Lentiviral Delivery of CRISPR Cas9 Prime Editing System for Cystic Fibrosis

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Dwiantari Satyapertiwi

Gene Medicine Research Group
Radcliffe Department of Medicine

and

Exeter College
University of Oxford

Supervisors: Prof Deborah R. Gill, Prof. Stephen C. Hyde,

and Dr Kamran M. Miah

Trinity Term 2024

Lentiviral Delivery of CRISPR Cas9 Prime Editing System for Cystic Fibrosis

Dwiantari Satyapertiwi

Radcliffe Department of Medicine and Exeter College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2024

Abstract

The field of gene editing is rapidly developing, particularly with advancements in gene editor technologies. Prime editing has been a significant breakthrough in gene editing, allowing precise, targeted correction of all types of DNA mutation, including insertion, deletion, and base substitutions, without inducing double strand breaks in the genome. Along with other gene editing tools based on the CRISPR-Cas9 technology, prime editing is being extensively researched for its potential to correct mutations causing monogenic disorders, such as cystic fibrosis (CF). Recombinant lentiviral vectors (LVs) were investigated as a delivery method for prime editors (PE), with the aim to correct the common F508del mutation in the *CFTR* gene that causes CF. The 'all-in-one' vector strategy utilised a PE-LV delivery vector, which was constructed to accommodate both the PE and the prime editor guide RNA (pegRNA), enabling efficient delivery via transduction of target cells with one vector. To target *CFTR*, the all-in-one vector was designed to transduce and express transgenes in lung cells, using a lentiviral genome backbone derived from Simian Immunodeficiency Virus (SIV) and a hybrid human CMV enhancer/EF1α (hCEF) promoter driving expression of PE. Insertion of CTT bases at the human HEK3 locus at the +1 position from the cut site (+1CTT) was attempted using the PE-LV

vector approach, but editing was not observed when HEK293T cells were used as a model, suggesting that further optimisation is required for the PE-LV construct.

An alternative 'two-vector' strategy was developed for delivery of PE (termed Split-PE), and as an alternative gene editing technology, Cas9-based Homologous Independent Targeted Integration (HITI). Upon optimisation for targeting HEK293T cells, the two-vector strategy utilised a lentiviral genome backbone based on Human Immunodeficiency Virus (HIV) and a reporter protein was incorporated in one of the vectors. Through transfection of Split-PE genome plasmids, +1CTT insertion was observed at a rate of (11.33%±4.04%), compared with (17.67%±5.86%) editing obtained through PE2 and pegRNA plasmid transfection. HITI was used to integrate a 'superexon' donor encompassing exons 11-27 of *CFTR*. However, only very low levels of HITI donor integration were observed, and this approach was not pursued further.

One notable challenge in using prime editing is identifying the optimal pegRNA to target specific mutations in different loci. A 'mini library' of pegRNA was developed to facilitate high throughput screening of over 400 pegRNA candidates targeting the F508del mutation in *CFTR*, using a synthetic target incorporated into the library construct. Next Generation Sequencing analysis did not indicate any pegRNA as having produced successful edits however, subsequently, the F508del mutation in *CFTR* has been shown to be particularly difficult to target. Despite this, the mini library remains a promising strategy for screening a large number of pegRNA candidates. The inherent adaptability of the strategy for targeting other disease-causing mutations further highlights the usefulness of this approach.

In conclusion, while prime editing shows great promise for precise correction of mutations, challenges remain in optimizing its use for therapeutic purposes, such as for treating cystic fibrosis. Recombinant lentiviral vectors have been explored as a delivery method, but initial attempts, including the all-in-one PE-LV and two-vector approaches, require further refinement. Despite difficulties in targeting the F508del mutation, the mini library of pegRNAs offers a valuable tool for screening potential pegRNA candidates. Continued research and sequence optimizations will be essential for using LVs as a delivery method for prime editing.

Acknowledgements

Alhamdulillahi Rabbil-'Alamin, I thank God Almighty, for without Him, I would not have had the strength to overcome all the difficulties I faced during my DPhil and throughout my writing process. I would like to thank my supervisor, Prof Deborah R. Gill, Prof Stephen C. Hyde, and Dr Kamran M. Miah for the guidance and support they have provided throughout my DPhil. I also would like to thank Dr Jakob Hansen Haldrup for his invaluable assistance, generously sharing his expertise in Prime Editing and bioinformatics throughout my research.

To Hanifi, my husband/best friend, for everything from the delicious meals you cooked and the comforting hugs when the future seemed bleak. To my beloved family, Mama, Itia, Madik, I wouldn't have even set foot in the UK without your unwavering support throughout this journey. To my bonus family, Mama Ema, Papa Fahmi, Hafis, Fina, Mbanad, thank you for coming into my life and supporting me. Tarmidzis and Tarmidzi cousins, thank you for filling our group chat with the most random stuff that warms my heart. My precious grandparents who passed during the time of my DPhil, Agus Tarmidzi, Tamsi, and Sutiyah, for all the love and encouragement, may you rest in peace.

To all past and present Gene Medicine Group members, (almost) Dr Aimee Ruffle, Omar Habib, Eoin Mac-Reamoinn, Gavin Turnbull, Marina Cerezuela, Galina Boskh, Dr Arlene Glasgow, Tina Garland, Dr Hamid Dolatshed, Dr Stephanie Jones, Dr Rosie Munday, Dr Helena Meyer-Berg, Dr Joost van Haasteren, Dr Dina Yue Du, Dr Jack Tan, Dr Toby Gamlen, Dr Altar Munis, for the friendship and shared memories. To Tanya Ren, for being such an amazing student and friend. The VP team, who always help me with technical questions about LVs: Mariana Viegas, Rebecca Dean, Catriona Conway, Emily Castells, Shaz Tariq, I cannot be more

grateful. Thank you to Dr James Davies, Dr Weijiao Zhang, and Roman Doll, thank you for your generous help with the NGS.

To Vita and Rafi, and S-TEAM, thanks for being the best friends I could ever ask for in my life, supporting me from when I was an awkward teenager to where I am now. To my Oxford-Indo friends, Noni, Lhuri, Rangga, Aisha, and the London gang, Aulia Syakhroza, Adhi Kuncoro, Riesta, and Dikra, thank you for giving me a sense of home away from home. The lovely bioengineers: Arianna, Jiyoon, and Anna, for being the most adorable friends I have in the UK.

I would also like to extend my sincere thanks to the Jardine Foundation for their generous scholarship and the precious community with the scholars, which has provided crucial support throughout my studies.

Declaration

This thesis has been composed by myself and has not been used in any previous application for a degree. The results presented were obtained by myself and contributions made by others have been acknowledged within the text and legends. All sources of published or unpublished information have been specifically acknowledged by means of a reference or personal communication notation, respectively.

List of Abbreviations

16HBEo14-	Human bronchial epithelial cell line
A549	Adenocarcinomic human alveolar basal epithelial cells
AAV	Adeno-associated virus
ABE	Adenine base editor
AP	Ammonium persulfate
BE	Base editing
bp	Basepairs
BSA	Bovine serum albumin
CAR-T	Chimeric antigen receptor T-cell
Cas9	CRISPR associated system 9
Cas9n	Cas9 nickase
CBA	Chicken β-actin
CBE	Cytosine base editor
CDNA	Complementary DNA
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
ChIP-SEQ	Chromatin immunoprecipitation followed by sequencing
Cl-	Chloride ion
CMV	Cytomegalovirus
CpG	Cytosine guanine dinucleotide
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
DAPI	4',6-diamidino-2-phenylindole
dCas9	Deactivated Cas9
ddPCR	Droplet digital PCR
DECODR	Deconvolution of Complex DNA Repair
DISCOVER-SEQ	Discovery of in situ Cas off-targets and verification by sequencing
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribose nucleic acid
dNTP	deoxynucleotide triphosphate
DSB	Double strand break
dsgRNA	catalytically inactive ("dead") sgRNA
dsODN	double-stranded oligonucleotides
E.coli	Escherichia coli
EF1α	Human elongation factor-1 alpha
EGFP	Enhanced GFP
epegRNA	Enhanced pegRNA
EvopreQ1	prequeosine1-1 riboswitch aptamer
EXO1	Human endonuclease 1
F/HN	Fusion / haemagglutinin-neuraminidase protein
FACS	Fluorescence activated cell sorting
FANCF	FA Complementation Group F

FBS	Foetal Bovine Serum
FEN1	Flap endonuclease 1
gDNA	Genomic DNA
GFP	Green Fluorescence Protein
gRNA	Guide RNA, consisting of crRNA and tracrRNA
GUIDE-SEQ	Genome-wide Unbiased Identification of DSBs Enabled by Sequencing
hCEF	Hybrid human CMV enhancer/EF1α
HDR	Homology directed repair
HEK	Human embryonic kidney
HEK3	site 3 locus in HEK293T cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HITI	Homology independent targeted integration
hU6	Human U6 promoter
ICE	Inference of CRISPR Edits
ICLV	Integrase-competent LV
IDLV	Integrase deficient LV
Indel	Insertion and deletion
kb	kilobases
KRAB	Krüppel-associated box
LB	Lysogeny broth
LNP	Lipid nanoparticles
LTR	Long terminal repeats
LV	Lentiviral vector
LVNP	Lentiviral derived nanoparticle
MFI	Median fluorescence intensity
MLH1dn	MutL protein homolog 1 dominant negative
MMEJ	Microhomology mediated end joining
MMLV	Moloney murine leukaemia virus
MMR	Mismatch repair
MOI	Multiplicity of infection
mpknot	MMLV pseudoknot
NBD1	Nucleotide binding domain 1
NEB	New England Biolabs
NFW	Nuclease free water
NGS	Next Generation Sequencing
NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
Npu	*Nostoc punctiforme*
P2A	Porcine teschovirus-1 2A
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer-adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Prime editor
PE2	Prime editor 2
PEmax	Prime editor max
PenStrep	Penicillin-Streptomycin

PFA	Paraformaldehyde
PGK	human phosphoglycerate kinase 1
qPCR	Quantitative polymerase chain reaction
rHIV	Recombinant HIV
RIPA	Radioimmunoprecipitation assay
rLV	Recombinant LV
RNA	Ribonucleic acid
RNF2	Ring finger protein 2
RNP	Ribonucleoprotein
RRE	Rev responsive element
rSAP	Recombinant shrimp alkaline phosphatase
rSIV	Recombinant SIV
RT	Reverse Transcriptase
SaCas9	*Staphylococcus aureus* Cas9
SDM	Site directed mutagenesis
SDS	Sodium dodecyl-sulphate
SeV	Sendai Virus
sgRNA	Single guide RNA
SOC	Super Optimal broth with Catabolite repression
SORT	Selective organ targeting
SpCas9	*Streptococcus pyogenes* Cas9
StCas9	*Streptococcus thermophilus* Cas9
T2A	Thosea asigna virus 2A
T7EI	T7 Endonuclease I
TBS	Tris buffered saline
TBS-T	Tris buffered saline-Tween
TEMED	N,N,N',N' -Tetramethylethylenediamine
TG	Tris Glycine
Tm	Melting temperature
tracrRNA	Trans-activating CRISPR RNA
VSV-G	Vesicular stomatitis virus G protein
WPRE	Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1. CRISPR-Cas9

Clustered regularly interspaced short palindromic repeats, or CRISPR, is a family of deoxyribonucleic acid (DNA) sequences adapted from a bacterial defence mechanism, where the bacteria capture viral DNA sequences to recognize and destroy it in future infections. Together with the CRISPR-associated system 9 (Cas9) nuclease, CRISPR-Cas9 has been adapted to create a revolutionary gene editing technology that allows precise DNA modification[1,2]. The most commonly used Cas9 variants are isolated from *Streptococcus pyogenes* (SpCas9)[1] and *Staphylococcus aureus* (SaCas9)[3]. To function as a gene editing tool, CRISPR-Cas9 editing requires a synthetic guide ribonucleic acid (gRNA) comprising CRISPR RNA (crRNA), which contains the spacer sequence that complements the target DNA sequence and trans-activating crRNA (tracrRNA) which acts as a binding scaffold to the Cas9 nuclease protein. Even though these two components exist as separate RNA molecules in nature, they are commonly fused into a single guide RNA (sgRNA) to simplify the guide RNA format for CRISPR based editing[1]. In this format, the Cas9 protein binds to the sgRNA to form a ribonucleoprotein (RNP) complex, and the sgRNA guides the Cas9 to make a cut at the preferred DNA target site.

In addition to sgRNA, the Cas9 system also requires a protospacer adjacent motif (PAM) at the DNA target site to function. Interestingly, the sequence of the PAM is specific to each Cas9 protein variant. The commonly used SpCas9 requires an NGG PAM sequence next to the spacer sequence on the target site, while the SaCas9 variant requires an NNGRRT sequence. Other Cas9 variants with more flexible PAM requirements have been isolated

and developed, such as SpCas9-NG, which requires an NG PAM[1]; SpG, which requires NGN; SpRY, which requires NRN and, to a limited extent, NYN PAMs[4]; xCas9, which has flexible NG, GAA, or GAT PAM requirements[5]; and ScCas9, isolated from *Streptococcus canis*, which requires an NNG PAM[6]. When specificity is crucial, Cas9 variants with longer PAM requirements may be preferred, such as StCas9 variants isolated from *Streptococcus thermophilus*, with St3Cas9 requiring an NGGNG PAM and St1Cas9 requiring an NNAGAAW PAM[7].

Cas nucleases, like Cas9, trigger DNA modifications by producing a double strand break (DSB) at a location specified by the spacer sequence of the gRNA[1,8]. This DSB is then repaired via cellular repair pathways including non-homologous end joining (NHEJ)[9,10] and microhomology mediated end joining (MMEJ)[11–13], typically resulting in various, somewhat random, insertion and deletions (indels)[14,15]. Due to this, the resulting sequences are not easily predictable, therefore in some experimental configurations, a template is provided to direct the result of editing. Through homology directed repair (HDR)[10,16,17], or homology independent targeted integration (HITI)[18,19], a DNA donor template may be inserted into the generated DSB and allow precise, predictable editing outcomes. However, the efficiency of insertion especially in non-dividing cells, is very low for the HDR method[20–22] . With HITI, the insertion efficiency is higher in quiescent cells as it relies on NHEJ for integration pathway[19]. HITI will be described in more detail in the section 1.1.2 below.

Despite its usefulness for gene editing, there are concerns about inducing DSBs in cells. The DNA Damage Response (DDR) is induced when DSBs occurs, which halts the cell cycle

progression and if the DSB is not repaired properly, it could lead to cell cycle arrest and ultimately cell death[23–25]. Other studies have also reported large deletions, inversions, duplications, chromosomal rearrangements, and recombination between homologous chromosome arms associated with Cas9 induced DSBs[26–29]. Therefore, a gene editing technology that does not require a formation of DSB may avoid this toxicity to the targeted cells. The Cas9 nickase (Cas9n) is an engineered version of Cas9 nuclease which has one of the cutting domains, either RuvC-like or HNH domain inactivated, resulting in a single strand cut (often referred to as a "nick") rather than a double strand break[30]. To inactivate the HNH domain and cleave only the target strand, a D10A mutation is incorporated into the sequence[2], whereas a H840A is mutation is used to inactivate the Ruv-C like domain and cleave the non-target strand[1]. The latest additions to the rapidly evolving Cas9-derived gene editing toolbox such as Base Editing (BE) and Prime Editing (PE) use Cas9n to facilitate specific base changes without inducing a DSB. There are currently two potent base editing enzymes, Adenine Base Editor (ABE) converts A•T base pairs to G•C[31], while Cytosine Base Editor converts C•G base pairs to T•A[32,33]. Prime Editing, which was extensively used for this study is described in section 1.1.1.

When both the D10A and H840A mutations are incorporated in the Cas9 nuclease, it creates deactivated Cas9 (dCas9) which can still bind to genomic DNA (GDNA) without its cleaving function. Before the efficiency was improved in the later versions with a Cas9n, the earlier version of the CBE uses a dCas9[34]. Deactivated Cas9 is regularly used in CRISPR activation (CRISPRa) or CRISPR interference (CRISPRi)[35–37] approaches where CRISPR targeting is used to influence other DNA metabolic events. The usage of CRISPRa can be in

conjunction with VP64 activator only[38], or with additional p65 and RTA protein in the VPR system[39,40] to recruit transcription factors and activate endogenous coding and non-coding sequences. On the other hand, CRISPRi can be used by itself to block the progression of RNA polymerase[41], or with transcriptional repressors such as the Krüppel-associated box (KRAB) domain[42].

Despite the enormous potential of CRISPR-Cas9, off-target effects still pose a significant challenge, especially in translational settings[43–46]. The sgRNA used to guide the Cas9 nuclease can have up to 3 nt mismatches in the PAM distal region[43,44,46–48], allowing it to bind to a non-target genomic sites and create DNA cleavage in unintended locations. It is therefore highly important to thoroughly investigate the potential off-target sites associated with the use of certain gRNAs, both through *in silico* screening and *in vitro* detection and validations prior to further development for clinical use. *In silico* screening, through alignment-based methods such as Cas-OFFinder[49], or combined alignment-based and experimental based methods such as DeepCRISPR[50] or the CRISPOR[51,52] website, can provide an initial assessment of sgRNAs while accounting for off-target effects for each candidate. Experimental detection, such as Genome-wide Unbiased Identification of DSBs Enabled by Sequencing (GUIDE-SEQ)[53] rely on the delivery of double-stranded oligonucleotides (dsODNs) with known sequences alongside Cas9 and sgRNA to cell culture models. These dsODNs integrate at double-strand break (DSB) sites through non-homologous end joining (NHEJ) mechanisms. The integrated dsODNs are then amplified and sequenced using next-generation sequencing (NGS), providing reliable detection even at off-target sites with indel frequencies as low as 0.03%. Though useful for cell line models,

the requirements of dsODN makes this approach not suitable for *in vivo* applications. Another method, discovery of in situ Cas off-targets and verification by sequencing (DISCOVER-SEQ), relies on the recruitment of the DNA repair protein Meiotic Recombination 11 (MRE11) to sites of DNA damage[54,55]. The interaction between DNA and the protein is then detected and quantified using chromatin immunoprecipitation followed by sequencing (ChIP-Seq). Unlike GUIDE-SEQ, DISCOVER-Seq does not require dsODN templates and has been demonstrated to work in induced pluripotent stem cells (iPSCs) as well as in a mouse liver model[55]. This advancement potentially opens new pathways for off-target detection in preclinical and clinical research.

1.1.1. Prime Editing

Prime editing is a novel CRISPR-Cas9-based editing approach which uses the combination of Cas9n with H840A mutation and a reverse transcriptase (RT) enzyme from Moloney Murine Leukaemia Virus (MMLV-RT)[56], named as prime editors (PE) The use of Cas9n enables editing without inducing a DSB, therefore avoiding this potential toxicity. Another advantage of PE over the regular CRISPR-Cas9 system is the ability to install any desired edit using the MMLV-RT coupled with the PE guide RNA (pegRNA) without the need of an additional donor template, unlike HDR-mediated repair. The pegRNA is similar to CRISPR-Cas9 sgRNA, with a spacer region that directs PE to the target DNA locus and a scaffold region that binds the guide RNA to the Cas9 protein, but with an added 3' extension region downstream of the scaffold[56] (Figure 1.1). The 3' extension region contains a primer binding site and an RT template sequence used by the MMLV-RT to install the desired edit. The primer binding site sequence is complementary to the target sequence, whereas the RT

template contains the desired edits that are to be installed in the target site. This configuration of pegRNA allows PE to install any desired types of edits, from substitution, insertion, and deletion at a user-controllable distance from the spacer/PAM target sequence, which makes PE a highly useful tool in the gene editing field.

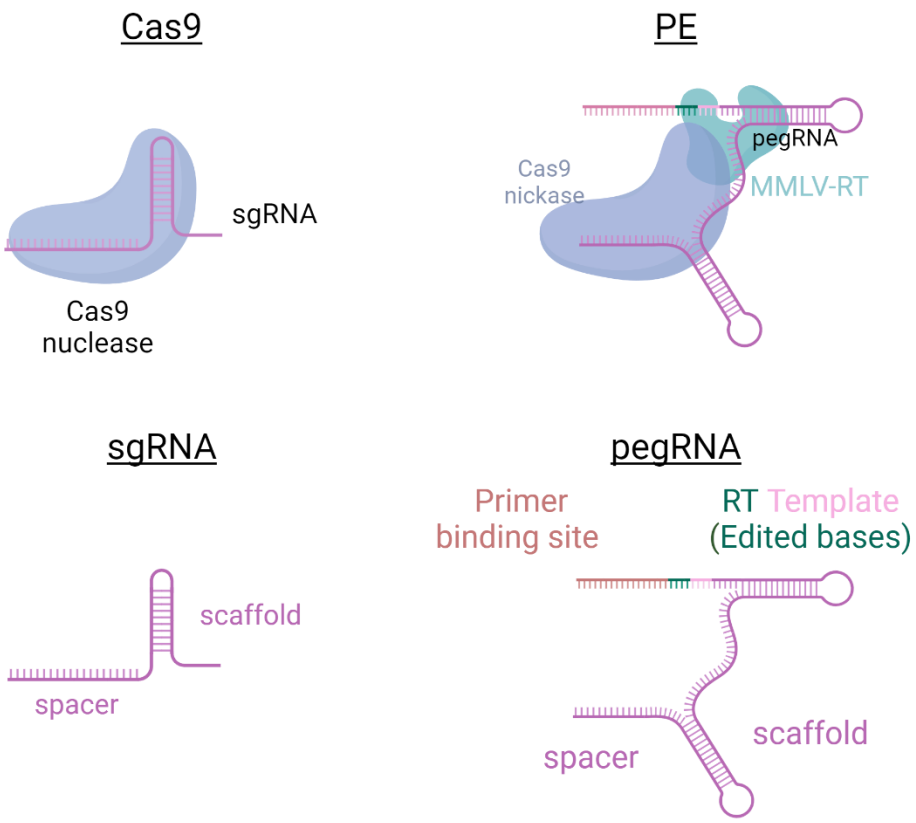

Figure 1.1. Comparison between Cas9 and sgRNA, and PE and pegRNA.

PE utilises Cas9 nickase and pegRNA which consist of spacer, scaffold, and 3' extension which includes the primer binding site and RT template. Image created in BioRender.com.

The mechanism of action of prime editing involves four main steps: guiding, nicking, template hybridisation and DNA synthesis, and incorporation of the desired edits[56,57]. First, as in the regular CRISPR-Cas9 system, the spacer sequence of the pegRNA directs the PE protein to the target site, where Cas9n then nicks the DNA strand opposite the target strand (Figure 1.2A). The 3' extension of the pegRNA then serves as a template for RT to synthesize the new, desired DNA sequence through hybridization with the opposite strand of the target strand (Figure 1.2B). This process generates a 3' flap, which equilibrates with a 5' flap that lacks the correct edit. The 5' flap is excised by structure-specific enzymes such as Flap Endonuclease 1 (FEN1)[58] or human Endonuclease 1 (EXO1)[59]. This excision promotes the incorporation of the 3' flap containing the desired edit into the target strand. DNA repair mechanisms then resolve the heteroduplex of the edited and non-edited (and therefore non-complementary) strands by copying the information from the edited strand to the original non-edited strand, permanently installing the desired edit into the target locus (Figure 1.2C).

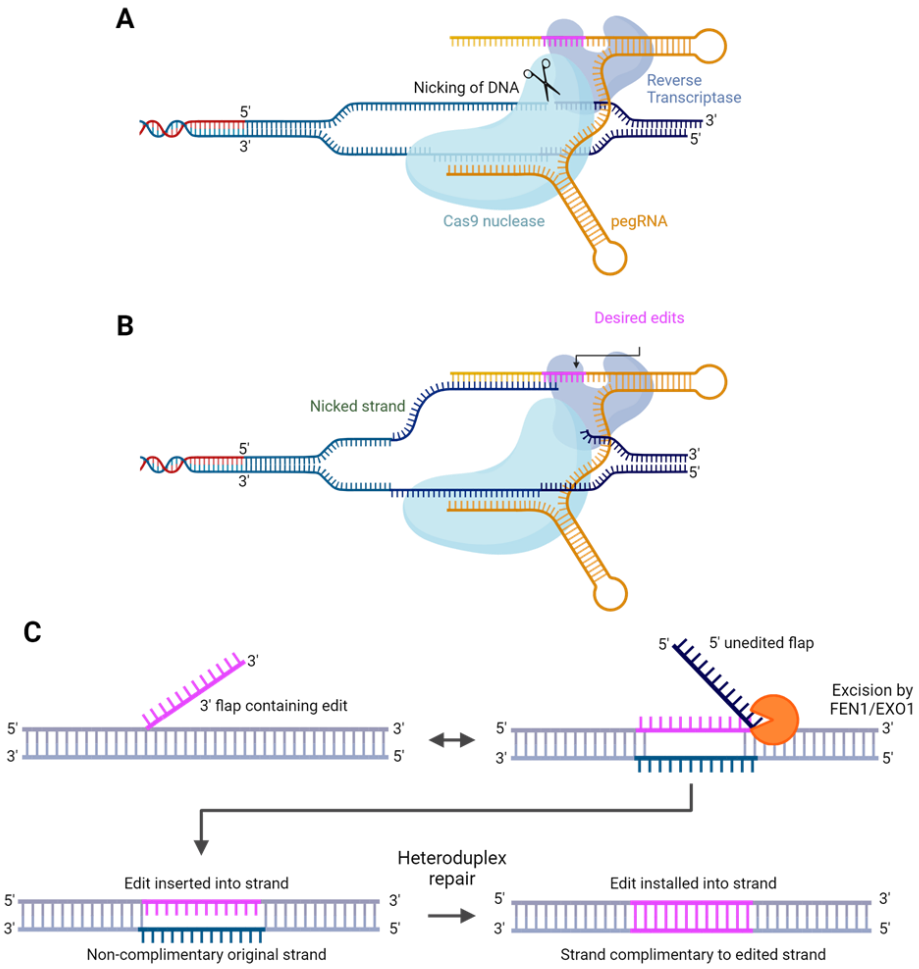

Figure 1.2. Mechanism of action of prime editing.

A) Spacer sequence of pegRNA guides PE complex to the target sequence and Cas9n nicks the opposite of the target strand. B) Hybridisation of the 3' extension sequence to the opposite strand of the target strand, new DNA sequence is synthesised using RT enzyme. C) Formation of 3' flap containing the desired edit which equilibrates with the 5' flap, however FEN1 and EXO1 activity promotes the excision of 5' flap and shifts the equilibrium towards the incorporation of the 3' flap containing the desired edits. The DNA repair mechanism then resolves the heteroduplex of the edited and non-complementary or strands by copying the information from the edited strand to the original strand, installing the desired edit into the target locus. Image created in BioRender.com.

At the time of writing, several published improvements of PE protein and pegRNA have been developed over the original configuration and these are listed in Table 1.1. These PE variants showed increased editing efficiency and minimised off-target effects[60–63]. The first improvement of PE, termed PE2, introduced five different mutations (D200N, T306K, W313F, T330P, and L603W) in the MMLV-RT sequence which increased the editing efficiency from 1.6 to 5.1-fold compared with the first version, PE1[56]. The mutations in the MMLV-RT domain in PE2 were subsequently retained in PE3, PE4, and PE5. For many of the experiments in this study, PE2 was used, in particular for the 'all-in-one' and the 'two-vector' delivery approaches of PE described in later chapters.

Beyond mutating the Cas9 and the MMLV-RT protein to increase the efficiency, modulating the DDR after introducing the PE protein can significantly influence the editing efficiency of prime editing. In PE4 and PE5 systems, the dominant-negative MutL Homolog 1 (MLH1dn) protein was co-expressed with the PE protein to further improve the editing efficiency by inhibiting the mismatch repair (MMR) mechanism[60,64] . Inhibition of cellular DNA repair pathways, such as MMR, influences the fate of the edited strand by preventing the reversal of the editing processes, thereby allowing the prime editing-induced edit to be successfully installed at the target locus. In this instance, the MMR mechanism reduces PE efficacy by excising the heteroduplex formed by the PE-edited 3' flap and the target strand, leaving the unedited strand to reattach. Inhibition of the MMR mechanism by the introduction of MLH1dn prevents the reversal of the installed edits, thus increasing editing efficiency[65]. In the same study, it was also shown that incorporating additional silent edits into the RT template, which alter the final DNA sequence but do not affect the desired amino acid

sequence, could further prevent the reversal of PE edits by acting as a decoy for the MMR system[60]. By incorporating silent edits into the RTT, the DNA sequence is altered in a way that reduces MMR recognition, thereby preventing the excision of the edited strand.

The next iteration of the PE protein was PEmax, which incorporates two additional Simian virus 40 (SV40) nuclear localization signals (NLS)—one in the linker sequence between the Cas9n and RT domains and another in the C-terminus of the PE construct. The addition of NLS increases the transport of PE complexes to the nucleus, and therefore improves the likelihood of the complexes binding to and ultimately editing the target site[60]. The Cas9n region of the PEmax was also further optimised with the addition of R221K and N394K mutations, while the expression of the MMLV-RT domain was enhanced by optimising the codon-usage for human-derived cells. PEmax demonstrated a 1.2-fold and 2.3-fold increase in editing efficiency in HEK293T and HeLa cells, respectively, compared with regular PE2 for substitution edits[60]. The PEmax version of PE was employed in the mini-library approach discussed in Chapter 5 of this thesis.

Improvements to the pegRNA have further enhanced the efficiency of the PE system. In a study by Nelson et al.[66], the prequeosine$_1$-1 riboswitch aptamer motif (evopreQ$_1$), the pseudoknot from MMLV (mpknot), or truncated versions of both were added to the 3' end of pegRNA, resulting in an enhanced pegRNA (epegRNA). These modifications were aimed at boosting PE efficiency by increasing pegRNA stability and preventing 3' pegRNA degradation by exonucleases. Optimising the scaffold sequence for sgRNA of SpCas9

promotes on-target activity of the sgRNA[67,68] and there is a possibility that it could similarly improve pegRNA, however more studies are needed to confirm this.

One notable challenge is the delivery of PE complex. The coding sequence of PEmax totals ~6.3 kb in size, encoding a complex ~240kDa-sized protein. The size and complexity of the PE system limits the choice of delivery vehicle/vector to efficiently deliver the necessary components into the target cells or tissues. Strategies used to deliver complex genome editors are described in section 1.2.

Table 1.1. Table of Prime Editing (PE) Variants and Descriptions of Modifications

No.	PE Type	Description	Mutations in Cas9n and/or MMLV-RT Domain	pegRNA Modifications	Reference
1	PE1	First version of the prime editing system and includes a fusion of a Cas9 nickase (H840A mutation) and an engineered reverse transcriptase (M-MLV-RT) to introduce targeted edits without double-strand breaks.	Cas9n: H840A; M-MLV-RT: none	None	[56]
2	PE2	Codon optimised reverse transcriptase domain for higher editing efficiency, increased editing accuracy and efficiency. The Cas9 nickase (H840A) remains the same as the PE1 version.	Cas9n: H840A; M-MLV-RT:D200N, T306K, W313F, T330P, and L603W		
3	PE3	PE3 uses the PE2 system with an additional sgRNA added to nick the non-edited strand. This increased the editing efficiency by promoting repair using the edited strand as a template.	Same as PE2	Additional sgRNA for nicking	
4	PE3b	A variation of PE3 which uses a second sgRNA that only targets the complementary strand after the edit has been installed designed.		Additional sgRNA specific to post-edit complementary strand	
5	PE4	Enhances editing efficiency by including a dominant-negative MutL Homolog 1 (MLH1dn) to PE2 system to suppress mismatch repair, thereby reducing indels. It also includes mutations in the reverse transcriptase for improved activity.		None	[60]
6	PE5	PE3 with an addition of MLH1dn.			

Table 1.1. (cont'd). Table of Prime Editing (PE) Variants and Descriptions of Modifications

No.	PE Type	Description	Mutations in Cas9n and/or MMLV-RT Domain	pegRNA Modifications	Reference
7	PEmax	PEmax uses human codon-optimized RT, a 34-aa linker containing a bipartite SV40 NLS, an additional C-terminal c-Myc NLS, and added R221K N394K mutations in SpCas9n	Cas9n: R221K, N394K, H840A; M-MLV-RT: human codon optimised	None	60
8	PE6a	The smallest prime editor to date, using evolved *Escherichia coli* retron, Ec48-RT	Cas9n: Same as PEmax; Evo-Ec48 RT: E60K, K87E, E165D, D243N, R267I, E279K, K318E, and K343N		62
9	PE6b	Suitable for short, unstructured RT templates, and to reduce indels and scaffold insertion products generated by PEmax; using evolved *Schizosaccharomyces pombe* retrotransposon, Tf1-RT	Cas9n: Same as PEmax; Tf1 RT: P70T, G72V, S87G, M102I, K106R, K118R, I128V, L158Q, F269L, A363V, K413E, and S492N		
10	PE6c	Suitable for pegRNA with a highly structured 3' extension or twin-PE strategy, using engineered and evolved Tf1 RT	Cas9n: Same as PEmax; Tf1-RT: P70T, G72V, S87G, M102I, K106R, K118R, I128V, L158Q, F269L, A363V, K413E, S492N, K118R, S188K, I260L, S297Q, and R288Q		
11	PE6d	Suitable for pegRNA with a highly structured 3' extension or twin-PE strategy, using RNaseH-truncated evolved and engineered MMLV-RT (M-MLVRTDRNAseHrt)	Cas9n: Same as PEmax;M-MLVRTDRNAseHrt: T128N, V223Y, and D200C		

Table 1.1. (cont'd). Table of Prime Editing (PE) Variants and Descriptions of Modifications

No.	PE Type	Description	Mutations in Cas9n and/or MMLV-RT Domain	pegRNA Modifications	Reference
12	PE6e	Evolved and engineered Cas9 domains to further improve editing efficiency	Cas9n: H840A, K918A, K775R; M-MLVRTDRNAseHrt	None	[62]
13	PE6f		Cas9n: H840A, E471K, H99R, I632V, H721Y, D645N, K918A; M-MLVRTDRNAseHrt		
14	PE6g		Cas9n: H840A, E471K, H99R, I632V, H721Y, R654C, D645N; M-MLVRTDRNAseHrt		
15	PE7	PEmax with an addition of La(1–194) fusion constructs at the C-terminal	Same as PEmax		[63]

1.1.2. Homology Independent Targeted Integration

The HITI is a genome editing technique that allows integration of donor DNA template via the NHEJ repair pathway [69,70], thus eliminating the need for sequence homology arms between the target and the donor template which is a requirement in HDR[71]. Also unlike HDR, which is active during the S and G2 phases of the cell cycle[72] and works primarily on cells that are actively dividing, HITI leverages NHEJ which is active in both dividing and non-dividing cells[9]. This gives an advantage to HITI over HDR in editing non-dividing cells.

The mechanism of HITI involves the creation of DSBs by Cas9 and sgRNA in both the DNA target and the donor template, which can be delivered as a plasmid, DNA minicircle or linear DNA template[70]. To design the donor template for HITI, Cas9 cut sites must be incorporated in the opposite orientation to the target sequence (Figure 1.3). After Cas9 cuts the target site, three possible outcomes can occur: first, no integration is observed; second, the donor DNA template is inserted, but in the undesired reverse orientation; and third, integration of the donor template in the correct orientation. In the first two cases, the Cas9 cut site is reconstituted and can be readily recut, allowing another cycle where the template to be reinserted in the correct orientation[19]. By allowing continuous, ongoing recutting following no integration or reverse integration, the balance ultimately shifts towards successful integration of the donor template in the correct orientation at the target site.

With this configuration of Cas9, sgRNA, and donor template, the types of edits that HITI can achieve are limited by the availability of a suitable PAM sequence near the target site, as well as the size of the donor template which might limit the efficiency of delivery. One

potential downside of the HITI approach is the difficulty in precisely controlling the insertion event, which could lead to concatemeric (repeated) insertions of the DNA template[73]. Additionally, since HITI relies on DSBs for donor template insertion, there is always the risk of unwanted insertions or deletions (indels) occurring at the integration site with or without an insertion event. Despite these challenges, the potential of HITI in genome editing is significant, especially in scenarios where HDR-mediated genome editing is inefficient.

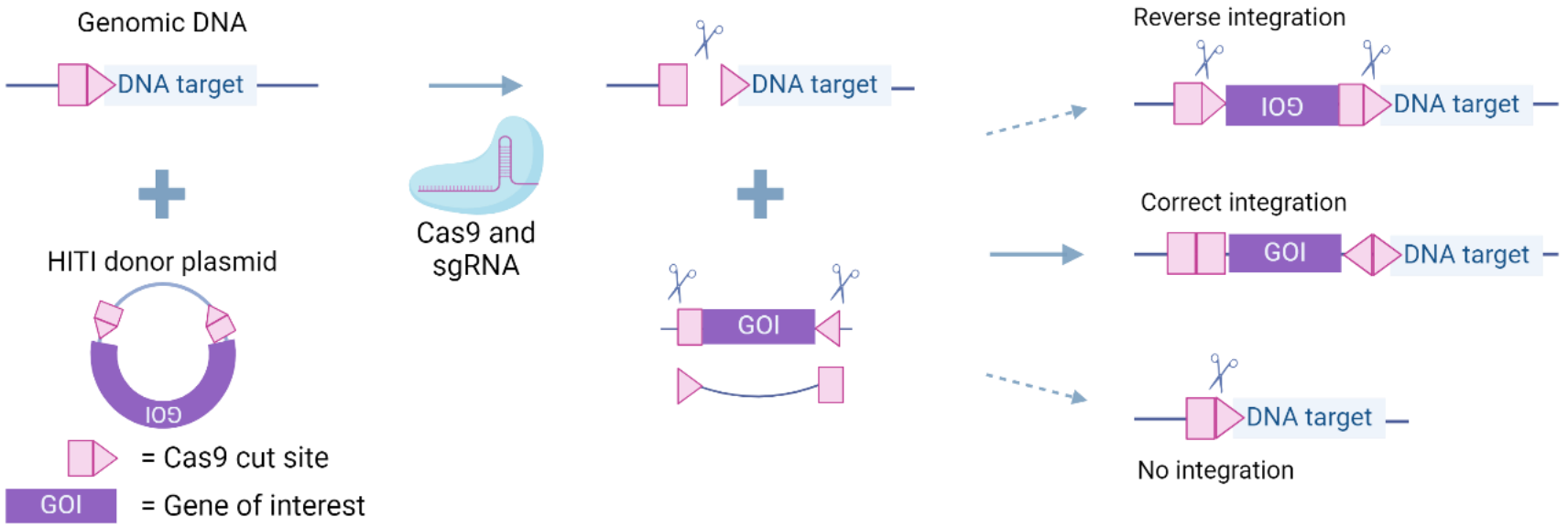

Figure 1.3. Schematic diagram of the HITI mechanism.

