.14. Electrophysiology study in primary epithelial BMI-1 cells. The short-circuit current (Isc) traces of CFBE, mixed population corrected CFBE cells (A and B) and NHBE (C and D), and their response to the treatment with amiloride (apical side), forskolin and IBMX (apical and basolateral sides) followed by CFTR Inhibitor-172 (apical side) were recorded by the Ussing chamber. Vertical deflections reflect the Isc responses to the application of 2 mV pulses across the epithelium.
5.4 Discussion
The aims of this chapter were, to optimise the transfection of CF primary cells, to optimise CFTR indel formation with the CRISPR system on these cells using Cas9 as mRNA and protein with chemically modified gRNA, and finally, to correct the ΔF508 mutation of the CFTR gene mediated by the CRISPR system. The effects of gene correction were than analysed at the mRNA and protein level, as well as performing an Ussing chamber assay to study the effects on the ion transport efficiency and CFTR function.

In the first part, we addressed the difficulty of pDNA transfection on CF primary BMI-1 cells (Figure 5.1.) where we achieved a maximum of 20% GFP expression. The rate of cell division in primary cells, in general is slower than in cell lines, which increases the barriers to nuclear uptake of pDNA and transcription (364). On the other hand, plasmid DNA is eliminated through more frequent cell divisions (365).

The mRNA molecules have an advantage over pDNA in transfections, that they do not require transport into the cell nucleus for translation, which is important in slow or non-dividing cells (366). Here we showed that transfections with GFP mRNA in RTN formulation were more efficient than pDNA (Figure 5.2.), achieving up to 80% transfection efficiency. On the other hand, mRNA transfections with formulations of liposomes or peptide alone, was not as efficient as when combined together. Here the RTN showed a four-fold increase over liposome transfection.

In our lab we have shown the same synergic effect in lipopolyplexes with pDNA and siRNA in various cells including epithelial cells (320, 368-370). The peptide in the RTN formulation (peptide E) has a dual function, first by targeting the ICAM-1 receptor on the airway epithelium through its SERSMNF ligand. Second, the oligolysine motif at the other end of the peptide chain mediates packaging of nucleic acid (371). The lipid component part of the formulation contains a cationic lipid and a neutral lipid. Cationic lipids (such as DOTMA) are essential for packaging the nucleic acid molecules, while a neutral lipid (DOPE) is promoting endosomal escape which allows trafficking of the nucleic acid to the cytoplasm then to the nucleus (372).
In GFP mRNA transfection we showed (Figure 5.2.) that the C14 (DTDTMA/DOPE) lipid and C18 (DOTMA/DOPE) are more efficient than C16 (DHDTMA/DOPE). Interestingly, each nucleic acid has a preferential lipid in terms of transfection efficiency. Tagalakis et al. 2011 demonstrated that the C16 lipid on 1HAEo- epithelium cells was more efficient for transfection of plasmids encoding luciferase compared to C14 and C18. On the other hand, C18 showed a significant difference when GAPDH siRNA were used compared to shorter alkyl tail lipid (C14 and C16) (294). The optimal liposome for mRNA transfection correlated with siRNA results as they are both RNA molecules. The reason for that could be related to less flexible alkyl tails in C16 liposomes, which affect the stability of nanoparticles and packaging (373).

The attempt to improve the transfection efficiency by adding cholesterol to the RTN formulation did not show an effect on these primary cells but in vivo applications might have an advantage over non-cholesterol RTN in terms of stability (374). Various ratios have been investigated between cholesterol and lipids to enhance the stability without affecting the efficiency, and it was found that the inclusion of 30% cholesterol was the best (315).

Moving to gene editing on the CF primary cells, which is the main aim for this chapter, we attempted to screen the indel efficiency mediated by CRISPR/Cas9 using different form of molecules (pDNA, gRNA/Cas9 mRNA, and ribonucleoprotein (RNP)). In the beginning we transfected the cells with pDNA (pX-330) encoding Cas9 and gRNA, however the indels were almost undetectable by TIDE assays, even with L2K (Figure 5.4. Q). In fact, it has been reported that gene editing using pDNA in primary cells such as human keratinocytes and other hard-to-transfect cells such as Human iPSCs, was unsuccessful unless an electroporation delivery method was used (312). On the other hand, transfecting the Cas9 as mRNA or protein with chemically modified gRNA made a huge difference (Figure 5.5. Q). The chemical modification of the gRNA molecule has been reported to increase the intracellular stability and as a consequence to enhance the editing effect (375). While the in vitro synthesis of gRNA in the lab using kits might be cost effective, the purity outcome, the probability of increasing the error rate during synthesis especially at the 5′ end of
the molecule (376), and inducing the innate immunity (377) makes it less preferable for gene editing in primary cells. Different base chemical modifications incorporated in the gRNA backbone have been studied for effects on gRNA stability as well as their ability to reduce off-target effects. Using three chemical modifications at both the 5′ and 3′ ends of the gRNA, Ayal Hendel and colleagues showed that the 2′-O-methyl, 2′-O-methyl 3′phosphorothioate or 2′-OMe 3′-thioPACE modifications with Cas9 mRNA and Cas9 protein, caused ten-times and three-times higher editing respectively compared with the unmodified gRNA in primary cells (Figure 5.15.) (375).

![Chemical Modifications](image)

**Figure 5.15. Examples of chemical modification in gRNA synthesis.** Adapted from (375).

The transfection of primary cells at ALI when polarized with tight junctions is challenging, even for viral transduction (371). Here we managed to detect GFP expression after mRNA transfection (Figure 5.3.) but the efficiency was very low compared to submerge transfection (almost 20-fold less with RTN formulation) while editing was undetectable after transfection on ALI. The reason for that could be related to the size of Cas9 molecules which is four times larger than GFP mRNA, which makes the transfection harder. The similar RTN formulations to those used here have recently shown effective transfections of CF cells on ALI with ENaC siRNA, with up to 50% silencing and restoration of mucociliary properties after triple doses (72). The transfection and editing of differentiated cells on ALI, therefore, requires further optimisation as no editing was detected.

In the last part of this chapter we assessed the correction of the ΔF508 mutation in primary cells through HDR mediated by CRISPR/Cas9. The donor plasmid that
was used in CFBE41o-cells in the previous chapter (Figure 4.9) was not suitable for primary cells as they have the pac gene from lentivirus BMI-1 transduction. For that reason, we replaced pac in the donor plasmid with the neomycin resistance gene (neo). However, we noticed that the selection of corrected cells was less effective with neomycin than puromycin. In a study conducted by Amanda Lanza and colleagues, where they compared various selection markers to establish mammalian cell lines with stable expression of GFP, they found that puromycin was three times more efficient than neomycin for the selection of HEK293T cells (378).

Here we attempted to grow the corrected primary cells as clones to perform the functional evaluation on pure corrected cells (either homozygous or heterozygous correction), however, the cloned cells stopped proliferating and expanding. The transduction of primary NHBE and CFBE cells with BMI-1 extends the culturing up to 20 passages (93, 379), however, the proliferation before reaching senescence depends on the capability for the population doubling rather than passing numbers. Despite starting from a low passage number of BMI-1 cells (passage 5) the cells went through many doublings in their population as the cells were first transfected with the CRISPR system followed by G418 selection, then expanding them and finally performing single cell cloning. Therefore, we decided to perform the functional evaluation on a mixed population of corrected cells where the percentage of correction was estimated as 10% (Figure 5.9B).

The electrophysiological studies for the CFBE primary BMI-1 cells after correction showed a reduction in ENaC activity as a response to the amiloride drug (Figure 5.14.). This result also correlates with the qRT-PCR results where the mRNA expression for SCNN1 genes decreased, particularly for SCNN1B (Figure 5.12.). In this study, we evaluated the CFTR and ENaC channel activity by the Ussing chamber technique. However, this technique has some limitations, such as the time taken for cells to grow on the snapwells and forming a tight junction. The patch clamp assay is another technique used to measure the opening and closing rates of the CFTR channel, where the function study was evaluated on patches of the cell membrane (380). However, the above technique requires special equipment and skills for operation. Another strategy to assess the CFTR function
is by developing cells with a yellow fluorescent protein (YFP) based halide sensor integrated in their genome, where the CFTR function is assessed by quenching the YFP in response to CFTR-mediated iodide entry after activation by forskolin. The mutated ΔF508 cells are incapable of transporting the chloride ion where the CFTR channel is not functional and as a result the cell fluorescence does not respond to the activator (381).

Here we managed to correct the ΔF508 mutation in almost 10% of the whole population of cells [at the molecular level (Figure 5.9B)]. This caused a small increase at the CFTR mRNA level (Figure 5.10.) with no detectable effect at the protein level (Figures 5.13 and 5.14A). A concern here is that the non-CF cells (NHBE BMI-1) (Figures 5.13 and 5.14C), did not show a response to the CFTR activator and inhibitor in the Ussing chamber study, nor was CFTR protein detected in the western blot. The primary BMI-1 cell model may need more investigation in future work to determine the causes for these observations, especially in the case of the electrophysiological study. This could be related to an overexpression of BMI-1 or to the cells’ passage number.

In terms of future work, more optimisation is needed for correcting the ΔF508 in the primary CFBE BMI-1 cells such as using a smaller donor plasmid such as minicircle (382) or mini-intronic plasmid (383), which shows a higher transfection efficiency compared to conventional plasmid. In addition, alternative selection markers for neomycin such as zeocin or hygromycin may show better selection results (378), or one can use a fluorescent gene such as GFP. Finally, using different approaches to enhance the CFBE primary proliferation by growing the cells on irradiated fibroblast feeder layer in the presence of rho-associated protein kinase (ROCK) inhibitors (91) or using dual SMAD inhibitor (DMH-1 and A-83-01) as a free feeder layer culture technique (92).