To introduce desired genetic edits, HITI uses a donor template, Cas9, and sgRNA to target specific Cas9 cut sites (represented by pink rectangles and triangles). The donor template, which contains the gene of interest (GOI), has one Cas9 cut site for the minicircle format or two sites for the plasmid format at both the 5' and 3' ends. Cas9 and sgRNA cleave both the donor template, releasing the GOI, and the target sequence in the genome. The process can result in three possible outcomes: (i) successful integration of the GOI in the correct orientation, (ii) incorrect reverse orientation insertion, which Cas9 can re-cut, allowing for correction, and (iii) no integration, where the target site is re-cut to facilitate further

attempts at insertion, An additional possible outcome is indel formation following double-strand break (DSB) formation, which prevents integration from occurring. Image created in BioRender.com.

1.1.3. Current therapeutic application of PE and HITI

CRISPR-Cas9 has rapidly moved from scientific discovery to a powerful technology in medicine. Its versatility and efficiency in editing genes have propelled rapid advancements and entering the market in just a little over a decade. The first CRISPR-based therapeutic to reach the market, Casgevy, obtained FDA approval for the treatment of sickle cell disease in late 2023[74]. This highlighted a significant milestone in the application of gene editing for therapeutic use.

Both prime editing and HITI are extensively explored in translational research with numerous ongoing pre-clinical trials, and one clinical trial for PE scheduled to start in 2024[75]. One of the notable trials involving prime editing is for Chronic Granulomatous Disease[75] led by Prime Medicine. They also have other genetic diseases in their pipeline, such as Retinitis Pigmentosa/Rhodopsin, Friedreich's Ataxia, and cystic fibrosis (CF)[76]. These efforts demonstrate the therapeutic potential of prime editing for correcting point mutations and small insertions or deletions in various diseases. Similarly, HITI has shown promise in animal models, specifically for correcting genetic disorders. It has been used to treat Mucopolysaccharidosis type VI (MPS VI) and Haemophilia A in mouse models[77]. Additionally, HITI has been employed to correct autosomal dominant retinitis pigmentosa[78], and it has also shown potential in the clinical-scale manufacturing of CAR-T cells for cancer therapy[79].

1.2. Viral and non-viral vector for delivery of CRISPR-Cas9 based technology

With its versatility and ease of use, CRISPR-Cas9 based gene editing technology has emerged as a leading candidate for therapeutic applications, offering unprecedented precision in correcting genetic defects. However, despite its promise, several challenges remain to be addressed. Aside from safety concerns associated with the formation of DSBs and potential off-target effects, one of the main issues with Cas9-based gene editing for therapeutic purposes is the limitation of effective delivery methods. Delivering Cas9 and the gRNA to the target cells without causing unintended effects is critical for successful gene editing. The ideal delivery method would avoid integrating Cas9 and/or the gRNA into the host genome, which could lead to long-term expression and increased risk of off-target effects. At the same time, it should ensure that Cas9 and the gRNA persist long enough in the cell to carry out the desired edit accurately. Current delivery strategies include viral vectors such as lentiviral vectors and adeno-associated virus (AAV), non-viral DNA or mRNA nanoparticles, and direct delivery of Cas9 protein and gRNA as RNP complexes. Of course, each strategy has its own limitations. Balancing these factors to ensure efficient delivery while minimizing risks is a critical barrier for translating CRISPR-Cas9 into viable gene therapies.

1.2.1. Lentiviral vectors

1.2.1.1. Lentiviral vector biology

Lentiviral vectors (LV) are derived from a subtype of retrovirus called lentiviruses, which contain a single-stranded RNA genome that is converted into DNA in the transduced cell using viral reverse transcriptase[80]. Unlike simple retroviruses such as gamma retrovirus,

lentiviruses have the ability to transduce both dividing and non-dividing cells[81,82]. One key feature of lentiviruses is that they can integrate their genetic material into the host cell's genome, allowing the inserted genes to be inherited by the cell's progeny and therefore resulting in more sustained gene expression[83]. Lentiviral vectors also have a large packaging capacity, up to 9.6 kb[84], allowing them to carry larger or multiple transgenes. The most commonly used LVs are based on Human Immunodeficiency Virus type-1 (HIV-1), though other viral species including Simian Immunodeficiency Virus (SIV) are also used.

There are several important components in the lentiviral genome which are crucial in LV production. Firstly, *gag*, which encodes structural proteins including the matrix, capsid, and nucleocapsid proteins; and *pol*, which supplies reverse transcriptase (RT), protease, and integrase viral enzymes[85]. The second group includes regulatory elements essential for post-transcriptional viral mRNA transport from nuclei to cytoplasm which is encoded in *rev* gene. Rev facilitates the transport of viral mRNA by binding to a specific RNA structural element known as the rev responsive element[86] (RRE). Lentiviral genomes are flanked by long terminal repeats (LTR) at the 5' and 3' ends (5' LTR and 3' LTR), and a packaging signal located 3' of the 5' LTR (Ψ)[87]. The expression of lentiviral proteins encoded in the genome is regulated by the *tat* gene, which transactivates the promoter in the 5' LTR to enhance transcription of the proviral genome into full-length viral RNA[88].

Lentiviral vectors also require an envelope glycoprotein, encoded by the *env* gene, which enables the virus to attach to and enter host cells[85]. The use of heterologous envelope proteins, a process known as pseudotyping[89], enhances the tropism of LVs. For example,

Vesicular Stomatitis Virus-Glycoprotein (VSV-G), is widely used in the gene therapy field due to its broad tropism, robust stability, and high vector titres[90]. In this study, in addition to VSV-G, two co-acting envelope proteins from Sendai Virus (SeV) – the Fusion (F) and Haemagglutinin-neuraminidase (HN) proteins comprise the F/HN pseudotype with known tropism for targeting lung cell types[91,92].

For use in gene therapy, LVs are carefully modified to enhance their safety profile by removing genes required for viral replication and developing self-inactivating LTRs which hinder the production of viral RNA. Replication deficient LVs cannot replicate and generate more virus particles. The required components are also split into multiple plasmids, with second generation LVs dividing the components into three plasmids and third generation LVs dividing the components into four plasmids (Table 1.2). Splitting the components into different plasmids reduces the potential for recombination events that could lead to generation of replication-competent lentiviruses.

In second generation LVs, the required viral packaging components *gag, pol, tat* and *rev*, are packaged into one plasmid while in third generation LVs, the necessary components are spread across two plasmids, with one containing *gag* and *pol*, and the other containing *rev*. Third generation LVs do not require *tat* for RNA transcription, as the *tat*-dependent native promoter in the 5' LTR region is replaced with an heterologous promoter such as the immediate/early cytomegalovirus (CMV) promoter. Self-inactivating (SIN) LVs were developed to further improve the safety profile of LVs by deleting 400-nucleotide in the 3' LTR, abolishing the LTR promoter activity without affecting vector titres or transgene

expression *in vitro*[93,94]. The ablation of LTR promoter activity necessitates SIN vectors to have an internal homologous promoter to drive the transgene expression[95], where CMV or Human elongation factor-1 alpha (EF1-α) are commonly used. A comparison between first, second, and third generation LVs [80,96,97] is summarized in Table 1.2.

Table 1.2. Comparison between first, second, and third generation LVs

Feature	First-Generation LV	Second-Generation LV	Third-Generation LV
Number of plasmids	3 packaging plasmid (includes *gag-pol-rev*), *env*, transfer vector)	3 (packaging plasmid (includes *gag-pol-tat-rev*), *env*, transfer vector)	4 (Two packaging plasmids (*gag-pol*, and *rev* on separate plasmid), *env*, transfer vector)
Accessory Genes (*vif*, *vpr*, *vpu*, *nef*)	Present	Removed to enhance safety	Removed to enhance safety
Tat Dependency	Requires *tat* for efficient vRNA transcription	Requires *tat* for vRNA transcription	Does not require *tat* for vRNA transcription, uses modified 5' LTR with for example a CMV promoter
Safety Profile	Lowest (due to the presence of accessory genes)	Improved (due to accessory genes removed)	Highest (no accessory genes, and enhanced modifications)
Self-Inactivating (SIN) LTRs & Transgene Expression	Contains intact LTRs. Transgene expression typically achieved from LTR-derived mRNA	SIN LTRs optional. When SIN LTR present, transgene expression requires internal promoter – e.g. CMV or EF1α promoter.	SIN LTRs present. Transgene expression requires internal promoter – e.g. CMV or EF1α promoter.
Risk of Insertional Mutagenesis	Higher (due to intact LTRs)	Lower (with optional SIN LTRs)	Lowest (SIN LTRs further reduce risk)
Tropism (ability to infect various cell types)	Broad (pseudotyping possible)	Broad (pseudotyping possible)	Broad (pseudotyping possible)
Typical Uses	Early research, no longer used in clinical trials	Preclinical studies	Current standard for most clinical applications

1.2.1.2. CRISPR-Cas9 delivery using Lentiviral Vectors

In gene therapy, lentiviral vectors are widely used due to their large packaging capacity, ability to transduce both dividing and non-dividing cells, broad tropism, and reliable safety profile. Lentiviral vectors have been used to deliver corrective genes for various disease targets, including CF[98], adenosine deaminase deficiency[99], and beta thalassemia[100].

With respect to delivering CRISPR-Cas9 based gene editing system, one notable concern using LV as a delivery method that the transgene would integrate into its host genome, potentially allowing constitutive expression of Cas9 and its gRNA. However, this could be addressed by using Integrase-Deficient LVs (IDLV)[101], which carry a D64V mutation in the integrase[102,103] limiting the duration of effect in dividing cell populations. Importantly, this approach circumvents the risks from integrating Cas9 components into the genome, without sacrificing the large LV packaging capacity. This strategy of using IDLVs to deliver Cas9 has been used in studies to correct age-related macular degeneration[104] and myocilin glaucoma[105] in mice, and for *ex vivo* correction of haematopoietic stem cells[106].

1.2.2. Adeno-associated virus

Adeno-associated virus (AAV) is a small, non-enveloped, single-stranded DNA virus that is extensively used in the field of gene therapy as a delivery vector[107]. The packaging capacity of typical AAV vectors is around 4.8 kb, which is about half of that of LVs. However, unlike LVs, DNA delivered by AAVs typically remains episomal and the frequency of AAV integration into the host genome is quite low, approximately only 0.1% [108]. While the small

packaging capacity limits the use of AAV vectors for larger genes, it is sufficient for many gene therapy applications, especially when smaller genes are involved.

To overcome the size constraint, strategies such as dual AAV vector systems are being explored, allowing the delivery of larger therapeutic constructs by splitting the cargo into two parts. Multiple studies have used two AAV vectors for delivery of the components necessary for HITI and prime editing. For example, two AAV vectors have been used to deliver HITI constructs targeting the liver in a mouse model[109] and to correct full-length dystrophin expression in a humanised mouse model of Duchenne muscular dystrophy[110]; and delivery of PE constructs to target brain, liver and heart[111], correct inherited retinal degeneration[112] and correction of *Pcsk9*[113] in mouse. Another approach is to attempt to reduce the size of the genetic cargo. Due to its smaller size, SaCas9 can be more suitable than SpCas9 when using AAV as a delivery method[114]. In the case of PE, development of truncated variants of PE and the use of a split-intein system has been pursued to make the system compatible with AAV delivery[113,115].

Similar to pseudotyping LVs, different AAV serotypes have different tropisms that can be used for targeted delivery[116]. For example, AAV4, AAV5, AAV6, AAV9, AAV10, AAV12 serotypes can be used to target lung epithelia. However, as each serotype might have tropism to variety of tissues, there is a possibility of unintended delivery to other organs.

1.2.3. Non-viral delivery methods

With concerns about viral vector safety in patients, arising from genomic integration and immunogenicity associated with viral vector delivery, non-viral delivery methods serve as an alternative. Different types of non-viral delivery vectors have been developed for gene therapy approaches, such as lipid nanoparticles[117] (LNP). Notable use of LNPs was shown during the COVID-19 pandemic for delivering mRNA-based vaccines such as BNT162b2 from *Pfizer-BioNtech*[118] and mRNA-1273 from *Moderna*[119]. Delivery using LNPs is suitable especially for systemic delivery. However, when targeted delivery is required, viral-derived particles such as lentivirus-derived nanoparticles[120] (LVNP), or engineered virus-like particles[121] (eVLP) need to be considered as it is possible to include a pseudotype to add target specificity to these particles. Targeted delivery to the lungs through intravenous delivery has also been achieved using selective organ targeting (SORT) LNPs[122,123] where the charge of the nanoparticles is modified by adding different components.

1.3. Cystic Fibrosis

Cystic fibrosis is a rare autosomal recessive genetic disease that is caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene[124,125]. This disease mainly affects people with European ancestry[125], however there also have been many reports of CF affecting other races and ethnicities[126–128].

The *CFTR* gene is located on chromosome 7 and is responsible for expression of a 1480 amino acid CFTR protein. The CFTR protein consist of two transmembrane domains (TMD1 and TMD2) which forms the channel of CFTR protein, two cytosolic nucleotide binding

domains (NBD1 and NBD2) which regulates the opening and the closure of the channel, and a single regulatory R domain[129,130]. The main function of CFTR is to act as a chloride (Cl^-) and bicarbonate ion (HCO_3^-) channel, and it is found in the apical surface of epithelial cells in the airway, pancreatic duct, sweat duct, intestine, biliary tract, and vas deferens[131]. In 2018, a subset of a rare cell population in human lungs was identified as forkhead box I1 positive ($Foxi1^+$) pulmonary ionocytes, accounting for less than 1% of the epithelial cells in mouse airway and between 0.5% –1.5% in human airway[132]. Although they made up only a small portion of the cell population, ionocytes are considered responsible for the majority of CFTR expression in the lungs[132,133].

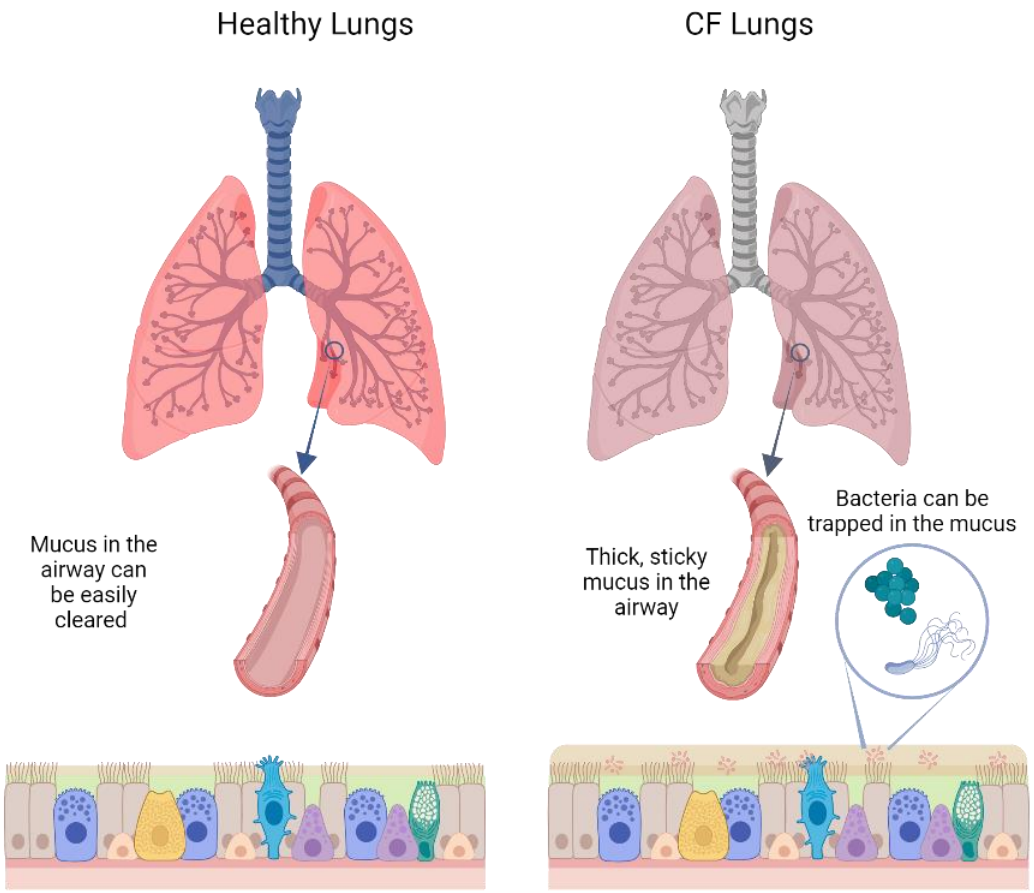

Figure 1.4. Comparison between healthy and Cystic Fibrosis lungs.

Lung airways of people with CF are covered by sticky mucus that is not easily cleared, trapping bacteria and increasing the possibility of lung infections. Image created in BioRender.com.

The mutation in *CFTR* results in production of defective CFTR protein with disrupted ion transport function, particularly chloride and bicarbonate ions, leading to dehydration of airway surface liquid[134], which in turn causes mucus buildup in respiratory, digestive, and reproductive systems. The lungs are often the most seriously affected organ, with mucus building up in the airway (Figure 1.4), impairing the mucociliary clearance and providing an ideal space for bacteria such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* to grow and therefore causing repeated infections and inflammation and ultimately respiratory failure[135].

There are over 2000 known mutations in *CFTR* (http://www.genet.sickkids.on.ca/cftr), yet not all these mutations lead to CF. The mutations have recently been categorized into seven distinct classes with different levels of severity[136] (Figure 1.5). Class I mutations prevent protein translation from the mRNA due to early stop codon, with examples including G542X and W1282X. Class II mutations affect protein trafficking into the cell membrane, with over 80% of people with CF having this mutation at least in one allele. One prominent class II mutation is the deletion of Phenylalanine in position 508 of the amino acid (F508del), affecting about 90% of people with CF globally[137]. The deletion occurs in the NBD1 domain which causes CFTR protein to misfold and is thus detected as non-functional and kept inside the endoplasmic reticulum rather than trafficked to the cell membrane [138].

Class III and IV mutations cause impairment of channel activity, with the former caused by impaired gating and the latter by decreased conductance of the channel. Class V mutations have less protein produced due to incorrect splicing, and class VI mutations produce less

stable protein. Lastly, class VII mutations result in no mRNA transcript and consequently no protein is produced, examples of this mutation class include 1717-1G→A and dele2,3(21 kb)[136]. A summary of this information is provided in Figure 1.5.

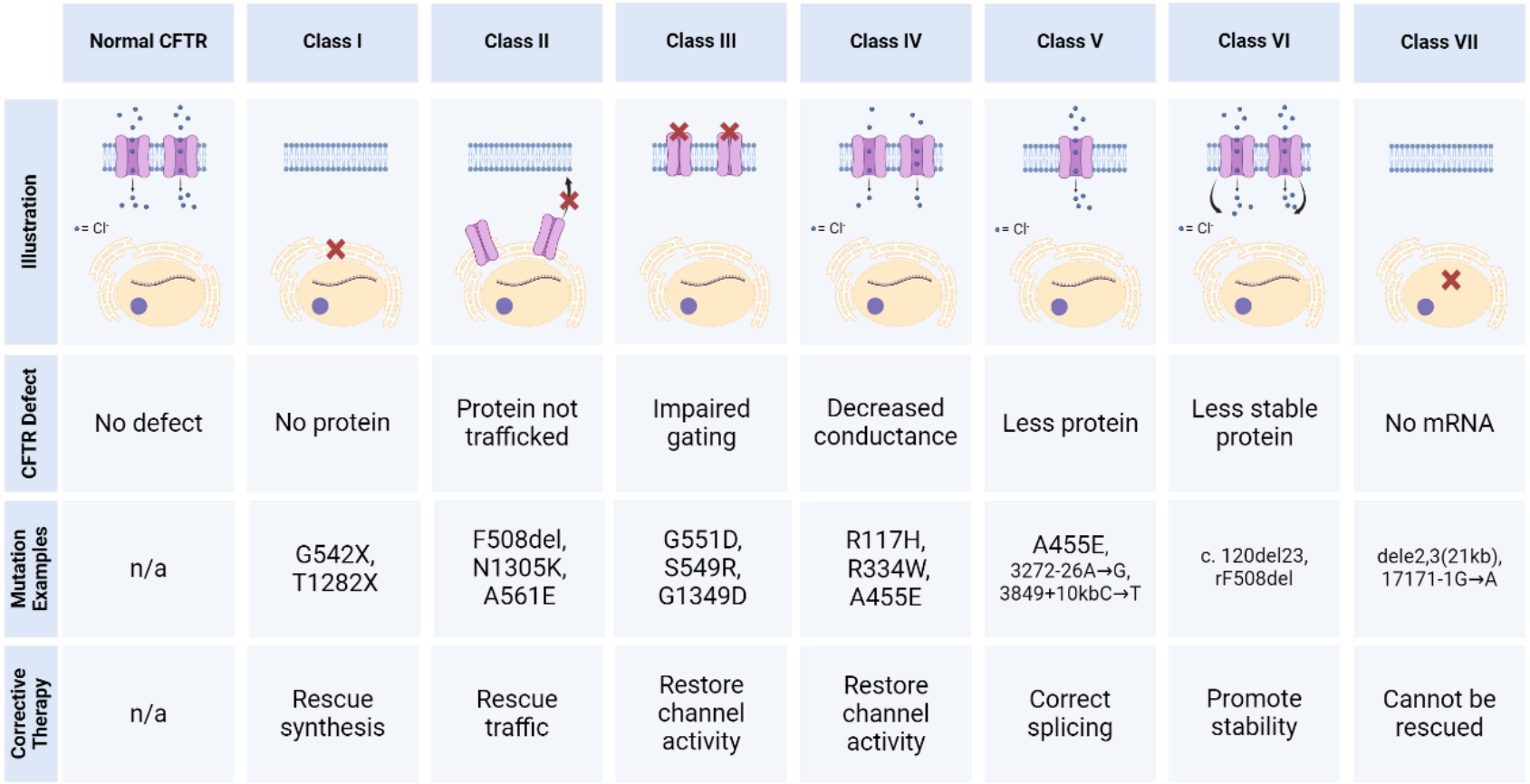

Figure 1.5. Classification of Cystic Fibrosis mutations.

Mutations in CFTR are classified into seven distinct categories, each varying in severity and its specific impact on the function of the CFTR protein.

Image created in BioRender.com.

1.3.1. Current therapies for Cystic Fibrosis

Current therapy for people with CF includes the use of small molecule CF modulators, designed to correct or to modify dysfunctional CF protein. The modulators include 'potentiators' or 'activators', such as ivacaftor; and 'correctors', such as lumacaftor, tezacaftor, and elexacaftor. However, treatment with these CF modulators is only prescribed to patients with specific CFTR mutations. Drugs containing ivacaftor (Kalydeco®) have been approved for patients harbouring the G551D mutation in at least one allele, and also 96 other mutations on the list (https://www.kalydeco.com/who-kalydeco). Lumacaftor/ivacaftor combination (Orkambi®) has been approved for patients with both alleles having the F508del mutation, while tezacaftor/ivacaftor combination (Symdeko®) have been approved for patients homozygous for the F508del mutation, plus 154 other mutations listed on the manufacturers' website (https://www.symdeko.com/who-symdeko-is-for). Lastly, the combination of elexacaftor/tezacaftor/ivacaftor (Trikafta®) has been approved for patients with at least one allele having the F508del mutation[139], plus 177 other approved mutations (https://www.trikafta.com/who-trikafta-is-for). Although these mutation lists are extensive, they are not exhaustive as there are mutations which cannot be treated with these small molecule modulators.

These drugs have been life-changing for many people with CF by increasing FEV_1 and reducing pulmonary exacerbations[139–141], improving the life quality and expectancy for patients. However, they do not address the underlying problem which is the actual genetic mutation. Also, the small molecule therapies do not help a small subset of patients with mutations outside of the approved list, especially people with Class I and Class VII CF

mutations where there is no CFTR protein to be rescued. In addition to potential ineligibility for these modulator drugs, there are also some people with CF who are unable to tolerate the side-effects[142]. These highlight the unmet clinical need for this subset of people with CF, where gene editing and gene therapy strategy could be employed. However, all CF patients could ultimately benefit from an effective gene therapy regardless of the types of mutations that they have but, so far, there is no approved gene therapy or gene editing CF treatment available.

Restoration of about 10%-35% of functional CFTR protein activity is expected to lead to clinically meaningful improvements, particularly in lung function and ion transport [143,144]. This level of correction can help move chloride ions effectively, improving the hydration of mucus in the airways. This sets the baseline for the required correction level in the genome to bring therapeutic benefits to patients.

1.3.2. Gene Therapy for cystic fibrosis

People with CF who still do not benefit from available modulator therapies, include those with Class I and Class VII mutations. For this patient population, which is estimated to be around 10% of total patients[131,145], gene therapy would be one feasible option to treat their condition. Even for other patients, gene therapy offers the possibility of a mutation-agnostic, 'one-and-done' treatment so that affected individuals would no longer depend on taking medications for their whole life. There are several gene therapy approaches for CF, including gene addition therapy, and gene editing.

Gene addition therapy works by supplementing a functional copy of the defective gene, (*CFTR*) to the lungs of the patients. The supplemented *CFTR* can be in the form of mRNA or complementary DNA (cDNA). Various approaches to deliver the CFTR transgene, including using non-viral particles, AAV, and LV have been described[146]. Clinical trials using AAV have proven safe for delivering CFTR-cDNA with AAV2[147], but the clinical trial in 2007 did not show improved lung function (measured by forced expiratory volume (FEV_1)) 30 days after administration of the CFTR cDNA, compared with placebo[148]. However, *4D Therapeutics* recently reported positive interim results showing improvements in FEV_1 from two out of three patients in a phase I/II trial of 4D-710 AAV-based gene therapy[149]. Preclinical studies have also used a mutated version of AAV2 (termed AAV2H22) that has been shown to correct anion transport and restore bacteria killing in pig models[150], indicating a promising future for AAV-based gene therapy for CF.

On the LNP delivery approach, a phase 1/2 trial from *Translate Bio* (Sanofi) titled RESTORE-CF[151,152] uses a codon-optimized CFTR mRNA (MRT5005) delivered by aerosolised LNPs to the lung for mutation-agnostic treatment of CF. The trial establishes the safety of the product, however no improvement in FEV_1 was measured. In 2024, *ReCode Therapeutics* started a phase 1b clinical trial for product RCT2100, an inhaled mRNA therapy delivered using proprietary SORT-LNPs[153]. So far, no result has been announced for this trial.

A lentiviral vector approach for delivery of CFTR cDNA has been reported, that uses F/HN pseudotyped SIV (rSIV.F/HN) with a synthetic promoter that is a cytosine guanine dinucleotide (CpG)-free, hybrid of the human CMV enhancer and the EF1α (hCEF) promoter

driving a CpG-free CFTR cDNA. A first-in-human clinical trial using this vector is anticipated to commence in 2024[145]. Preclinical data so far have shown promising results where around 14% airway cell transduction was achieved that was not affected by pre-existing and acquired immune responses[98]. The vector was also third-generation (SIN and replication-incompetent, Table 1.2) further mitigating safety risks associated with LVs. In an air-liquid interface (ALI) model, rSIV.F/HN vector restored CFTR-mediated chloride currents to physiological levels in primary human bronchial-epithelial cells homozygous for F508del, as well as in cell models with class I mutation of *CFTR*, further supporting this technology to progress to clinical trial.

1.4. Prime editing for cystic fibrosis

In this thesis, my approach to tackle the root cause of this problem involves using prime editing to correct the F508del mutation. Repairing the F508del mutation using prime editing is challenging for multiple reasons. The first reason is that the efficiency of prime editing differs depending on the type of edit performed, with insertions being the most difficult, followed by deletions, and then substitutions[154]. Secondly, one of the most important factors for successful editing with prime editing is the GC content of the PBS[61,154], which is tied directly to the target sequence. As GC content upstream and downstream of the F508del mutation is less than 40%, it was anticipated that a robust and high throughput method was required to screen multiple pegRNAs to find the most optimal for correction of F508del mutation in *CFTR*.

In a patient-derived intestinal organoid model, correction of the F508del mutation using PE was much lower than with the traditional HDR correction method; the HDR-mediated editing resulted in functional repair of more than 100 patient-derived organoids, whereas PE-mediated editing achieved functional repair of only four organoids[155], further highlighting the difficulties in targeting this region. The latest effort in using PE to correct the F508del mutation showed that combining epegRNA, incorporating silent edits, using PE6c, and supplementing catalytically 'dead' sgRNA[156,157] to increase the accessibility of the target site, all boosted the editing efficiency up to 28-fold when compared with using just PEmax and pegRNA in the human bronchial epithelial cell line model (16HBE14o-)[158]. In conclusion, significant optimisation may be necessary before prime editing is ready for clinical use for CF.

1.5. Aims

Prime editing is a gene editing tool with high potential to correct various disease-causing mutations, including the F508del mutation in *CFTR*. However, the size of PE is much larger than the regular CRISPR-Cas9 system, and so requires a different delivery approach. The aims of this thesis are:

1. To develop an 'all-in-one' lentiviral vector for delivery of PE and to validate the approach by inserting CTT bases into the *HEK3* locus (Chapter 3)
2. To develop a 'two-vector' delivery for PE, for validation at the *HEK3* locus, and to explore two-vector delivery for HITI as an alternative, testing the approach in *CFTR*. (Chapter 4)

3. To generate a pegRNA mini library for high throughput screening of pegRNA targeting F508del mutation in *CFTR*. (Chapter 5)

Chapter 2: Materials and Methods

2.1. Common Reagents and Solutions

All nucleic acid preparations were dissolved in nuclease free water (NFW) (Ambion®, Thermo Fisher Scientific), and all buffers were made in Milli-Q reverse osmosis ultrapure water (Merck Millipore). Where necessary, solutions were sterilised by autoclaving for a minimum of 20 minutes at 121°C or by filtration method using 0.22 µm surfactant-free cellulose acetate filter (Thermo Fisher Scientific). Where concentrations are stated as percentages [%], these refer to weight of solids per volume (w/V) ratios at room temperature unless otherwise specified.

Lysogeny broth (LB) used for liquid culture of bacteria was prepared using 1% NaCl (Sigma Aldrich), 1% Tryptone (Merck), and 0.5% Yeast Extract Granules (VWR) in Milli-Q water. LB-Agar was made by adding 1.5% Select Agar (Invitrogen, Fisher Scientific) to LB. Both LB and LB-Agar were sterilized by autoclaving. Kanamycin (50 mg/mL, filter sterilised and stored as one-shot aliquots at -20°C) (Sigma) or Ampicillin (100 mg/mL, filter sterilised and stored as one-shot aliquots at -20°C) (Sigma) was added to a final concentration of 50 µg/mL or 100 µg/mL respectively to LB or LB-Agar before use.

2.2. Cell culture

Adherent Human Embryonic Kidney 293T/17 (from here onwards referred to HEK293T/17) (ATCC®, CRL-11268) were maintained in High glucose Dulbecco's Modified Eagle's Medium (DMEM) (Gibco™, Fisher Scientific) containing 10% foetal bovine serum (FBS) (Scientific

Laboratory Supplies), 100 U/mL penicillin and 100 µg/mL streptomycin (PenStrep) (Gibco™). CFTR-cDNA expressing HEK293T cells (obtained under MTA from Carlon group, KU Leuven, Belgium) were maintained in the same medium as above. PEmax-EGFP expressing HEK293T cells generated in this study were also maintained in the same medium as above. Cells were cultured in Falcon Tissue Culture Flask with vented cap (Scientific Laboratory Supplies) at 5% CO_2, 37°C and maintained at 50%-70% confluency with regular passaging every 3 to 4 days.

Suspension HEK293T/17 SF cells were adapted to suspension growth with sequential serum reduction in the growth medium (ATCC®, ACS-4500) and were maintained in Erlenmeyer Shake Flask (Corning) with FreeStyle™ 293 Expression Medium (Gibco™) without any serum or antibiotics, in an 8% CO_2 humidified agitating incubator (125 RPM). Cell growth and count was determined using 0.2% Trypan Blue Dye (Sigma) and a Neubauer Improved Haemocytometer. Cells were passaged twice a week by diluting to 3.5×10^5E5 live cells per mL media.

Adherent 16HBE14o- cells harbouring the F508del mutation, with either M470 or V470 variant in *CFTR* (obtained under MTA from Dieter Grünert, UCSF, USA) were maintained in Minimum Essential Medium (Gibco™) containing 10% FBS (Gibco™) and 100 U/mL penicillin and 100 µg/mL streptomycin. Cells were cultured in Falcon Tissue Culture Flask with vented cap (Scientific Laboratory Supplies), coated with LHC-8 basal medium containing 7.5% bovine serum albumin (BSA) (Gibco™), 1% bovine collagen solution type I (Advanced BioMatrix), and 1% fibronectin from human plasma (Thermo Fisher Scientific);

at 5% CO_2, 37°C and maintained at 50%-70% confluency with regular passaging every 3 to 4 days. All the cell lines used in this study are listed in Table 2.1.

Table 2.1. Cell lines used in this study

No	Cell Line	Type	Transgene	Source
1	HEK293T/17	Adherent	n/a	ATCC®, CRL-11268
2	16HBE14o-, *CFTR* F508del M470	Adherent	n/a	Obtained under MTA from Dieter Grünert (UCSF, USA)
3	16HBE14o-, *CFTR* F508del V470	Adherent	n/a	Obtained under MTA from Dieter Grünert (UCSF, USA)
4	CFTR-cDNA HEK293T	Adherent	CFTR cDNA	Obtained under MTA from Carlon group (KU Leuven, Belgium)
5	HEK293T-PEmax-P2A-EGFP	Adherent	EF1a.PEmax.EGFP.W3	This study
6	HEK293T/17 SF	Suspension	n/a	ATCC®, ACS-4500

2.3. Microscopy

All cell images from plates and/or flasks were taken using EVOS FL Auto 2 (EVOS, Thermo Fisher Scientific) widefield fluorescence microscope with 4X magnification, unless stated otherwise. Images were then processed using Fiji image processing software (ImageJ) when necessary.

2.4. Molecular cloning techniques

2.4.1. Polymerase chain reaction

All polymerase chain reaction (PCR) was performed using Phusion® High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs, NEB) according to manufacturer's protocol, with reaction volume scaled up or down as necessary. Primers were designed manually using SnapGene or using Primer3 (https://primer3.ut.ee/). Applied Biosystems™ ProFlex™ PCR System (Thermo Fisher Scientific) was used for all thermal cycling unless specified otherwise. For amplification of GDNA, 50 ng-200 ng template was used depending on the reaction scale (for example, 100 ng of DNA in 50 µL reaction). Where PCR product was sent for sequencing, purification steps were performed as described in section 2.4.12.

2.4.2. Site Directed Mutagenesis

To produce pGM935 and pGM936, suitable restriction sites were added to pGM899 (pCMV.PE1), pGM900 (pCMV.PE2), and pGM357 (pSIV.hCEF.EGFP) plasmid for cloning purposes, using Q5 Site-Directed Mutagenesis Kit (NEB). Primers used are shown in Table 2.2.

Table 2.2. Primer sequences for Site Directed Mutagenesis

Name	Sequence	Length	TM (°C)
PE2 NotI Insert F	ccgcGGTCATCATCACCATCAC	22-mer	55
PE2 NotI Insert R	ccgcGGTTAGACTTTCCTCTTC	22-mer	52
pGM357 NotI Insert F	ggccgCACCATGGTGAGCAAGGG	23-mer	58
pGM357 NotI Insert R	GCTAGCAAACTCACAGGAG	19-mer	54

PCR was performed with 25 ng plasmid DNA template, 0.5 µM each forward and reverse primer, 1× final concentration of Q5 Hot Start High Fidelity 2× Master Mix (NEB), and NFW to final volume of 25 µL. PCR was performed according to the manufacturer's protocol (E0554), however, the annealing temperature of the primers was set to 58°C, with 5 minutes extension time. Kinase, ligase, and DpnI treatment was performed according to manufacturer's protocol. Five microliters of the final SDM PCRs were transformed using One Shot™ Stbl3™ Chemically Competent *E. coli* (Invitrogen) according to section 2.4.13.

2.4.3. Blunt TOPO cloning

Blunt TOPO cloning technique was used to ligate all DNA fragments synthesised and ordered from GeneArt (Thermo Fisher Scientific) used in this study and amplify the fragments through bacterial liquid culture according to section 2.4.14. Zero Blunt™ TOPO™ PCR Cloning Kit (Invitrogen) was used according to manufacturer's protocol. For larger fragments, an incubation time of 16 hours at 4°C were used to increase the chance of fragment insertion into the TOPO backbone. The ligation product was processed further as described in section 2.4.14.