In summary, in this chapter we have optimised the transfection of CFBE BMI-1 cells where mRNA is used as an alternative approach delivered with targeting nanoparticles (RTN). In addition, we optimised the CRISPR/Cas9 transfection in CF primary cells using Cas9 mRNA/gRNA and RNP delivered with RTN to be as effective as the commercial reagents. Finally, we showed the correction of ΔF508 at the molecular level.
CHAPTER 6

Results

Homology-independent targeted integration (HITI) as alternative approach for gene editing
6. Homology-independent targeted integration (HITI) as alternative approach for gene editing

6.1. Introduction

The CRISPR/Cas system has emerged as an effective technique for gene editing and gene therapy. There are two main pathways for editing after targeting the region of interest with this system, as discussed in Chapter 1: the error-prone pathway (Non-homologous end joining (NHEJ)), which results in indels after DSB. The second pathway, homology-directed repair (HDR), is more precise for editing in the presence of a donor homology sequence. Various strategies have been investigated to improve HDR, such as cell cycle synchronisation techniques; however, these techniques are not usable for in vivo applications or for non-dividing cells (384). Proper design and improved delivery of the CRISPR/Cas system enhances the efficiency of HDR (354). In addition HDR also can be improved directly by enhancer molecules, such as RS-1 (RAD51-stimulatory compound 1) (385), or indirectly, by targeting the essential proteins involved in NHEJ (386). However, the efficiency of HDR is relatively low in comparison with NHEJ as it is a cell-cycle-dependent process, where the donor DNA serves as a repair template at S and G2 phases (387). The consequence of this is that the editing of non-dividing cells by HDR is unlikely to occur.

In this chapter, we investigate another approach for inserting a DNA fragment, or removing a piece of DNA and replacing it with a new sequence, using the NHEJ pathway, which, unlike the HDR pathway, is efficient and can proceed in non-dividing cells. It is based on using a donor DNA flanked with the same sequence of the targeted gRNA, and PAM motif at the genome targeting site, but in an inverted direction. The integration could occur in forward or reverse orientation; however, in the reverse integration, the PAM site will be intact and will be targeted again by the CRISPR system, which will lead to enrichment of the forward integration (Figure 6.1.). This technique is termed homology-independent targeted integration (HITI) (388).
Figure 6.1. **Homology-independent targeted integration (HITI) strategy.** The donor template has the same gRNA and PAM sequence but in inverted orientations. The integration can be in the forward direction (loss of intact gRNA) or reverse direction, where the gRNA sequence remains intact and which might undergo additional cutting by the CRISPR system.

6.2. **Aims**

In this chapter, we aim to:

1. Develop the HITI protocol using the Neuro-2A Ai9 cell model for single double-strand break (DSB);
2. Develop the HITI protocol for GFP knock-in with double DSBs in both the donor plasmid and the Neuro-2A Ai9 cell model;
3. Investigate the replacement of *CFTR* exon 10 by HITI in CFBE cells.
6.3. Results

6.3.1. HITI in neuro-2A Ai9 cell model
The Neuro-2A Ai9 cell model is useful for determining the efficiency of knock-in using the HITI strategy as it has its own promoter (CAGGs) with a promoterless donor construct (Figure 3.14.). Here, we are using two approaches to knock-in GFP, either with single cut at the front of the stop codons or with double cuts (one before and one after the stop codons).

6.3.1.1. Integrating GFP at Ai9 with single DSB
The aim of the experiment is to integrate the GFP gene at the Ai9 construct of Neuro-2A cells by making a DSB before the stop codons, using the target 3 gRNA. In parallel, the cells will be transfected with plasmid donor as the GFP gene without promoter is flanked with the same gRNA and PAM sequences, but in an inverted direction at one site or two sites (Figure 6.2.). There are two possibilities for integration (forward and reverse). The reverse orientation recreates an intact gRNA sequence which may undergo additional cutting by Cas9 and so enrich for the forward integration.

![Diagram](image)

**Figure 6.2. Schematic of GFP knock-in at gRNA T3 site by HITI.** The donor template either has one cutting site or two cutting sites before and after the GFP gene. The integration can be in the forward orientation, where the GFP will be expressed, or the reverse orientation, which causes the intact gRNA sequence to remain, which might subject it to additional cutting by Cas9.
6.3.1.1.1. Constructing the donor plasmid and transfection

To construct the donor plasmid, we designed the oligonucleotides containing the gRNA and PAM flanked by extra bases in order to proceed with cloning which was cloned by linearising the GFP plasmid with AattII and BamHI restriction enzymes, followed by ligation step. The successful cloning was first confirmed by colony PCR, followed by Sanger sequencing. To create a donor template that has two cutting sites, the previous donor plasmid was linearised by Sacl and PspOMI restriction enzymes then inverted gRNA T3 was cloned (Figure 6.3.).

![Figure 6.3. Schematic of constructing GFP HITI donor template with Ai9 target 3 gRNA. A) the original parental pMC plasmid where the GFP gene is flanked with two sites for cloning. B) The plasmid cloning strategy for inverted gRNA T3 at the front of the GFP gene, while in C) the same sequence of gRNA and PAM region is cloned in addition after the polyA to create a donor with double cutting sites.](image)

The first step was to compare the knock-in efficiency with single gRNA cutting site in front of the GFP gene, where the whole plasmid (5kb in size) is expected to integrate at the DSB, with another plasmid which has double cutting sites, which results in the integration of the GFP gene and polyA (approximately 1kb in size). The Neuro-2A Ai9 cells were transfected with either of the two HITI donors together with the pX-330 plasmid encoding the gRNA T3 and Cas9. Flow cytometry analysis of GFP expression suggested that the donor plasmid with single gRNA was slightly more efficient for knock-in (3%±0.3) than the GFP knock-in construct with donor plasmid has double T3 gRNA sites (2.7%±0.4) (Figure 6.4.).
Figure 6.4. GFP knock-in efficiency with HITI strategy at gRNA target 3 site with a donor which has one or two cutting sites. The Neuro-2A Ai9 cells were transfected with the HITI donor template and gRNA T3. The knock-in is represented by the GFP expression on the x-axis, while the y-axis (tdTomato) represents the NHEJ where the fragment containing the stop codons was deleted. A) The knock-in and NHEJ efficiency after transfection with the HITI donor template (one gRNA target site) and pX330 encoding Cas9 and gRNA T3. B) The knock-in and NHEJ efficiency after transfection with the HITI donor template (two T3 gRNA sites) and pX330 encoding Cas9 and gRNA T3. The transfections were performed with L2K transfection reagent. The results represent single well transfection analysed by flow cytometry.

The next step was to study the improvement in knock-in efficiency enhanced after multiple transfections. The transfections were performed either with L2K or RTN (C18 peptide Y) formulation. The GFP knock-in after the first transfection using L2K was 4.5% on average, which increased by almost three-fold after the second transfection (16.3% ±1.9) (Figure 6.5. B and C). On the other hand, the transfection with the targeting nanocomplex shows 3.9% of total cells have GFP expression, which is an almost three-fold improvement and reaching 11.6%±1.6 after the second transfection (Figure 6.5. A).
Figure 6.5. The GFP knock-in efficiency with HITI strategy at the gRNA target 3 sites with multiple transfections. The Neuro-2A Ai9 cells were transfected with the HITI donor template and gRNA T3. A) The knock-in and NHEJ efficiency after transfection with the HITI donor template plasmid and pX330 encoding Cas9 and gRNA T3 using the RTN formulation. The HITI transfection was performed two times where the cells were passaged and re-seeded after the first transfection. B) The knock-in and NHEJ efficiency after two plasmids were transfected with the HITI donor template and p-X330 encoded Cas9 and gRNA T3 using an L2K transfection reagent. The transfection is performed twice. The knock-in is represented by the GFP expression on the x-axis, while the y-axis (tdTomato) represents the NHEJ when the stop codons have been removed. The results represent a single well transfection analysed by a flow cytometer. C) Fluorescent imaging for GFP knock-in and NHEJ (tdTomato) after the first transfection with L2K.
6.3.1.1.2. RNP transfection with donor plasmid

In the previous experiments the GFP knock-in was almost 4% after a single transfection using a donor plasmid and pX-330 plasmid encode Ai9 gRNA T3 as well as Cas9. Here we attempt to investigate the possibility to improve the efficiency by delivering the CRISPR system as ribonucleoprotein (RNP). The transfections with the RNP and plasmid donor showed, surprisingly, a lower knock-in efficiency (0.64%) (Figure 6.6. A) than with pX-330 (4.5%) (Figure 6.5. A). It was thought the reason for the low efficiency could be the result of transfecting all the components at the same time, where Cas9 protein has a short half-life (329). In order to possibly improve the knock-in, the HITI donor was transfected to the cells first, then after 24 hours, the RNP was delivered. However, that did not show an improvement at the knock-in (figure 6.6. B).

![Figure 6.6. The GFP knock-in efficiency with the HITI strategy at the gRNA target 3 site using Cas9 protein and donor plasmid.](image)

The knock-in is represented by the GFP expression on the x-axis, while the y-axis (tdTomato) represents the NHEJ when the stop codons have been removed. A) The knock-in and NHEJ efficiency after using the HITI donor plasmid and Cas9 protein and synthetic gRNA T3 simultaneously. B) The knock-in and NHEJ efficiency where the HITI donor plasmid is transfected first, then after 24 hours the Cas9 protein and synthetic gRNA T3 are delivered to the cells. The transfection is performed by an L2K transfection reagent. The results represent single well transfection analysed by a flow cytometer.
6.3.1.1.3. GFP clone analysis by Sanger sequencing

In order to confirm the knock-in at the expected locus and investigate the pattern of indels at the ligation site, we first sorted the GFP expressing Neuro-2A cells from pX-330 plasmid and single cutting plasmid (Figure 6.4.A) using a flow cytometry cell sorter. We then seeded the sorted cells in 96 well plates at a density of one cell per well. After two weeks, we managed to obtain 18 GFP clones. The DNA extracted from these clones was used to perform junction PCR at the 5'prime end (HITI_Ai9_5'end HDR primer), with the forward primer located in the Neuro-2A Ai9 genome, while the reverse primer was located in the GFP gene (Figure 6.7 A). All clones were positive for the PCR, with a product of 0.6kb. The PCR products from the clones were sent for Sanger sequencing. The expected sequence pattern at the ligation site was a PAM sequence (three bases) followed by the last three bases from the targeting sequence (18,19 and 20) from the Ai9 cells, then the PAM motif, and the last three bases of the gRNA sequence (from HITI donor plasmid). This pattern was found in four clones, which represented 22.2% of all clones. In the same number of other clones (4 out of 18), we saw extra bases at the ligation site, which came as a result of cutting between positions 16 and 17 in the Ai9 gRNA locus. The most abundant pattern was found in seven clones, where four bases from each end of the gRNA sequence were found next to each other (17 to 20). Finally, in three clones we found a deletion of the PAM, and the end of the gRNA sequence from the HITI construct at the integration site (Figure 6.7B). In summary, the sequencing results from the HITI integration site indicate no intensive insertion or deletion after the DSB at the donor template as well as Ai9 cells.
6.3.1.2. Integrating GFP at Ai9 with double cuts

The aim of this experiment was to integrate the GFP gene at the Neuro2A Ai9 cells by making two DSBs: one before the stop codons (target 1 gRNA), and one after (target 4 gRNA). In parallel, the cells were transfected with plasmid donors with a promoterless GFP gene flanked by the same gRNAs and PAM sequences, but in an inverted direction (Figure 6.8.). The two gRNAs (T1 and T4) have been optimised previously on these cell models (chapter three) where they gave the highest NHEJ by removing the stop codons fragment. The advantage for this model of knock-in with two DSBs was that it enabled an estimate of the NHEJ efficiency from tdTomato expression, as well as the HITI efficiency from GFP expression, while the previous model with only one DSB (T3 gRNA) showed only the knock-in. In addition, the stategy will used later for replace CFTR exon 10.
Here, there are two possibilities for integration (forward and reverse) as seen in single DSB with gRNA T3, however, the reverse integration is not going to be targeted again by Cas9 as the PAM sequence and part of the gRNA sequence are destroyed at the time when the donor is linearised.

**Figure 6.8.** Schematic of the GFP knock-in at the stop codons site by the HITI technique. The donor template has two cutting sites (T1 gRNA and T4 gRNA), before and after the GFP gene, respectively. The integration can be at the forward direction, where the GFP will be expressed, or the reverse orientation, where GFP will be in an inactive state.

### 6.3.1.2.1 Constructing the donor plasmid and transfection

To construct the donor plasmid, the Ai9 gRNA T1 (inverted sequence) was cloned first, by linearising the donor plasmid with the AattII and BamHII restriction enzymes, followed by ligation step, where the correct clones were confirmed firstly by colony PCR, followed by Sanger sequencing. Then the same steps were used to clone Ai9 gRNA T4 (inverted sequence) at the Sacl and PspOMI site (Figure 6.9.).
Figure 6.9. Schematic of constructing the GFP HITI donor template with target 1 and 4 gRNAs. A) shows the original parental pMC plasmid, where the GFP gene is flanked with two sites for cloning. B) shows the plasmid cloning strategy for inverted gRNA T1 with the PAM sequence at the front of the GFP gene, while the inverted gRNA T4 and PAM region is cloned after the polyA.

The transfections were performed on Neuro-2A Ai9 cells in 24 well plates where 500ng of each plasmid had been used (T1 plasmid, T4 and HITI donor plasmid). The transfections were performed either with L2K (ratio 1:3 of nucleic acids to L2K) or with an RTN (peptide Y and C18 lipid). The cells were imaged under a fluorescent microscope (figure 6.10. C), then analysed by flow cytometry. The knock-in with L2K was on average 0.84% of GFP+ve cells, with 8.2% of the cells undergoing NHEJ (tdTomato). The GFP knock-in (HITI) increased almost twice to reach 1.7% after the second transfection and reached 4.9% ±0.3 of all cells after the third dose, which was approximately half of the NHEJ event (11%±0.3) (Figure 6.10. B). On the other hand, transfecting the cells using the nanoparticles was able to knock-in around 1% of the cells, while the NHEJ (tdTomato) percentage was higher by fivefold. The GFP knock-in increased after the second transfection to an average of 1.3% of the cells. The third transfection almost doubled the number of the cells (2.41±0.7) with 6.8%±0.4 expressing tdTomato, indicating NHEJ (Figure 6.10. A).
Figure 6.10. GFP knock-in efficiency with HITI strategy at the stop codons site. The Neuro-2A Ai9 cells were transfected with the HITI donor template and two gRNAs (T1 and T4). A) The knock-in and NHEJ efficiency after three plasmids were transfected (two encoded Cas9 and gRNAs) and the HITI donor template, using the RTN formulation. The HITI transfection was performed three times. B) The knock-in and NHEJ efficiency after three plasmids were transfected (two encoded Cas9 and gRNAs) and the HITI donor template, using an L2K transfection reagent. The transfection is performed three times. The knock-in represented by the GFP expression on the x-axis, while the y-axis (tdTomato) represents the NHEJ when the stop codons have been removed. The results represent single well transfection analysed by a flow cytometer. C) Fluorescent imaging for GFP knock-in and NHEJ (tdTomato) after the first transfection with L2K.
6.3.1.1.2. GFP clone analysis by Sanger sequencing

In order to confirm the knock-in at the Ai9 locus and investigate the pattern of indels at the ligation site, we performed a junction PCR at the 3’ prime end (HITI_Ai9_3’end HDR primer) where the forward primer was located at the GFP gene, while the reverse primer was at Neuro-2A Ai9 (Figure 6.11. A) followed by cloning the PCR product into the pJET 2.1 plasmid. 12 colonies were sent for sequencing, which showed half of the sequence results had the expected pattern (after direct orientation) at the ligation site, where the polyA sequence was followed by an inverted T4 gRNA (1 to 17), then the T4 gRNA from the Ai9 cells (1 to 17), as the three bases and PAM motif were removed after the DSB. The second abundant pattern was found in five colonies where three bases at the end of gRNA on Ai9 cells were deleted (the remaining gRNA sequence running from 1-14). Interestingly, we found a PCR product with a relatively big deletion (17 bases), where four bases were deleted from T4 gRNA at the HITI construct (from 14 to 17) and 13 bases from gRNA, located in the cells (from 5 to 17) (Figure 6.11. B).

![Image](image-url)

**Figure 6.11. Sanger sequence at the 3’ prime end of HITI T1+4.** The PCR from the mixed population is amplified, then cloned into pJET 2.1 plasmid. A) shows the junction primers site where the forward primer is located at the GFP gene, while the reverse primer is at Neuro-2A Ai9. B) The sequence results in 12 colonies, where the T4 at HITI construct is shown in purple, while the T4 from Ai9 cells is in brown.
6.3.1.1.3. HITI and HDR techniques together

The other approach we attempted was to try to improve the GFP knock-in by using a donor template that had homology arms in the area before the gRNA T1 (446bp) and after gRNA T4 (448bp), as well as using both of the gRNAs with the PAM as an inverted sequence. So, the idea here was to use both the NHEJ and HDR pathways for integrating the GFP. The homology arms and gRNA were synthesised as DNA fragments flanked by appropriate restriction enzyme sites for cloning (Figure 6.12). The cloning was confirmed by colony PCR, then Sanger sequencing.

![Figure 6.12. Schematic of constructing GFP HITI HDR donor template with target 1 and 4 gRNAs. A) shows the original parental pMC plasmid where the GFP gene is flanked with two sites for cloning. B) shows the plasmid cloning strategy for inverted gRNA T1 with the PAM sequence following with the left homology arm upstream of the GFP gene, while the right homology arm is followed by the inverted gRNA T4 and PAM region, cloned downstream of the polyA.](image)

Here there are three possibilities for integration: forward or reverse integration of the GFP gene, depending on the NHEJ pathway (HITI). The other possibility is the GFP knock-in, using the HDR pathway enhanced by the homology arms at the donor template (Figure 6.13.).
Figure 6.13. Schematic of GFP knock-in at stop codons site by HITI and HDR technique.
The donor template has right and left homology arms with Neuro-2A Ai9 cells, followed by two cutting sites (T1 gRNA and T4 gRNA) before and after the GFP gene, respectively. The integration can be at the forward orientation, where the GFP will be expressed as result of NHEJ or HDR. The reverse integration could also happen where the donor plasmid is linearised, then integrated at the cutting site at the opposite orientation, which would result in GFP not being expressed.