2.4.4. HITI Donor template design

HITI donor DNA was designed for HITI gRNA spacer 1 and 3 by adapting the design from Bednarski[159] by including the gRNA target sequence and PAM, CFTR exon 11-27 and intein splicing signal. Complete sequences are available in the Appendix section. Both donor constructs were synthesized and ordered from GeneArt (Thermo Fisher Scientific) and TOPO cloned for bacterial amplification according to section 2.4.3. HITI donor templates were then cloned into the plasmids carrying the hU6 promoter and its respective gRNA (see Table 2.3 in section 2.4.6).

2.4.5. Split PE Design

Split PE plasmids were designed based on a paper by Liu et al.[160]. Firstly, Cas9 nickase was split into residue 1-713 corresponding to the N-terminus, and 714-1368 corresponding to the C-terminus. Then, both fragments were then linked to intein fragments so that the full-length protein could assemble after translation of each component. The Split PE fragments were ordered as GeneArt constructs (Thermo Fisher Scientific) and cloned into an HIV genome plasmid backbone with CMV promoter (pGM1105 and pGM1106 respectively for N-terminus and C-terminus construct). The pegRNA that encompasses the template for CTT insertion in *HEK3* site, driven by hU6 promoter, was also ordered as a Genestring construct with appropriate restriction sites, to be cloned into the HIV genome plasmid expressing the N-terminus of split PE (pGM1151).

2.4.6. Guide RNA cloning for HITI and Split PE plasmids

Guide RNAs targeting the F508del mutation in CFTR were designed based on F508del-containing CFTR sequence as the reference. The gRNAs were ordered as oligo fragments from Sigma Aldrich (Table 2.5). Plasmid phU6-gRNA (Table 2.7) was used as the backbone for all the gRNA constructs. The plasmid backbone was digested using BbsI, and small fragments and residual buffers were subsequently removed using the QIAGEN PCR purification kit. Oligos were re-suspended in NFW to a concentration of 100 µM. The forward and reverse oligos were then annealed in a PCR tube containing 1 µL each of forward and reverse oligos (100 µM), 1 µL of 10× T4 ligase buffer, and 7 µL of NFW. The mixture was incubated at 95°C for 5 minutes, followed by a ramp down to 25°C at a rate of 5°C per minute. The annealed oligos were ligated into the digested plasmids by incubating a mixture containing 1 µL of annealed oligo (10 µM), 100 ng of the digested backbone, 1 µL of 10× T4 ligase buffer (NEB), 1 µL of T4 ligase (NEB), and NFW was added up to a total volume of 10 µL. The reaction was incubated overnight at 4°C.

The ligation mixture was then transformed into One Shot™ Stbl3™ Chemically Competent *E. coli* (Invitrogen) according to section 2.4.13, then plated onto Kanamycin selection plates. Integration of the oligos was confirmed with plasmid sequencing (section 2.9) using M13 forward primer (5'GTAAAACGACGGCCAGT3') (primer was supplied and sequencing service was done by Source Biosciences). For HITI, all spacers were included in the screening except for HITI 8 to HITI 12 (see Table 2.3) due to a common variant in the 470 amino acid residue of *CFTR* (M470V) which might alter the binding efficiency of the spacers.

Table 2.3. Cas9 spacers for HITI and PE

Spacers	Sequence	Length (nt)
HITI 1	Gaagaatatacacttctgctt	21
HITI 2	Gcttctgcttaggatgataat	21
HITI 3	Gctgcttaggatgataattgg	21
HITI 4	gtcattatcaaatcacgctc	20
HITI 5	Gtgataatgacctaataatga	21
HITI 6	gataatgacctaataatgat	20
HITI 7	Gaaataaaacccatcattatt	21
HITI 8	Gtttccagacttcacttctaa	21
HITI 9	Gtcaccattagaagtgaagtc	21
HITI 10	Gttcacttctaatggtgatta	21
HITI 11	Gtcacttctaatggtgattat	21
HITI 12	Gaatggtgattatgggagaac	21
HITI 13	Gtgcttaattttaccctctga	21
HITI 14	gagggtaaaattaagcacag	20
HITI 15	gggagaactggagccttcag	20
HITI 16	ggagaactggagccttcaga	20
PE 1	Gaccattaaagaaaatatcat	21
PE 2	Gtctgtatctatattcatcat	21
PE 3	Gatattttctttaatggtgcc	21
PE 4	Gcagttttcctggattatgcc	21
PE nick 2	Gaatggtgccaggcataatcc	21
PE nick 5	Gattaaagaaaatatcatctt	21

2.4.7. Golden Gate pegRNA cloning

Cloning for pegRNA was performed as per the protocol for cloning 3'-extended pegRNAs into mammalian U6 expression vectors by Golden Gate assembly[56] Briefly, pegRNA acceptor plasmid was digested with BsaI and using gel purification, the 2.2 kb fragment containing the origin of replication, hU6 promoter, poly T sequence, and Ampicillin resistance gene was isolated. The pegRNA sequence was ordered as separate oligos (Table 2.6) from Merck (Sigma Aldrich) consisting of two fragment pairs (top and bottom strand) for spacer, scaffold, and 3' extension sequences. For all spacers, on top of the desired target 5'-3' sequence, 5' CACC and 3' GTTTT overhangs were added to the top strand, while the bottom strand contains the reverse complement of the target sequence and 5' CTCTAAAAC

overhang. The scaffold oligos was ordered phosphorylated with 5' AGAG overhang on the top strand and 5' GCAC overhang on the bottom strand. The 3' extension region was ordered with 5' GTGC overhang and 5' AAAA overhang on the bottom strand. For each oligo pair, the top and bottom oligos were was annealed in separate reaction and then the Golden Gate assembly was performed by combining 1 µL (30 ng/µL) purified and digested pegRNA acceptor plasmid backbone with 1 µL each of the three annealed pegRNA components (1 µM), BsaI enzyme (NEB, 0.25 µL), T4 DNA ligase (NEB, 0.5 µL), 10× T4 DNA ligase buffer (NEB, 1 µL), and H_2O (4.25 µL). The reaction was then incubated in a thermocycler at 16°C for 5 minutes and at 37°C for 5 minutes for 8 cycles. The reaction was further incubated at 37°C for 15 minutes, then 80°C for 15 minutes, and finally held at 12°C until the transformation step. The resulting plasmids (pegRNA plasmids, Table 2.7) were then transformed and processed further as described in sections 2.4.13, 2.4.14, and 2.4.15. Integration of correct pegRNA sequence was confirmed by sanger sequencing (section 2.9) using phU6 forward primer (5'GAGGGCCTATTTCCCATGATT3').

2.4.8. Restriction enzyme digest

All restriction digests were performed using enzymes purchased from NEB. Where two or more enzymes were used, the NEBcloner website (https://nebcloner.neb.com/#!/redigest) was used to determine optimal conditions for the digestion reaction. Generally, for 1 µg of template DNA, 1 hour of incubation time at the recommended temperature was used with total enzyme not exceeding 10% of the final volume. For diagnostic digests, 100 ng of template was used instead with total enzyme not exceeding 10% of the final volume. Where the digested plasmid was to be used as a backbone for cloning, recombinant Shrimp Alkaline Phospatase (rSAP, NEB) was added directly after the incubation step for digestion was finished, with 1 unit of rSAP used per pmol of DNA ends (calculated using

https://nebiocalculator.neb.com/#!/dsdnaends). The reaction was further incubated at 37°C for one hour, and then at 65°C for 5 minutes to stop the dephosphorylation. The digested plasmid was then electrophoresed through agarose and when necessary the required fragment was extracted from the gel as described in section Gel extraction of DNA fragments

2.4.9. Ligation of DNA fragments

Ligation reaction was performed using T4 DNA Ligase (NEB). The reaction mixture consisted of the DNA vector and insert in a 1:3 molar ratio, 1× T4 DNA ligation buffer (NEB), and T4 DNA ligase. Nuclease-free water was added to adjust the final reaction volume to 20 µL. The reaction was incubated at 16°C overnight or at room temperature for 1 hour to facilitate ligation. Following incubation, the ligation product was used directly for transformation into competent cells as described in section 2.4.13.

2.4.10. Agarose Gel Electrophoresis

Separation of DNA fragments based on molecular weight was performed by agarose gel electrophoresis. Unless stated otherwise, gel for electrophoresis was made by adding 1% w/V agarose (Life Technologies) to 1× Tris-acetate-ethylenediaminetetraacetic acid (TAE) buffer (VWR), and then heating it to dissolve the Agarose powder, and after it has cooled down slightly adding 1× SYBR Safe gel stain diluted to working concentration (1:10000) (Thermo Fisher Scientific) and pouring the gel into the mould set. DNA samples (PCR product or plasmid) were mixed with 6× Purple Loading Dye (NEB) to a final 1× concentration and electrophoresed at 12V/cm for 45 minutes to 1 hour. Samples were run

against a DNA molecular weight ladder (NEB, 1 kb or 100bp as appropriate), to determine the approximate molecular weights of sample DNA amplicons/fragments. Gels were imaged at 302 nm using iBright FL1500 imager (Thermo Fisher Scientific).

2.4.11. Gel extraction of DNA fragments

DNA fragments from gel electrophoresis were visualised using the iBright FL1500 imager (Thermo Fisher Scientific), and the expected DNA fragments were excised from the gel using a scalpel and placed into a clean 2 mL tube (Eppendorf). The gel fragments were then weighed, and the DNA extracted from the gel using QIAquick Gel Extraction Kit (QIAGEN) according to manufacturer's protocol and DNA eluted using NFW at the final step.

2.4.12. Purification of PCR Products

Where appropriate, PCR amplicons were purified directly using QIAquick PCR Purification Kit (QIAGEN), following the manufacturer's protocol, with the final elution performed in 30 µL of NFW. Where a higher final concentration of the PCR product was desired, the MinElute PCR Purification Kit (QIAGEN) was used instead, and the purified PCR product was eluted in the final step with 10 µL of NFW.

2.4.13. Bacterial transformation of plasmid DNA

Unless indicated otherwise, One Shot™ Stbl3™ Chemically Competent *E. coli* (Invitrogen) was used for all plasmid transformations according to the manufacturer's protocol, with the exception that one vial containing 50 µL of *E. coli* was split into two tubes of 25 µL each.

Approximately 5 µL of the ligation product, or 1 ng of plasmid DNA, was added to 25 µL of competent cells. The tubes were then gently flicked to mix the DNA and the competent cells, followed by incubation for 30 minutes on ice. Subsequently, the cells were heat-shocked in a water bath at 42°C for 45 seconds and then placed back on ice for 2 minutes. SOC medium (Invitrogen; 700 µL) was added to the competent cells, and the mixture was placed in a shaking incubator at 225 RPM for 1 hour at 37°C. Then, 100 µL of the transformed cells were spread with a sterile L-shaped spreader on 10-cm petri dishes containing LB-Agar with either Kanamycin (50 µg/mL) or Ampicillin (100 µg/mL), where appropriate, and incubated in an inverted position in a static incubator at 37°C overnight. The remaining transformation reaction was kept at 4°C in case the colonies formed on the plates were too crowded, and further dilution was needed. Single colony transformants were selected and processed as per section 2.4.14.

2.4.14. Plasmid DNA amplification in liquid bacterial cultures

Single colonies from the transformation agar plates were picked with a sterile inoculating loop, streaked on a new agar plate with the appropriate antibiotic to make a "copy", and also used to inoculate 5 mL of LB medium containing either Kanamycin (50 µg/mL) or Ampicillin (100 µg/mL), for what will be referred to as the 'Miniprep' process,. The cultures were then placed in a shaking incubator (225 RPM overnight at 37°C). When further amplification of plasmid DNA was required, 1:500 to 1:1000 dilution of the Miniprep culture was added to a final volume of 200 mL of LB medium (for 'Maxiprep') or 400 mL (for 'Megaprep'), with appropriate antibiotics.

2.4.15. Plasmid DNA verification

Plasmid DNA was extracted from the cells according to section 2.4.16. The resulting plasmid DNA was then subjected to diagnostic restriction digest (section 2.4.8) and agarose gel electrophoresis (section 2.4.9) for molecular weight evaluation to confirm the correct plasmid size, and where appropriate, the region of interest was sequenced using Sanger Sequencing (Source Bioscience).

2.4.16. Purification of plasmid DNA from liquid bacterial cultures

Plasmid DNA from Miniprep culture was extracted from the cells using the Wizard plus SV Miniprep DNA purification system (Promega) according to the manufacturer's protocol. The resulting plasmid DNA was then subjected to diagnostic restriction digest and agarose gel band size evaluation to confirm the correct plasmid size, and where appropriate, the region of interest was sequenced using Sanger Sequencing (Source Bioscience). When further amplification of plasmid DNA was required, 1:500 to 1:1000 dilution of the Miniprep culture was added to a final volume of 200 mL of LB medium (for 'Maxiprep') or 400 mL (for 'Megaprep'), with appropriate antibiotics.

Plasmid DNA from Maxiprep and Megaprep cultures (section 2.4.14) was processed using the appropriate EndoFree® Plasmid Maxi or Mega Kit (QIAGEN), according to the manufacturer's protocol, with exceptions made for the centrifugation steps. To collect the bacterial pellet from the liquid culture, 500 mL of the culture was centrifuged at 6,000*g* for 15 minutes and then resuspended in P1 buffer as outlined in the protocol. To isolate the plasmid DNA, the isopropanol-DNA mixture was centrifuged at 15,000*g* for 30 minutes.

After completely removing the isopropanol, endotoxin-free ethanol was added to the DNA pellet, and the ethanol-DNA mixture was centrifuged again at 15,000*g* for 15 minutes. Subsequently, the ethanol was discarded, and the DNA pellet was allowed to dry for about 15 minutes before being resuspended in an appropriate volume of NFW (Promega). The concentration of plasmid DNA was measured using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific).

2.4.17. Lipofectamine plasmid transfection

Unless otherwise specified, all transfections were performed using Lipofectamine 3000 (Invitrogen) following the manufacturer's protocol. Briefly, adherent cells were seeded onto plates (ranging from 24-well to 6-well) the day before transfection and allowed to grow overnight to achieve 60%-70% confluence by the day of transfection. For transfections in a 24-well plate, 1.5×10^5 cells were seeded the day before and transfected with up to 1.8 µg of plasmid DNA; the total volume for the transfection reaction was 50 µL. The protocol was scaled up accordingly for different plate sizes. For all transfections, a ratio of 1 µg of plasmid DNA to 1 µL of Lipofectamine 3000 to 2 µL of P3000 supplement was used. The mixture of plasmid DNA and P3000 supplement with Lipofectamine 3000 was incubated for 15 minutes before being added directly to the cells.

2.4.18. DNA Quantification

Quantification of DNA was performed using Nanodrop Spectophotometer 2000 (Thermo Fisher Scientific) except for Next Generation Sequencing samples where Qubit® Flex Fluorometer (Thermo Fisher Scientific) was used. Approximately 1.5 µL of sample was used

for determining the concentration of DNA using NanoDrop Spectrophotometer 2000. The purity of the preparation was measured by calculation of the OD260/OD280 ratio, and values of < 1.8 was considered to be indicative of protein contamination.

The Qubit® Flex Fluorometer and Qubit® dsDNA HS Assay Kits were used when low DNA concentrations were expected, following the manufacturer's protocol. In brief, DNA samples were diluted with varying dilution factors (e.g., 3×, 10×, 30×). The Qubit working solution was prepared by diluting the assay reagent in a 1:200 ratio with the dilution buffer. Then, 10 µL of each standard (S1 and S2) was added to 190 µL of the working solution (creating a 20× dilution of S1 and S2), followed by vortexing to mix. Next, 2 µL of each sample was added to 198 µL of the working solution and vortexed to mix. All reactions were incubated for 2 minutes before being measured using the Qubit® Flex Fluorometer.

2.5. Primers, oligos, and plasmids

Primers, oligos, and plasmids used in this study are listed in the tables below:

Table 2.4. Primers for cloning; and PCR and sequencing of *HEK3* and *CFTR*

No	Purpose	Name	Forward Sequence (5'-3')	Reverse sequence (5'-3')
Cloning primers				
1	Adding BglII site to pU6.Sp.pegRNA.HEK3_CTT_in	BglII-pU6.Sp.pegRNA.HEK3_CTT_in	AGATCTgacgtcgctagctgtaca	AGATCTagagaggtacctcgagcg
2	Adding NotI site to PE2	PE2-NotI Insert F & R	ccgcGGTCATCATCACCATCAC	ccgcGGTTAGACTTTCCTCTTC
3	Adding NotI site to pGM357	pGM357 NotI Insert F & R	ggccgCACCATGGTGAGCAAGGG	GCTAGCAAACTCACAGGAG
PCR Primers				
4	*HEK3* site amplification	HEK3 CTT PCR F & R	agtgctggagaatgggtc	catgcaggtgctgaaagc
5	*CFTR* F508del amplification	CFTR F508 PCR F & R	ccttctctgtgaacctctat	ccttctctgtgaacctctat
6	HITI integration	Geurts F&R	tggaggcaagtgaatcctgagcgt	tctgctggcagatcaatgctca
Sanger sequencing primers				
7	*HEK3*	HEK3 CTT sequencing F & R	atgtgggctgcctagaaagg	cccagccaaacttgtcaacc
8	*CFTR* F508del	CFTR F508 sequencing F & R	gtacctgaaacaggaagtat	gtacctgaaacaggaagtat

Table 2.5. Oligos for HITI and PE guide RNA

Name	Sequence (5' - 3')
HITI 1 pos	caccGaagaatatacacttctgctt
HITI 1 neg	aaacaagcagaagtgtatattcttC
HITI 2 pos	caccGcttctgcttaggatgataat
HITI 2 neg	aaacattatcatcctaagcagaagC
HITI 3 pos	caccGctgcttaggatgataattgg
HITI 3 neg	aaacccaattatcatcctaagcagC
HITI 4 pos	caccgtcattatcaaatcacgctc
HITI 4 neg	aaacgagcgtgatttgataatgac
HITI 5 pos	caccGtgataatgacctaataatga
HITI 5 neg	aaactcattattaggtcattatcaC
HITI 6 pos	caccgataatgacctaataatgat
HITI 6 neg	aaacatcattattaggtcattatc
HITI 7 pos	caccGaaataaaacccatcattatt
HITI 7 neg	aaacaataatgatgggttttatttC
HITI 8 pos	caccGtttccagacttcacttctaa
HITI 8 neg	aaacttagaagtgaagtctggaaaC
HITI 9 pos	caccGtcaccattagaagtgaagtc
HITI 9 neg	aaacgacttcacttctaatggtgaC
HITI 10 pos	caccGttcacttctaatggtgatta
HITI 10 neg	aaactaatcaccattagaagtgaaC
HITI 11 pos	caccGtcacttctaatggtgattat
HITI 11 neg	aaacataatcaccattagaagtgaC
HITI 12 pos	caccGaatggtgattatgggagaac
HITI 12 neg	aaacgttctcccataatcaccattC
HITI 13 pos	caccGtgcttaattttaccctctga
HITI 13 neg	aaactcagagggtaaaattaagcaC
HITI 14 pos	caccgagggtaaaattaagcacag
HITI 14 neg	aaacctgtgcttaattttaccctc
HITI 15 pos	caccgggagaactggagccttcag
HITI 15 neg	aaacctgaaggctccagttctccc
HITI 16 pos	caccggagaactggagccttcaga
HITI 16 neg	aaactctgaaggctccagttctcc
PE 1 pos	caccGaccattaaagaaaatatcat
PE 1 neg	aaacatgatattttctttaatggtC
PE 2 pos	caccGtctgtatctatattcatcat
PE 2 neg	aaacatgatgaatatagatacagaC
PE nick 2 pos	caccGaatggtgccaggcataatcc
PE nick 2 neg	aaacggattatgcctggcaccattC
PE nick 5 pos	caccGattaaagaaaatatcatctt
PE nick 5 neg	aaacaagatgatattttctttaatC

Table 2.6. Oligos for CFTR pegRNA screening

No	Name	5' overhang	Spacer sequence
1	pegRNA 2-15-21 pos	gtgc	tcatctttggtgtttcgtatgatgaatatagataca
	pegRNA 2-15-21 neg	aaaa	tgtatctatattcatcatacgaaacaccaaagatga
2	pegRNA 2-15-20 pos	gtgc	catctttggtgtttcgtatgatgaatatagataca
	pegRNA 2-15-20 neg	aaaa	tgtatctatattcatcatacgaaacaccaaagatg
3	pegRNA 2-15-19 pos	gtgc	atctttggtgtttcgtatgatgaatatagataca
	pegRNA 2-15-19 neg	aaaa	tgtatctatattcatcatacgaaacaccaaagat
4	pegRNA 2-14-27 pos	gtgc	aaaatatcatctttggtgtttcgtatgatgaatatagatac
	pegRNA 2-14-27 neg	aaaa	gtatctatattcatcatacgaaacaccaaagatgatatttt
5	pegRNA 2-13-22 pos	gtgc	atcatctttggtgtttcgtatgatgaatatagata
	pegRNA 2-13-22 neg	aaaa	tatctatattcatcatacgaaacaccaaagatgat
6	pegRNA 2-13-21 pos	gtgc	tcatctttggtgtttcgtatgatgaatatagata
	pegRNA 2-13-21 neg	aaaa	tatctatattcatcatacgaaacaccaaagatga
7	pegRNA 2-13-20 pos	gtgc	catctttggtgtttcgtatgatgaatatagata
	pegRNA 2-13-20 neg	aaaa	tatctatattcatcatacgaaacaccaaagatg
8	pegRNA 2-13-19 pos	gtgc	atctttggtgtttcgtatgatgaatatagata
	pegRNA 2-13-19 neg	aaaa	tatctatattcatcatacgaaacaccaaagat

Table 2.7. Plasmids generated and used in this study

No	Plasmid	Promoter	Transgene	Source	Internal code
	Prime Editing Plasmids				
1	pCMV.PE1	CMV	PE1	Addgene #132774	pGM899
2	pCMV.PE2	CMV	PE2	Addgene #132775	pGM900
3	pU6.Sp.pegRNA.HEK3_CTT_ins	hU6	n/a	Addgene #132778	pGM902
4	pU6.Sp.pegRNA.GG_acceptor	hU6	n/a	Addgene #132777	pGM901
5	pCMV.PE2.P2A.GFP	CMV	PE2-P2A-GFP	Addgene #132776	pGM1300
	SIV Lentivirus production plasmids				
6	pSIV.hCEF.EGFP	hCEF	EGFP	GMG plasmid	pGM357
7	pSIV.CMV.EGFP	CMV	EGFP	GMG plasmid	pGM310
8	pSIV.gagpol	CBA	gagpol	GMG plasmid	pGM691
9	pSIV.rev	CBA	Rev	GMG plasmid	pGM299
	HIV Lentivirus production plasmids				
10	pHIV.CMV.EGFP	CMV	EGFP	GMG plasmid	pGM378
11	pHIV.gagpol	CMV+	gagpol	GMG plasmid	pGM281
12	pHIV.rev	RSV	Rev	GMG plasmid	pRSV.rev
	Lentiviral pseudotypes				
13	Fusion protein pseudotype	CAG	Fct4	GMG plasmid	pGM301
14	Haemagglutinin-Neuraminidase pseudotype	CAG	Envct-HN	GMG plasmid	pGM303
15	VSV-G pseudotype	CMV	VSV-G	GMG plasmid	pMD2-G

	All-in one PE LV				
16	pSIV.hCEF.PE1.hU6.CTTpegRNA	hCEF	PE1	This study	pGM948
17	pSIV.hCEF.PE2.hU6.CTTpegRNA	hCEF	PE2	This study	pGM949
18	pSIV.hCEF.PE1.hU6Rev.CTTpegRNA	hCEF	PE1	This study	pGM950
19	pSIV.hCEF.PE2.hU6Rev.CTTpegRNA	hCEF	PE2	This study	pGM951
20	pSIV.hU6.CTTpegRNA.hCEF.EGFP	hCEF	EGFP	This study	pGM935
21	pSIV.hU6Rev.CTTpegRNA.hCEF.EGFP	hCEF	EGFP	This study	pGM936
	Split PE				
22	pHIV.CMV.Cas9n(1-713).P2A.GFP	CMV	Cas9n N-term	This study	pGM1105
23	pHIV.CMV.Cas9n(714-1368).MMLV-RT	CMV	Cas9n C-term	This study	pGM1106
24	pHIV.CMV.Cas9n(1-713).P2A.GFP.hU6.CTTpegRNA	CMV	Cas9n N-term	This study	pGM1151
	Split HITI				
25	pCMV.Cas9.T2A.GFP	CMV	Cas9-T2A-GFP	Merck	pGM1101
26	pHIV.CMV.Cas9.T2A.GFP	CMV	Cas9-T2A-GFP	This study	pGM1100
27	pHIV.hU6.HITI_gRNA1_donor_1	hU6	n/a	This study	pGM1115
28	pHIV.hU6.HITI_gRNA3_donor_3	hU6	n/a	This study	pGM1116
29	pHIV.hU6.HITI_scr_gRNA_donor_1	hU6	n/a	This study	pGM1117
30	pHIV.hU6.HITI_scr_gRNA_donor_3	hU6	n/a	This study	pGM1118
	PE mini library				
31	pEF1a.mCherry.hU6.pegRNA acc	EF1a	mCherry	Gift from Dr Jakob Haldrup	
32	LentiPEmax.EGFP.WPRE	EF1a	PEMax-P2A-GFP	Gift from Dr Jakob Haldrup	
33	LentiPEmax.EGFP.W3	EF1a	PEMax-P2A-GFP	Gift from Dr Jakob Haldrup	
34	pEF1a.mCherry.hU6.HE3_CTT_ins_pegRNA	EF1a	mCherry	This study	pGM1301

35	pEF1a.mCherry.hU6.CFTR_pegRNA_library	EF1a	mCherry	This study	pGM1302
	pegRNA plasmids				
36	pU6.Sp.pegRNA.GG_15.21	hU6	n/a	This study	
37	pU6.Sp.pegRNA.GG_15.20	hU6	n/a	This study	
38	pU6.Sp.pegRNA.GG_15.19	hU6	n/a	This study	
39	pU6.Sp.pegRNA.GG_14.27	hU6	n/a	This study	
40	pU6.Sp.pegRNA.GG_13.22	hU6	n/a	This study	
41	pU6.Sp.pegRNA.GG_13.21	hU6	n/a	This study	
42	pU6.Sp.pegRNA.GG_13.20	hU6	n/a	This study	
43	pU6.Sp.pegRNA.GG_13.19	hU6	n/a	This study	
	gRNA plasmids				
36	phU6-gRNA	hU6	n/a	Addgene #53188	
37	phU6-gRNA.HITI-1	hU6	n/a	This study	
38	phU6-gRNA.HITI-2	hU6	n/a	This study	
39	phU6-gRNA.HITI-3	hU6	n/a	This study	
40	phU6-gRNA.HITI-4	hU6	n/a	This study	
41	phU6-gRNA.HITI-5	hU6	n/a	This study	
42	phU6-gRNA.HITI-6	hU6	n/a	This study	
43	phU6-gRNA.HITI-7	hU6	n/a	This study	
44	phU6-gRNA.HITI-13	hU6	n/a	This study	
45	phU6-gRNA.HITI-14	hU6	n/a	This study	
46	phU6-gRNA.HITI-15	hU6	n/a	This study	
47	phU6-gRNA.HITI-16	hU6	n/a	This study	
48	phU6-gRNA.PEspc1	hU6	n/a	This study	

49	phU6-gRNA.PEspc2	hU6	n/a	This study	
50	phU6-gRNA.PEspc3	hU6	n/a	This study	
51	phU6-gRNA.PEspc4	hU6	n/a	This study	
52	phU6-gRNA.PEnick2	hU6	n/a	This study	
53	phU6-gRNA.Pnick5	hU6	n/a	This study	

2.6. Lentiviral vector production

Viral vector production steps are described in several sections below:

2.6.1. Polyethyleneimine (PEI) transfection in suspension HEK293T cells

For small scale lentiviral production, suspension HEK293T cells were seeded at 1×10^6 cells/mL in 20 mL FreeStyle media (Gibco) per reaction. Transfection of the suspension HEK293T cells was performed using Polyethyleneimine pro (PEIpro, Polyplus) with genome plasmid, *gagpol*, *rev*, and pseudotype plasmids added in 20:9:6:12 molar ratio, or 20:9:6:6:6 when two pseudotype plasmids were used. The total amount of plasmid for transfection was constant at 6.6 µg (except when the genome plasmid size used were large, and therefore to keep the copy number consistent the amount was adjusted accordingly), and Freestyle media was added to a final volume of 1000 µL. After adding all the plasmids, the mix was vortexed for 10 seconds and incubated for 15 minutes at room temperature. In a separate tube, 2 µL PEIpro was added per µg plasmid DNA and Freestyle media was added to 1000 µL and vortexed for 10 seconds. PEIpro mix was then added to plasmid DNA mix, vortexed for 20 seconds, and the mix was further incubated for 20 minutes. Transfection mix was then added to the suspension HEK293T cells, and after overnight incubation, 8.3 mM Sodium Butyrate in final volume of 30 mL of Freestyle media was added to the transfected cells. After 72 hours, the transfected cells were moved into a 50 mL centrifuge tube (Corning) and centrifuged at 500*g* for 10 minutes. The supernatant containing the lentiviral vectors was collected into a new 50 mL centrifuge tube and the cell pellet discarded. Then, 25 U/mL Benzonase, 10 µL of $MgCl_2$, and when F/HN pseudotypes were used, also 50 µL of 10× TrypLE Express (Thermo Fisher Scientific) were added to the supernantant and incubated for 1 hour at 37°C. Where TrypLE Express was added, 1 mL of 1× Defined Trypsin Inhibitor (DTI, Gibco) was added to inactivate the TrypLE Express.

Afterwards, the lentiviral vectors were then clarified through a 0.45 µm syringe filter, and concentrated according to section 2.6.2. or aliquoted and kept at -80°C until use.

2.6.2. Overnight centrifugation for concentration of lentiviral vectors

Clarified lentiviral vectors were centrifuged at 4800*g* for 22h-24h in clean 50 mL centrifuge tubes. After centrifugation, the supernatant was gently discarded and 250 µL of Freestyle media was added to the bottom of the centrifuge to concentrate the lentiviral vectors between 100× to 200× the original concentration. The concentrated vectors were then aliquoted and kept at -80°C.

2.6.3. Lentiviral transductions

For transduction in 24-well plates, adherent cells were seeded at 1.5×10^5 cells/well 1 day before the planned transduction, in the appropriate maintenance media as described in section 2.2. Next, the appropriate volume of lentiviral vectors (as determined by the target multiplicity of infection (MOI) or required dilution factor) was added to OptiMEM media (Gibco) with 8µg/mL Polybrene, in a total volume of 250 µL to create the transduction mix. The medium from cells was then aspirated, and replaced with the transduction mix (250 µL per well). After 6-8 hours of transduction, maintenance media containing 20% FBS was added in a 1:1 ratio with the transduction mix and cells were further incubated at 37°C with 5% CO_2 for 72 hours. Cells were imaged as described in section 2.3 and harvested. When required, gDNA was extracted according to section 2.8 Where cells were to be subjected to flow cytometry, cells were processed according to section 2.7 after imaging and harvest.

Table 2.8. Lentiviral vectors generated in this study

No	Back bone	Transgene	Transgene Plasmid code	Pseudotype	Internal code
All-in-one PE-LV					
1	SIV	hCEF.PE2.hU6.CTTpegRNA	pGM949	VSV-G	rSIV.949.VSV-G
2	SIV	hCEF.PE2.hU6Rev.CTTpegRNA	pGM951	VSV-G	rSIV.951.VSV-G
3	SIV	hCEF.EGFP	pGM948	VSV-G	rSIV.357.VSV-G
Two-vector LV					
4	SIV	hCEF.EGFP	pGM357	F/HN	rSIV.F/HN.hCEF.EGFP
5	SIV	CMV.EGFP	pGM310	VSV-G	rSIV.VSV-G.CMV.EGFP
6	HIV	CMV.EGFP	pGM378	VSV-G	rHIV.VSV-G.CMV.EGFP
Mini Library					
7	HIV	EF1α-PEmax-P2A-EGFP	LentiPEmax.EGFP.W3	VSV-G	rLentiPEmax.EGFP.W3
8	HIV	pEF1a.mCherry.hU6.HE3_CTT_ins_pegRNA	pGM1301	VSV-G	rLV.HEK3_pos
9	HIV	EF1a.mCherry.hU6.CFTR_pegRNA_library	pGM1302	VSV-G	rLV.minilib

2.6.4. Determination of lentiviral titre

Determination of lentiviral titre was performed by quantitative polymerase chain reaction (qPCR), or where appropriate flow cytometry, using serial dilution method. First, adherent HEK293T cells were seeded in 24-well plates at 1.5×10^5 cells/well. On the next day, lentiviral vectors were added to the cells with a series of dilutions, (for example 100×, 300×, 900×, 2700x, and 8100×) in triplicate for each of the dilutions, and the rest of the transduction process followed the method in section 2.6.3. For qPCR assay, the vector copy number was measured by amplification of the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE) region, with *CFTR* used as a housekeeping gene for the qPCR. The integrated WPRE copies were quantified against a standard curve containing DNA sequence of WPRE and CFTR in predetermined copy numbers, starting from the highest copy number of 81920, and diluted 4× until the lowest copy number was 20 (gift from Emily Castells). For the PCR, the primers and probes (see Table 2.9) with a cycling profile of 50°C for 2 minutes, 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 s and 60°C for 1 minute. Functional transducing units per mL of virus (TTU/mL) was determined from the slope of a linear regression curve generated from the dilution series.

Table 2.9. qPCR primers and probes for lentiviral titration

Primers and Probes	Sequence (5'>3')
hCFTRint1-9987F	CTT CCC CCA TCT TGG TTG TTC
hCFTRint1-10100R	TGA CAG TTG ACA ATG AAG ATA AAG ATG A
hCFTRint1-10022T	**VIC**-TGT CCC CAT TCC AGC CAT TTG TAT CCT-**TAMRA**
DNAVEC_WPRE_203F	TGG CGT GGT GTG CAC TGT
DNAVEC_WPRE_281R	CCC GGA AAG GAG CTG ACA
DNAVEC_WPRE_223T	**6FAM**-TTG CTG ACG CAA CCC CCA CTG G-**TAMRA**

2.7. Flow cytometry and FACS

For flow cytometry, cells were washed once with 1× phosphate buffered saline (PBS), or 10 times if the cells had been transduced with a lentiviral vector, and then harvested with TrypLE Select Enzyme (Life Technologies) and centrifuged at 500*g* for 5 minutes. Cells were then washed once more with 1× PBS and then fixed in 2% Paraformaldehyde (PFA) for 15 minutes and washed twice with 1× PBS. Afterwards, cells were resuspended in approximately 300 µL 1× PBS /1×10^6 cells. Cells were examined for EGFP and/or mCherry signals with the filters as shown in Table 2.10. Flow cytometry data were obtained with LSRII flow cytometer (BD Biosciences) and analysed with FlowJo X software (BD Biosciences).

Table 2.10. Flow cytometry parameters

	Excitation (nm)	**Emission (nm)**	**Dichroic mirror / Bandpass filter**
EGFP	489	511	505LP – 515/20
mCherry	587	610	600LP – 610/20

Cells were harvested and washed as above for Fluorescence Activated Cell Sorting (FACS), however for the final step, cells were resuspended in 4% FBS in 1× PBS solution with 4',6-diamidino-2-phenylindole (DAPI) (Thermo Fisher Scientific) added as per manufacturer's protocol. Post FACS, when applicable, cells were grown in High Glucose DMEM with 20% FBS, 1% Peniciilin-Streptomycin (PenStrep) (5,000 U/mL), 2mM GlutaMax Supplement (GibcoTM), 20mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) Buffer Solution (GibcoTM) before gradually transitioning to regular cell maintenance media.

2.8. Genomic DNA Extraction

Genomic DNA extraction from mammalian cell culture was performed using the QIAamp DNA Blood Mini Kit (QIAGEN), following the manufacturer's protocol. For the final step, the GDNA was eluted with 50 µL to 100 µL of NFW (Promega), depending on the initial number of cells.

2.9. Sanger sequencing

Sanger sequencing was performed by Source Biosciences. All samples were provided at 10 ng/µL concentration for plasmids or 100 ng/µL for purified PCR products. All primers used for sequencing that were not provided by Source Biosciences were provided at 3.3 ng/µL concentrations. Where sequencing was performed to check for editing, the sequencing result was processed with ICE Analysis (Synthego, https://ice.synthego.com/) or DECODR (https://decodr.org/).

2.10. T7 Endonuclease I assay

The T7 Endonuclease I (T7EI) assay was performed on PCR amplicons corresponding to region surrounding the F508 amino acid on CFTR products from cells transfected with plasmids expressing Cas9 and gRNA spacers for both HITI and PE gRNA selection experimentsas a template. Briefly, gDNA from the cells was extracted (section 2.8), forward primer (5'CCTTCTCTGTGAACCTCTAT3') and reverse primer (5'CTTAAAGCATAGGTCATGTG3') were used to amplify the region where the Cas9-induced double-strand break was expected to occur. Then, 200 ng of the PCR product was mixed with 2 µL of 10× NEBuffer 2 (New England Biolabs, NEB) and water was added to a

final volume of 19 µL. The hybridization of the PCR product was performed in a thermal cycler by denaturing the PCR product at 95°C for 5 minutes, followed by a temperature ramp down to 85°C at a rate of -2°C per second, and further down to 25°C at a rate of -0.1°C per second. After the hybridization step, 1 µL of T7EI enzyme (NEB) was added to each reaction and incubated for 15 minutes at 37°C. The final product was then electrophoresed on a 2% agarose gel and visualized using an iBright FL1500 imaging system (Thermo Fisher Scientific). Gel images and band intensities were analysed using Fiji software and the cutting efficiency was calculated using the following equation[161]:

$$\%\ gene\ modification = 100\ \times (1 - (1 - total\ fraction\ cleaved)^{\frac{1}{2}})$$

2.11. Western Blot

For separation of proteins, samples were subject to Western Blotting using SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) followed by membrane transfer and antibody detection.