In order to complete comparison, we synthesised another donor plasmid had only homology arms, without inverted gRNA sequences (Figure 6.14.), which can be used to knock-in a GFP gene through the HDR pathway only (Figure 6.15.).

Figure 6.14. Schematic of constructing GFP HDR donor template. A) shows the original parental pMC plasmid where the GFP gene is flanked with two sites for cloning. B) shows the
plasmid cloning strategy for the left homology arm at the front of the GFP gene, while the right homology arm is cloned after the polyA, without the target 4 gRNA sequence.

![Diagram](image)

**Figure 6.15.** Schematic of the GFP knock-in at the stop codons site by the HDR pathway. The donor template has right and left homology arms, with Neuro-2A Ai9 cells after the gRNA T1 and T4 cutting sites before and after the GFP gene. The integration can happen at the forward direction only, where the GFP will be expressed as a result of the HDR.

The Neuro-2A Ai9 cells were seeded in a 24 well plate, then transfected with the donor plasmid only, or with the donor plasmid and the pX-330 plasmid encoding the paired gRNAs (T1 and T4). The knock-in efficiency with the HITI HDR strategy was 3.3% ±2.7, while 5.05% of the total cells on average expressed tdTomato (NHEJ) (figure 6.16 B). On the other hand, the GFP knock-in as result of the HDR only was 1.3%±0.3, where the efficiency of stop codons removal was 5.4% on average (Figure 6.17 B).

In conclusion, the HITI technique using NHEJ pathway for knock-in GFP at Neuro-2A Ai9 cells was encouraging as alternative technique for HDR. The cell model with Ai9 construct has advantage of fast screening by flow cytometry with ability to measure HITI and NHEJ events in single transfection.
Figure 6.16. GFP knock-in efficiency with HITI HDR strategy at stop codons site analysed by flow cytometry. The Neuro-2A Ai9 cells transfected with HITI HDR donor template and two gRNAs (T1 and T4). The knock-in represented by the GFP is expressed on the x-axis, while the y-axis (tdTomato) represents the NHEJ when the stop codons have been removed. A) Negative control where the donor plasmid is transfected only. B) The knock-in and NHEJ efficiency after three plasmids were transfected (two encoded in Cas9 and gRNAs) and the HITI HDR donor template. The results represent single well transfection analysed by flow cytometry.

Figure 6.17. GFP knock-in efficiency with HDR strategy at stop codons site analysed by flow cytometry. The Neuro-2A Ai9 cells were transfected with the HDR donor template and two gRNAs (T1 and T4). The knock-in represented by the GFP is expressed on the x-axis, while the y-axis (tdTomato) represents the NHEJ when the stop codons have been removed. A) Negative
control where the donor plasmid is transfected only. B) The knock-in and NHEJ efficiency after three plasmids were transfected (two encoded in Cas9 and gRNAs) and the HDR donor template. The results represent single well transfection analysed by flow cytometry.

6.3.2. Replacement of CFTR exon 10 by HITI technique

6.3.2.1. Targeting CFTR intron 9 and 10 by CRISPR/Cas9

The previous experiments were performed on an Ai9 model to remove the stop codons by two gRNAs and knock-in the GFP gene. Here, in this part, we move to the main goal to remove exon 10 of CFTR. This is removed by two flanking gRNAs making intronic DSBs then, using a donor template, a wild type of exon 10 is introduced flanked with inverted gRNAs and PAM sequences. The first step was to screen for efficient gRNAs, cutting before and after CFTR exon 10. Using the Benchling website for screening, we chose three gRNAs before the exon 10 (at intron 9) and another three after it (at intron 10) (Table 6.1 and Figure 6.18.). In order to screen them, six gRNAs were cloned into the pX-330 plasmid, then confirmed by Sanger sequencing. To screen the indels efficiently, we transfected HEK293 cells in 24 well plates, with 500ng of each pX-330 encoding the gRNAs (gRNA: intron 9-1, intron 9-2, intron 9-3, intron 10-1, intron 10-2, intron 10-3) using L2K. The genomic DNA was extracted after 48hours, followed by the PCRs (CFTR_outside donor primer,1.8kb in size), then a T7 endonuclease I assay was performed. The results showed that the most efficient gRNA that cuts before exon 10 was intron 9-1 gRNA with 24% indels efficiency (Figure 6.19.) which was confirmed by the TIDE assay (31.6% total efficiency) (Figure 6.20. A). The intron 9-2 and intron 9-3 gRNAs were less efficient, with 17.5% and 8% indels respectively, as assessed by T7 endonuclease I assay (Figure 6.19.). On the other hand, the three gRNAs we chose for targeting the CFTR intron 10 did not show detectable indels by T7 endonuclease I assay (Figure 6.19.), while only 10-3 gRNA showed 18.3% measured by the TIDE assay (figure 6.20. B). Therefore, we screened additional gRNAs in order to find a more efficient gRNA; intron 10-4 and 10-5 were cloned into pX-330 plasmid (Table 6.1. and Figure 6.18.). The TIDE analysis after the transfection shows that gRNA 10-4 achieved 37.1% indels, while the efficiency of 10-5 was very low (3.1%) (Figure 6.20. B and C).
From the above screening, we concluded that gRNA intron 9-1 and intron 10-4 were the most efficient. In order to remove exon 10, we transfected HEK293 cells with both gRNAs in the same well, using 500ng of each plasmid. PCR (CFTR_5’end outside donor primer, 1.45kb product) was performed where both primers were outside the gRNA targeting region. The PCR worked where we were able to detect the main band (1.45kb), and also another smaller product (around 0.87kb) as a result of the deletion of exon 10 after cutting by both gRNAs (0.58kb), where the efficiency of exon 10 deletion was approximately 21% (Figure 6.21.).

Figure 6.18. CFTR introns 9 and 10 with the eight SpCas9 gRNAs targeting the region. The direction of the arrow indicates the strand targeting (sense or antisense) while the PAM motif of each gRNA is indicated by same colour rectangles.
<table>
<thead>
<tr>
<th>Target</th>
<th>Strand</th>
<th>Sequence</th>
<th>PAM</th>
<th>Efficiency Score</th>
<th>Specificity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>intron 9-1</td>
<td>1</td>
<td>CTGCTTAGGATGATAATTGG</td>
<td>AGG</td>
<td>41.47</td>
<td>10.97</td>
</tr>
<tr>
<td>intron 9-2</td>
<td>-1</td>
<td>GATAGAGGTTCACAGAGAAG</td>
<td>GGG</td>
<td>22.55</td>
<td>60.70</td>
</tr>
<tr>
<td>intron 9-3</td>
<td>-1</td>
<td>GTCATTATCAATACACGCTC</td>
<td>AGG</td>
<td>46.02</td>
<td>20.37</td>
</tr>
<tr>
<td>intron 10-1</td>
<td>1</td>
<td>AATAATGCATTATATGAAA</td>
<td>TGG</td>
<td>23.16</td>
<td>8.66</td>
</tr>
<tr>
<td>intron 10-2</td>
<td>1</td>
<td>TCTGCCAGCAGAGAATTAGA</td>
<td>GGG</td>
<td>37.45</td>
<td>65.26</td>
</tr>
<tr>
<td>intron 10-3</td>
<td>-1</td>
<td>TTGCAAGCTTCTAAAGCAT</td>
<td>AGG</td>
<td>34.25</td>
<td>37.22</td>
</tr>
<tr>
<td>intron 10-4</td>
<td>1</td>
<td>GCATGCATATAAGTATATG</td>
<td>TGG</td>
<td>38.43</td>
<td>65.24</td>
</tr>
<tr>
<td>intron 10-5</td>
<td>-1</td>
<td>TGCCATATAAATAACCATTG</td>
<td>AGG</td>
<td>38.82</td>
<td>11.75</td>
</tr>
</tbody>
</table>

**Table 6.1. List of SpCas9 gRNAs targeting the CFTR introns 9 and 10.** The table shows each gRNA sequence, targeting the DNA strand sense as 1 or antisense as -1, the PAM motif sequence, the predicting efficiency score (where a score of 100 is the most efficient), specificity score where a high score indicates potential more off-targeting. The on-target and off-target scores were calculated by https://benchling.com/crispr.

**Figure 6.19. T7 endonuclease I assay for CFTR introns 9 and 10 gRNAs.** The transfections were performed by lipectamine2000, using 500ng of plasmid (pX-330). The red and green arrows indicate the fragments after cutting by T7EI, while the blue arrow shows the main PCR product.
247

Figure 6.20. Indels analysis of targeting CFTR introns 9 and 10 by CRISPR/SpCas9. The analysis shows the efficiency of four gRNAs on HEK293 cells assessed by TIDE (Tracking of Indels by Decomposition) where the transfected sample sequence is compared to the control sample.

Figure 6.21. Removing the CFTR exon 10 using two gRNAs. The agarose gel shows the PCR fragment amplified when both primers are outside the targeting site. In the first lane, the 4.5kb PCR product was from an untransfected sample where there was only one fragment. In the second lane, two gRNAs (9-1 and 10-4) were used to remove the CFTR exon 10 (0.58kb).

6.3.2.2. Constructing CFTR exon 10 donor plasmid
The aim of the previous screening was to find two efficient gRNAs flanking CFTR exon 10. To construct the HITI donor plasmid, we ordered that as a double strand
DNA fragment (gBlock) from IDT (636bp) (Figure 6.22.), which has a wild-type exon 10 sequence flanked by gRNA target 9-1 and 10-4 with the PAM sequence in the reverse original direction. In order to facilitate screening, we introduced a silent mutation in the construct, creating a ClaI restriction site (Figure 6.22.). The DNA fragments cloned into the pJET 2.1 plasmid were confirmed by Sanger sequencing (Figure 6.23.).