2.11.1. Sample Preparation

Cells were harvested and washed with cold PBS, and centrifuged at 500*g* for 5 minutes. Then, PBS was aspirated, and 1×10^6 cells were lysed by adding 400 µL of 1× Radioimmunoprecipitation assay (RIPA) buffer (Merck Millipore) supplemented with cOmplete mini EDTA-free protease inhibitor cocktail tablet (Sigma Aldrich) and incubated on ice for 5 minutes and briefly vortexed. Then, the cells were further incubated for 30 minutes. Cells debris was removed by centrifugation at 21000*g* for 20 minutes at 4°C. The

protein lysate was then collected into pre-chilled 1.5mL Eppendorf tubes and stored at -20°C until use. To determine the concentration of the protein lysate, a detergent compatible (DC) protein assay (BioRad) was performed according to the manufacturer's protocol, except the concentration of the standards was adjusted to 2 µg/mL, 1.5 µg/mL, 1 µg/mL, 0.75 µg/mL, 0.5 µg/mL, 0.25 µg/mL, 0.125 µg/mL, and 0 µg/mL. Samples were diluted as appropriate, typically between 1:3 to 1:5, and absorbance was measured at 750 nm using a SpectraMax i3x plate reader and SoftMax Pro software v7.0 (Molecular Devices).

2.11.2. Polyacrylamide Gel Electrophoresis

For separation of proteins, samples were normalised to total protein content of 15 µg per well and incubated at 37°C with Laemmli sample buffer (5× Laemmli buffer was made with 250mM Tris-HCl [pH 6.8], 10% sodium dodecyl sulphate (SDS), 25% Glycerol, 0.01% Bromophenol Blue, 20% β-mercaptoethanol, and diluted to 1× working concentrationwith the protein lysate) for 30 minutes. Polyacrylamide separating gels were made by mixing H_2O, Acrylamide/Bis-acrylamide (30%/0.8% w/v), 1.5M Tris (pH=8.8) 10% (w/v), SDS 10% (w/v) ammonium persulfate (AP), N,N,N',N' -Tetramethylethylenediamine (TEMED), with detailed volume for different concentration listed on Table 2.11. Stacking gels were made by mixing 2.975 mL of H2O, 1.25 mL of 0.5 M Tris-HCl, pH 6.8, 0.05 mL of 10% (w/v) SDS, 0.67 mL of Acrylamide/Bis-acrylamide (30%/0.8% w/v), 0.05 mL of 10% (w/v) ammonium persulfate (AP), and 5 µL of TEMED. Separating gels were casted onto the SDS-PAGE casting glass plates, and after it has solidified, followed by the stacking gels with appropriate gel combs.

Table 2.11. Reagent volumes for making polyacrylamide gels

Acrylamide percentage	8%	10%
H_2O	4.6 mL	3.8 mL
Acrylamide/Bis-acrylamide (30%/0.8% w/v)	2.6 mL	3.4 mL
1.5M Tris(pH=8.8)	2.6 mL	2.6 mL
10% (w/v)SDS	0.1 mL	0.1 mL
10% (w/v) ammonium persulfate (AP)	100 µL	100 µL
TEMED	10 µL	10 µL

Samples were subjected to SDS-PAGE in 10% (GAPDH) or 8% (CFTR, Cas9) gels together with Kaleidoscope protein marker (BioRad), initially at 80V until the Laemmli buffer dye had entered the resolving gel, then at 150V in 1× SDS PAGE running buffer (diluted from 10× SDS-PAGE running buffer containing Tris-base 30.3g [250 mM], Glycine 144.1g [1.9 M], 50mL 20% SDS [0.22u f/s] (final 1%)). After completion of SDS-PAGE, proteins were transferred from the gels to 0.45 µm nitrocellulose membranes in 1× Tris Glycine (TG) Buffer (diluted from 10× TG buffer containing Tris-base 30.3g [250 mM], Glycine 144.1g [1.9 M], made to 1L H_2O with 20% methanol) at 150V for 1 hour and 10 minutes using a Trans-Blot Cell system (BioRad) placed on a box of ice to reduce buffer heating. Membranes were then trimmed as needed and blocked with TBS-Tween (0.1%) + 5% powdered milk solution and rocked slowly for about 1 hour.

2.11.3. Immunoblotting

For detection of proteins, primary antibodies were applied to the membranes in 10 mL of blocking buffer (made using Tris Buffered Saline (TBS) with 0.1% Tween and 5% milk) using the dilution factor specified in Table 2.12. Membranes were then rocked slowly overnight while making sure that the vessel was appropriately covered to prevent evaporation. The

next day, the blocking solution containing the primary antibody was removed and the membranes were washed with Tris buffered saline-Tween (TBS-T) three times for 5 minutes. Next, anti-mouse IgG secondary antibody was applied for 1 hour diluted as in Table 2.12 and washed again with TBS-T three times for 5 minutes. The blot was developed using the Clarity ECL substrate (BioRad) and imaged using iBright FL1000 system (Thermo Fisher Scientific).

Table 2.12. List of antibodies for Western blot

Target	Antibody	Dilution factor	Species	Source
CFTR	UNC596 anti CFTR	1:1000	Mouse	UNC
Cas9	7A9-3A3 anti CRISPR Cas9	1:1000	Mouse	Abcam
GAPDH	CB1001 anti GAPDH	1:10000	Mouse	Sigma Aldrich
Mouse IgG	Anti Mouse IgG H&L (HRP)	1:10000	Goat	Thermo Fisher Scientific

2.12. Mini Library approach

2.12.1. Generation of PEmax-GFP expressing HEK293T cell line

Adherent HEK293T cells were seeded in a 6 well plate (7×10^5 cells per well), 24 hours prior to transduction. Approximately 300 µL of lentivirus (rHIV-VSVG-EF1a- PEmax-P2A-EGFP-W3) were added to each well, with an MOI of 4. Cells were imaged 72 hours post transduction, and bulk sorting of GFP positive cells was done using FACS, gating for medium or high GFP expression level. Bulk sorted cells with high GFP expression were then re-sorted into single cell clones, from which 12 clones were selected and expanded further, and then frozen down until use.

2.12.2. Mini Library design and cloning

To design the pegRNAs for the mini library, pegIT software (https://pegit.giehmlab.dk/) was used to generate all possible pegRNA candidates to correct F508del mutation in *CFTR*. Then, the limit of the design scope was determined as primer binding site length ranging from 11-20 bp, and RT template length not longer than 50 bp. Thus, 445 candidates of pegRNA from three different spacers was selected. A further three positive controls targeting *HEK3*, *FANCF*, and *RNF2* was added, and also six negative controls for each of the *CFTR* spacers and the positive controls were added making the total of the mini library elements to 454 pegRNAs. The mini library was ordered as one single stranded DNA oligo pool (Opools, IDT DNA), which was resuspended with nuclease free DNA to 1000nM concentration upon arrival. PCR amplification to create oligo duplex was performed with 10ng of the single stranded DNA with primers (described in Table 2.13) and PCR conditions as described in Table 2.14. The amplified fragments were then purified with MinElute PCR Purification Kit (Qiagen) as per manufacturer's protocol, and cloned into pEF1a.mCherry.hU6.pegRNA acc (Table 2.7) at the BsmBI restriction site (section 2.4.8 and 2.4.9).

Table 2.13. Primers for duplexing mini library oligos

Internal code	Name	Sequence (5' to 3')
GM22-1682	MiniLib-duplex-PCR For	CAATCCGCCCTCACTACAACCG
GM22-1683	MiniLib-duplex-PCR-Rev	CTACTCTGGCGTCGATGAGGGA

Table 2.14. PCR conditions for duplexing mini library oligos

Cycle step	3-step protocol		Cycles
	Temp.	Time	
Initial denaturation	98°C	5 mins	1
Denaturation	98°C	10 s	16
Annealing	70°C	30 s	
Extension	72°C	30 s	
Final Extension	72°C	10 mins	1
	4 °C	hold	

2.12.3. Bacterial electroporation for mini library plasmids

For bacterial electroporation, Lucigen Endura ElectroCompetent cells (Lucigen) were thawed on ice for 15 minutes, while 0.1 cm electroporation cuvettes (USA Scientific) on ice. Subsequently, 2 µL of 25 ng/µL library plasmid DNA or up to 4 µL of heat-inactivated ligation reaction was combined with ~50 µL of Lucigen cells (1 vial), mixed by flicking, and incubated on ice for 5-10 minutes. Half of the mixture (~25 µL) was transferred into a pre-chilled cuvette, ensuring no bubbles were formed, for a total of two cuvettes. Electroporation was performed using the Bio-Rad Gene Pulser Xcell at 1.8 kV, 600 Ω, and 10 µF.

Immediately following electroporation, 1 mL of Lucigen recovery media (Lucigen) was added to the electroporated cells, which were then transferred into a 14 mL round-bottom culture tube. The culture was shaken at 37°C for 1 hour at 225 rpm. To assess transformation efficiency, 2-3 dilution were prepared, with dilutions ranging from 5,000× to 200,000×. The dilution plates were prepared by taking 10 µL from the 2 mL transformation reaction and diluting it into 5 mL LB (1:500 dilution). From this, 200 µL was used to create a 1:5,000 dilution, and 20 µL and 5 µL were used for 1:50,000 and 1:200,000 dilutions, respectively, by spreading them onto agar plates with ampicillin resistance. The remaining recovery volume was used to inoculate 0.5 L of LB media with ampicillin in a 2L flask.

The culture was shaken at 37°C for 12-14 hours at 225 rpm. Transformation efficiency was then evaluated from the dilution plates, and if the coverage was at least 30×, purification

of plasmid DNA was performed according to section 2.4.16 using Maxiprep kits. For example, for a library of ~15,000 elements, at least 10 colonies on a 50,000x dilution plate would indicate sufficient coverage.

2.12.4. Next Generation Sequencing

Samples for Next Generation Sequencing were first PCR amplified with primers listed in Table 2.15, using 5 µg of gDNA template per sample. PCR conditions are described Table 2.16. The PCR product was electrophoresed to ensure that the product bands matched the expected product length. Samples were purified using the column purification method as described in section 2.4.12.

Table 2.15. NGS PCR primer

Internal code	Name	Sequence (5' to 3')
GM22-1684	MiniLib-Amp1-For-1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGAAAAAGTGGCACCGAGTCG
GM22-1685	MiniLib-Amp1-For-2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCTTGAAAAAGTGGCACCGAGTCG
GM22-1686	MiniLib-Amp1-For-3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTTGAAAAAGTGGCACCGAGTCG
GM22-1687	MiniLib-Amp1-Rev-1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAATACTGCCATTTGTCTCAAGA
GM22-1688	MiniLib-Amp1-Rev-2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAATACTGCCATTTGTCTCAAGA
GM22-1689	MiniLib-Amp1-Rev-3	GTGACTGGAGTTCAGACGTGTGTCTTCCGATCTAATACTGCCATTTGTCTCAAGA

Table 2.16. NGS PCR condition - first round

Cycle step	3-step protocol		Cycles
	Temp.	Time	
Initial denaturation	98°C	5 mins	1
Denaturation	98°C	10 s	30
Annealing	60°C	30 s	
Extension	72°C	30 s	
Final Extension	72°C	10 mins	1
	4 °C	hold	

After purification, the PCR samples were subjected to a second round of PCR for indexing purposes using NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs Set 2). Each PCR reaction used 20ng of the purified PCR as the template, with PCR conditions as described in Table 2.16 except with 63°C annealing temperature, and 11 cycles for denaturation, annealing, and extension steps. The product of the second PCR round was pooled and purified using AMPure XP Beads (Beckman Coulter) according to manufacturer's protocol with 0.8µL AMPure XP Beads added per 1µL of sample. Fragment lengths of the pooled samples fragment lengths were estimated using TapeStation 4200 system (Agilent Technologies) by running 2 µL of the pooled sample. Sample concentrations were measured using Qubit® Flex Fluorometer (Thermo Fisher Scientific) as described in section 2.4.18 and samples were diluted to 4nM concentration before NGS Sequencing (Weatherall Institute of Molecular Medicine). Analysis of NGS data was performed using batch analysis in CRISPResso2 (Pinello lab) with Python (code written by Dr Jakob Haldrup, available in Appendix).

2.13. Statistical analyses

Statistical analyses were performed using Prism 8 software for Mac (GraphPad). Where the data were discontinuous, or the sample size was three or fewer, non-parametric tests were used. A significant difference was defined as having a p-value of <0.05 and indicated in figures where appropriate (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, and **** = $p < 0.0001$). Error bars in figures represent the standard deviation of the mean.

Chapter 3: Prime Editing: All-in-one lentiviral vector delivery approach

3.1. Introduction

When prime editing technology was first introduced in 2019[56], there was a very high expectation of its potential to create seamless and accurate correction of genetic mutations that are causative in many debilitating and fatal genetic diseases, such as sickle cell disease and Tay-Sachs disease, as shown in an *in vitro* model in the first study. However, despite this potential, there remains some reservations concerning prime editing; for example, the efficacious delivery of the prime editor is of high concern. This concern is largely due to the size of the prime editors (6.3 kb) and the pegRNA, ~150 nt – 200 nt), which are much larger than the conventional SpCas9 (4.1 kb) and its guide RNA (gRNA, ~100 nt). The PE system is composed of Cas9 nickase and MMLV-RT domain, connected with a flexible linker. The pegRNA, while also containing a spacer and a scaffold sequence similar to regular Cas9 gRNA, has a 3' extension region which has a primer binding site and RT template which will provide the template for editing.

The increased size of the PE system restricts the choice in vectors that could be used to deliver both components into target cells that require the genetic editing for genetic correction. Adeno-associated viral vectors (AAV), one of the most widely used vectors for gene editing, cannot easily accommodate the size of PE as the packaging capacity is limited at 4.8 kb. The approach using AAV requires the use of two-vector split PE approach and thus requiring both vectors to effectively transduce the cells. Therefore, to overcome this

delivery concern I looked at adopting recombinant lentiviral vectors (rLVs) instead, given that these vectors have a substantially larger packaging capacity (up to 9.8 kb) as a potential candidate vector to deliver PEs for the correction of genetic mutations related to Cystic Fibrosis (CF) lung disease.

Successful prime editing depends on the delivery of both the PE and pegRNA to the target locus. In this chapter I aimed to simplify the delivery of the PE and pegRNA by using an 'all-in-one' lentiviral vector delivery approach, meaning that both PE effector and pegRNA are encoded in the same rLV construct. This would mean that cells would only need to uptake a single rLV vector to transfer both essential PE components into the same cell to perform the intended edit. To achieve this, and as a proof of concept, I designed, produced, and assessed the capability of an all-in-one prime editing lentiviral vector to insert a CTT sequence (as would be necessary to correct the common F508del mutation in CF) into the *HEK3* locus of HEK293T cell line.

3.2. Results

3.2.1. Validating *HEK3* Site Editing with PE plasmids

At the initiation of these studies, prime editing had been only very recently published and was considered a highly novel approach to mediate accurate and precise genome edits without introducing DSBs in the target GDNA. It was, therefore, critical to ensure that the levels of prime editing described by Anzalone et al.[56] were reproducible before proceeding to the next step to adopt prime editing to correct genetic lung diseases including F508del mutation in *CFTR*.

Using the plasmids described by Anzalone et al.[56] and deposited in the plasmid repository, Addgene (Table 2.7), I aimed to reproduce the precise editing reported at the HEK293T cell *HEK3* site by inserting three bases in the +1 position (+1CTT) directed by the HEK3_CTT_ins pegRNA as previously described[56]. Here, as a first pilot experiment, PE plasmid (pGM899 or pGM900 encoding PE1 and PE2 respectively; Table 2.7) and *HEK3* site targeting pegRNA (pGM902; Table 2.7) were introduced into HEK293T/17 (from here onwards written as HEK293T) cells per well via Lipofectamine™ 3000 mediated transfection (section 2.4.17). After 72 hours, GDNA was harvested from treated cells (Section 2.8), and the *HEK3* locus, where the +CTT edit was intended, was PCR amplified and sequenced using the primers (HEK3 CTT PCR/Sequencing F & R; Table 2.4). resultant edits were analysed using ICE Analysis (Synthego) by comparing the sequencing result of the transfected samples to the naïve control (2.9).

Figure 3.1 Confirming PE1 and PE2 efficacy to edit HEK3 locus.

HEK293T cells were transfected with PE1 or PE2 with pegRNA for *HEK3* +1 CTT editing (n=12 each) to compare editing efficiencies between PEs.

A) Editing efficiency of PE1 and PE2 for +1 CTT insertion in *HEK3* site (n =12). PE1 showed an average of 2.92 ± 0.9% while PE2 showed an average

of 28.33 ± 1.44% for editing efficiency in HEK293T cells. Two representative images showing sequencing chromatograms of pU6.Sp.pegRNA.HEK3_CTT_ins and pCMV.PE1 (B) or pCMV.PE2 (C) transfected samples compared to the wild-type sequence in the control sample. Dotted line indicated the nicking site of PE, and dotted arrows showed where the insertion has occurred with the colour blue indicating Cytosine (C) and red indicating Thymine (T). In the sample transfected with pCMV.PE2, a clear frameshift can be observed, and three bases insertion (CTT) on the desired site was also observed indicated by the blue-red-red arrow.

As reported in Anzalone, et al.[56] the on target editing efficiency for PE1 was about 10% and PE2 about 30%. In this experiment, I obtained an editing efficiency of 2.92 ±0.9% for PE1, whereas PE2 showed an average of 28.33 ±1.44% (Figure 3.1). Even though the result for PE1 obtained here was much lower compared with the published result[56], these results gave an early indication that the experiment was reproducible, and from the two PE versions, PE2 performed better (Figure 3.1). Therefore, I decided to use PE2 for all subsequent editing experiments.

3.2.2. PE-LV all-in-one plasmid construction

After confirming that prime editing in HEK293T cells (*HEK3* +CTT insertion) could be reproduced, especially with the PE2 effector molecule, the subsequent step was to clone the PE2 cDNA into an SIV lentiviral vector genome backbone (pSIV.hCEF.EGFP.mWPRE genome plasmid, pGM357; Table 2.7) to construct PE-LV genome plasmids for the production of rLV encoding the PE and a targeting pegRNA of choice. As the end goal for the all-in-one vector is for translational purposes for treatment of CF, the SIV LV genome backbone was chosen as it has shown efficient transduction of the airway[162]. In the first instance, I opted to target the *HEK3* site for the well-established +CTT insertion[56].

To generate the all-in-one PE-LV construct, the hU6 promoter, the *HEK3* +CTT pegRNA, as well as the PE sequence, were all cloned into a lentiviral genome plasmid backbone (Figure 3.2A). In order to clone the PE2 cDNA fragment into the LV genome plasmid backbone, site directed mutagenesis was first performed. This was to generate a NotI restriction enzyme site that 5' flanked the EGFP reporter sequence in pGM357, and also to add a NotI restriction enzyme site that 3' flanked the PE1/PE2 cDNA sequence (described in section

2.4.2). This facilitated the ease of cloning of the PE cDNA as a NotI-NotI fragment into the SIV LV genome plasmid backbone. Next, the hU6 promoter with the pegRNA targeting *HEK3* site was PCR amplified from pU6.Sp.pegRNA.HEK3_CTT_ins plasmid (BglII-pU6.Sp.pegRNA.HEK3_CTT_in primer, Table 2.7) with added 5' and 3' BglII restriction sites. Then, the PCR fragment was purified and cloned into the pGM357 backbone according to sections 2.4.12, 2.4.8 and 2.4.9. Two resulting intermediate cloning constructs were obtained – the pGM935 (pSIV.hU6.CTTpegRNA.hCEF.EGFP) with hU6 promoter and pegRNA in the forward orientation, and pGM936 (pSIV.hU6Rev.CTTpegRNA.hCEF.EGFP) (Table 2.7) with hU6 promoter and pegRNA in the reverse orientation.

Afterwards, PE1 or the PE2 cDNA was subcloned downstream of the hCEF promoter into pGM935 or pGM936 as a NotI-NotI fragment to generate pGM948, pGM949, pGM950, and pGM951 (Figure 3.2B), where only constructs with the PE cDNA in the forward orientation per hU6-pegRNA configuration were selected (Table 2.7). All the resulting plasmids were sequenced to ensure that the final construct have correct sequences (data not shown). Each construct contains either PE1 or PE2 under hCEF promoter, and hU6 driven CTT-pegRNA in either a forward or reverse direction. With the plasmid having two different orientations of the hU6 promoter and the pegRNA, these alternate configurations allowed evaluation of the impact of the directionality of the pegRNA on the editing efficiency at the target site.

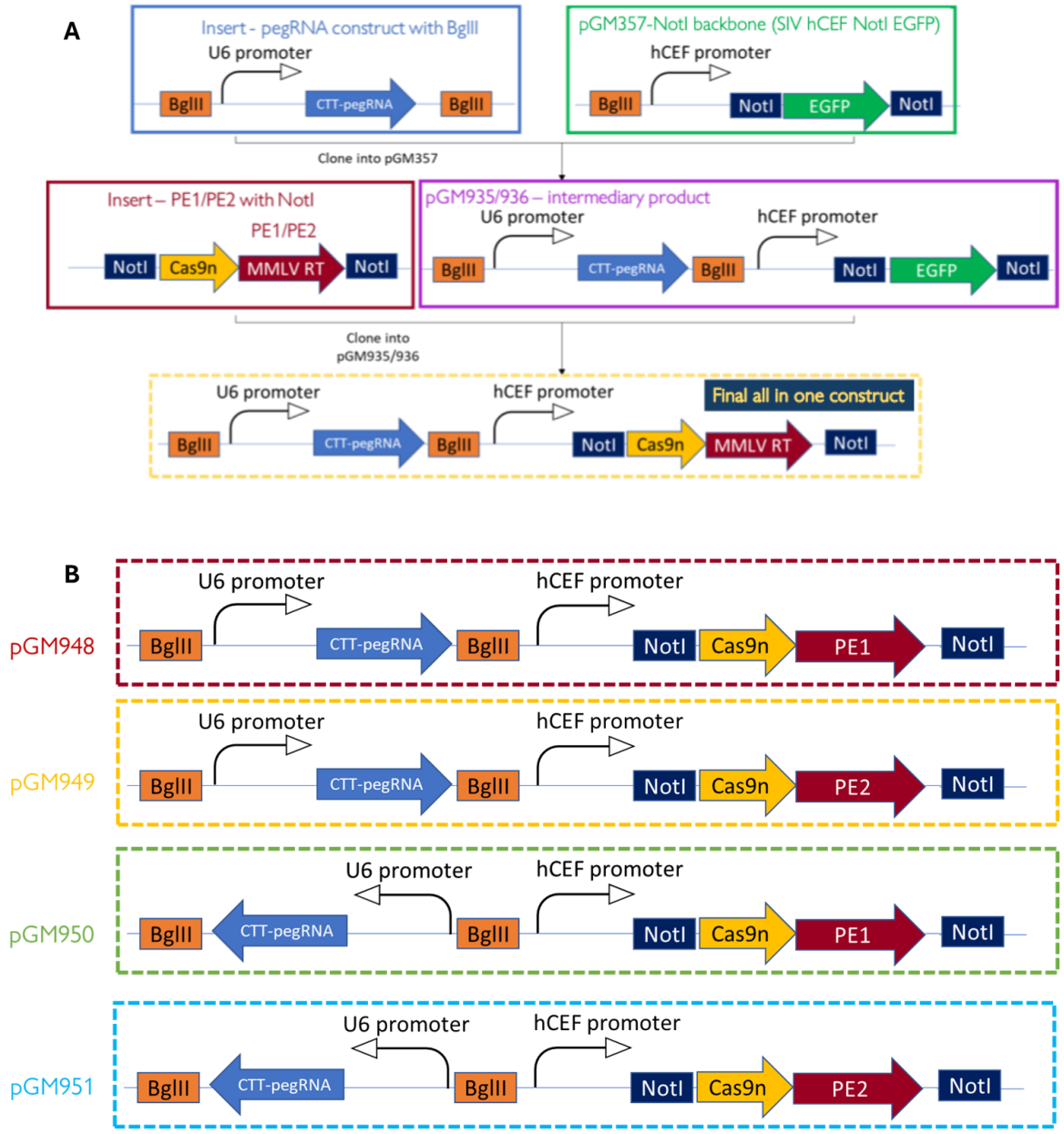

Figure 3.2. Cloning of the PE2 and *HEK3* +CTT pegRNA into pSIV LV genomes for PE2-LV manufacture.

A) Schematic describing the cloning steps performed to construct PE2-LV genome plasmids for subsequent rSIV PE2-LV manufacture. The hU6 pegRNA and PE2 sequences were cloned into lentiviral backbone (pSIV.hCEF EGFP; internal code pGM357; Table 2.7). First, using site directed mutagenesis, NotI restriction sites were added to pGM357 and the PE2 plasmids (pCMV.PE2) for ease of cloning. Then the hU6 promoter as well as the pegRNA targeting

HEK3 site was PCR amplified with added BglII restriction sites on both ends and cloned into the pGM357-NotI, resulting in pGM935 and pGM936 with hU6-pegRNA in either orientation. The PE2 fragment, flanked with NotI sites, was then cloned into pGM935/pGM936. B) Schematics of the final constructs generated for all-in-one PE2-LV that were generated in this study.

After generating the all-in-on PE2-LV genome plasmids, I next confirmed that the PE2 and pegRNA sequences were not compromised by the protracted cloning process. To examine this, I tested if these novel plasmids could mediate the intended +CTT insertion edit at the *HEK3* locus in HEK293T cells. To do this, the pSIV PE2-LV genome plasmids (1.5 µg each pGM949 and pGM951) were delivered separately into HEK293T cells by transient transfection using Lipofectamine 3000® (Section 2.4.17). As a positive control for *HEK3* +CTT prime editing, 1 µg of pCMV.PE2 and 500 ng of pU6.Sp.pegRNA.HEK3_CTT_ins (Table 2.7) was transfected into HEK293T/17 cells as described in section 2.4.17. Seventy-two hours post- transfection, the GDNA and *HEK3* locus were analysed by sequencing and ICE analysis (section 2.9).

Sequencing and Synthego ICE analysis indicated that both all-in-one PE2-LV constructs managed to edit the *HEK3* site, though the efficiency was lower when compared with the two-plasmid transfection strategy. The pGM951 construct, which has reverse orientation of the hU6-pegRNA appeared to have slightly higher editing efficiency of the two novel LV vector genome configurations at 15.0 ± 7.1% compared with pGM949 at 9.0 ± 9.9%. However, as the standard deviation was quite large, this result needs to be interpreted with caution. After confirming the functionality of the construct, the production of the all-in-one PE-LV was initiated.

Figure 3.3. Editing of CTT insertion at the *HEK3* site using all-in-one PE2-LV plasmid.

Figure 3.3. Editing of CTT insertion at the *HEK3* site using all-in-one PE2-LV plasmid.

A) Percentage editing efficiencies of CTT insertion using all-in-one PE2-LV plasmid transfections for pGM949 was (9.0 ± 9.9)% and pGM951 was (15.0 ± 7.1)% (n=2). The on target +CTT editing efficiency was found to be lower using PE2-LV plasmids compared with the positive control, which uses two-plasmid transfection (pCMV.PE2 and of pU6.Sp.pegRNA.HEK3_CTT_ins) that has 36% editing efficiency (n=1). B) Sequencing showed a CTT insertion at the expected editing site, (reflected as a +AAG in the ICE analysis readouts because sequencing was read from the complementary DNA strand. Sample transfected with pGM949 and pGM951 were compared using Welch's t-test; $P > 0.05$.

3.2.3. PE- LV production and titration

The next step was to produce the all-in-one PE-LV, using the method that was described in section 2.6. Briefly, five plasmids carrying *gag pol*, *rev*, F protein pseudotype, HN protein pseudotype, and relevant genome plasmid were transfected into HEK293T suspension cells for LV production. This resulted in two PE2-LV vectors: rSIV.949.VSV-G which carries hU6-pegRNA in forward orientation and hCEF-PE2, and rSIV.951.VSV-G; Table 2.8), and transduction control rLV, rSIV.357.VSV-G which carries hU6-pegRNA in reverse orientation and hCEF-PE2; and rLV, rSIV.357.VSV-G which carries hCEF-EGFP and act as a mock editing control and a transduction control, where successfully transduced cells express EGFP (Table 2.8). The VSV-G pseudotype[163] was chosen for the lentiviral vectors as it is commonly used for transduction, has broad tropism, and is able to transduce HEK293T cells[90,164], which was used as the cell line for editing. To determine the transducing unit titre of the vectors, serial dilutions of each PE-LV were used to transduce HEK293T cells (section 2.6.3) and vector integration was measured using qPCR according to section 2.6.4.

The result showed that both PE-LV vectors were produced at a low titre (7.66E6 ± 1.68E6 TTU/mL for rSIV.949.VSV-G, 8.38E6 ± 2.61E6 TTU/mL for rSIV.951.VSV-G) compared with the rSIV.357.VSV-G control vector (4.31E7 ± 9.73E6 TTU/mL; Figure 3.4). The reduction in titre was unsurprising as the size between the LTR of the PE-LV vectors is approximately 9.9 kb, which is approaching the packaging limit of a lentiviral vector. Nevertheless, the titres were still adequate for transduction purposes, as discussed in the next section.

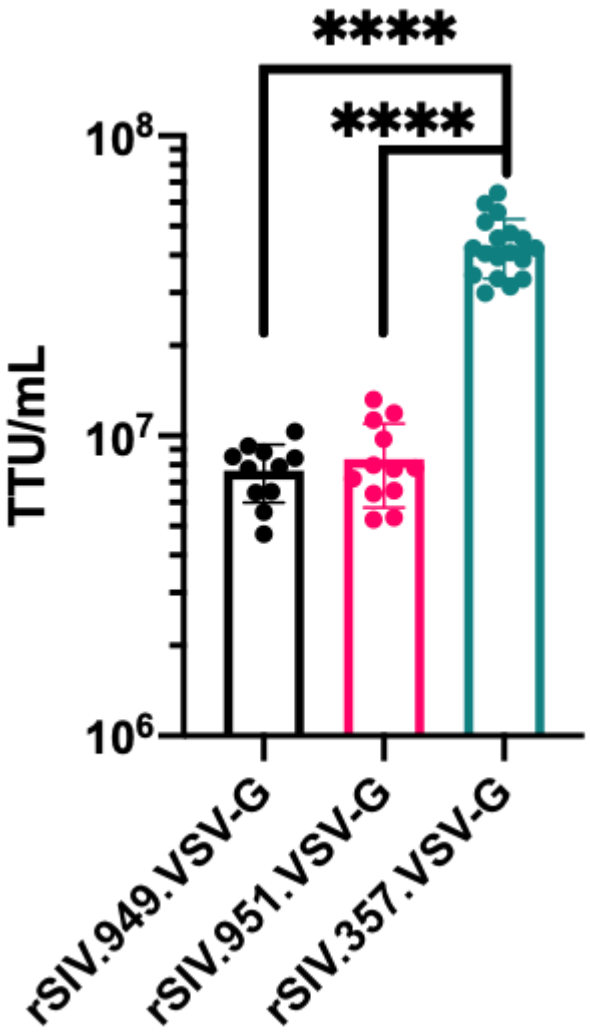

Figure 3.4. Titres of all-in-one PE-LV vectors.

Each dot represents each titration point and error bars represent standard deviation. The titre for each vector are rSIV.949.VSV-G (7.66E6 ± 1.68E6) TTU/mL, rSIV.951.VSV-G (8.38E6 ± 2.61E6) TTU/mL, and rSIV.357.VSV-G 4.31E7 ± 9.73E6 TTU/mL. Statistical analysis was performed using one-way ANOVA with Tukey's post hoc test, $P<0.0001$.

3.2.4. Editing of HEK3 site in HEK293T cells line

The all-in-one PE2-LVs were used to transduce HEK293T cells with an MOI of 3 in triplicates to mediate +CTT insertion at the *HEK3* site as described in section 2.6.3. As a transduction and mock editing control, rSIV.357.VSV-G was also included. To check whether HEK293T cells had been transduced with rSIV.VSV-G vectors, cells that were transduced with rSIV.357.VSV-G were examined for EGFP expression using fluorescence microscopy 72 hours post transduction (Figure 3.5). The transduction control using rSIV.357.VSV-G would help suggest that, at the same MOI, transduction with rSIV.949.VSV-G and rSIV.951.VSV-G have likely occurred, indicating that the PE constructs have been delivered to the target cells to mediate expression and editing. Fluorescence microscopy showed low EGFP expression, with only a slight increase visible compared with naïve (non-transduced) cells. This indicated that hCEF promoter might have low activity in HEK293T cells. Despite this, the GDNA from cells that were transduced with the PE-LV were harvested and prepared for sequencing (section 2.8) with HEK3 CTT PCR and sequencing primers (HEK3 CTT sequencing F & R; Table 2.4). The sequencing result showed that in all PE-LVs treated samples, the traces were comparable to the samples that were transduced with rSIV.357.VSV-G, suggesting that no detectable editing has occurred, and the all-in-one PE-LV strategy was not highly potent.

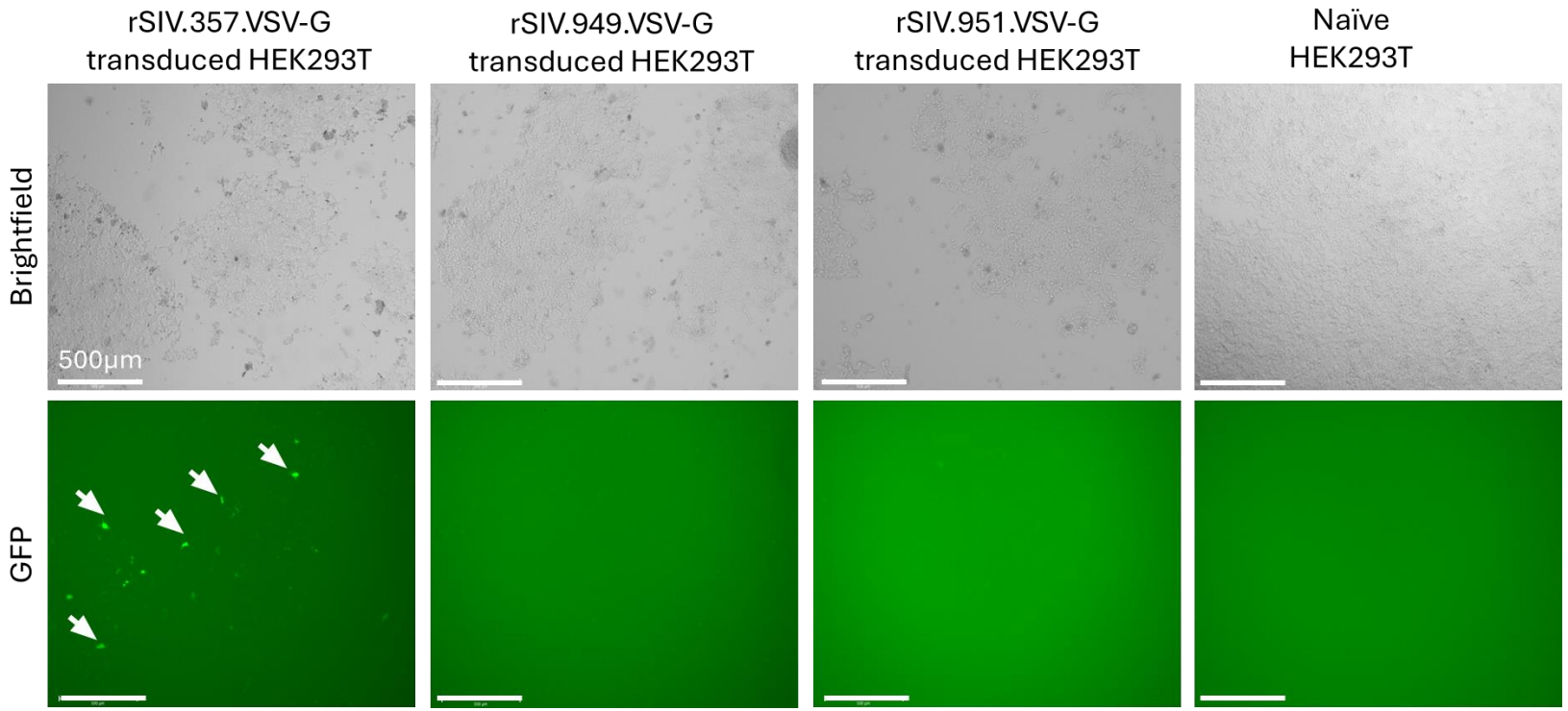

Figure 3.5. Fluorescence microscopy images of HEK293T cells.

HEK293T cells were transduced with rSIV.357.VSV-G, rSIV.949.VSV-G, or rSIV.951.VSV-G. Naïve refers to non-transduced cells. Lower activity of hCEF promoter indicated by low GFP expression from rSIV.357.VSV-G transduced HEK293T cells (MOI=3). Images were taken 72 hours post-transduction. Compared with naïve cells, there were only a few cells expressing GFP (indicated with white arrows) with low fluorescence intensity. Scale bar = 500 µm.

	FILENAME	KO	R^2	INDEL	%	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
rSIV.357.VSV-G	Rep 1	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
	Rep 2	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
	Rep 3	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
rSIV.949.VSV-G	Rep 1	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
	Rep 2	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
	Rep 3	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
rSIV.951.VSV-G	Rep 1	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
	Rep 2	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA
	Rep 3	0.0	1.00	0	100.0	TGGGGCCCAGACTGAGCACG | TGATGGCAGAGGAAA

Figure 3.6. ICE analysis of HEK3 site editing with all-in-one PE-LV.

ICE analysis result for editing in HEK3 site of samples transduced with either rSIV.357.VSV-G, rSIV.949.VSV-G, or rSIV.951.VSV-G (n=3)**.** No editing was observed in samples that were transduced with rSIV.949.VSV-G or rSIV.951.VSV-G.

3.2.5. Comparison of hCEF promoter activity in HEK293T and A549 cells

In section 3.2.4, based on the EGFP expression level post transduction using rSIV.357.VSV-G, it was suspected that hCEF promoter has low activity in HEK293T. To test this hypothesis, I repeated the transduction in both HEK293T and adenocarcinomic human alveolar basal epithelial (A549) cells with MOI 3 or MOI 10 according to section 2.6.3. As A549 cells are lung-derived, it was expected that hCEF promoter will have high activity in these cells, and therefore it was used as a benchmark to HEK293T cells. Fluorescence microscopy images were taken (section 2.3), and flow cytometry was performed 72 hours post transduction (section 2.7).

From the fluorescence microscopy images, the level of transduction and also the mean fluorescence intensity (MFI) seemed to be higher in A549 cells, in particular for the cells transduced with MOI 10 (Figure 3.7). This was confirmed by flow cytometry, where MFI for cells transduced with MOI 3 was (19.43±0.85) for HEK293T cells, and (208.00±9.54) for A549 cells. Meanwhile, the MFI for cells transduced with MOI 10 was (23.03±0.32) for HEK293T cells, and (239.33±13.32) for A549 cells (Figure 3.8). The percentage of EGFP positive cells were also higher in A549, with (17.40±1.56)% and (46.50±3.53)% of total cells were positive in MOI 3 and MOI 10, respectively. In comparison, only (6.11±0.48)% of total cells in MOI 3, and (25.03±1.85)% of total cells in MOI 10 were positive for EGFP. This confirmed the hypothesis that hCEF promoter has reduced activity in HEK293T cells compared to A549.

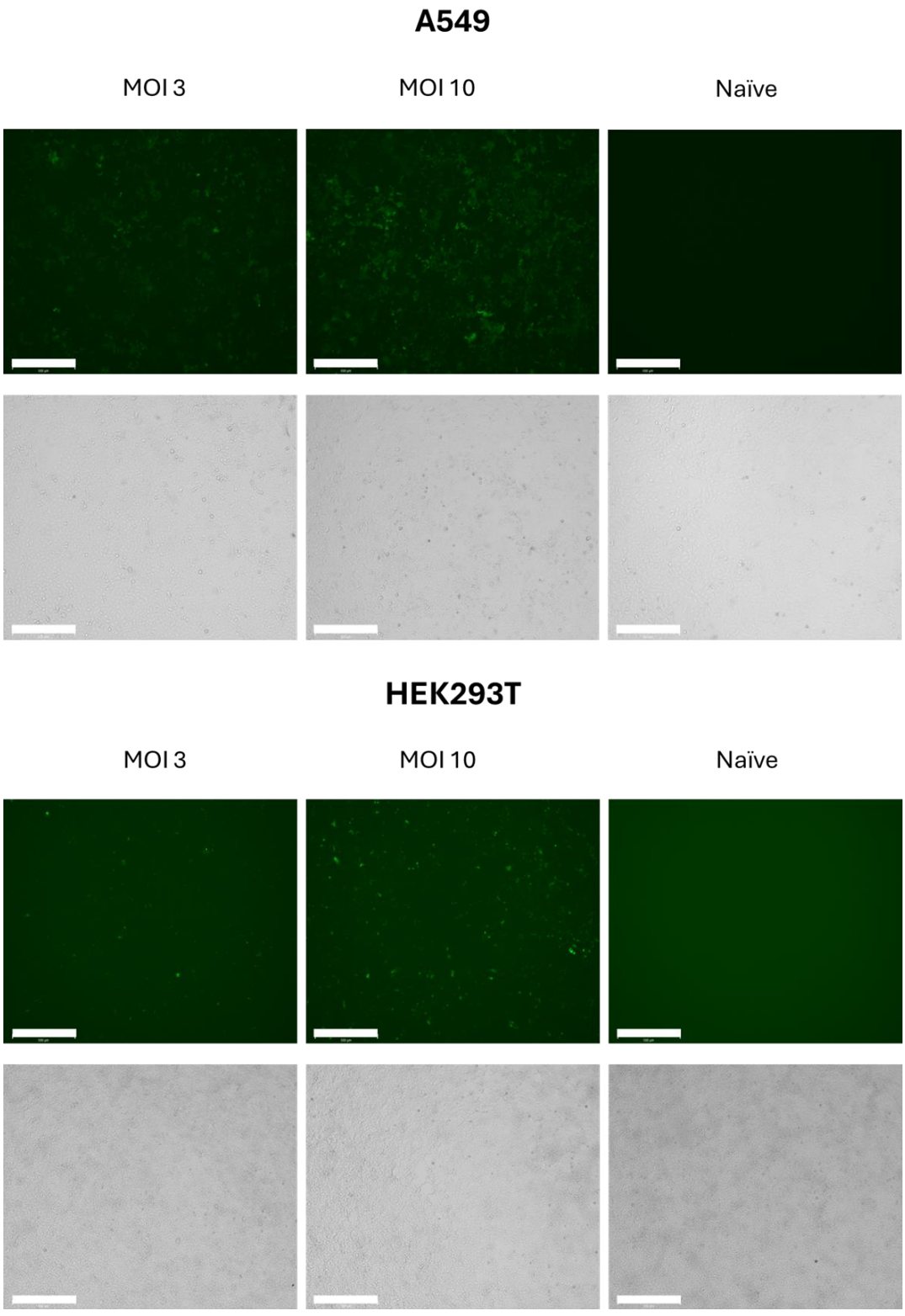

Figure 3.7. Fluorescence microscopy of HEK293T and A549 cells.

HEK293T and A549 cells were transduced with rSIV.hCEF.EGFP.VSV-G with MOI 3 or MOI 10, and images were taken 72 hours post transduction**.** Scale bar = 500 µm.

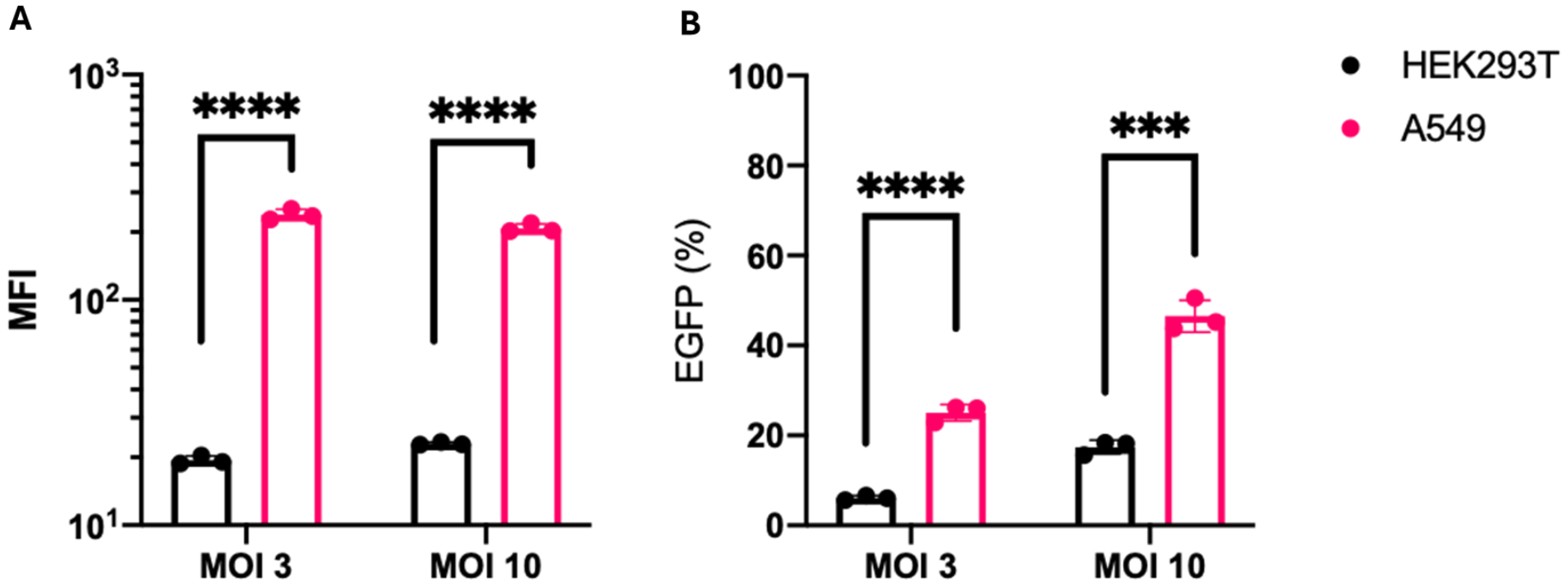

Figure 3.8. Quantification of Mean Fluorescence Intensity and EFGP positive cells.

A) Comparison of MFI (n=3 each) and B) Percentage of EGFP positive cells (n=3 each) in HEK293T and A549 cells transduced with rSIV.hCEF.EGFP.VSV-G using MOI 3 or MOI 10. Statistical analysis was conducted using an unpaired t-test to compare each group; $P < 0.001$ for all comparisons except for percentage of EGFP positive cells transduced with MOI 10, where $P < 0.001$.

3.3. Discussion and Future plans

All-in-one lentiviral delivery systems offer a neat solution for the delivery problems associated with large and complex PEs. However, in this chapter the all-in-one delivery approach was shown to pose quite a significant challenge namely the limited cargo capacity of the lentiviral backbone. In total, the size of the all-in-one construct was almost 10 kb, which is larger than the size of wild type SIV genome of around 9.5 kb[165] and HIV genome around 9.2 kb-9.6 kb[84]. There have been studies showing that genome plasmids larger than that of the wildtype virus will result in reduction of titres in lentiviral vector production[166–168]. In addition to vector genome size, it is also clear that expression of a lentiviral vectors transgene(s) can have a negative impact on vector titre[169] thus it also remains possible that unhindered expression of the PE machinery might have affected manufacturing vector titre.

Therefore, it was perhaps unsurprising that an all-in-one vector PE system would also suffer from reduced titres and compromised overall yields, when compared with the manufacture of (for example) an EGFP control vector that has vector genome size of 3.7 kb. Overall, there was ~1 log reduction of titres for both rSIV.949.VSV-G and rSIV.951.VSV-G, which was in line with what was described in previous publications[167]. In the small-scale production performed in this study, the resulting titre was still in the range of 1E6 TU, and was still usable for a small-scale transduction of cell lines. However, when considering the possibility of taking this strategy further for *in vivo* animal studies or even into human clinical trials, the reduced titre might pose a problem for scaling up vector production. If control of vector transgene expression would increase titre, the use of the transgene repression in vector production system[169] might enhance yield.

For experiments in this study, the hCEF promoter was selected as the ultimate aim was to deliver PE2-LV to the airway and/or lung-derived cell lines to evaluate prime editing for lung disease correction, such as to correct F508del mutation in *CFTR*. The hCEF promoter has been well characterised and performs well compared with the widely used CMV or EF1α promoters in murine airways *in vivo*[162]. Preliminary experiments were also planned in a cell line that may be more relevant as an airway cell model, for example the lung alveolar type II cell line, A549, or the human broncho-epithelial cell line, 16HBE14o-. Therefore, a promoter that expresses well in airway cells was selected. As more papers were published describing the challenges of prime editing in other cell lines[60,64], the strategy then changed to use human kidney epithelial HEK293T cells instead. Unfortunately, the activity of the hCEF promoter in the context of SIV vectors appears to be far from optimal in non-airway cell lines, and therefore the result that was obtained in this study with the all-in-one PE-LV approach in HEK293T cells did not show any editing of the *HEK3* locus.

In retrospect, there are several improvements that could be applied to the all-in-one PE-LV system. The first would be to incorporate smaller versions of prime editors. When these experiments were initiated, the first generation of prime editors (PE1, PE2) were the only ones available. As the field has moved ahead, new, smaller, prime editors have been developed by truncating the MMLV-RT domain[115,170] that might be more suitable for this approach. More importantly, a smaller prime editor would mean that reporter protein such as GFP or RFP could be added to the construct to ease visualization or selection of the transduced cells to enrich the PE transduced population.

In addition, the use of constitutive promoters such as CMV or EF1α needs to be considered especially for the preliminary testing in HEK293T cells. While hCEF has been predominantly developed for lung airway expression, its low activity in HEK293T cells likely resulted in the insufficient expression of the prime editor and therefore affected the editing efficiency. Thus, for future studies using the all-in-one strategy, it might be worth testing the system using a CMV or an EF1α promoter in HEK293T cells, before adapting the system to target the airways.

In conclusion, the all-in-one PE-LV system is an interesting strategy and with several improvements might be optimised for success.

Chapter 4: Prime Editing and HITI: Two-vector delivery approach

4.1. Introduction

One considerable limitation of the all-in-one lentiviral delivery approach is the packaging constraint in the lentiviral vectors. There is little capacity for a reporter protein or antibiotic resistance gene to be incorporated into the all-in-one construct in addition to the prime editor, without severely reducing the viral titre during virus production. As editing in the target site is not always observed, the addition of reporter gene is helpful to verify that transfection of the plasmid, or transduction of the lentiviral vector has been successful, and that the lack of editing can therefore be attributed to other factors. It is appreciated that while inclusion of such markers is helpful during development studies, it would be inappropriate to retain them in products destined for human use. Therefore, as a proof of concept, I devised a two-vector delivery approach to overcome this limitation, as well as to test an alternative gene editing technique, namely HITI. In a previous study[171], superexon donor integration was used as a strategy to integrate exon 11 to exon 27 of *CFTR* (termed a superexon) using a zinc finger nuclease. I adapted this strategy by using a Cas9-HITI approach to insert the superexon donor template into *CFTR* intron.

In this chapter, I describe the selection process for choosing the best lentiviral backbone for this approach and the gRNA. I also describe the design of the dual vector approaches for both HITI and PE of the *CFTR* F508del mutation, the proof-of-concept experiments using two-vector delivery of PE to edit the *HEK3* site, as well as the two-vector delivery of HITI to insert superexon donor template into the *CFTR* locus in HEK293T cells.

4.2. Results

4.2.1. Determining the best lentiviral vector for two-vector delivery

In Chapter 3, the SIV-based lentiviral vector with the hCEF promoter was used for development of the all-in-one PE delivery strategy. Considering the low transduction efficiency achieved with the F/HN pseudotype and the low activity of the hCEF promoter especially in HEK293T cells, I performed an experiment to compare the SIV based lentiviral vector with the hCEF or more commonly used CMV promoter, as well as the HIV based lentiviral vector with the CMV promoter. A lung-relevant, human bronchial-epithelial cell line (16HBE14o-) was used as the transduction model. The aim was to choose the best lentiviral vector / promoter combination with which to generate the two-vector system for both PE and HITI delivery.

Three different vectors expressing EGFP were compared (Table 2.8): rSIV.F/HN.hCEF for the all-in-one approach, rSIV.F/HN.CMV and rHIV.VSV-G.CMV using MOI ranging from 125 to 1500 (Figure 4.1). From these three vectors, rHIV.VSV-G.CMV.EGFP performed the best in terms of transduction efficiency even at the lowest MOI of 125. For the SIV-based vectors, rSIV.F/HN.CMV.EGFP performed slightly better than rSIV.F/HN.hCEF.EGFP as indicated by the increased EGFP signal. However, the rHIV.VSV-G.CMV.EGFP outperformed the rest, therefore a comparison study with r.SIV.VSV-G.CMV.EGFP vector (Table 2.8) was conducted.

From the comparison of rHIV.VSV-G.CMV.EGFP and r.SIV.VSV-G.CMV.EGFP, MOI of 12.5, 25, and 75 were used for rHIV.VSV-G.CMV.EGFP, and MOI of 312.5, 625, and 1250 were used for rSIV.VSV-G.CMV.EGFP (Figure 4.2). The contrast in the MOI used for rHIV and rSIV was

based on the previous experiment where rHIV performed significantly better even in the lowest MOI, and it was shown that cells transduced with rHIV.VSV-G.CMV.EGFP had significantly higher EGFP expression even at the lowest MOI of 12.5. Therefore, the HIV-based lentiviral vector together with the VSV-G pseudotype and CMV promoter was the preferred configuration for the two-vector approach.

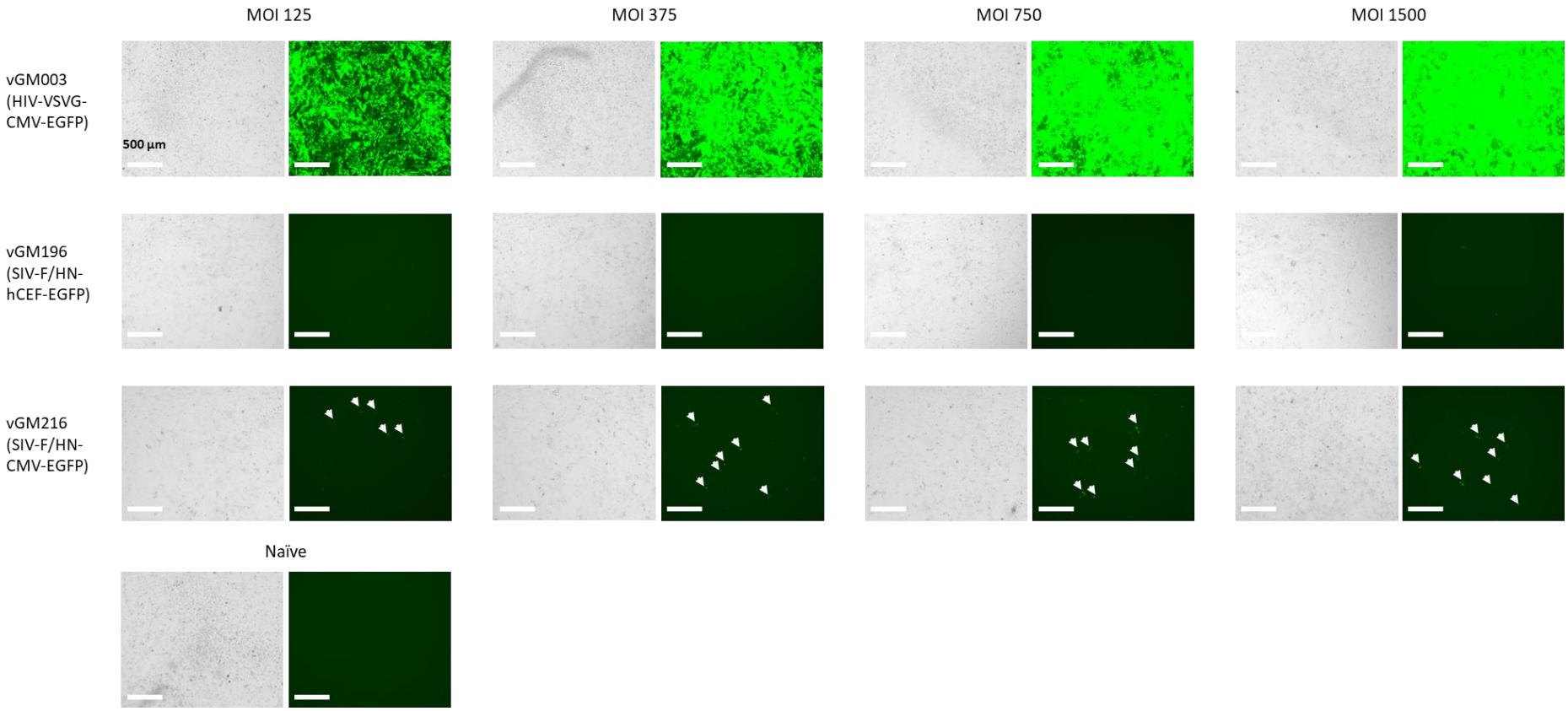

Figure 4.1. Comparison of three lentiviral vectors for transduction of 16HBE14o- cells.

Figure 4.1. Comparison of three lentiviral vectors for transduction of 16HBE14o- cells.

Three vector configurations were used to transduce 16HBE14o- cells: rSIV.F/HN.hCEF.EGFP, rSIV.F/HN.CMV.EGFP, and rHIV.VSV-G.CMV.EGFP. At the lowest MOI of 125, rHIV.VSVG.CMV.EGFP produced efficient transduction and a high level of EGFP expression. The SIV based lentiviral vectors could not transduce efficiently and a very low EGFP expression was detected. White arrows indicate EGFP signals. Scale bar is 500µm.

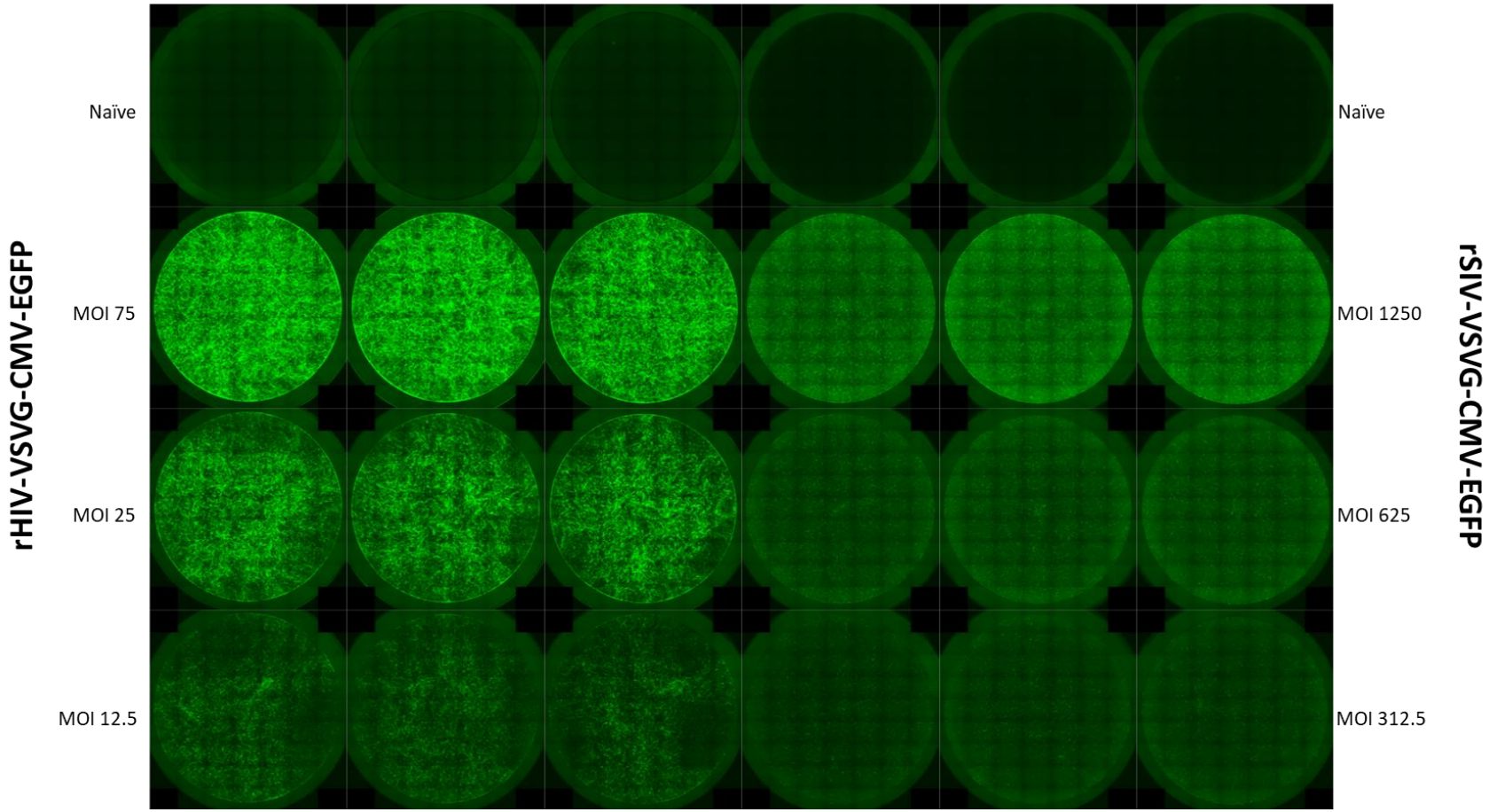

Figure 4.2. Transduction of 16HBE14o- cells with rHIV.VSV-G.CMV.EGFP or rSIV.VSV-G.CMV.EGFP.

Figure 4.2. Transduction of 16HBE14o- cells with rHIV.VSV-G.CMV.EGFP or rSIV.VSV-G.CMV.EGFP

Imaging of a 24-well plate of 16HBE14o- cells transduced with either rHIV.VSV-G.CMV.EGFP (triplicates on the left-hand side) or rSIV.VSV-G.CMV.EGFP (triplicates on the right-hand side) at the MOIs indicated. Cells transduced with rHIV.VSV-G.CMV.EGFP showed significantly higher EGFP expression even at the lowest MOI of 12.5.

4.2.2. Guide RNA selection for HITI and PE

The editing efficiency of genome editors, including Cas9-dependent HITI and PE, is highly dependent on the efficiency of the gRNA spacer sequence to bind to its complementary sequence in the target genome[1]. Therefore, potent and efficient gRNA spacer sequences are necessary for both HITI and PE. To identify optimal and efficient gRNA spacer sequences for HITI and PE of the *CFTR* F508del mutation, six PE gRNAs and sixteen HITI gRNAs were designed using CRISPOR[51] (http://crispor.gi.ucsc.edu/), using the F508del-containing *CFTR* sequence as the reference. These gRNAs are graphically represented in Figure 4.3.

As the HEK293T cell line model lacks the F508del mutation, one spacer candidate for PE (PE spacer 1) could not be tested as it was dependant on the presence of the F508del mutation itself as part of its spacer sequence. Another five spacer candidates for HITI (spacers 8-12) were also not tested as these spacers bind to a region where a common single nucleotide polymorphism confers either the *CFTR* variant M470 or V470. Therefore, the PE spacers 8-12 were deemed not compatible for all the *CFTR* variants. The remaining spacer candidates were tested in HEK293T cells to check for their respective editing efficiency after transient transfection with a Cas9-encoding plasmid (pCMV.Cas9.T2A.GFP, pGM1101) and the guide RNA encoding plasmids (gRNA plasmids, Table 2.7). Seventy-two hours post transfection, GDNA was extracted, and the relevant regions targeted by the guide RNAs were PCR amplified. Editing efficiency was then determined using the T7E1 assay (section 2.10), and the editing efficiency was calculated using densitometry. Briefly, quantification of the bands observed after gel electroporation was determined using FIJI, and the editing efficiency was estimated as described in section 2.10.

From the quantification, two HITI gRNA spacers (HITI gRNAs 1 and 3) and one pegRNA spacer (PE spc 2) had the highest percentage of gene modification: for each HITI 1, HITI 3, and PE spc 2 these are 5.50 % ± 4.86%, 4.83% ± 4.94%, and 8.73% ± 10.17%, respectively, (Figure 4.4). However, as the editing efficiency was not consistent between replicates, the standard deviations were considerable and therefore there was no statistically significant difference between the candidates. For HITI, the two spacer sequences were selected and the respective donor templates for each gRNA were designed and is described in section 4.2.5. For PE, the lack of mutation in the HEK293T cell line prevented evaluation of the widely utilised spacer in publications[155,172,173] for correction of the F508del mutation, which is the PE spacer 1 (5'ACCATTAAAGAAAATATCAT3', Table 2.5). Therefore, PE spacer 2 (5'TCTGTATCTATATTCATCAT3', Table 2.5) will be used for screening in targeting the F508del region in *CFTR*.

.

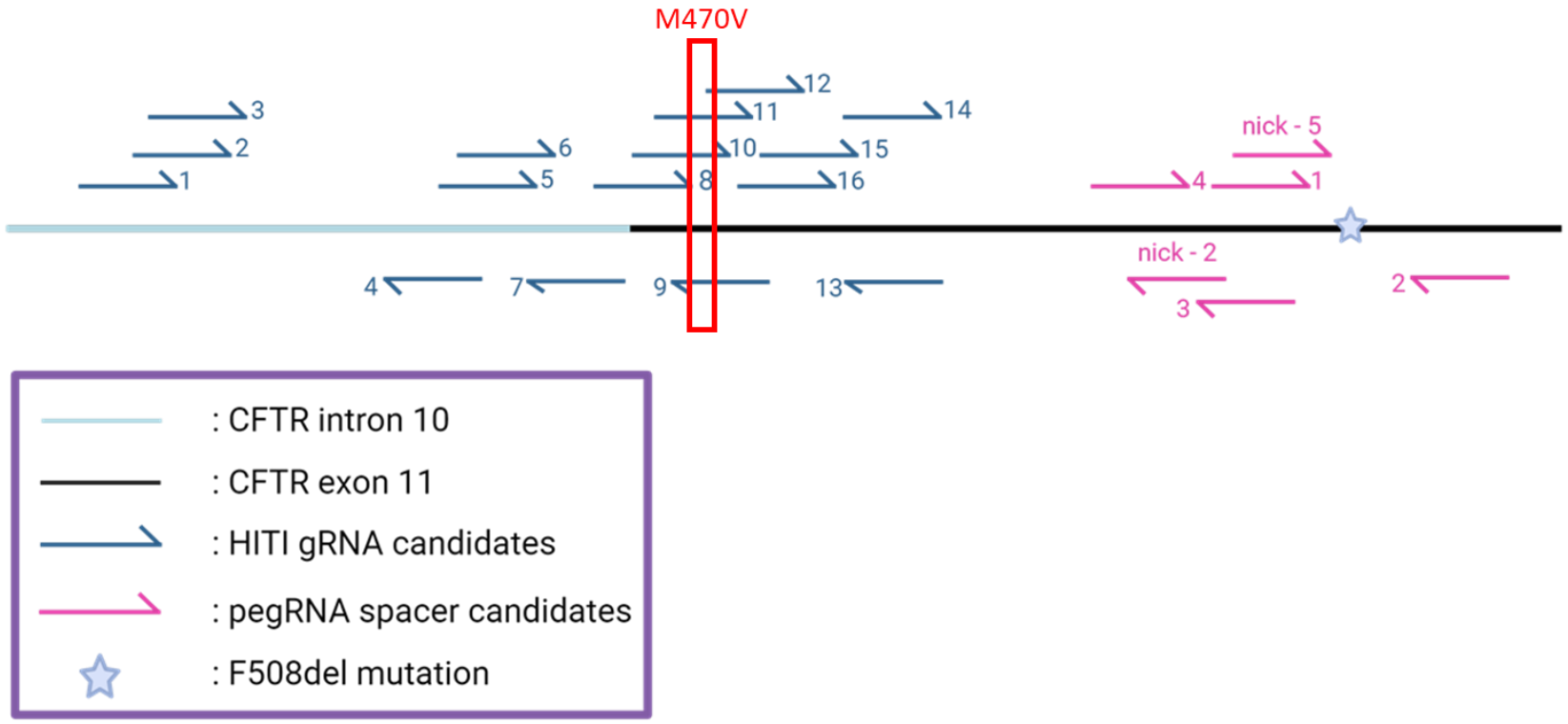

Figure 4.3. Diagram of candidate gRNA spacer sequences for HITI and PE of the *CFTR* F508del mutation.

Figure 4.3. Diagram of candidate gRNA spacer sequences for HITI and PE of the *CFTR* F508del mutation.

Guide RNA spacer sequences are shown for HITI (blue arrows) and PE (pink arrows) against the human *CFTR* gene for the editing of the *CFTR* F508del mutation. Positions are shown relative to the CFTR F508del mutation in the genome (indicated by a star on black line); distance in figures are not to scale. HITI spacer candidates span from intron 10 (light blue line) to exon 11 (black line) of the *CFTR* gene, while PE spacer candidates are located close to the mutation sites. Due to the lack of the relevant F508del mutation, some of the gRNAs listed above such as the PE spacer 1 could not be tested. For HITI, the sequences of spacers 8-12 partially overlaps to amino acid 470, which confers either the M470 or V470 variant of *CFTR* (red box). Ultimately, this compromises the design of HITI spacers 8-12, which were not further evaluated as they would not be considered universally compatible for all CF variants.

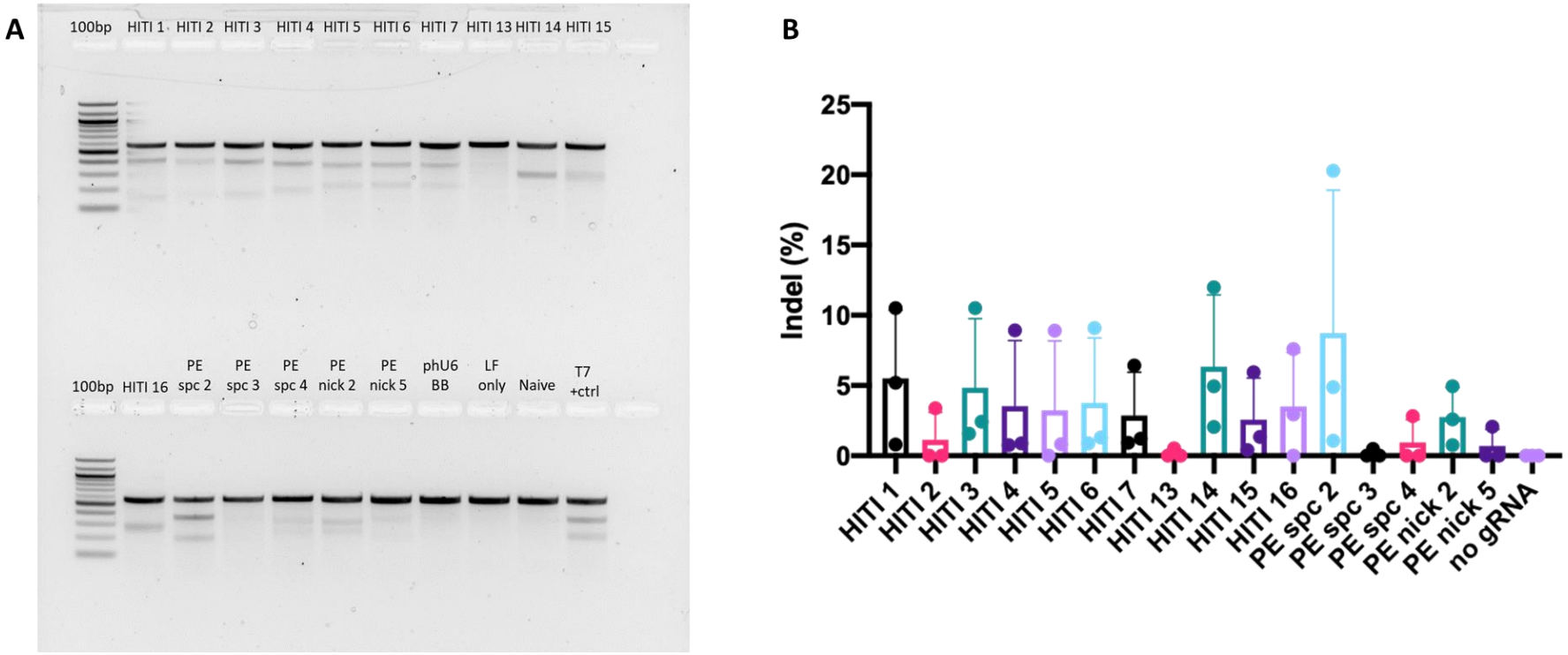

Figure 4.4. T7E1 assay-based quantification of indels after transfection of gRNA and Cas9 plasmid in HEK293T cells.

Twelve gRNA candidates for HITI and five gRNA candidates for PE targeting the regions close to the F508del mutation in *CFTR* were co-transfected with Cas9 plasmid. Based on the agarose gel electrophoresis image (A) post-PCR, percentage of indels were quantified for each gRNA candidate (n = 3) (B). Ordinary one-way ANOVA was performed to compare the means of percentage indel from each sample; no significant difference was found (P>0.05).

4.2.3. Testing pegRNA candidates in the F508del region

Based on the results of the guide RNA selection, I designed eight pegRNA candidates using pegIT[174] (https://pegit.giehmlab.dk/) to correct the F508del mutation using PE spacer 2 as the spacer sequence. Since HEK293T cells, which were used as the model for testing the pegRNAs, do not carry the mutation, I opted to introduce a silent edit (changing TC**C** to TC**G**) in the codon for serine at amino acid position 511 (S511) of CFTR to measure the editing efficiency. A total of eight pegRNA candidates (Table 4.1) were cloned into pU6.Sp.pegRNA.GG_acceptor (Table 2.7), and used for transfection with PE2-GFP encoding plasmid (pCMV.PE2.P2A.GFP, Table 2.7). Each pegRNA candidate is referred to by the length of the primer binding site and the RTT, for example, pegRNA 15.21. Seventy-two hours post transfection, cells were imaged, GDNA was extracted, and the relevant regions targeted by the guide RNAs were PCR amplified (Table 2.4). Editing efficiency was then determined using ICE and DECODR analysis.

Initial analysis with ICE indicated that pegRNA 15.21 successfully introduced the correct C→G edit at the target site with an efficiency of 10% (Figure 4.5). Although the initial result seemed promising, upon repeating the experiment, no editing was observed Figure 4.6. Even the positive control, involving transfections with pCMV.PE2 or pCMV.PE2.P2A.GFP and pU6.Sp.pegRNA.HEK3_CTT_ins, failed to show any editing (data not shown). To not add more complexity to the experimental design, I then decided to move forward using *HEK3* locus as a target for the two-vector approach.