Figure 6.22. DNA fragment for HITI CFTR exon 10 replacement. The DNA fragment that was consisted of the wildtype CFTR exon 10 (grey) flanked with inverted gRNAs and PAM targeting sequence (9-1 and 10-4).
Figure 6.23. Cloning the HITI CFTR exon 10 donor. The DNA fragment of wildtype CFTR exon 10 donor is cloned into the pJET 2.1 plasmid, then screened by ClaI enzyme digestion. The agarose gel shows a positive colony after the transformation, where the plasmid was linearised after ClaI digestion in the second lane.

6.3.2.3. Transfection of HEK293 cells to replace CFTR exon 10

The next step after optimising the gRNAs surrounding the CFTR exon 10 was to transfect them (intron 9-1 and intron 10-4) into cells in parallel with the donor plasmid (Figure 6.24.). So, 500ng of each plasmid was used for the transfection of HEK293 cells using L2K. The replacement of exon 10 could happen either in a forward direction or reverse direction (Figure 6.24), as we have seen in the stop cassette at Neuro-2A Ai9 model above (section 6.3.1.2). Following the transfection, the genomic DNA was extracted from the cells, then the PCR reaction was performed where the forward primer was outside the donor template, while the reverse primer was inside it (CFTR_primer, 0.62kb in size). The PCR products from the transfected cells with only the donor plasmid, or with both donor plasmid and two gRNAs, were digested with ClaI restriction enzyme to investigate the correction by the HITI technique. The gel electrophoresis showed no extra band for the control sample (only donor plasmid) in the first lane, while the second lane, where two gRNAs were used, showed an extra fragment.
(0.45kb) indicating that the exon 10 replacement occurred by the HITI technique with replacement efficiency 4.5% (Figure 6.25).

Figure 6.24. Schematic of replacement of the mutated exon 10 by the HITI technique. The donor template has two cutting sites (T9-1 gRNA and T10-4 gRNA) before and after the exon 10, respectively. The integration can be in a forward direction or in the reverse direction where the donor plasmid is linearised, then integrated at the cutting in the opposite direction. A silent mutation introduced in the ClaI restriction site is used in the construct in order to determine the efficiency after HITI transfection.
Figure 6.25. Replacement of the CFTR exon 10 on HEK293 cells using the HITI technique. A) Agarose gel for PCR samples where cells were transfected with either donor plasmid only or with two gRNAs (9-1 and 10-4). The red and green arrows indicate the fragments after cutting by Clal, while the blue arrow shows the main PCR product. B) Sanger sequencing from HEK293 cells transfected with donor plasmid only (top) or with both donor plasmid and 2 gRNAs (bottom) while the middle sequence is from the donor plasmid. The highlighted base shows where the expected Clal cut site lies (not existing in HEK293 cells).

6.3.2.4. Transfection on CFBE41o- to replace CFTR exon 10
We demonstrated the replacement of CFTR exon 10 on HEK293 cells at the molecular level in the previous experiments. However, these cells do not express
CFTR protein. Consequently, we cannot assess functional restoration. To do that we repeated the same experiment on CFBE41o- cells which are homozygous for the ΔF508 mutation at CFTR exon 10. The cells were transfected with three plasmids (two of pX-330 encode 9-1 gRNA and 10-4 gRNA and HITI donor) or with a HITI donor plasmid only. Seventy-two hours after the transfection, the DNA was extracted from the samples and PCR (CFTR_5'end outside donor primer) was performed to confirm the removal of exon 10 as a result of targeting using two gRNAs. The results showed that the efficiency of removing exon 10 from all cells was approximately 19% (Figure 6.26. A). To confirm the replacement of the mutated exon 10 with wild type, another PCR was performed (CFTR_primer, 0.62kb in size) followed by an RFLP assay where a Clal restriction enzyme was used. The results did not show extra bands (0.17kb and 0.45kb). This could indicate a low replacement percentage not detected by this assay (Figure 6.26. B). To further assess the effect of the replacement of exon 10 on CFTR mRNA expression, qRT-PCR was performed on samples transfected with a donor plasmid as a control, and samples were transfected with paired gRNAs and a HITI donor plasmid. The analysis showed that the CFTR mRNA increased on the HITI corrected cells by 60.15±15.5-fold, which was statistically significant (p=0.019, unpaired Student’s t-test) compared to the control samples.

To conclude, we demonstrated that the replacement of exon10 with the HITI technique could be an alternative approach for correcting the ΔF508 mutation by HDR, which is challenging when using a gRNA making DSB far from the mutation.
Figure 6.26. Replacement of the CFTR exon 10 on CFBE41o- cells using the HITI technique. A) The agarose gel shows the amplified PCR fragments, where both primers are outside the targeted site. In the first lane, the 1.45kb PCR product was from a transfected sample with HITI plasmid only, where there was only one fragment. In the second lane, CFBE41o- cells were transfected with two gRNAs (9-1 and 10-4) that were used to remove the CFTR exon 10 and HITI donor plasmid (two PCR products 1.45kb and 0.87kb). B) the agarose gel for the PCR samples from cells that were transfected with either HITI plasmid only or with paired gRNAs and HITI donor. The red and green arrows indicate the expected fragments after cutting by ClaI, while the blue arrow shows the main PCR product.

6.3.2.5. CFTR intron 9-1 and 10-4 off-target screening

As we screened early for on-target activity for the gRNA target intron 9-1 and intron 10-4 on HEK293 cells, here we attempted to screen the specificity for both gRNAs, as performed previously on CFTR gRNAs (section 4.3.2.6). We used the Benchling engine to search for the potential off-targeting on the human genome. So, we chose five possible off-targets for each guide RNA with the highest off-target scores (Table 6.2 and Table 6.3). To have the optimal condition for the screening we used the sample from the HEK293 transfection with Cas9 encoded by pX-330 plasmid. In order to perform the screening, we amplified the region of interest by PCR then sent it for Sanger sequencing. The results did not show alternation on the chromatogram trace at the end of each off-target site, except the off-target number 1 for gRNA intron 10-4 which showed almost 14% indels using TIDE analysis (Figure 6.27. and Figure 6.28.). These
data suggest, the CRISPR/SpCas9 has a potential off-targeting effect even when the PAM sequence is not NGG (it was GAG in off-target 1 for gRNA 10-4).

Table 6.2. Potential off-target for 9-1 gRNA. The table shows the top 15 off-target and PAM sequences on the human genome with chromosomal location and strand direction (sense or anti-sense) where the red letter indicates the nucleotide difference with the on-target sequence. The blue highlighted sequence indicates the on-target sequence, while the grey highlights those chosen for screening. https://benchling.com/crispr is used for the off-target screening.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>PAM</th>
<th>Score</th>
<th>Gene</th>
<th>Chromosome</th>
<th>Strand</th>
<th>Mismatches</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTGCTTAGGATGATAATTGG</td>
<td>AGG</td>
<td>100</td>
<td>chr7</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>TTGAATAGCTGTGATAATTGG</td>
<td>AGG</td>
<td>1.433</td>
<td>chr11</td>
<td>-1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>CTCTTTAGGTTTATAATTGG</td>
<td>AGG</td>
<td>1.328</td>
<td>chr6</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CACCCTAGGATGTTAATTGG</td>
<td>AGG</td>
<td>1.227</td>
<td>chr10</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CAGGCTTAGGAGATAATTGA</td>
<td>TAG</td>
<td>1.040</td>
<td>chr9</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CTAATTGCGATGATAATTGG</td>
<td>GAG</td>
<td>0.964</td>
<td>chr1</td>
<td>-1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>CTCTTTGTAGTATGATAATTGG</td>
<td>AGG</td>
<td>0.917</td>
<td>chrX</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GTGATAATGATGATAATTGG</td>
<td>AGG</td>
<td>0.902</td>
<td>chrX</td>
<td>-1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GTGCCGAGGCTGTGATAATTGG</td>
<td>AAG</td>
<td>0.851</td>
<td>chr10</td>
<td>-1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CTTCTAAGGATGATATGTTGG</td>
<td>AAG</td>
<td>0.841</td>
<td>chr11</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ATATTAGAGTTAATTTGG</td>
<td>GAG</td>
<td>0.796</td>
<td>chr9</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>TTCTTAGAATGATAATTAG</td>
<td>AGG</td>
<td>0.766</td>
<td>chr21</td>
<td>-1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>AACCGT TAGGATGATAATGG</td>
<td>AGG</td>
<td>0.749</td>
<td>chr8</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GTGTGAAAGGTATGATAATTGG</td>
<td>AAG</td>
<td>0.718</td>
<td>chr13</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GTGCTTAGTTTTAAATTTGG</td>
<td>GAG</td>
<td>0.672</td>
<td>chr13</td>
<td>-1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ATTTTAGGGTGATAATTAG</td>
<td>AGG</td>
<td>0.625</td>
<td>chr17</td>
<td>-1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>CTTCATAAGATGCTATAATTGG</td>
<td>TGG</td>
<td>0.621</td>
<td>chr4</td>
<td>-1</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.27. Sanger sequence screening for five off-target sequences for intron 9-1 gRNA on HEK293 cells. The highlighted sequence indicates the off-targeting sequence and the red rectangle shows the PAM sequence for each potential off-target.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>PAM</th>
<th>Score</th>
<th>Gene</th>
<th>Chromosome</th>
<th>Strand</th>
<th>Mismatches</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCATGCATATAAGTGTATG</td>
<td>TGG</td>
<td>100</td>
<td>chr7</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>CCAATGCATATAAGTGTATG</td>
<td>CAG</td>
<td>1.715</td>
<td>chr13</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GAGATGCATAGTGTATG</td>
<td>GAG</td>
<td>1.458</td>
<td>chr7</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GATCTCATATAAGTGTATG</td>
<td>AAG</td>
<td>1.326</td>
<td>chr12</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GATTTGATGAGTGTATG</td>
<td>GAG</td>
<td>1.008</td>
<td>chr8</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GCATGCATATAAGTGTATG</td>
<td>TGG</td>
<td>0.953</td>
<td>chr15</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ACATGAATATAAGTGAATG</td>
<td>GAG</td>
<td>0.940</td>
<td>chr6</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GAAAACATATAAGTGTATC</td>
<td>TAG</td>
<td>0.894</td>
<td>chrX</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GCCATGCATATAAGTGTATG</td>
<td>TGG</td>
<td>0.849</td>
<td>chrX</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>ATATGCAATATAAGTGAATG</td>
<td>GGG</td>
<td>0.775</td>
<td>chr11</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GCATTCAGCAGTGTATG</td>
<td>AGG</td>
<td>0.743</td>
<td>chr2</td>
<td>-1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GCATCCACAAAGTGTATTT</td>
<td>TGG</td>
<td>0.716</td>
<td>chr3</td>
<td>-1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GTACACATATAAGTGAATG</td>
<td>TGG</td>
<td>0.711</td>
<td>chr3</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GCAGGCATTATGGAATG</td>
<td>TAG</td>
<td>0.703</td>
<td>chr4</td>
<td>-1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.3. Potential off-target for intron 10-4 gRNA. The table shows the top 15 off-target and PAM sequences on the human genome with chromosomal location and strand direction (sense or anti-sense), where the red letter indicates the nucleotide difference with the on-target sequence. The blue highlighted sequence indicates the on-target sequence while the grey highlights those chosen for screening. https://benchling.com/crispr is used for the off-target screening.