Table 4.1. List of pegRNA candidates for silent edit introduction in S511

No	Spacer-PE spc 2	Primer Binding Site	RT Template	PBS length	RT length
1	tctgtatctatattcatcat	atgaatatagataca	tcatctttggtgtttcgtatg	15	21
2	tctgtatctatattcatcat	atgaatatagataca	catctttggtgtttcgtatg	15	20
3	tctgtatctatattcatcat	atgaatatagataca	atctttggtgtttcgtatg	15	19
4	tctgtatctatattcatcat	atgaatatagatac	aaaatatcatctttggtgtttcgtatg	14	27
5	tctgtatctatattcatcat	atgaatatagata	atcatctttggtgtttcgtatg	13	22
6	tctgtatctatattcatcat	atgaatatagata	tcatctttggtgtttcgtatg	13	21
7	tctgtatctatattcatcat	atgaatatagata	catctttggtgtttcgtatg	13	20
8	tctgtatctatattcatcat	atgaatatagata	atctttggtgtttcgtatg	13	19

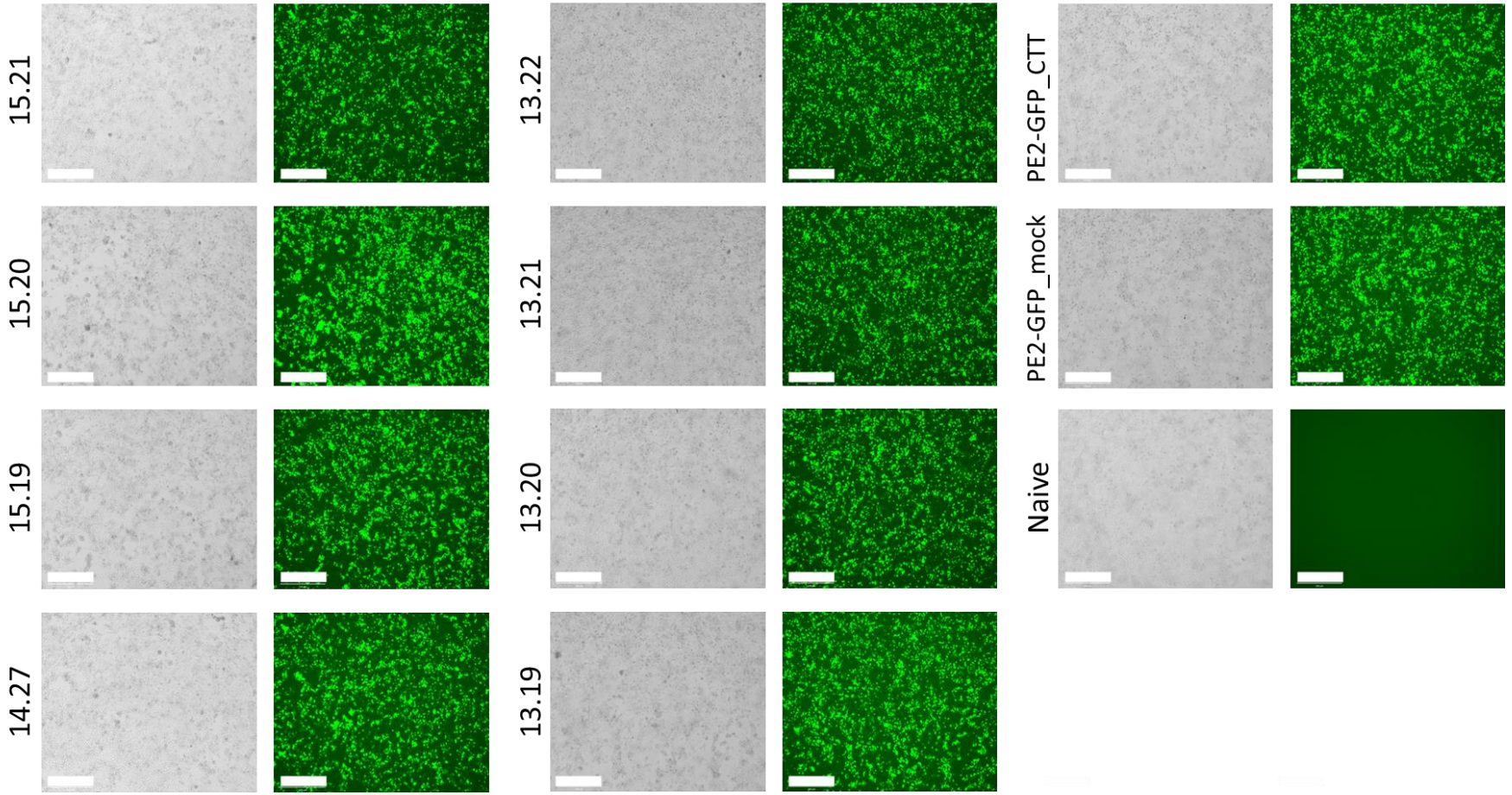

Figure 4.5. Fluorescence microscopy image of HEK293T cells transfected with pCMV.PE2.GFP and pegRNA.

Based on the level of GFP expression, **e**fficient transduction was observed in all of the transfected samples

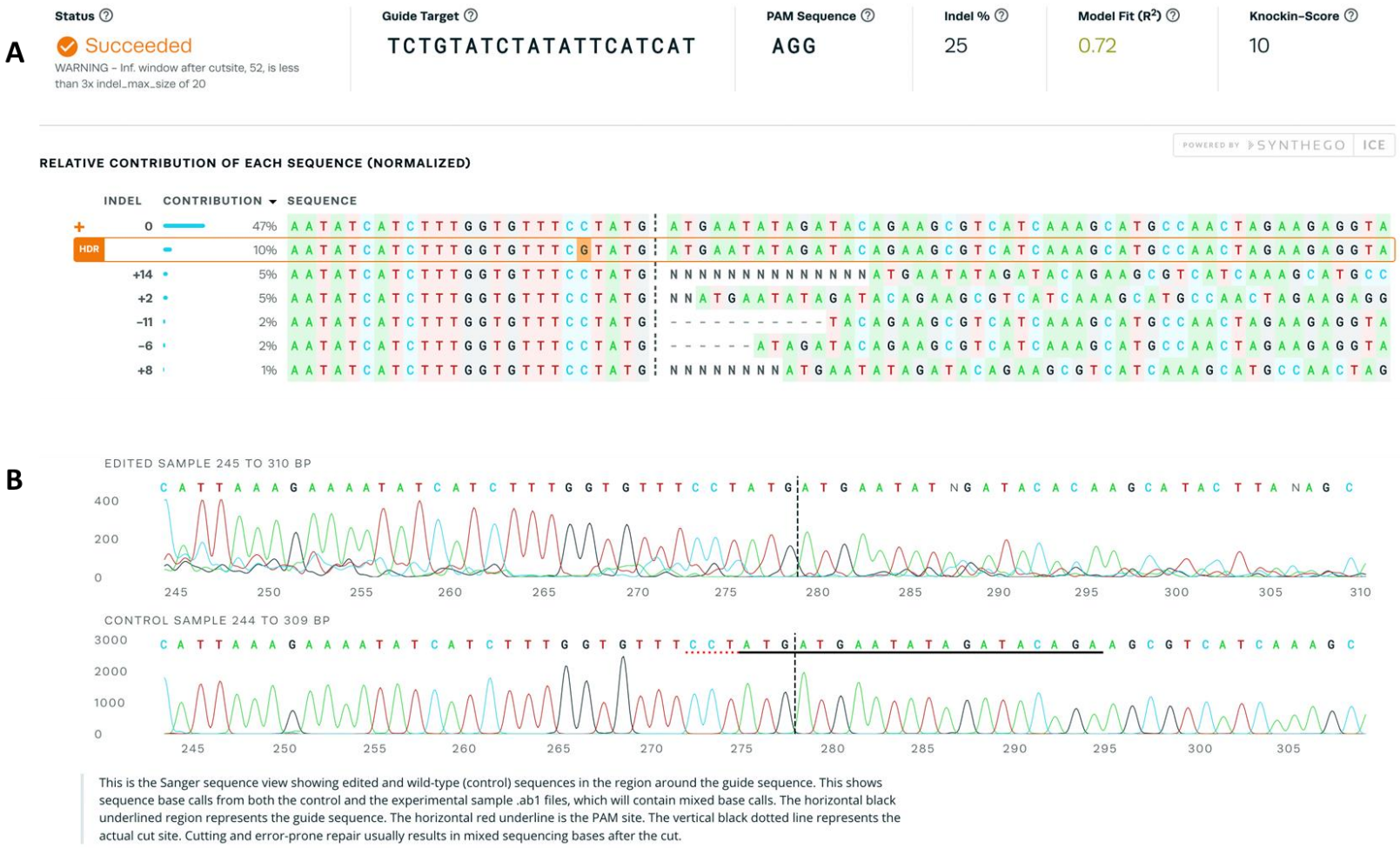

Figure 4.6. ICE Analysis of silent edit introduction in S511.

A) Sample transfected with pegRNA 15.21 showed 10% silent edits of TC**C** →TC**G.** B) Sequencing traces of sample transfected with pegRNA 15.21.

4.2.4. Designing the two-vector approach for PE

To generate the plasmid constructs for the two-vector approach for PE, the PE components, pegRNA, and additional reporter protein were cloned into separate lentiviral genome plasmid backbones, from here onwards referred to as 'split PE plasmids'.

The hurdle for delivery is the size of the PE encoding gene, which is approximately 6.7 kb. Using a reference from a previously published study[160,175] the Cas9n component of PE was split into the N-terminus from amino acid 1-713 (termed Cas9n N-term), and the C-terminus from amino acid 714-1368 (termed Cas9n C-term). To facilitate the trans-splicing of these two halves of Cas9n protein, intein from *Nostoc punctiforme* (Npu) was used[176]. The sequence of N-terminus Npu intein was added at the 3' end of the Cas9n N-term followed by the P2A sequence and EGFP (Figure 4.5A, top). The C-terminus Npu intein was added at the 5' end of the Cas9n C-term, with linker and MMLV-RT added to its 5' end (Figure 4.7A, bottom).

Using an in-house lentiviral genome plasmid as a backbone, pGM378 (pHIV.CMV.EGFP, Table 2.7), the components were then cloned into two separate plasmids to construct the split PE plasmids. The first plasmid (pGM1151, Table 2.7) encodes the CMV promoter driving the expression of the N-terminus of the Cas9 nickase, linked to the N-terminus of intein, followed by P2A self-cleaving peptide sequence and GFP reporter. On the 5' end of the CMV promoter, the hU6 promoter is driving the expression of pegRNA (Figure 4.7B). The total size of this plasmid is 9854 bp, with 7130 bp in between the 5' and 3' LTRs. The second plasmid encodes the CMV promoter driving the expression of C-terminus intein and C-terminus of Cas9 nickase, linked to the MMLV-RT (Figure 4.7C). The total size of this

plasmid is 10506 bp, with 7878 bp in between the 5' and 3' LTRs. To test that the Split PE two-vector approach is a feasible strategy, I first opted to use the pegRNA design that targets the *HEK3* site described extensively in Chapter 3 (section 3.2.1). The two-vector plasmid was then used for transfection of HEK293T cells to determine whether the desired +1CTT insertion could be achieved.

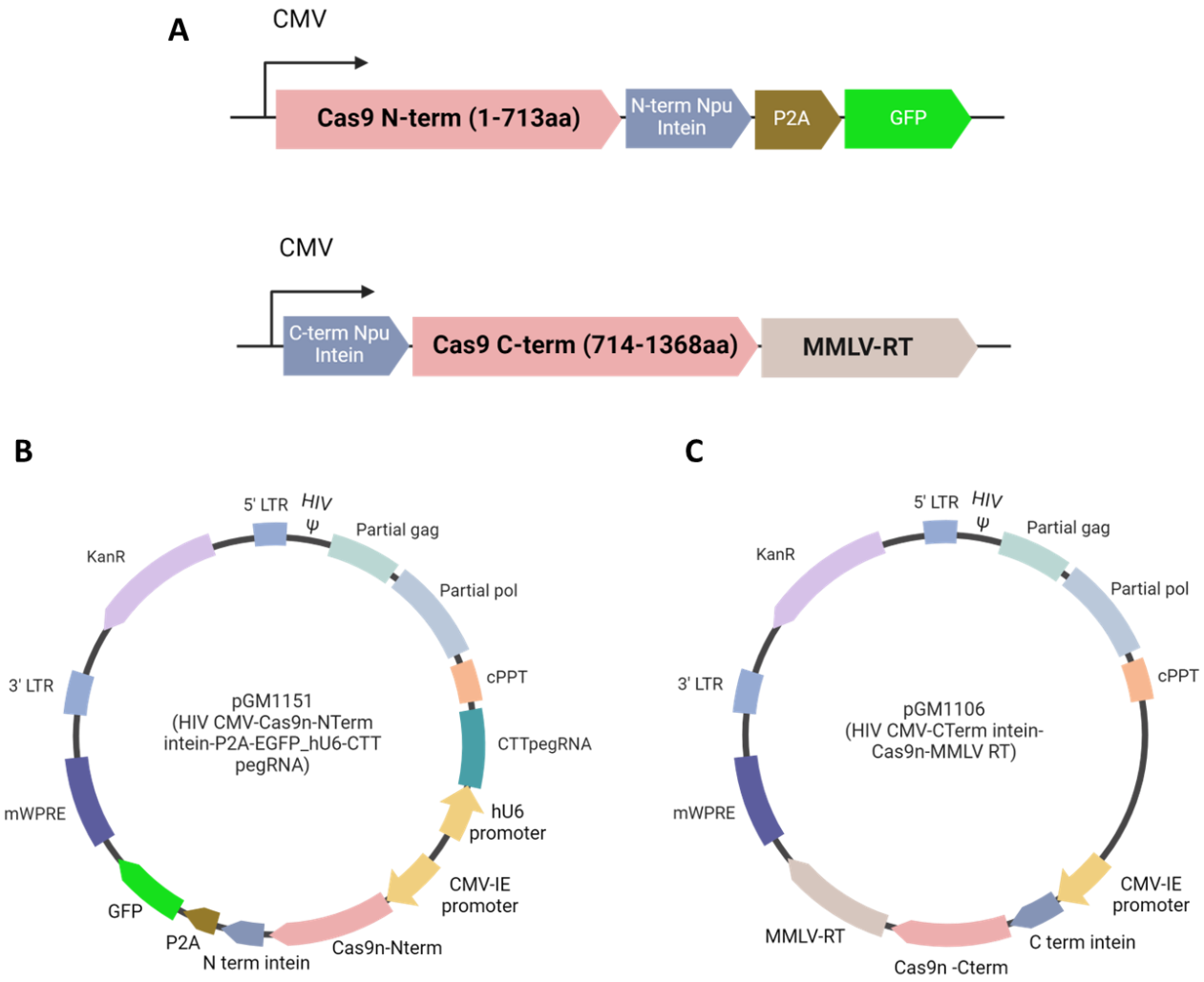

Figure 4.7. Two components of the split PE plasmids.

Figure 4.7. Two components of the split PE plasmids.

A) The first component of split PE includes Cas9 N-term (1-713 aa), N-terminus of Npu intein, P2A, and EGFP (top). The second component of split PE contains C-terminus of Npu intein, Cas9 C-term (714-1368 aa), linker sequence (not depicted), and MMLV-RT (bottom). B and C) Two-vector delivery approach for prime editing using HIV vector genome backbone for both plasmids. B) The first plasmid, pGM1151 encodes CMV promoter driving the expression of the first component of split PE. On the 5' end of the CMV promoter in the first plasmid, hU6 promoter is driving the expression of the pegRNA targeting the *HEK3* site. C) The second plasmid, pGM1106, encodes CMV promoter the expression of the second component of split PE. Image created in BioRender.com.

4.2.4.1. Intein mediated Split PE trans-splicing

The components of the split PE in the two-vector delivery system are linked by intein sequences to mediate trans-splicing of the two halves of the effector PE molecule after their peptide translation in the same cell. The intein catalyses a splicing reaction ligating the N-terminus and C-terminus of split PE fragments and cleaving itself out in the process[176]. This results in the formation of a full-length protein, consisting of full-length Cas9n linked to MMLV-RT, while the intein is removed as a byproduct. It is important to ensure that the split PE components are capable of assembling correctly in the cell line post-transfection. To examine this, HEK293T cells were co-transfected with the LV genome components of the split PE two-vector system, following which western blotting was used to evaluate the intein-mediated assembly of full-length PE using antibody raised against the N-terminus of the original SpCas9 molecule, therefore, the C-terminus component of split PE could not be detected with this antibody (Figure 4.6A).

Western blot confirmed the expression and assembly of the full-length PE effector molecule (~270 kDa; red box in Figure 4.6B) in HEK293Ts transiently transfected with LV genome components encoding both halves of the intein-containing PE sequence. In the same sample, a band of a lower molecular weight (~100 kDa) corresponding to the size of the Cas9 N-terminus was also detected at a much higher intensity. When compared with the control that was transfected with a plasmid encoding PE2, however, the band from split PE trans-splicing was less intense. This indicated that the trans-splicing process was not very efficient and could potentially affect the editing efficiency in the two-vector system.

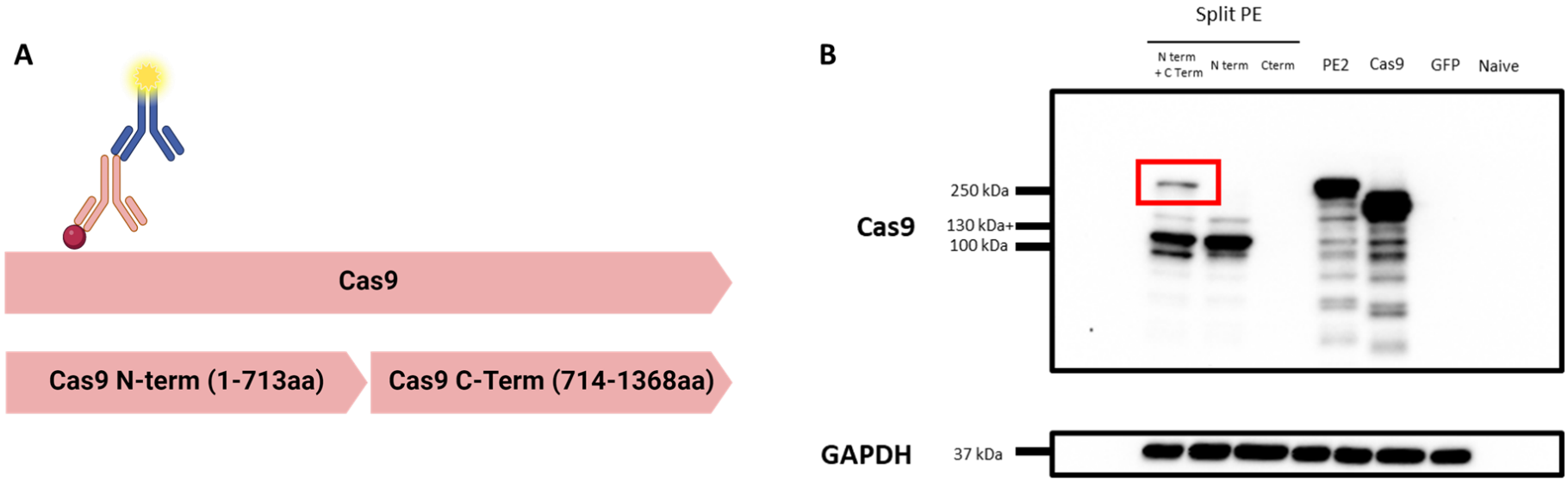

Figure 4.8. Western blotting of split PE trans-splicing in HEK293T cells.

A) Schematic showing the binding site of the antibody used in the western blotting experiment. B) Western blotting of split PE trans-splicing post transfection in HEK293T cells. The resulting band from ligated split PE (lane 1 from left, Split PE N term + Cterm) is less dense compared with the band from PE2 transfected cells (lane 3), suggesting that the trans-splicing of the two fragments is not very efficient. Figure 4.8A was created in BioRender.com.

4.2.4.2. Transfection of Split PE two-vector plasmids into HEK293T cells

Having confirmed that the intein-containing split PE sequences can ligate to assemble the expected full-length PE effector molecule, the activity of the assembled PE effector molecule in mediating the desired +CTT edit at the *HEK3* site was assessed. To measure this, HEK293T cells were transfected with the split PE plasmids, pGM1151 (encoding the N-terminus of split PE and the *HEK3* +CTT pegRNA driven by the human U6 promoter) and pGM1106 (encoding the C-terminus of N-terminus of split PE) (Table 2.7). As a control for the desired +CTT insertion at the *HEK3* locus, cells were also transfected with both the PE2-GFP and *HEK3* +CTT pegRNA expressing plasmids (pCMV.PE2.P2A.GFP and pU6.Sp.pegRNA.*HEK3*_CTT_ins plasmids respectively, Table 2.7) (Figure 4.9).

From fluorescence microscopy, it appears that the level of transfection is similar between the samples transfected with pGM1151 and pGM1106, and samples transfected with PE2-GFP and HEK3 +CTT pegRNA expressing plasmids (Figure 4.9A). Encouragingly, ICE analysis of Sanger sequencing data showed that both the split PE and the control (written as PE2-GFP + CTT pegRNA in the figure legend) approaches resulted in similar levels of editing efficiency, with Split-PE plasmid transfection achieving (11.33%±4.04%) of editing, compared to (17.67%±5.86%) of editing obtained through PE2-GFP and *HEK3* +CTT pegRNA plasmid transfection (Figure 4.9B). The traces from the samples transfected with the split PE plasmids also showed the expected CTT insertion at +1 position (Figure 4.9C). Thus, despite the apparent low efficiency of intein-mediated split PE trans-splicing (Figure 4.8), this did not appear to be limiting the rate of editing observed. Together, these observations

support the use of the two-vector approach and suggest this would be a viable strategy for effecting PE in target cells.

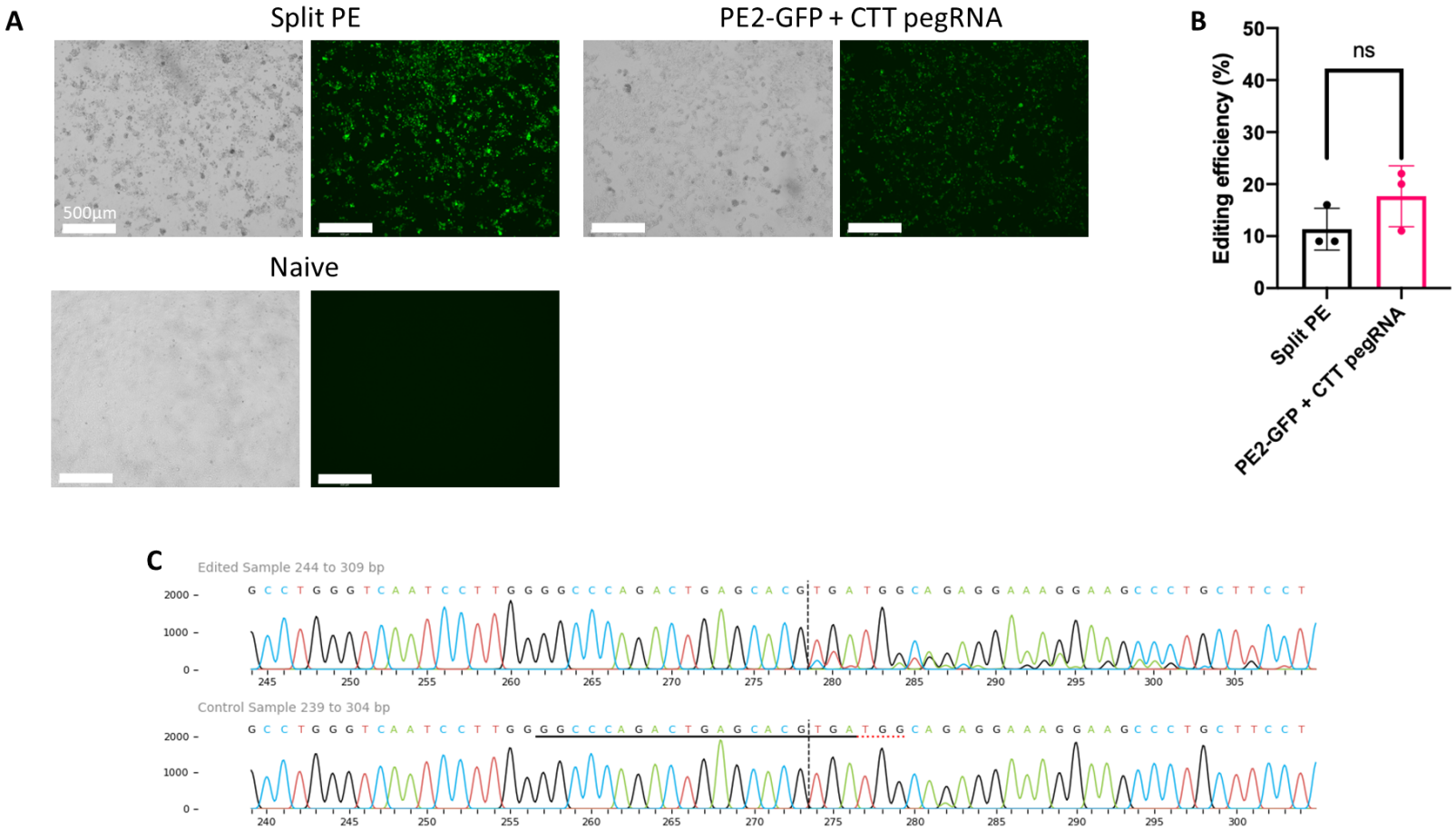

Figure 4.9. Transfection of Split-PE plasmids

Figure 4.9 Transfection of Split-PE plasmids

A) Brightfield and fluorescence microscopy for EGFP fluorescence between split PE and PE2-GFP + CTT pegRNA transfections of HEK293T cells, relative to naïve cells as control. B) Editing efficiency in HEK3 site from split PE and PE2-GFP + CTT pegRNA transfected cells (n = 3) analysed by Sanger sequencing and Synthego ICE analysis. There are no significant differences ($P>0.05$) in editing efficiency between split PE and PE2-GFP + CTT pegRNA transfected cells. C) Representative trace image from Split PE transfected sample.

4.2.5. Two-vector delivery approach for HITI

The two-vector delivery approach for HITI was devised as an alternative to the two-vector PE approach. HITI involves inserting a donor sequence, via NHEJ, into a Cas9-mediated, gRNA-targeted double strand break. In this study, I adapted the strategy from Bednarski et al.[171] whereby a superexon donor template encompassing exons 11-27 of the wildtype *CFTR* gene could be incorporated to correct the underlying CF mutation. The basis for incorporating *CFTR* superexon 11-27 as the donor template for HITI, rather than just exon 11 where the F508del mutation is located, is that 86.47% of CF patients have more than one mutation identified in the *CFTR* gene[177], with the majority of the UK CF population possessing one or more *CFTR* mutations downstream of exon 10. (https://www.cysticfibrosis.org.uk/sites/default/files/2023-10/CFT_2022_Annual_Data_Report_FINAL_v8.pdf). Therefore, the superexon integration strategy would be useful for many CF patients, without the need for customisation for every patient.

From the gRNA selection experiment (section 4.2.2), HITI gRNA 1 and HITI gRNA 3 were selected as the best performing gRNA for HITI. Therefore, the donor templates used in this study was designed for use of gRNA 1 and gRNA 3. The sequence in the HITI donor templates includes the *CFTR* intron 10 sequence, intron splicing site signal, and the superexon 11 to 27 of *CFTR*. These sequences were then flanked with the respective gRNA target site for HITI gRNA 1/Donor 1 and HITI gRNA 3/Donor 3 with PAM sequence included, placed in reverse orientation (Figure 4.10A). Upon delivery of the HITI components, the correct insertion of the HITI donor will occur in the intron 10 region of *CFTR* (Figure 4.10B).

The minor difference of the cut site of the HITI 1 gRNA and HITI 3 gRNA affects the length of the intron 10 sequence included in the template, where donor 1 includes 14 bp more of intron 10 compared with donor 3 (Figure 4.10C). Complete sequences for HITI donor 1 and HITI donor 3 are attached in the Appendix.

In order to achieve HITI of the *CFTR* superexon donor template to correct CF mutations, I first designed and produced the LV genome plasmids required for the HITI approach (Figure 4.11). The first LV genome plasmid contains Cas9-T2A-GFP expression cassette driven by the CMV promoter (pHIV.CMV.Cas9.T2A.GFP, internal code pGM1100) with total plasmid size of 11058 bp, and 8348bp between the 5' and 3' LTRs. I also produced two LV genome plasmids containing the HITI gRNA and the HITI donor template, namely pGM1115 which has HITI gRNA 1 and HITI donor template 1; and pGM1116 which has HITI gRNA 3 and HITI donor template 3 (Figure 4.11; plasmid Table 2.7). The total size of pGM1115 is 8814 bp, with 6104 bp in between LTRs and the total size of pGM1116 8800 bp, with 6090 bp in between LTRs.

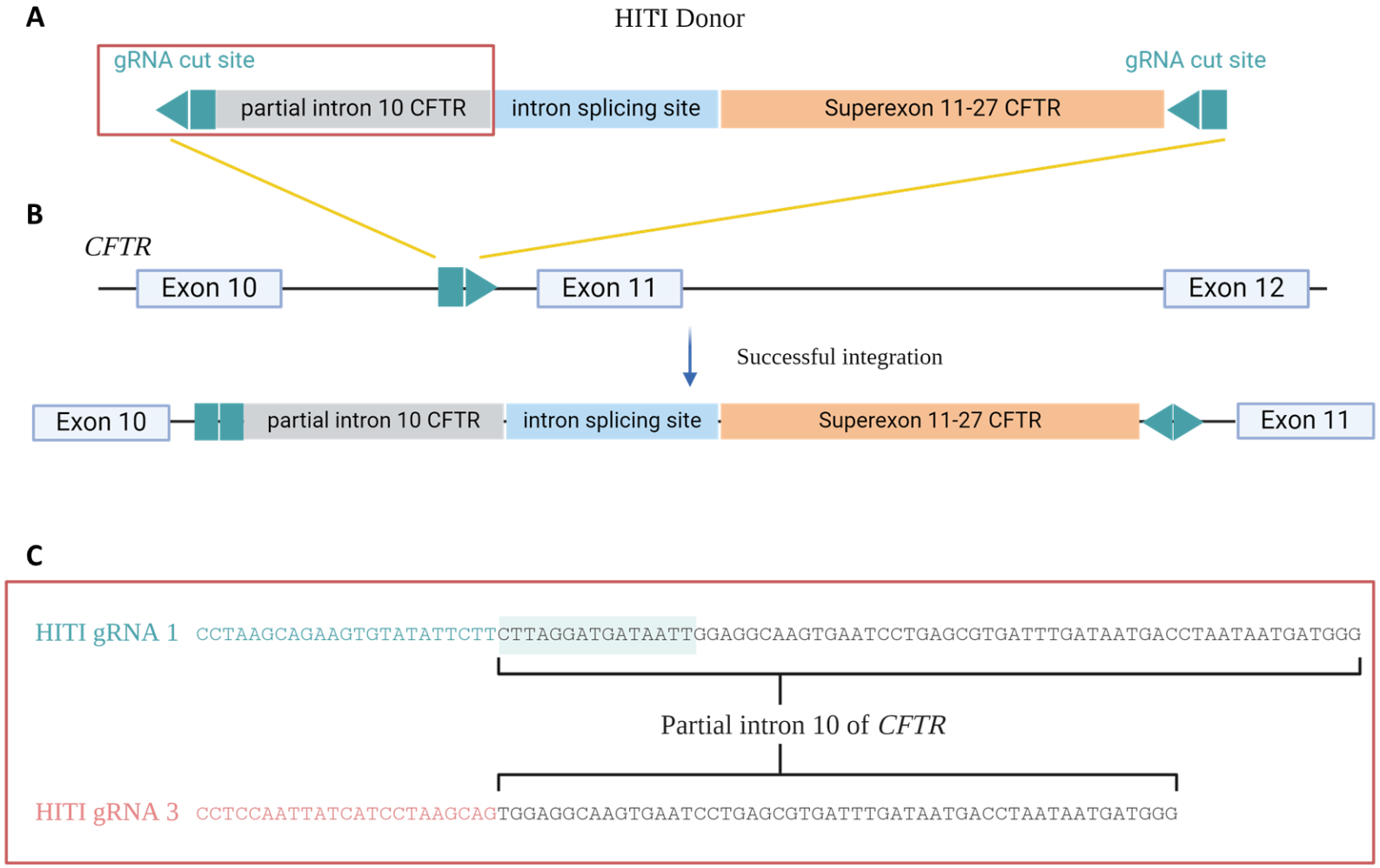

Figure 4.10. Schematic of HITI donor configuration and integration into *CFTR* target locus.

Figure 4.10. Schematic of HITI donor configuration and integration into CFTR target locus.

A) Configuration of HITI donor template including gRNA cut site on both 5' and 3' ends, with partial *CFTR* intron 10 sequence, intron splicing site, and superexon 11-27 of *CFTR*. B) Position of HITI integration site in the genome. The target site for both HITI gRNA 1 and HITI gRNA 3 is located in the intron 10 of *CFTR*. When successful integration occurs in the correct orientation, the gRNA cut site is ablated ensuring that the construct is permanently inserted into the genome. C) Details of the sequences highlighted with a red box in Figure 4.8A. Image created in BioRender.com.

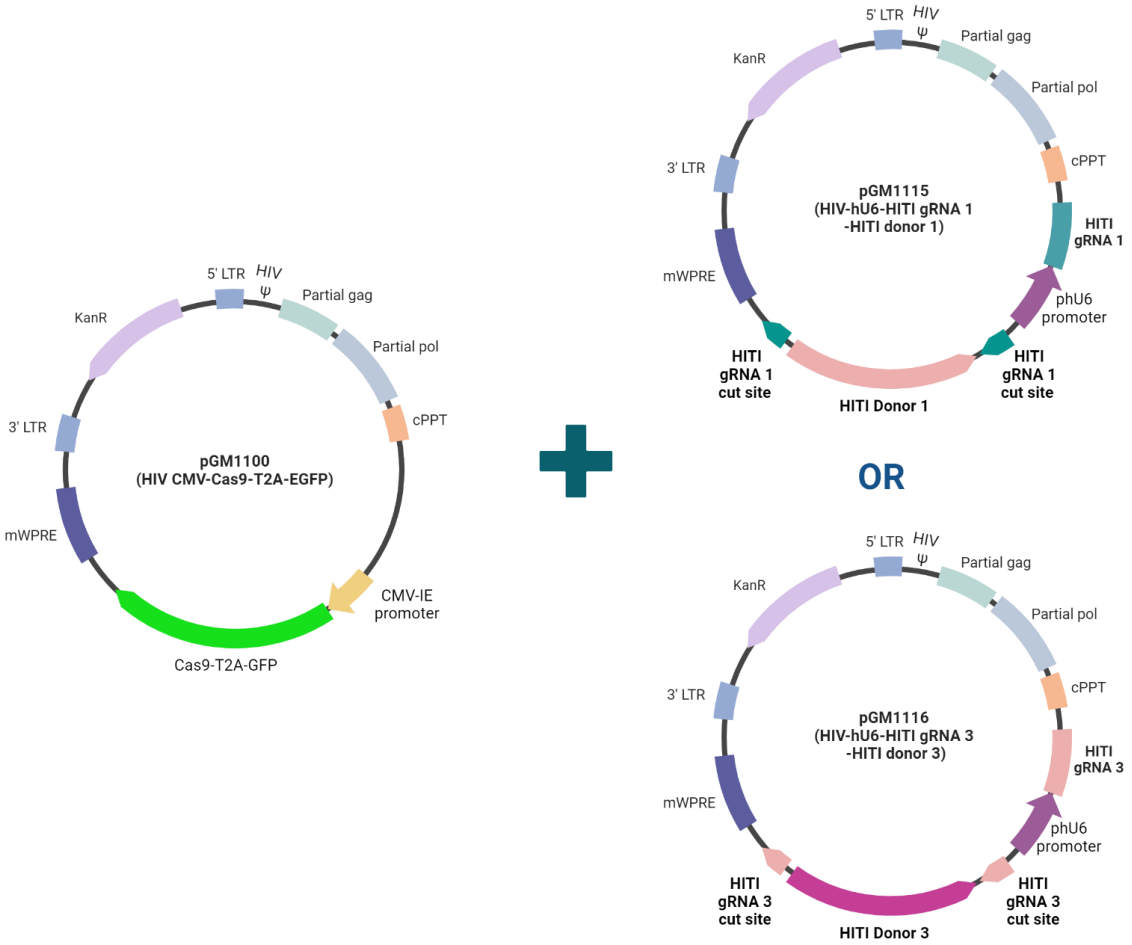

Figure 4.11. LV genome plasmids generated for the HITI approach.

Figure 4.11. LV genome plasmids generated for the HITI approach.

The first plasmid contains Cas9-T2A-GFP under CMV-IE promoter (pGM1100, left side of image). This plasmid was used in combination either with pGM1115, carrying HITI gRNA 1 under hU6 promoter and HITI donor 1 (top right), or with pGM1116 (bottom right) carrying the HITI gRNA 3 under hU6 promoter HITI donor 3. Image created in BioRender.com.

.