Figure 6.28. Sanger sequence screening for five off-target sequences for intron 10-4 gRNA on HEK293 cells. The highlighted sequence indicates the off-targeting sequence and the red rectangle shows the PAM sequence for each potential off-target. The indels for off-target are analysed by TIDE (Tracking of Indels by Decomposition) where needed.
6.4. Discussion

The main aim of this chapter was to explore an alternative technique for precise gene editing that explores a Non-Homologous End Joining (NHEJ) pathway (homology-independent targeted integration), using a donor template with opposite directions of gRNA and PAM sequences. We started by performing that in a cell line model (Ai9), either with one cut in the cells or with two cuts by replacing the stop codons cassette. Finally, we moved on to investigate how to correct the ΔF508 mutation of the CFTR gene by replacing the exon 10. The Neuro-2A Ai9 model was a useful reporter to study the efficiency of knock-in using the HITI technique. The gRNAs we used here have been optimised previously in chapter three for NHEJ. In addition, the HITI constructs were promoterless and it depended on the Ai9 promoter (CAGGs) to drive the GFP expression, as well as tdTomato, which give the advantage of being able to monitor the efficiency without waiting for the expression from transient transfection to dilute when there is a promoter inside the HITI donor. Furthermore, monitoring both the knock-in (GFP) and NHEJ (tdTomato expression) at the same transfection was feasible when two gRNAs were used to replace the stop cassette with GFP. As seen in Figure 6.10, the NHEJ event frequency was 10 times higher than the HITI knock-in when L2K was used for the transfection, while when the plasmids were delivered by the nanocomplex this was approximately 5 times higher.

Despite the advantages, we mentioned above for the Ai9 model, there are some limitations. The inverted knock-in events cannot be detected as the donor plasmid is promoterless. To investigate that, a screening by PCR for non-fluorescent clones is needed to determine the possibility for inverted knock-in, or the use of a donor that has its own promoter, then confirming the direction of integration by a junction PCR. The first approach was examined by Suzuki, et al. 2016, where they showed inverted insertion happened in almost 2% from all non-GFP clones (388). Furthermore, a donor with its own promoter could be useful for screening for off-target insertion.

This technique has been investigated previously through using various types of donor templates and gene editing tools. The first report to use the NHEJ pathway
goes back to 2013, when Maresca et al. 2013 studied knock-in of an eGFP gene at an AAVS1 site, driven by a CAG promoter using zinc finger nucleases (ZFNs). The donor plasmid which was used had inverted ZFNs targeting the DNA sequence to facilitate the integration after cutting. They named this technique Obligate Ligation-Gated Recombination (ObLiGaRe). Despite having proved the principle, the efficiency was very low (0.6%), which was six-fold higher than the HDR and which could be as a result of inefficient DSB at that locus by ZFNs (389). On the other hand, He et al. 2016 constructed the donor plasmid have a targeting site at the front of the reporter gene, different from the targeting site in the cells. By using the CRISPR system, they reported that the efficacy of knock-in somatic cells was 20% at the GAPDH locus, while the efficacy in human embryonic stem cells (ESCs) was 1.7% (390). The principle was also proved for non-dividing cells (primary mouse neurons cells), where the GFP knocked-in at Tubb3 gene, with 0.58% efficiency (388).

The HITI technique is based on linearising the donor plasmid inside the transfected cells rather than use already linearised donor. We have not performed this comparison; however, transfection of a circular plasmid for knock-in using the NHEJ pathway has been reported as more efficient and less toxic. Thomas Auer and colleagues showed that in an in vivo experiment they found the circular plasmid was more efficient for knock-in in comparison to the linearised plasmid by restriction enzyme (almost seven times higher). In addition, the linearised plasmid was more toxic for the zebrafish as they observed that the death rate increased three times (391).

Multiple transfection dosage is an efficient technique to maximise gene editing by NHEJ as we shown in Chapters Three and Five. Here, we confirm that for the HITI knock-in increases threefold after the second transfection, either with L2K or RTN, when the Ai9 model is targeted with single gRNA, as seen in (Figure 6.5.). On the other hand, the GFP knock-in by dual DSB at Ai9 increased two-fold after the second dosage with L2K, while the knock-in delivered by the nanocomplex increased by 30% (Figure 6.10.). This technique is useful, especially for in vivo application, to increase the gene editing effect, which might be very low from the first dose. To achieve this we need the RTN formulation,
which does not show immunogenicity or toxicity for in vivo work after one dose (392). The undesirable effect after multiple dosing for in vivo with RTN needs more investigation in future.

The other approach we used to maximise the knock-in efficiency was using both ways for gene editing (NHEJ and HDR) where the donor template has homology arms, as well as two inverted gRNA target sequences and PAM. The knock-in by this approach increased by two-fold (Figure 6.16.). A similar approach was used by He et al. 2016, where they compared the knock-in efficiency using the NHEJ pathway alone, or in combination with HDR, and they found when the donor template has homology arms, as well as a site for targeting by gRNA, this increased the knock-in efficiency by 20% to 30% (390).

Sequencing the integration ends for HITI constructs was useful for investigating the pattern of the ligation. The sequence results show the ligation locus did not have indels in most of the clones. Interestingly, we found one PCR product where 17 bases had been deleted (Figure 6.11.) on the gRNA target sequence without extending to the reporter genes. These findings are confirmed by other papers using the NHEJ pathway for knock-in; for example, Suzuki et al. 2016, found 80% on average, at both ends of the integration (5’ prime and 3’ prime), did not have indels (388).

Interestingly for GFP knock-in at Neuro-2A Ai9 gRNA target 3 site, the donor plasmid which has only one cutting site has been more efficient than the same plasmid with double cutting sites in having a shorter inserting fragment (1 kb), despite the one causing integration of the whole plasmid (almost 5kb) (Figure 6.4.). The reason for that is that the double cleavage on the donor plasmid might cause bigger indels, which affect the gene transcription. To study that, more screening is needed on non-GFP clones when double cleavage plasmids are used. The same finding has been reported by He et al. 2016 (390) and Suzuki, et al.2016 (388). On the other hand, this result suggested that the replacement of the stop cassette in the Ai9 model, where we used donor plasmid and has an GFP gene flanked with two cutting sites, could be improved if minicircle plasmid was used, which needs only one cut site as the bacteria backbone has been removed.
Using a single strand of DNA as a donor template for replacing a fragment of the genome targeted by two gRNAs could be another approach for the correction. However, the main limitation for that is the difficult synthesis of long single strand DNA. Quadros, et al. 2017 described an easy method to synthesise the long ssDNA by converting the double strand DNA (which has homology arms) to RNA molecules, firstly by in vitro transcription facilitated by a T7 promoter, then converting that to cDNA molecules, followed by digestion of the RNA strand by RNase H to get ssDNA (393). We investigated this technique for knock-in of the GFP gene at the stop cassette in the Ai9 cell model (Figure 6.29. A). To perform that we used the same HDR donor template (Figure 6.15.), but we added a T7 promoter by PCR to enable the conversion into RNA (Figure 6.29. B). However, the knock-in efficiency was very low (almost 0.1% when the Cas9 was delivered as plasmid, and 0.36% with RNP) (Figure 6.29. C). The problem could arise during the cDNA conversion as the quality of RNA in the beginning was optimal.
Figure 6.29. Replacement of the stop cassette at Neuro-2A Ai9 by GFP using single strand DNA (ssDNA) template. A) Schematic of GFP knock-in using ssDNA where it has right and left homology arms flanked with inverted Ai9 gRNA T1 and T4. B) methodology for synthesising ssDNA for GFP knock-in. C) flow cytometry analysis represents single well transfection with donor ssDNA and T1 and T4 gRNA. The results represent single well transfection analysed by flow cytometry.
The donor plasmid for the HITI technique can be optimised by using a minicircle plasmid instead of a normal plasmid. The minicircle plasmid is a supercoiled DNA molecule, but lacking all the prokaryotic parts required for replication, which makes it smaller in size and less immunogenic as the bacteria elements are removed which are rich in unmethylated CpG-motifs (394). In addition, minicircle plasmids are devoid from antibiotic resistance genes which cause sensitivity in certain individuals (395). The minicircle shows an improvement of the knock-in using the NHEJ pathway (around 58%) in comparison with a plasmid which has two cut sites (37.5%) or one cut site (43.5%). Interestingly, they found the gene expression (eGFP) delivered by a minicircle as template dropped by 25% after 80 days from the transfection (388). This result shows that, despite the efficiency of minicircle for expression with devoid CpG-motifs, they do not always prevent transcriptional silencing (396, 397).

In our lab, we have investigated the outcome of minicircle transfection for in vitro and in vivo experiments in terms of improving the gene expression as well as reducing the inflammatory response. The transfection on the epithelium cell line (16HBE14o-), using targeted nanoparticles (RTN) with minicircle encoding luciferase genes, shows a higher expression by five to ten-fold greater than plasmid DNA when the equivalent dose is used. On the other hand, the expression in the mouse lung was higher by two to four fold (382). These findings are encouraging for future application with the CRISPR system for gene editing.

All donor HITI templates in this chapter were cloned into parental plasmids of minicircle for future work optimisation.

Mini-intronic Plasmid (MIP) is another approach to devoid the antibiotic resistance gene in the plasmid, which has been developed by Jiamiao Lu and colleagues. However, they kept the essential plasmid DNA sequences required for bacterial replication within the expression cassette to avoid silencing. This modified plasmid shows improvement of the transgene expression between 2 to 10-fold in comparison with the minicircle in both in vitro and in vivo transfection (383, 398).

In Chapter Four and Five, we attempted to correct the ΔF508 mutation of the CFTR gene by homologous recombination (HDR), which was successful with the
enhancement of the selection marker. However, the other approach to the correction without selection was undetectable using Restriction Fragment Length Polymorphism (RFLP). The efficient correction of the mutation is quite difficult as there is a relatively big distance between the DSB and the correction site, even when the donor template or CRISPR system are properly designed.

There are various approaches which could be feasible for correcting the ΔF508 mutation using NHEJ for gene editing. Here we attempted to replace the CFTR exon 10 where the ΔF508 mutation existed. This approach has some advantages such as the relatively small size of the insertion (less than 1kb), which enhances the transfection efficiency and consequently, the knock-in. In addition, we are using the endogenous promoter for the CFTR gene without inserting an exogenous promoter.

There are other approaches which could be used for the correction; for example, inserting the whole CFTR gene (cDNA) either at a safe harbour locus, such as the Adeno-Associated Virus Integration Site 1 (AAVS1), or at intron 1 of the CFTR gene, using the endogenous promoter to drive the transcription. In addition, a super-exon for the CFTR could also provide another way of correcting the ΔF508 mutation, by making a double-strand break (DSB) at intron 9, then inserting a super-exon encompassing cDNA of the CFTR exons 10 to 27, which is 3051bp and smaller than the whole CFTR cDNA, which is 4443bp. This approach can be extended to include other mutations occurring in exon 10 and further downstream, such as G542X (exon 11), G551D (exon 11) and 3120+1G->A (splice mutation at intron 16). The super-exon approach has been applied successfully for correcting the ΔF508 mutation by Christien Bednarski and colleagues on the CFBE41o- cell line using the HDR technique. In this approach, zinc finger nucleases have been used to introduce DSB at exon 10, where a donor template encompasses the super-exon sequence. To facilitate the clonal screening, a puromycin resistance gene is incorporated inside the donor template (267). In this Chapter, we have noticed that the gRNAs have different efficiency for DSB, as seen also in Chapter Four where we screened for gRNAs targeting near the ΔF508 mutation. Despite the Benchling software showing most of the gRNAs at
intron 10 have a similar on-target efficiency score (Table 6.1), the \textit{in vitro} transfection shows the intron 10-4 was most efficient.

The off-target screening of the intron 10-4 gRNA showed an off-target effect (Figure 6.27.). Despite the off-target effect happening at a non-gene locus, it might be that another gRNA at intron 10 can be used for future work, to avoid the undesirable off-target effect. Interestingly, the PAM motif for the off-target indels was NAG, which is a non-canonical target by SpCas9 (Table 6.3.).

Here we managed to correct the ΔF508 mutation on HEK293 cells at a certain level (almost 4.5%), while the editing on CFBE41o- was undetectable by RFLP assay. However, the \textit{CFTR} mRNA expression increased significantly after the correction. In future work multiple transfection doses might be needed to increase the corrected cells. In addition, reverse integration needs to be evaluated, either by cloning the cells then sequencing them, or performing next-generation sequencing on the mixed population of the cells.

In summary, we demonstrated that knock-in (HITI) through the NHEJ pathway is an efficient technique as an alternative to the cell cycle dependent HDR pathway. In addition, we established a cell model (Neuro-2A Ai9) for all three gene editing pathways: HDR, NHEJ and HITI. Finally, we applied this technique as an approach to replace the \textit{CFTR} exon 10 where the ΔF508 mutation occurs.
CHAPTER 7

General discussion and conclusion
7. General discussion and conclusion

In this study we aimed to investigate the potential of genome editing to develop a genetic therapy for CF using the CRISPR/Cas system. This allows the gene-specific, targeted correction of disease-related mutations to be introduced at the chromosomal level, with a therapeutic potential for the ΔF508 CFTR mutation.

In the third chapter we started by optimising the CRISPR/Cas9 transfection delivered by a commercial reagent (L2K) and receptor-targeted nanocomplexes (RTN) on cell models, where the indels created by gene editing were monitored by the loss of GFP expression at the Neuro-2A cell model or the restoration of tdTomato expression following the excision of the stop cassette by paired gRNAs at the Neuro-2A Ai9 cell model. The CRISPR system was delivered to the cells in three formats: plasmid DNA that encoded both Cas9 and gRNA, Cas9 mRNA with gRNA, or Cas9 protein with gRNA (RNP), where we showed that gene editing was achieved successfully with our nanoparticle formulations. In addition, assessment of in vivo gene editing in the Ai9 mouse model was investigated where we managed to edit almost 0.5% of epithelial lung cells by NHEJ following three repeat doses by oropharyngeal instillation using RNP.

The correction of ΔF508 is investigated in the rest of this thesis where initially we screened for efficient gRNAs close to the ΔF508 mutation. We demonstrated that the most efficient gRNAs (T5+ and T7) resulted in 40% and 25% indels in CFBE41o- cells, respectively. By using a donor plasmid for wild type exon 10 CFTR and a puromycin selection marker, we managed to isolate the corrected cell clones which demonstrated response for CFTR activator and inhibitor at the electrophysiological studies. In addition, the correction led to an increase in CFTR mRNA expression which, to our knowledge, has not been previously reported (98, 268, 270). The work has also been extended to evaluate the correction of ΔF508 mutation on CF primary cells which have been transduced with the BMI-1 gene to delay senescence (93). We showed that the RTN formulated with RNP were able to induce gene editing in the primary cells to the same extent as the commercial transfection reagent with
RTN having the added advantage of repeat dosing. Finally, in Chapter Six, we investigated homology-independent targeted integration (HITI) as an alternative technique for gene editing (388). This approach was demonstrated effectively on the Ai9 cell model, then applied for the replacement of exon 10 in the CFTR gene.

One of the main impediments for gene editing with HDR or HITI is the inefficiency with regard to delivering the donor template to cells. This can be seen in terms of the transfection of primary cells with pDNA (section 5.3.1.1). Using a smaller size plasmid such as the minicircle plasmid (382) or a mini-intronic plasmid (MIP) (383) where the bacteria backbone has been removed, is a potential approach to improving donor delivery. Using a non-integrating virus such as AAV to deliver the donor template could be an option in terms of enhancing editing efficiency. In order to improve HDR in iPSC, Martin and colleagues used Cas9 RNP and AAV6 vector for donor delivery. Using this technique they managed to correct up to 63% of the single-nucleotide polymorphism (SNP) mutation in the HBB gene which causes sickle cell disease (399). On the other hand, an AAV vector was used successfully to deliver the donor template in non-dividing cells (neurons) in order to correct the cells through homology independent pathways (HITI) (388).

ssDNA is an alternative approach to dsDNA as a donor template, and might be applied for correcting CFTR mutation. The use of ssDNA shows promising results, especially when only a few bases require editing. In primary human T cells the co-delivery of RNP and ssDNA to replace 12 nucleotides at the CXCR4 locus, with homology arms containing 90 nucleotides each, showed up to 20% editing efficiency by HDR (400). In addition, rationally designing the ssDNA has an effect on the editing efficiency, in that Richardson et al.2016 showed that an asymmetric ssDNA (36 bases in the homology arm of the PAM-distal side and 91 bases in that of the PAM-proximal side) enhanced HDR efficiency compared with symmetric ssDNA by 50% (401). Furthermore, long ssDNA can be used to achieve large gene editing with low off-target integration events compared to the use of dsDNA (402) as is needed to replace the CFTR.
exon 10. In addition, in future work, hybridising the donor ssDNA with RNP delivered by the targeting nanoparticles (RTN) might be needed to enhance gene editing, where the ssDNA will be immediately available at the editing locus after DSB by the RNP (326).