4.2.5.1. Transfection of HITI vectors into HEK293T cells

After the construction of the plasmids required for HITI correction of *CFTR* mutations past exons 10, the LV genome plasmid pairs were then used to transiently transfect HEK293T cells, at 1:1 copy number ratio of Cas9-expressing construct and donor plasmid, to check whether integration of the donor template was evident. This would in turn suggest that delivery of the HITI components results in their functional expression and could also result in the expected editing/integration. To evaluate editing and HITI integration of the donor sequence, the gDNA from transfected HEK293T cells was used as template for PCR using the primers in Table 2.4. to amplify the HITI integrant sequence 72 hours post-transfection with the primer positions described in Figure 4.12A. Given the design of PCR primers, when integration has occurred a 3848 bp band is expected to be amplified. The agarose gel electrophoresis image from the PCR product showed a very faint band around the 3 kb region for HEK293T cells transiently co-transfected with Cas9-expressing construct (pGM1100) and HITI donor 1 construct (pGM1109), which encodes the HITI gRNA design 1 and corresponding HITI donor 1 sequence (Figure 4.12B and C, red box). No bands were observed for any other transfection condition, including HEK293T cells co-transfected with HITI components aiming to integrate the HITI donor 3 sequence using HITI gRNA 3, despite the similar indel frequency observed during gRNA selection process (Figure 4.12B). The faint 3 kb band that was observed on the gel electrophoresis image was not expected, as should the correct HITI integration occur, the resulting band should resolve closer to 4 kb. Regardless, I decided to investigate further if indeed any HITI integration has occurred. However, it was proven difficult to directly sequence the amplified GDNA region encompassing the HITI insert. To aid analysis, the region was re-amplified in a second round of PCR to obtain sufficient quantity and quality for reliable Sanger sequencing. After the

second round of amplification, the PCR amplicon was sequenced and compared with Sanger sequencing from unedited HEK293Ts as control. Sanger sequencing and Synthego ICE analysis provided initial evidence of the integration of the HITI donor 1 in the intron 10 of *CFTR* (Figure 4.13). Overall, Figure 4.13 suggests that the level of editing was not particularly high, and therefore, the approach using two-plasmid transfection of HITI components was not the most efficient way to introduce the *CFTR* superexon donor into the locus with the donor sequences and gRNA designs described in this study.

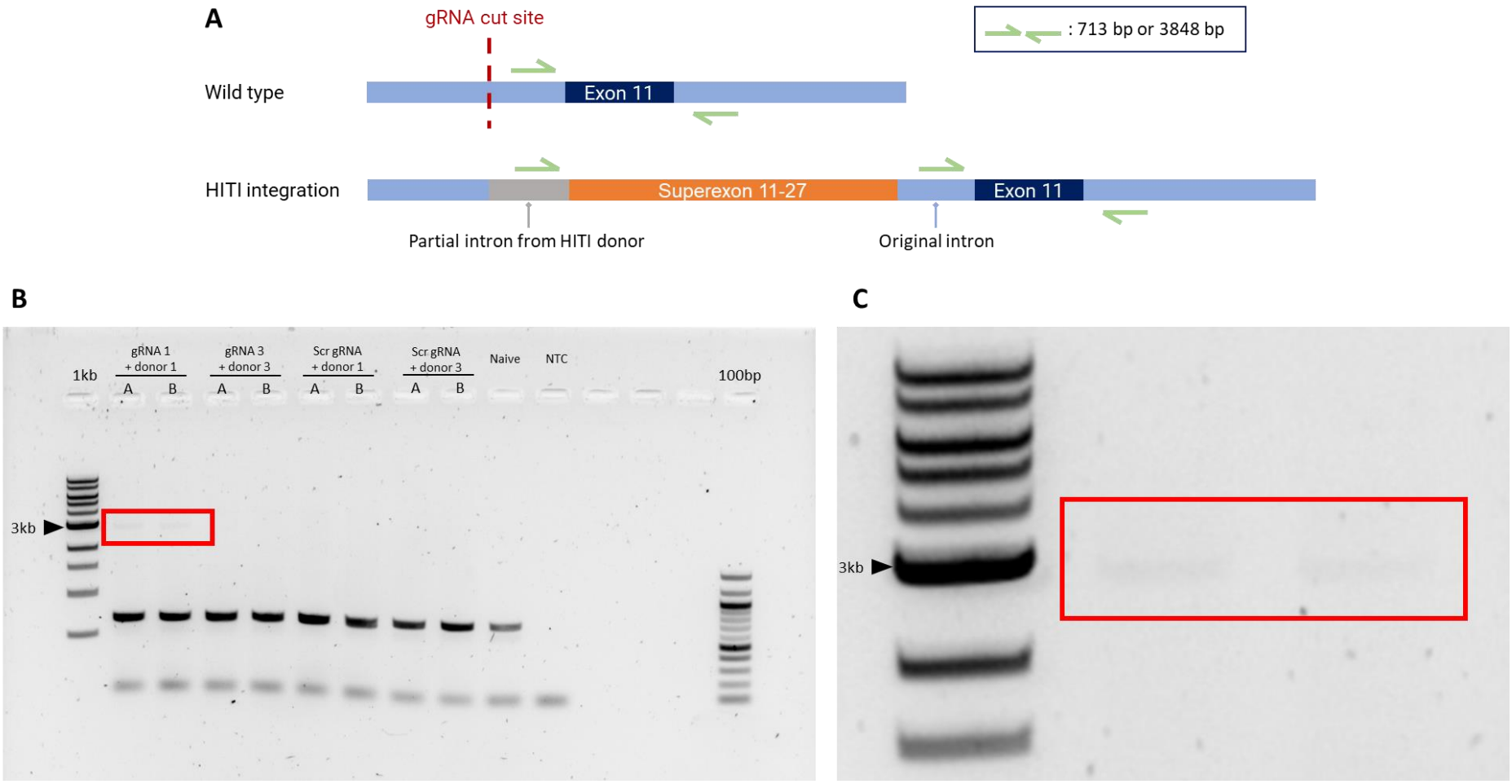

Figure 4.12. Evaluation of HITI Donor Integration.

Figure 4.12. Evaluation of HITI donor Integration.

A) PCR primer positions for amplification of GDNA post-transfection of HITI plasmids. The PCR primers bind to the intronic region of *CFTR*, specifically in the intron 10 (forward primer) and intron 11 (reverse primer), resulting in a 713 bp band. The forward primer binds to both the original intron 10, as well as the partial intron that was included in the HITI donor. Therefore, when HITI donor has integrated, the primer should amplify a longer region, resulting in a 3848 bp band, as well as a smaller wild type band around 713 bp. B) Low intensity band was observed in samples transfected with HITI gRNA 1 and donor 1 (red box). C) Zoomed in image from B).

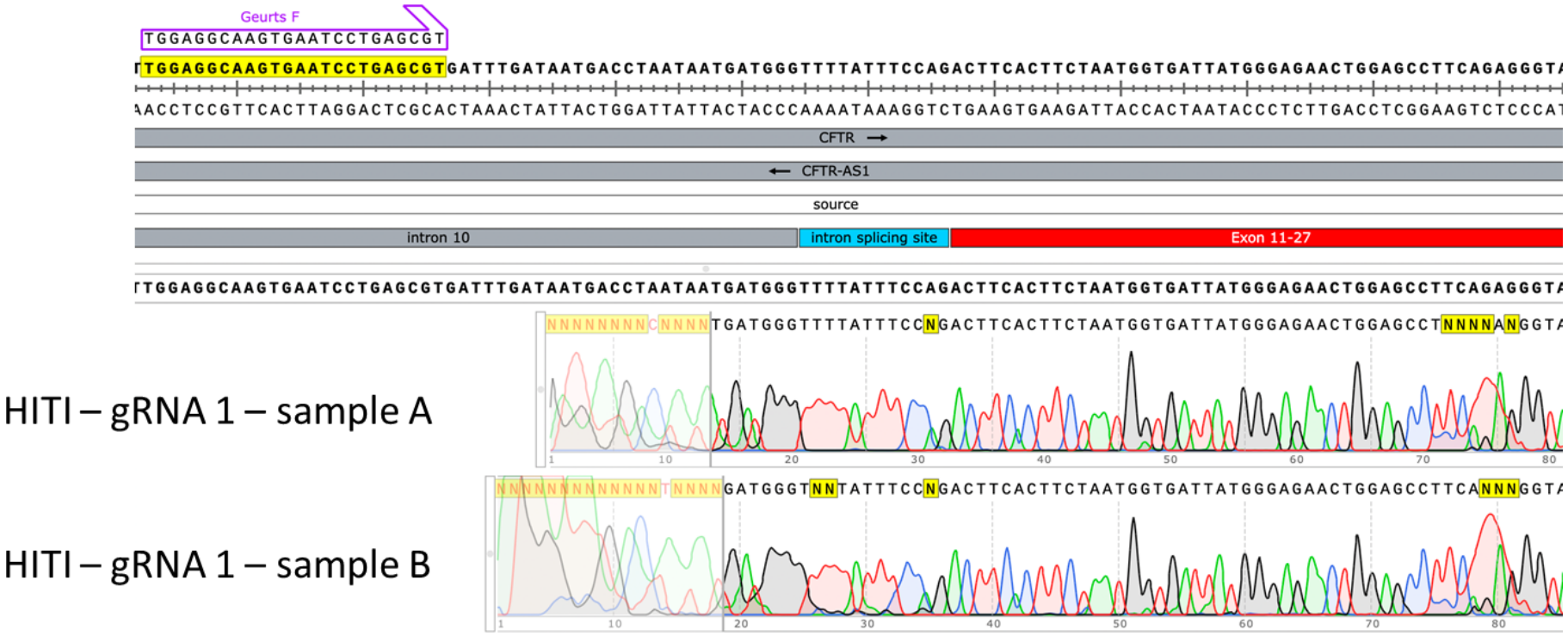

Figure 4.13. Sequencing traces for HITI integration.

Sanger sequencing showed integration of HITI donor in the target locus after transfection with HITI plasmids containing Cas9, HITI donor 1 and HITI gRNA 1. The full result of the sequencing confirmed integration up to exon 14 of *CFTR* (full traces not shown).

4.3. Discussion and Future plans

4.3.1. Overview of two-vector delivery approach

The all-in-one LV delivery systems has potential as a PE delivery system but appeared to be limited by the cargo capacity of the lentiviral backbone that in particular affects the lentiviral titre. The two-vector delivery approach is less restrictive in terms of cargo size, allowing the addition of reporter gene to at least one of the vectors. While inappropriate in therapeutic context, the addition of reporter gene was helpful for troubleshooting purposes during the development of the studies in this chapter. Due to time constraints, the studies only explored the usage of two-plasmid delivery and not yet two-vector delivery. One main advantage of using the vector delivery over plasmid delivery is the ability to control the amount of transgene delivered to the target cells. In future studies, it would be ideal to revisit the potential of two-vector delivery using rLVs and not only plasmid delivery.

The use of integrating LVs rather than integrase deficient LVs (IDLV)[103,178,179] in this chapter, and also for the all-in-one strategy described in Chapter 3, while good for proof-of concept studies, is also less ideal in a translational context. The integration of Cas9 components into the genome, resulting in prolonged expression of Cas9 machinery, is not desirable and should be avoided as it increases the frequency of off-target editing[180,181]. Therefore, future directions should look into the two-vector delivery system using IDLVs rather than LVs. Ultimately, the biggest challenge faced by the two-vector delivery, whether for PE or HITI, is that both components must be delivered into the cell, which might impact the editing efficiency compared to an all-in-one approach.

4.3.2. Splitting PE components to ease delivery

In the two-vector delivery approach, the PE components are delivered in two separate vectors, additionally allowing reporter genes such as EGFP to also be included in the overall approach. The first set of components in the split PE system contained the N-terminus of Cas9 nickase joined with a P2A sequence to an EGFP reporter, along with a suitable pegRNA. The second set of components included the C-terminus of the Cas9 nickase linker to an MMLV-RT. The addition of an EGFP reporter gene helps in visualising the successful delivery of the editing components using transduction or transfection. Another benefit from splitting the necessary components for PE into two separate vectors is that the smaller size of each vector may improve the ultimate titre following lentiviral production. As a comparison, the all-in-one PE-LV system has 9.9 kb between the 5' and 3' LTRs of the LV genome plasmid, whereas in split PE, the size of the first component was 7.1 kb and the second component was 7.8 kb between the 5' and 3' LTRs. This means that the split PE strategy reduced the LV cargo size up to 28% relative to the all-in-one PE-LV approach, and the LV titre might be expected to improve by up to 10-fold from this reduction in genome size[167].

Studies have shown similar dual vector delivery approaches, in particular using rAAV vectors due to its non-integrating nature, as opposed to LVs. However, as the packaging capacity of AAVs are much smaller than LVs, no reporter gene could be added to the construct. Regular split-PE approach using AAVs have been used to target murine brain, liver and heart[111], and to correct retinal degeneration[112]. Moreover, other studies have made efforts to further reduce the size of the PE components, for example, by truncating the MMLV-RT[115,170]. These studies further highlight the potential of split-PE delivery.

The use of intein to mediate the trans-splicing of the two PE components was proven to be another challenge of split PE delivery approach. A study by Liu et al.[160], has also shown low editing efficiency using intein mediated split-PE, with only less than 4% of editing observed in mouse liver after 10 weeks. A recent development in the field using "untethered" PE, where the Cas9n component is not linked to the MMLV-RT, could be a solution to this problem, particularly as the editing efficiency of the untethered PE was the same as standard "tethered" PE2[182].

From the experiments described in this Chapter, it appears that the two-vector approach has potential for delivering PE into target cells. Due to time constraints, further experiments to show delivery of PE into cells in using lentiviral transduction could not be pursued. Nonetheless, given the results shown here, these studies suggest that the two-vector system creates active PE components that can edit the *HEK3* locus. In Chapter 5, studies are focused on identifying potent pegRNAs targeting the F508del *CFTR* mutation which should aid adaptation of the system to include delivery via viral vectors.

4.3.3. HITI was not very efficient in delivering the superexon *CFTR* donor

As an alternative strategy to PE, HITI was used to deliver superexon (exon 11-27) of *CFTR*, this approach also used a two-vector strategy. From studies reported in this chapter, it was shown that the integration efficiency of the HITI superexon donor was low, similar to what was reported in the study by Bednarski et al.[171] and Aldossary[183]. In Bednarski et al.[171], superexon integration was performed using zinc finger nuclease as opposed to CRISPR-Cas9, and puromycin-resistance cassette under the expression of the human phosphoglycerate kinase 1 (PGK) promoter was added as a selection marker. It was reported that low level of integration was observed prior to enrichment using puromycin.

The addition of a selection marker, while beneficial for enrichment purposes, further stretch the capacity of the vector. The size of PGK promoter is 501 bp while the puromycin is about 600 bp, therefore it would add another 1.1 kb to the donor template making the total size about 7.2 kb between LTRs. Although still within the proposed capacity of LVs, the increase in genome size will likely reduce the viral titre in production. It was also reported by Aldossary[183], where HITI was used to replace exon 10 of *CFTR* in the CFBE41o- cell line, that the efficiency of the replacement was low, and only when qPCR was performed was an increase in mRNA expression observed. As HEK293T cells do not express *CFTR*, the qPCR assay would not be relevant in my study. Mention et al.[184] used HITI to integrate a smaller superexon consisting of exon 23-27 of *CFTR* into intron 22 to correct the W1282X mutation. Following enrichment with GFP selection for cells transfected with Cas9 and HITI superexon donor, they reported 5.8% (2736/46,555 reads) of GDNA amplicons contained the superexon sequence. The low efficiency of HITI reported in section 4.2.5.1 aligns with findings from other relevant studies. Coupled with the risks associated with double-strand breaks (DSBs), this strategy may not be the most suitable option for clinical applications.

The design of the two-vector HITI approach means that the donor should not be able to integrate unless it was liberated from the donor plasmid, requiring the Cas9 cutting both ends of the gRNA target site efficiently. Here, a 1:1 ratio was used for Cas9 and the HITI gRNA and donor which may have been sub-optimal for the desired integration outcome. In future studies, it was suggested that different ratios of Cas9 and gRNA are evaluated to enhance the efficiency of the desired HITI outcome[185–188]. Additionally, performing such studies in more relevant, though probably harder to transfect, cell line such as 16HBE14o- which harbour the F508del mutation, and performing functional testing of *CFTR*[189] would provide greater confidence for the feasibility of the approach.

Chapter 5: Mini Library approach for correcting F508del mutation in *CFTR*

5.1. Introduction

While prime editing allows correction of all 12 types of transition and transversion mutations as well as small insertions or deletions in the genome, it is critically restricted by the efficiency of the associated pegRNA. The efficiency of the pegRNA is determined by several factors including genome targeting by its 5' region and the reverse transcriptase template found in the 3' extension region[60]. Previous papers have reported guidance on how to design pegRNA and the range of optimal lengths for the 3' extension region[56,60]. A study by Kim et al.[60] utilized a library of 54,836 pegRNA-target sequence pairs to determine the key factors affecting the efficiency of pegRNAs. Two critical factors identified were the DeepSpCas9 score (https://deepcrispr.info/DeepSpCas9/) and the GC content of the primer binding site. However, these two factors are highly dependent on the target locus and thus it is challenging when the target site is naturally low in GC content and/or or is inherently difficult to target with SpCas9. Typically, these limitations necessitate testing a large number of pegRNA candidates for each desired prime editing outcome. Consequently, a more robust and higher throughput testing system was needed to address the challenge of selecting potent pegRNAs, particularly for therapeutically relevant levels of correction of the F508del mutation – the predominant mutation in the *CFTR* gene that affects the majority of CF patients.

To identify potent pegRNAs for correction of the F508del mutation, I designed a 'mini library' consisting of 454 pegRNAs utilizing a similar setup of Kim. et al (Figure 5.1). The mini library approach aimed to interrogate the functionality of all relevant pegRNA designs to edit a specific locus, and incorporate a cognate synthetic target downstream of the pegRNA expression cassette to aid evaluation. Studies have shown that HIV-based lentiviral vectors favour integration in actively transcribed genes, indicating the accessibility of the chromatin[190,191]. Therefore, adding the target sequence along with the library minimises one of the confounding factors, in this case the accessibility of the locus, which might otherwise decrease the pegRNA efficiency. Furthermore, to simplify screening, it was desirable to facilitate the readout of pegRNA efficacy via amplicon sequencing of the corresponding synthetic target locus, rather than different endogenous loci that would require custom-made capture libraries or whole genome sequencing, both of which are laborious and expensive. However, providing the synthetic template with the cognate pegRNA creates a potential caveat – that the pegRNA candidates might not perform in an identical fashion when targeting the endogenous loci.

In this chapter the mini library approach was used to screen around 400 pegRNA candidates designed to correct the F508del mutation in *CFTR*; to aid interpretation, a further six positive and three negative pegRNA controls were included. The positive pegRNA controls targeted well established prime editing target loci[56]: *HEK3*, *RNF2*, and *FANCF* with their relevant synthetic targets to provide an indication if editing has occurred. The aim for this chapter was to develop a high throughput screening method for selecting an efficient pegRNA, in particular to correct the F508del mutation in *CFTR*.

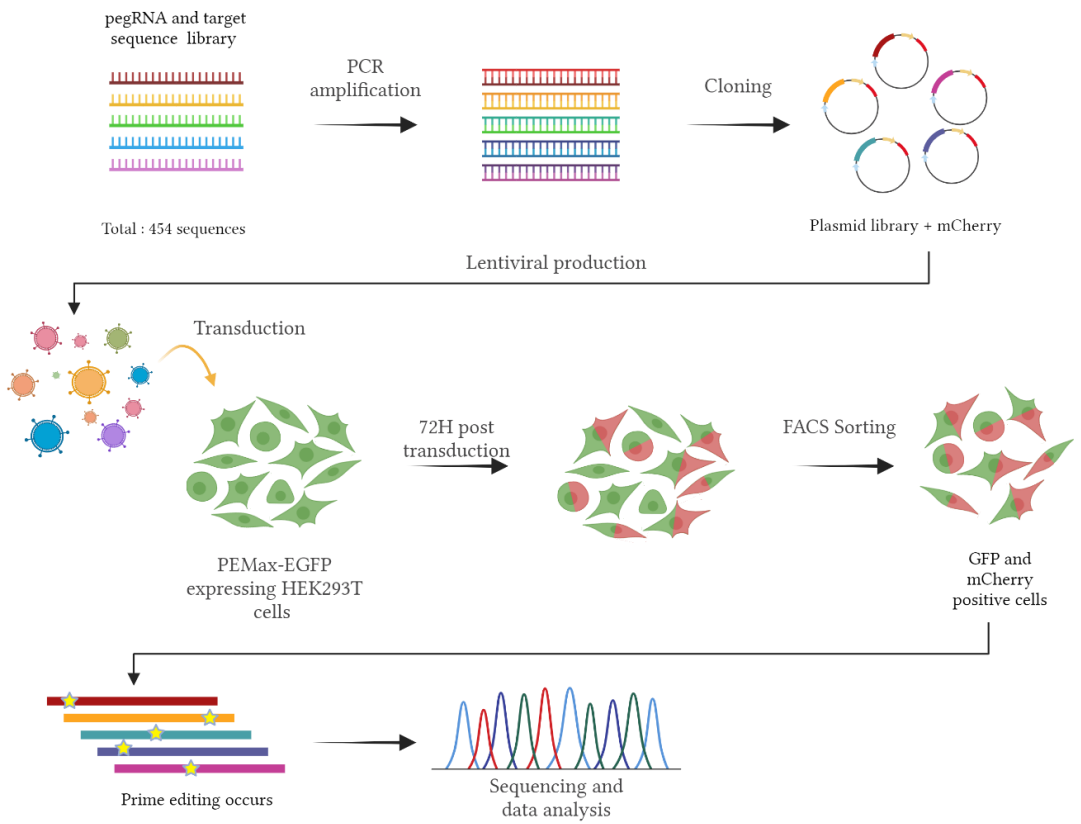

Figure 5.1. Schematics of mini library approach for correction of F508del.

Around 400 pegRNA candidates targeting F508del mutation in *CFTR* were screened using the mini library approach. Image created in BioRender.com.

5.2. Results

5.2.1. Generation and characterisation of HEK293T cell line expressing PEmax-EGFP

HEK293T cells stably expressing PEmax were generated by transducing adherent HEK293T cells (Table 2.1) with rHIV-1.VSV-G lentiviral vector encoding PEmax-P2A-EGFP under the transcriptional control of an EF1α promoter (Table 2.8) according to section 2.6.3. The lentiviral vector expressing PEmax-P2A GFP contains a WPRE3 sequence, a truncated version of WPRE to reduce the size of the genome plasmid[192] to increase viral titre. Following single cell sorting using FACS, collecting 4 × 96-well plates of single clone in each well, six clones were arbitrarily selected for expansion and characterization (Figure 5.2). The latter included measurement of Mean Fluorescence Intensity (MFI) using flow cytometry, and PEmax-P2A-EGFP protein expression levels using western blotting. Finally, to confirm the functionality of the PEmax expressing clones, the level of gene editing was assessed by insertion of +1CTT in the endogenous *HEK3* site after transient transfection with a construct encoding a suitable pegRNA. This comprehensive characterization provided the necessary information to determine which clone should be selected for the mini library transduction screening studies.

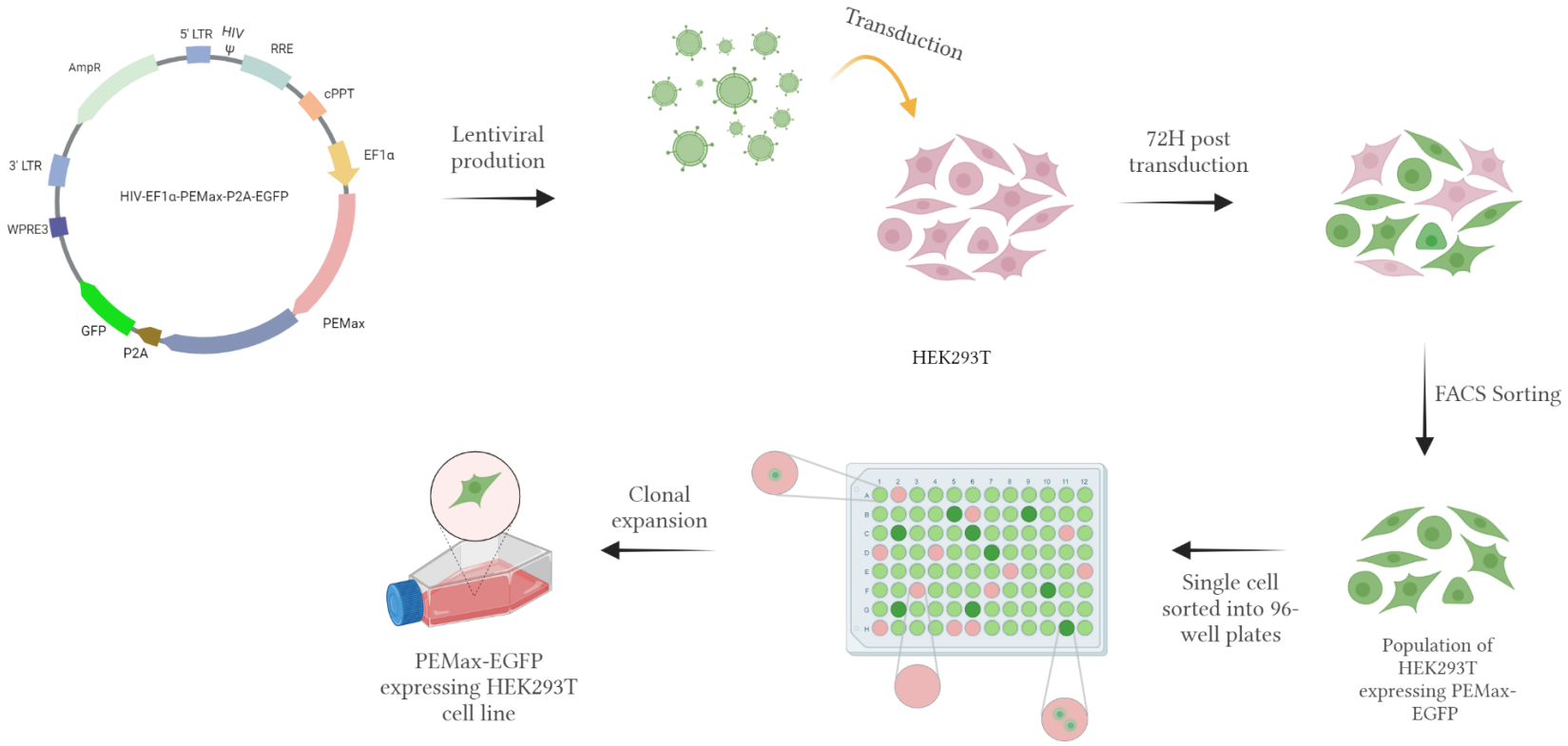

Figure 5.2. Schematic of the workflow for generation of HEK293T-PEmax-EGFP expressing cell line.

HEK293T cell line was transduced with HIV-EF1α-PEmax-P2A-EGFP (MOI=4) and after 3 days, FACS sorted for GFP positive cells. Each clone was expanded and characterised. Image created in BioRender.com

5.2.1.1. Mean Fluorescence Intensity level of 293T-PEmax-EGFP clones

To evaluate the clonal expression of PEmax-P2A-EGFP, the MFI was evaluated by flow cytometry. As PEmax and EGFP are linked by a P2A sequence in the construct[193], MFI was anticipated to provide a surrogate for PEmax expression crucial for evaluating the performance of the pegRNA mini library. Based on EGFP expression, three clones were assigned as medium EGFP (Med 2/B1, Med 5/B2, and Med 6/B3) whereas three other clones assigned as high EGFP (High 2/C1, High 3/A3, and High 4/B1) expressors (Figure 5.3). It was anticipated that different clones were likely to contain different copy numbers of integrated LV-PEmax-EGFP constructs, and likely to exhibit different levels of EGFP expression and median fluorescence intensity, which can be quantified using flow cytometry. Following flow cytometry analyses, clones designated as either medium or high level of EGFP displayed similar MFI values within each group (Figure 5.4), with no significant difference between clones in each designated group (average MFI of 25,544±1,857 for High EGFP clones, and average MFI of 9,175±1,005 for Medium EGFP clones $p<0.0001$)).

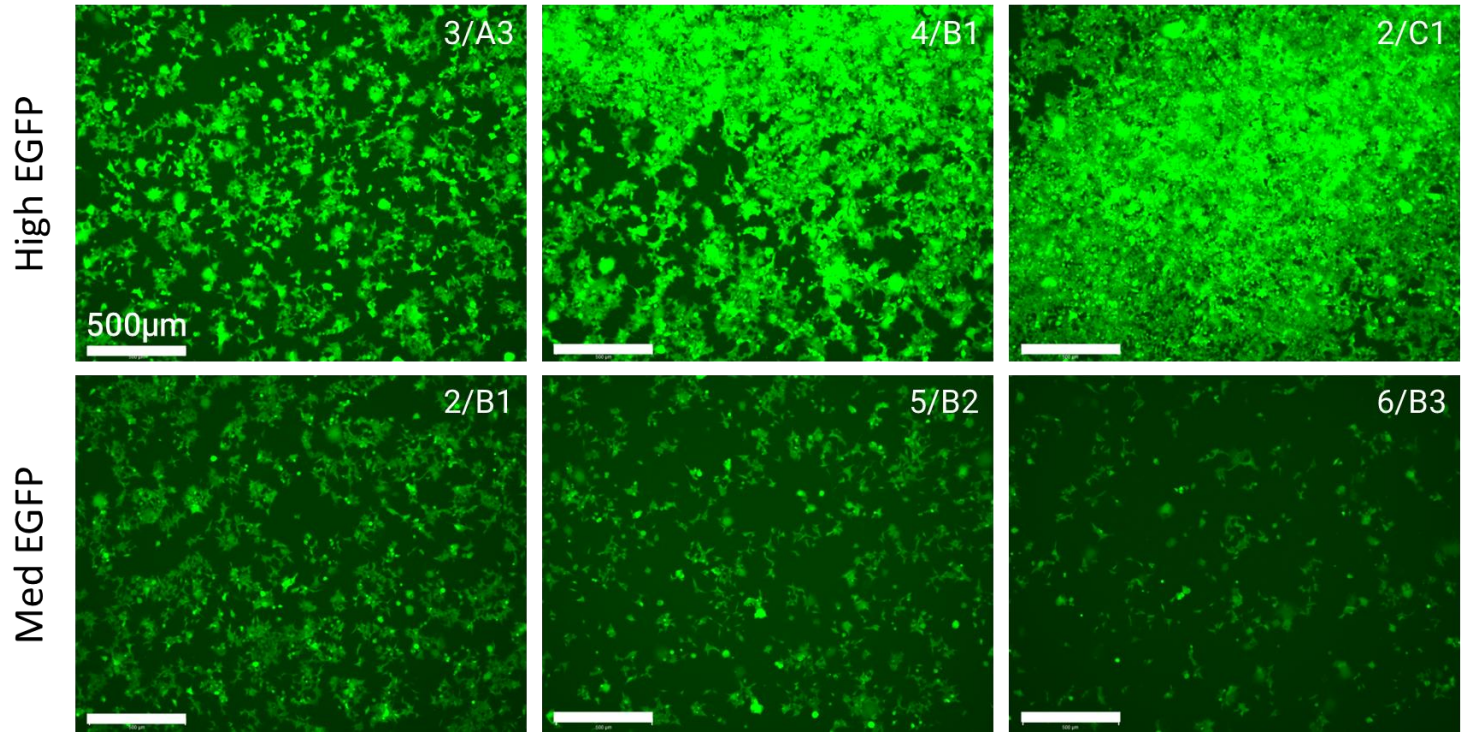

Figure 5.3. HEK293T cells transduced with rHIV1.VSV-G encoding PEmax-P2A-EGFP.

Cells were transduced (MOI 4), and single cell sorted based on high and medium EGFP fluorescence by FACS. Representative images of high (n = 3) and medium (n = 3) of EGFP expression in single cell clones (n=400) were visualised by fluorescence microscopy, and subsequently labelled as High 3/A3, High 4/B1, High 2/C1, and Med 2/B1, Med/b2, Med 6/B3, respectively. These clones were selected for further characterisation. Scale bar = 500 µm.

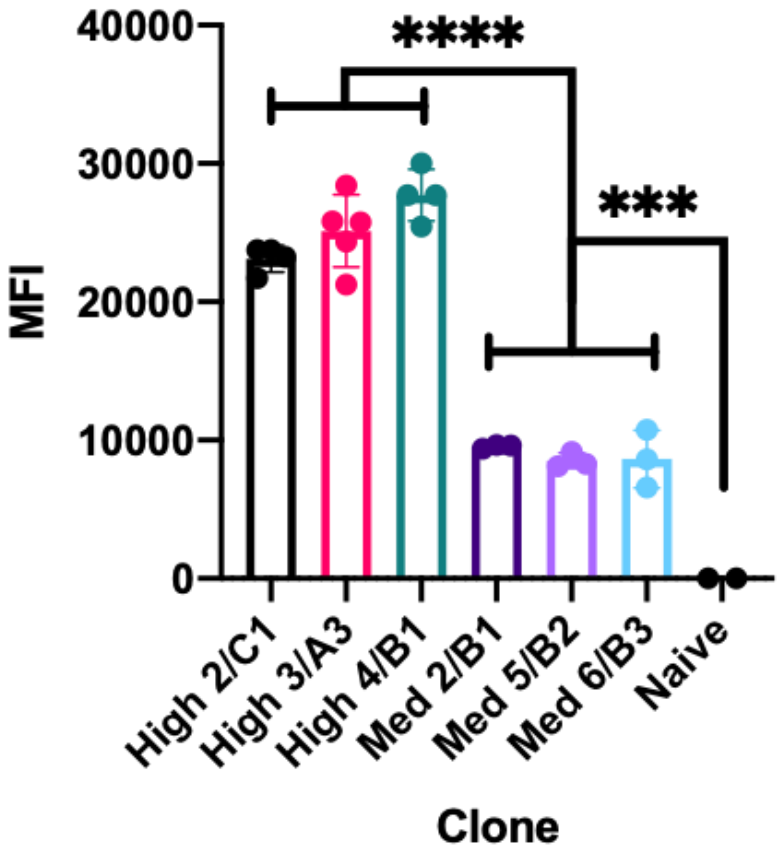

Figure 5.4. Mean fluorescence intensity value from six PEmax-EGFP expressing HEK293T clones.

Six clones generated from single cell sorting of LV-PEmax-EGFP transduced HEK293T cells were expanded and MFI of each clone quantified using flow cytometry. Three clones expressing high level EGFP have an average MFI of 25,544±1,857, while three clones expressing medium level of EGFP have an average MFI of 9,175±1,005. Data are presented as mean ± standard deviation (SD). Statistical significance was determined using Ordinary One-Way ANOVA followed by Tukey's Post-Hoc Test. No significant difference between each clone within high or medium level group, whereas significant difference was observed between high EGFP group and medium EGFP group ($p<0.0001$).

5.2.1.2. Characterisation of HEK293T-PEmax-EGFP clonal using Western Blot

To demonstrate protein expression of PEmax-EGFP in the HEK293T-PEmax-EGFP cell lines (Table 2.1), a Western Blot (WB) assay was performed against Cas9, using an antibody that binds to the N-terminal of the Cas9 protein (section 2.11). As controls, HEK293T cells were transiently transfected with (i) pHIV.CMV.Cas9.P2A.EGFP (Cas9-EGFP), (ii) pCMV.PE2.P2A.EGFP (PE2-EGFP), or (iii) pLentiPEmax.EGFP.W3 (PEmax-EGFP) plasmid (Table 2.7). The controls along the selected three clones (High 2/C1, High 4/B1, and Med 2/B1) and naïve HEK293T cells were included in the western blot experiments. In two separate attempts, I detected strong bands at about 160 kDa corresponding to Cas9n and about 230 kDa for PE and PEmax for overexpressing cells, but not HEK293T-PEmax-EGFP, High 2/C1, High 4/B1, and Med 2/B1 (Figure 5.5). Having confirmed the validity of the assay, it was evident that PEmax expression level in HEK293T-PEmax-EGFP was below the detection limit of Western Blotting.

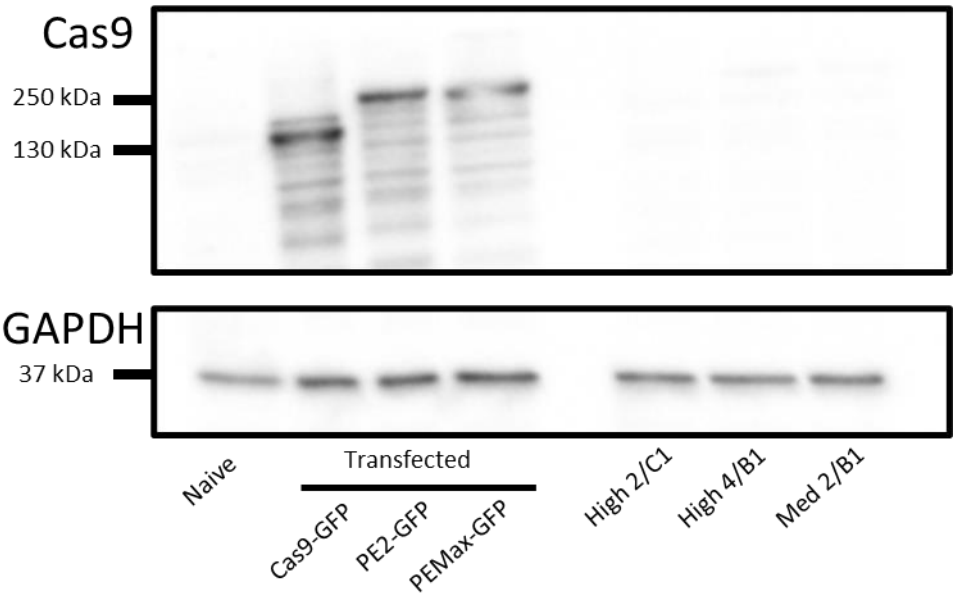

Figure 5.5. Western blot analysis of PEmax expression in HEK293T and HEK293T-PEmax-EGFP cell lines.