In this project we investigated the effect of correcting the ΔF508 mutation on the CFTR function using electrophysiology studies. In future work, other measurements can be used to confirm the restoration of the CFTR function, such as restoring the cilia beating frequency, the ASL depth and the localization of the CFTR at the apical side of the cells. The dysfunctional aspect of CFTR causes impaired homeostatic mechanisms in the epithelium cells. This leads to the dehydration of ASL, and mucus accumulation (403). By restoring the function of CFTR, we expect that the ASL depth will increase and will restore the cilia beating frequency. In addition, using the Z-stacking feature of the confocal microscope will help to determine the localization of CFTR on the corrected stained cells.

We investigated the effect of correcting the CFTR mutation on the ENaC channel as an indirect way to confirm gene editing, where we found that correcting the ΔF508 mutation causes a reduction in the ENaC expression. In future work, other channels interacting with CFTR such as the CaCC (Calcium-activated chloride channel) and the SLC26A9 (Solute Carrier Family 26 Member 9) might be investigated to determine the effect of correcting the CFTR mutation on them, either by inhibition or by activation (61, 404).

In this project we investigated the correction of ΔF508 mutation in the CFTR gene using HDR and HITI techniques, where both required a donor template to allow editing. In future work, correcting other CFTR mutations while not requiring a donor template for correction, might be investigated by precise editing using CRISPR base editing, or by removing the mutation by NHEJ. These might offer more efficient approaches. The base editing technique does not induce DSB when dSpCas9 is fused to the cytidine deaminase enzyme (convert C•G to T•A base pairs) or the adenosine deaminase enzyme (convert A•T to G•C base pairs). This approach shows up to 50% editing efficiency (253). Using this technique, CFTR nonsense mutation or other single base
mutations might be good candidates for correction. However, the main limitation of this technique is that it requires a targeted base for conversion, located at position 12-16 bases upstream of the PAM (253). To overcome this, the deaminase enzyme can be fused to other variants of dCas9 such as dSaCas9 (405) or dCpf1 (406). Other applications of the CRISPR system might be used to induce NHEJ as an efficient editing mechanism compared to HITI or HDR. Targeting the nonsense mutation (class I CFTR mutation) such as G542X using the CRISPR system might lead to the restoration of the reading frame and allow protein synthesis. On the other hand, exon skipping might be used to remove the affected exon through the use of paired gRNAs. Using modulator molecules might be necessary to restore the CFTR function, as they are most likely to be affected by indels after gene editing or exon removal. In addition, CFTR splice mutations could restore functionality by removing the mutation. In proving this principle, Sanz et al. 2017 edited three splice mutations (3849+10kbC>T, 1811+1.6kbA>G and 3272-26A>G) in HEK293T cells transfected by mini-gene plasmids containing the splice mutations. Paired gRNAs were used to remove the mutations which led to the restoration of normal splicing patterns (407). However, more investigation is needed to confirm the restoration of the function on CF cells having these mutations.

In Chapter Four we corrected the ΔF508 mutation in CFBE41o-cells where we showed increase of the CFTR mRNA expression and the CFTR channel function. In future work, creating a new cell line model involving homozygous ΔF508 mutation might be necessary, as the CFTR expression in CFBE41o-cells is very low. This makes the function evaluation challenging. 16HBE14o-cells demonstrated a high expression of mRNA compared to corrected CFBE41o- (>200-fold, Figure 4.16,) which showed a perfect response to CFTR activators and inhibitors (Figure 4.18.) The modified 16HBE14o- cell line could be created with HDR editing to create a ΔF508 mutation with regard to both alleles. However, the 16HBE14o- cells has limitations in terms of the function assessment of sodium channels where the cells do not respond to amiloride drug treatment (Figure. 4.18) compared to primary epithelium cells (Figure. 5.14). The reason for this has been investigated, and it was found that the
ENaC protein in the cells was produced, but did not reach the apical side of the cells (337). To overcome this limitation, feeding the cells with media containing dexamethasone or glucocorticoids to maximize the ENaC expression and trafficking might be needed (337, 408, 409).

Having more variants of Cas9 means that we could have more options for targeting certain loci, especially for performing HDR, where the efficiency of editing is correlated with the distance of the DSB from the mutation it is intended to correct. In Chapter Four we explored three Cas variants (SpCas9, SaCas9 and Cpf1) to create a DSB close to the ΔF508 mutation (less than 10 bp distance). However, a TIDE assay of the transfected cells did not demonstrate indels. Targeting this locus with other Cas9 variants such as CjCas9 (229) is expected to cut 9bp from the ΔF508 mutation (Figure 7.1.). The use of an expanded SpCas9 variant (xCas9) could be explored to target the ΔF508 mutation, where 5' GAT 3' and 5' CAA 3' PAM are located two bases from the mutation (Figure 7.1.) (410).

Figure 7.1 CFTR gene sequence near the ΔF508 mutation with the three possible Cas9 gRNAs (CjCas9 and xCas9 which is an expanded SpCas9 variant with broad PAM compatibility) targeting the region. The direction of the arrows indicate the strand targeting gRNA (antisense), while the PAM motif of each gRNA is indicated by a similar-coloured rectangle.

The use of CRISPR/Cas9 as a gene editing tool has a potential off-targeting effect. We explored such an effect by performing Sanger sequencing in Chapters 4 and 6, where we found the intron 10-4 gRNA caused off-targeting by creating indels at chromosome 6 (Figure 6.28.). In the Introduction section we discussed ways of reducing the off-targeting effect by optimising the two parts of CRISPR (gRNA and Cas9) (see sections 1.6.3.1 and 1.6.3.3).
needs more investigation and optimisation in future work. In addition, another potential drawback of the CRISPR system is the response of the immune system to the CRISPR components (gRNA and Cas9), an aspect which needs to be considered, especially with regard to repeat dosing in vivo. This potential immune response could relate to innate or adaptive immune systems. The innate immune system is the first immune barrier involving non-specific reaction to foreign bodies such as nucleic acids (411). In terms of the CRISPR system it has been reported that the IVT gRNA causes type I interferon release, as a result of the activation of DDX58 (DExD/H-box helicase) by the 5′-triphosphate group at the IVT gRNA. The removal of the triphosphate group by calf intestinal phosphatase (CIP) of IVT gRNA, or using chemically modified gRNA, does not show an innate immune response (412). On the other hand, a specific immune reaction (adaptive) could be another barrier for future application of the CRISPR system by inducing a humoral (antibody) or cell-mediated immune response. The adaptive immune response was explored in vivo, and it was found that the delivery of CRISPR by AAV9 evokes humoral and cellular immune response in mice, which the vector alone did not induce. This indicates a prior immune memory against Cas9 (413). In humans, screening of the donor’s serum showed that 79% have antibodies against SaCas9, while 65% of the samples gave a positive reaction to SpCas9. On the other hand, incubating SaCas9 with the donors’ peripheral blood mononuclear cells (PBMCs), activates T-cells and induces inflammatory cytokine interferon-γ (IFN-γ) in almost half of the donors (414). The immune response to CRISPR components might lead to a reduction in the therapeutic effect through neutralizing antibodies, or by eliminating the edited cells by a cell-mediated immune response. A severe systemic immune reaction could be the worst scenario as seen in a gene therapy clinical trial for ornithine transcarbamoylase (OTC) deficiency, which caused patient death in 1999 (415). Prior screening for the possibility of an immune response of the patient will be important before receiving CRISPR therapy.

The repair of DSB by NHEJ pathway on general associate with small indels at the targeting region (306). However, a large deletion or complex genomic
rearrangements can be happening by CRISPR/Cas9 or other gene editing technique which might have pathogenic effect. In this context, Tomberg et al. 2018 studied the possibility of large indels after CRISPR/Cas9 transfection in vitro and in vivo where they showed by using a long-read sequencing technique the repair at targeting site cause large deletion extended to many kilobases (416).

The barriers for in vivo gene editing as potential therapy are numerous which include the low editing efficiency specially in non-dividing cells, targeting the suitable cells and the possibility of inducing innate or adaptive immune reactions. In addition, completely avoiding the possibility of off-targeting is challenging in vivo. The future of gene editing in humans will be more likely an ex vivo approach where cells are removed from the patients then edited in the lab, followed by re-infusion to the patients. The ex vivo approach has the advantage that the corrected cells are enriched, followed by off-targeting screening. In addition, the immune reactions against the CRISPR components will not be an issue anymore. Figure 7.2 summarises the potential future therapy regime by the CRISPR system for cystic fibrosis using an ex vivo approach.
Figure 7.2 The potential future gene editing therapy for cystic fibrosis using CRISPR system using ex vivo approach.
8. References


Guimbellot JS, Leach JM, Chaudhry IG, Quinney NL, Boyles SE, Chua M, Aban I, Jaspers I, Gentzsch M. Nasospheroids permit measurements of CFTR-dependent fluid transport. JCI insight. 2017;2(22).


129. Aiuti A, Roncarolo MG, Naldini L. Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products. EMBO molecular medicine. 2017:e201707573.


264. Sargent RG, Suzuki S, Gruenert DC. Nuclease-mediated double-strand break (DSB) enhancement of small fragment homologous recombination (SFHR) gene modification in