Lysates from HEK293T cells transfected with plasmids encoding Cas9-EGFP, PE2-EGFP, or PEmax-EGFP show distinct bands corresponding to the expected molecular weight of PEmax. In contrast, no detectable bands were observed in lysates from the HEK293T-PEmax-EGFP cell line clones (High 2/C1, High 4/B1, and Med2/B1). This indicates that PEmax expression in the HEK293T-PEmax-EGFP cell line is below the detection limit of the western blot assay, suggesting the need for further functional characterization of the cell line.

5.2.1.3. Correlation of editing efficiency and amount of PEmax-EGFP expression

To evaluate the functional ability of the PEmax-EGFP cell lines to perform prime editing, the six HEK293T-PEmax-EGFP clones described above were transfected with pU6.Sp.pegRNA.HEK3_CTT_ins plasmid (Table 2.7) which encodes a pegRNA targeting the *HEK3* site, for +1CTT insertion. At 3 days post-transfection, the GDNA was extracted, and the PE efficacy was evaluated by Sanger Sequencing. Intriguingly, efficient genome editing (33% intended edit) was demonstrated by quantifying the insertion frequency of +1CTT in *HEK3* site using DECODR analysis in clone High 4/B1 (Figure 5.6, Figure 5.8). In contrast, we found high levels of unspecific editing for other tested clones. For example, in clone Med 5/B2, further sequence analysis revealed frequent insertion of CTA (20%) as opposed to the intended CTT insertion that was less frequent (Figure 5.7). Similarly, clone Med 6/B3 also showed CTA insertion, and the rest of the clones showed TGA or CGA insertion. These unexpected prime editing outcomes have not (to-date) been described in the literature and perhaps warrants further investigation. Nevertheless, the crucial ability of clone High 4/B1 to perform prime editing was confirmed.

As a proof-of-concept for the mini library approach, an initial pilot study was conducted using a single pegRNA. This experiment involved cloning the *HEK3*-targeting pegRNA construct, which serves as one of the positive controls in the mini library, into the pEF1a.mCherry.hU6.pegRNA plasmid backbone. This construct was then used to produce rLV.HEK3_pos (Table 2.8), and the functional titre of the lentivirus was measured as described in section 2.6.4. Following the transduction of the HEK293T-PEmax-EGFP cell line (clone High 4/B1) at MOIs of 3, 1, and 0.3, GDNA was harvested on Day 6. The *HEK3* target

at both the endogenous locus and the synthetic target site in the construct was PCR-amplified using the specified primers (Table 2.4). The purified PCR products were then analysed via Sanger sequencing and DECODR (Figure 5.9). The result suggested that the *HEK3* targeting pegRNA installed the desired edit (CTT insertion) at the synthetic target when analysed 6 days post transduction, even at the lowest MOI of 0.3 (21.5%). When MOI was increased to 1 and 3, the editing efficiency also increased to 23.2% and 34.9% respectively. Notably, the level frequency of +1CTT insertion was approximately doubled in the endogenous target (51.9% in samples transduced with MOI 0.3) when compared to the synthetic target. This suggests a targeting preference for the endogenous locus compared to the synthetic target. Nonetheless, as the level of intended prime editing in the synthetic target was readily observed with an MOI of 0.3, the mini library strategy appeared robust.

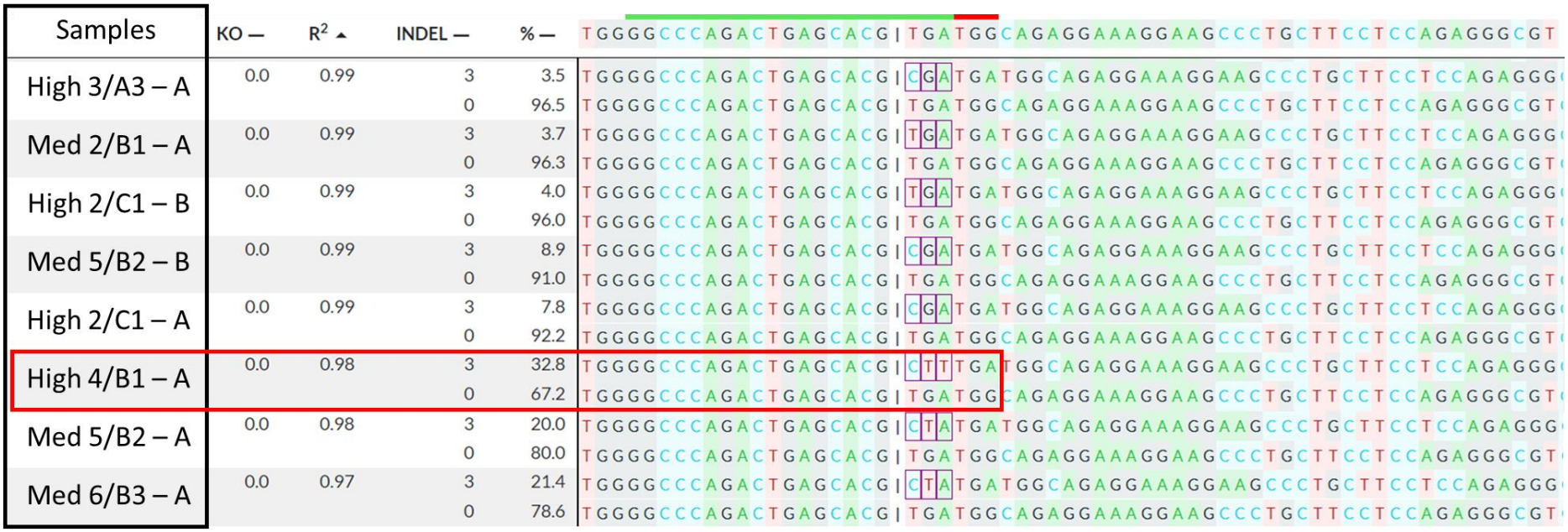

Figure 5.6. Prime Editing observed in the HEK293T-PEmax-EGFP clones transfected with pU6.Sp.pegRNA.HEK3_CTT_ins.

All clones displayed prime editing, but only the High 4/B1 clone showed the intended CTT insertion at 32.8%. Med 5/B2 clone showed 20% of unintended editing (CTA instead of CTT) by the DECODR algorithm, suggesting that the more prominent edit is the unintended CTA rather than the intended CTT insertion.

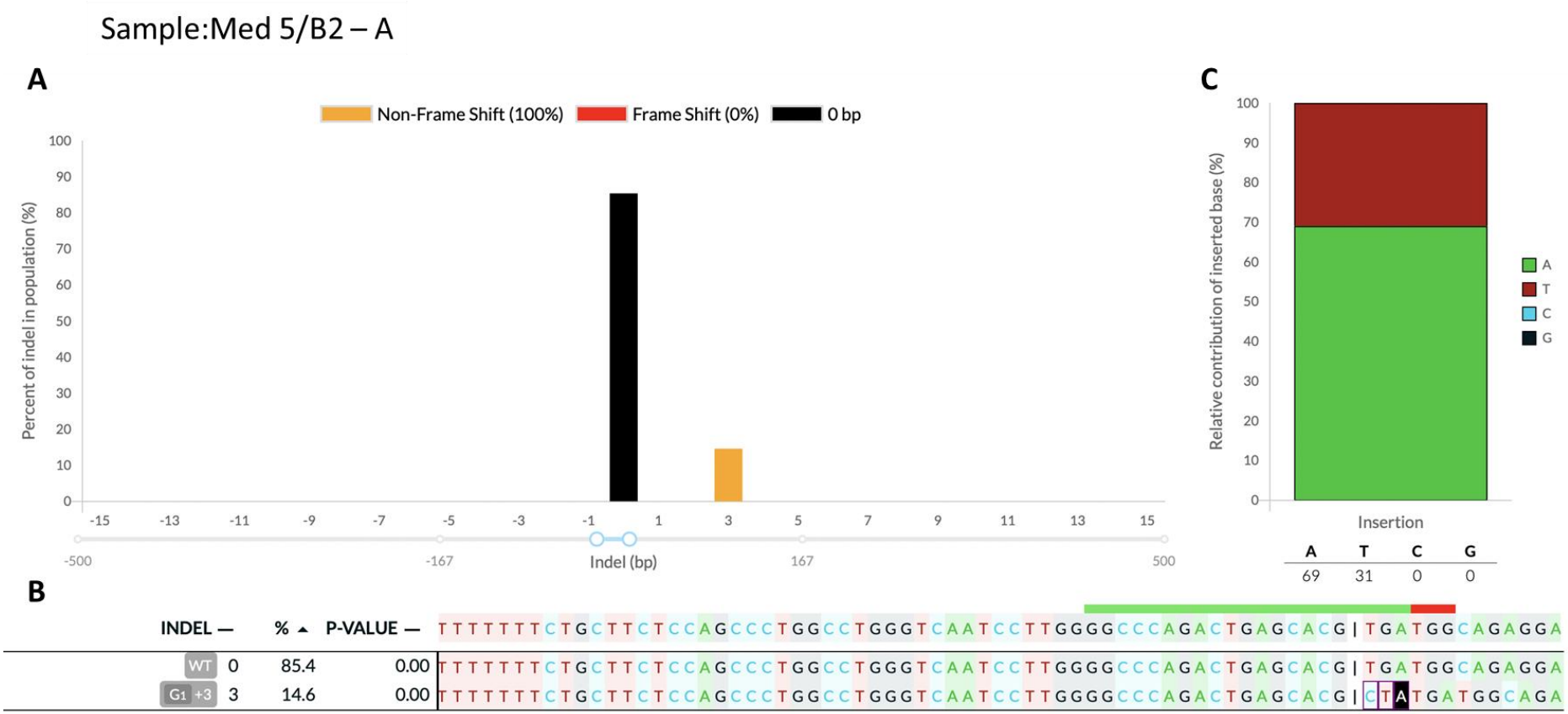

Figure 5.7. Quantification of prime editing efficacy using DECODR.

A) and B) Med 5/B2 clone displayed 14% unintended CTA insertion instead of intended CTT insertion. C) However, when probed further, the CTT insertion was detectable in 30% of total edits.

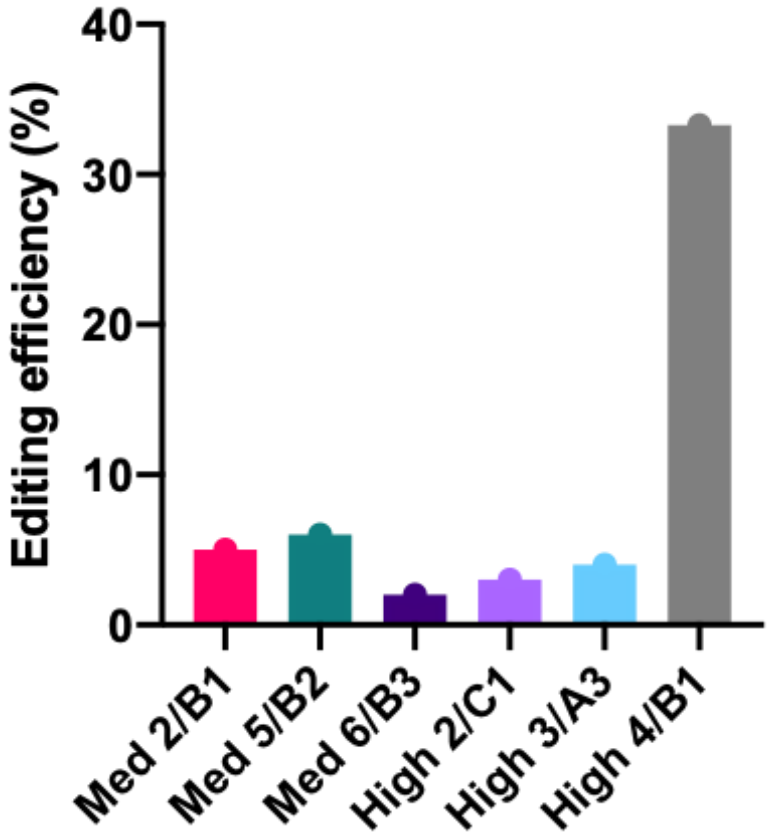

Figure 5.8. HEK293T-PEmax-EGFP clones evaluated for editing efficiency at the *HEK3* site.

Clone High 4/B1, exhibiting a total editing efficiency of 33% and precise insertion of CTT at the +1 position relative to the PAM, was selected for subsequent experiments.

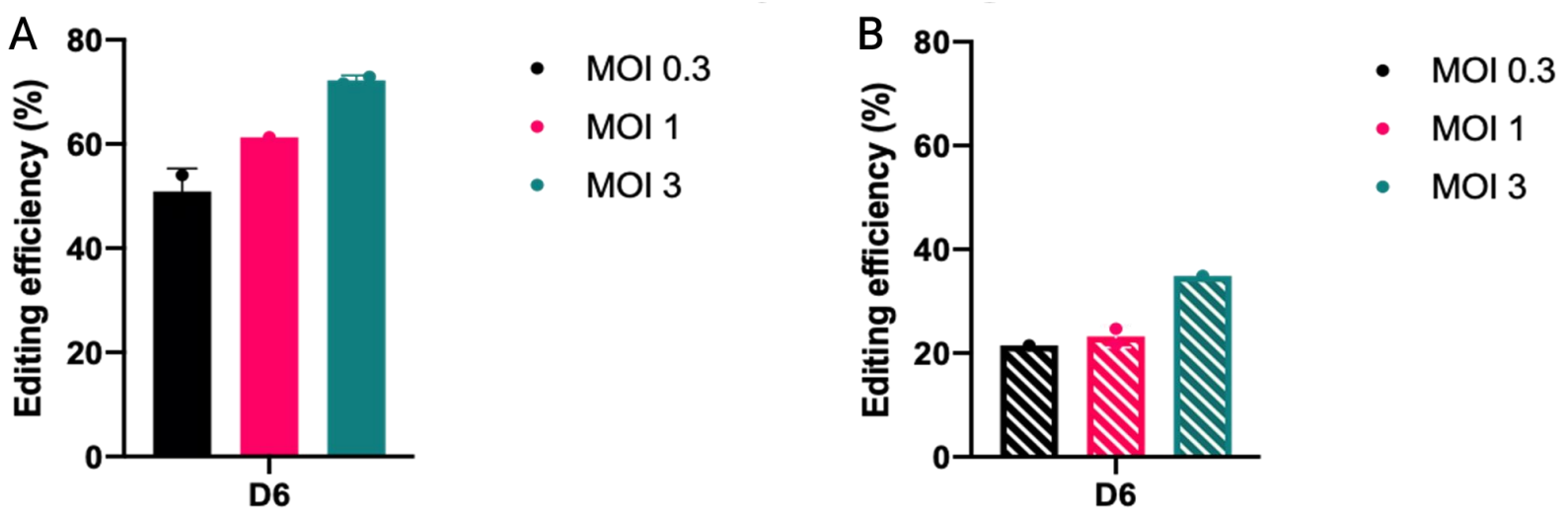

Figure 5.9. Comparison of editing efficiency at the *HEK3* locus in endogenous and synthetic targets.

Editing at both endogenous (A) and synthetic (B) targets was evaluated. HEK293T-PEmax-EGFP clone High 4/B1 was transduced with LV containing *HEK3* targeting pegRNA. Six days post-transduction the GDNA was harvested and the synthetic and endogenous *HEK3* targets were amplified. Intended editing of the *HEK3 loci* was observed in both the endogenous and synthetic targets, even at the low MOI of 0.3. (47.8% and 21.5%, respectively).

5.2.2. Design and production of pegRNA mini library

Using pegIT[174] (https://pegit.giehmlab.dk/), an initial *in silico* library based on three pegRNA spacer candidates (Figure 5.10) were generated to correct the F508del mutation in *CFTR,* each with different lengths of primer binding site and RT templates. This *in silico* library totalled over 1000 pegRNA candidates. For the mini library itself, the number of pegRNA candidates was reduced to 445 pegRNA configurations by restricting the primer binding site length between 11-20 nucleotides and using various RT length depending on the distance from the protospacer PAM to the edit site, with the shortest being 20 nucleotides and the longest 65 nucleotides. The reduction from over 1000 pegRNA candidates to 445 pegRNA candidates was done using a script in R (provided by Dr Jakob Haldrup, personal communication) adjusted according to the above design criteria. The design space for the length of for primer binding site and RTT was based on the pegRNA design recommendations established by Kim et al[164].

Each individual sequence in the mini library includes: 5' and 3' constant regions to allow unbiased PCR amplification and to allow Golden Gate assembly of the library; BsmBI restriction site to allow insertion into a specified plasmid backbone; the pegRNA construct including spacer, scaffold, and 3' extension region; poly T sequence to facilitate transcription termination; a unique barcode for each oligo sequence, and target sequences which accommodate the longest 3' extension for each pegRNA spacer. On the 5' end of the unique barcode, a constant 20bp CutAdapt region (5'GTGGACAGCACGATACACCG3') was also added for trimming of NGS reads. The individual sequences represented by the library

were commercially synthesised and pooled, and inserted into a lentiviral genome plasmid backbone expressing mCherry for downstream selection purposes (Figure 5.11).

The library of lentiviral genome plasmid backbones was electroporated into Stbl3 *E. coli* for plasmid amplification according to section 2.12.3. Total colony count was used estimate the coverage of mini library elements (Table 5.1). With estimated total colony number at 1.25×10^5, the coverage of each of the mini library elements was ~275 copies, which was deemed sufficient for lentiviral production and pooled library screening.

Lentiviral vector particles were generated from the library of lentiviral genome plasmid backbones (2.12.2), and the resultant lentiviral library was used to transduce HEK293T High 4/B1 cells that constitutively express PEmax-EGFP (Table 2.1) with MOI of 0.3 such that the majority of transduced cells contained <1 copy of unique pegRNA library element. Transduced cells were subsequently enriched by selection for mCherry expression and expanded as a pool for subsequent analysis. Following selection and expansion, the mini library transduced cells were harvested at different time points (3, 5, and 7 days post-transduction), the GDNA was prepared according to section 2.8. and the editing efficiency was measured with NGS (section 2.12.4). The aim of this chapter was to establish the design for pegRNA mini library and testing the corresponding prime editing capabilities in both endogenous and synthetic targets.

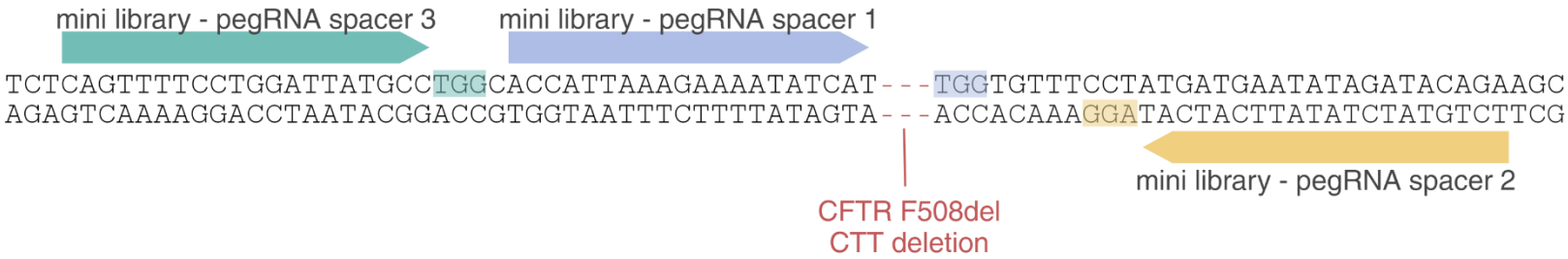

Figure 5.10. Schematic of pegRNA library construct.

This illustrates the positions of pegRNA spacers within the mini library relative to the F508del mutation in *CFTR*. The mini library includes three distinct spacers, with each spacer's corresponding PAM sequence highlighted in the same colour as the respective pegRNA spacer. The location of the F508del mutation is indicated by a red dashed line.

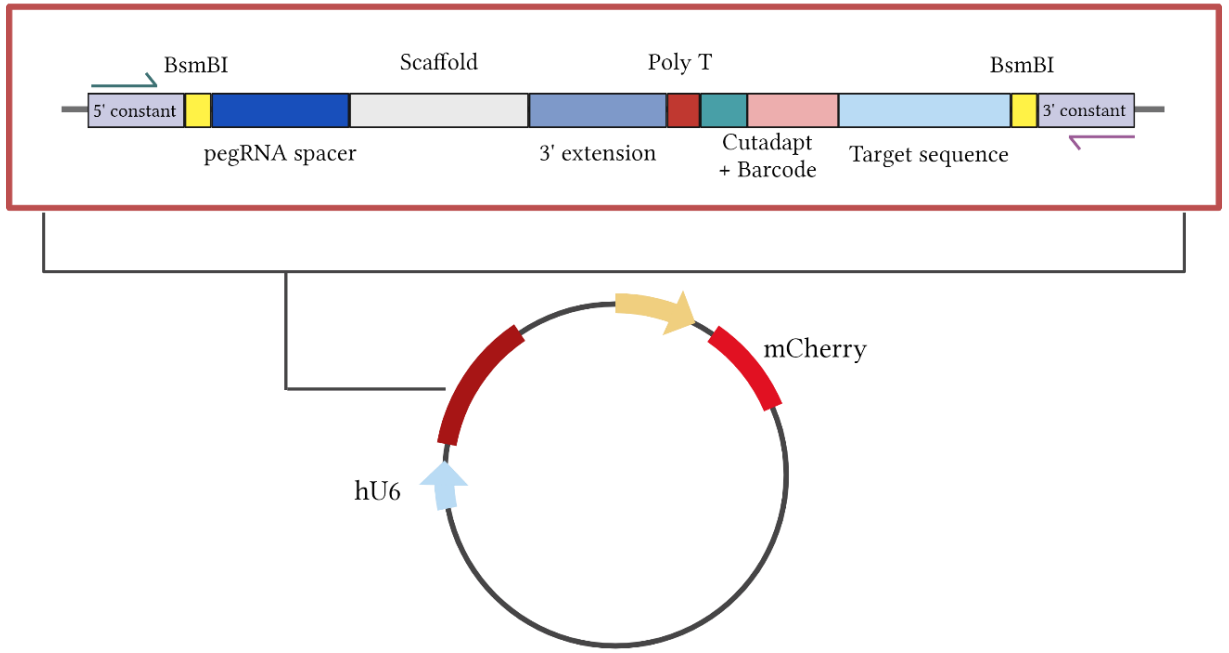

Figure 5.11. Configuration of the pegRNA constructs within the mini library.

The pegRNA constructs were synthesized as a single-stranded DNA oligonucleotide pool and then PCR amplified to generate double-stranded DNA for Golden Gate Assembly. Both the 5' and 3' ends of each construct were flanked by constant regions (5'CAATCCGCCCTCACTACAACCGGCGTCTCACACC3' at the 5' end and 5'CGTCTCAAAGCCTCCCTCATCGACGCCAGAGTAG3' at the 3' end) to facilitate amplification, followed by a BsmBI recognition sequence (CGTCTC). The primers used for amplification of the pooled oligonucleotides

are listed in Table 2.13. The scaffold sequence was uniform across all oligos. The library includes three different pegRNA spacer sequences designed to correct the F508del CFTR mutation, with varying lengths of RT templates and target sequences. It also contains three positive control sequences targeting the *HEK3*, *RNF2*, and *FANCF* sites, and six negative controls for each spacer, lacking primer binding sites and with consistent RT templates (5'TCTGCCATCT3'). In total, the oligo pool comprised 454 oligos, which were cloned into the recipient plasmid pEF1a.mCherry.hU6.pegRNA acc as detailed in section 2.12.2.

Table 5.1. Number of colonies from bacterial electroporation of mini library plasmid pool.

Dilution factor	Number of colonies	Estimated total colony number
5,000	24	1.20×10^5
50,000	3	1.50×10^5
200,000	1	2.00×10^5

To ensure that all the library elements were equally represented, the assembled, pooled, library of lentiviral genome plasmid backbones was characterized using NGS as described in section 2.12.4. The NGS data were analysed using custom-made Python scripts and CRISPResso2 (Pinello Lab) and plotted as number reads per pegRNA design in the mini library (Figure 5.12). The NGS analysis showed a significant overrepresentation of the positive and negative control pegRNAs in the pooled pegRNA-library (> 15000 read counts) whereas the CFTR pegRNAs was underrepresented (< 60 read counts on average). This trend is consistently observed in all the samples, as shown in Figure 5.13. This is likely attributable to unequal synthesis of the oligonucleotides because each unique library element (pegRNA) was PCR amplified using the same primers to minimize bias.

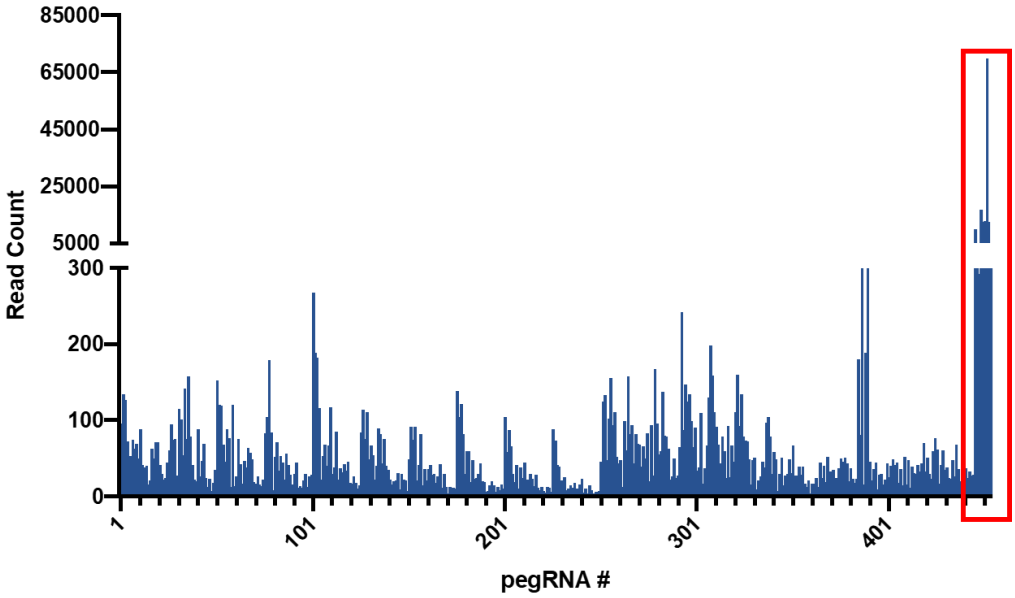

Figure 5.12. Read count of each pegRNA in the pooled plasmid pegRNA-library.

An uneven distribution of pegRNAs was observed and both positive and negative controls of the mini library were highly overrepresented (>5000 read counts, marked in red) compared to the rest of the pool (averaging 50 read counts). The experiment aimed to achieve approximately 500 × read count for each pegRNA, but the final average read count for each pegRNA in the pooled plasmid library was 367. However, due to the overrepresentation of the library's positive and negative controls, the average read counts of the library elements became highly skewed, leading to the underrepresentation of some library elements in the final plasmid pool. As a result, 397 library elements, or 87.4% of the library, were represented fewer than 100 times in the final pool.

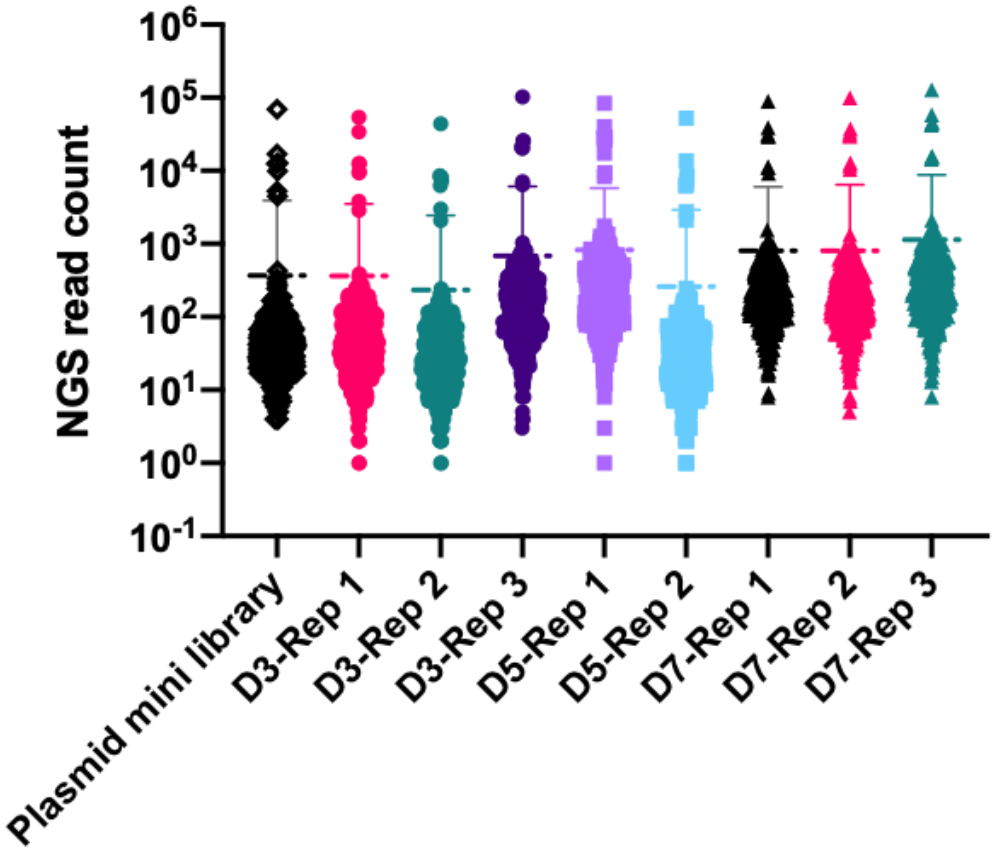

Figure 5.13. Trimmed read count distribution from NGS of the samples harvested at indicated time-points.

Plasmid mini library: plasmid mini library used to generate the lentiviral vector, D3 Rep1-3: NGS samples from gDNA following transduction with lentiviral vector mini library harvested at day 3 post transduction; D5 Rep 1-2: at day 5 post transduction; and D7 Rep1-3: at day 7 post-transduction.

5.2.3. Correction of F508del mutation in *CFTR* using mini library strategy

The pegRNA plasmid mini library detailed in section 5.2.2 was utilized for the production of the lentiviral vector LV.hU6.pegRNA_minilib.EF1a.mCherry, hereafter referred to as rLV.minilib (Table 2.8). Lentiviral mini library was produced in triplicate using small scale production method (section 2.6). The produced vectors were then titrated using functional titration method by transducing clone High/4B1, PEmax-EGFP expressing HEK293T cells using 20, 100, 500, 2500 and 12500-fold dilution factors. Using flow cytometry (Figure 5.14), the functional titre of each replicate ranged between 5E6/mL to 2.25E7/mL (Table 5.2). These titres indicated that each production contains over 10000 times the number of elements of the library, which in theory would maintain the initial library coverage of 500× following transduction of the HEK293T-PEmax-EGFP cell line. The rLV.minilib was then used to transduce PEmax-EGFP expressing HEK293T cells, clone High 4/B1 with an MOI of 0.3 as per section 2.6.3. and GDNA from D3, D5, and D7 post transduction were harvested for NGS analysis.

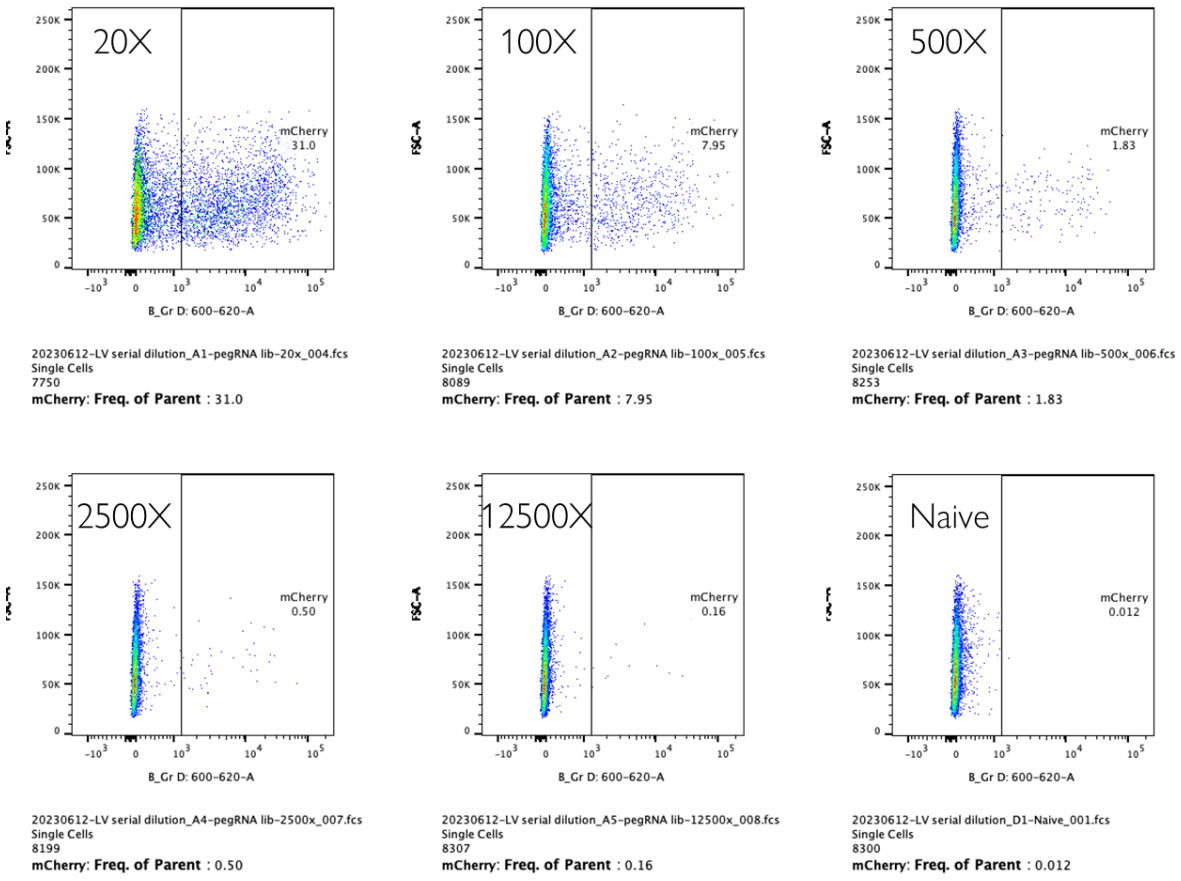

Figure 5.14. Flow cytometry of mCherry expression in HEK293T-PEmax-EGFP cell line, clone High 4/B1, transduced with rLV.minilb.

Flow cytometry was performed to measure mCherry expression in HEK293T-PEmax-EGFP cell line, clone High 4/B1, transduced with varying dilutions of rLV.minilb from three independent productions. The representative data shown are from replicate 1. As expected, mCherry expression decreases with increasing dilution of the lentiviral vector. These measurements will be utilized to calculate the functional titre of each production batch.

Table 5.2. Functional titre of rLV.minilib from three independent productions

	Percentage of mCherry positive cells		
Dilution Factor	**pegRNA lib-1**	**pegRNA lib-2**	**pegRNA lib-3**
20	31.00%	62.80%	76.40%
100	7.95%	20.90%	32.50%
500	1.83%	4.98%	9.63%
2500	0.50%	0.89%	2.10%
FTU (IU/mL)	n/a	9.36E+06	2.16E+07
	5.96E+06	1.12E+07	2.35E+07
Average FTU	5.96E+06	1.03E+07	2.25E+07

5.2.4. Next Generation Sequencing Analysis of editing efficiency in *CFTR* synthetic target

Next generation sequencing (NGS) was performed to quantify the editing efficiency of each pegRNA targeting F508del mutation in the *CFTR* synthetic target of the mini library. NGS was selected for its high sensitivity and accuracy for short sequences (< 300 bp). The NGS samples were sequenced from GDNA harvested on days 3, 5, and 7 post-transduction of the LV mini library. Each sample was indexed using NEBNext® Multiplex Oligos for Illumina®, 96 Unique Dual Index Primer Pairs Set 2 (NEB). After indexing and prior to sequencing, the samples were purified using magnetic beads as described in section 2.12.4 to remove impurities and smaller, unwanted fragments. The purity of the pooled samples was then checked using TapeStation. The pooled sample (lane B1) only has bands with 300bp length, confirming the purification was successful (Figure 5.15).

The NGS data were analysed using CRISPResso2 (Pinello lab) in batch mode using a Python code (provided by Dr Jakob Haldrup, personal communication). First, the data were de-multiplexed and compartmented based on the unique barcode for each pegRNA sequence, and the corresponding pegRNA number was assigned. Reads were trimmed using the constant region in CutAdapt to remove non-essential library elements for quantification of pegRNA efficacy, such as the scaffold region, in order to minimize computational burden. The positive controls targeting the *HEK3*, *FANCF*, and *RNF2* and the negative controls were excluded from the batch analysis as the target sequence is different from the *CFTR* targeting pegRNAs, and instead were processed individually on the CRISPResso2 website (http://crispresso2.pinellolab.org/submission) by selecting the prime editor as the editing tool. The *CFTR* F508del-targeting pegRNA candidates were further processed using a

python script (provided by Dr Jakob Haldrup) and analysed by the CRISPResso2 package in Jupyter.

Analysis using CRISPResso2 indicated that prime editing was efficient with the control pegRNA targeting *RNF2* at days 3, 5, and 7 post-transduction (Figure 5.16). No editing was observed in the other controls, contrary to prior experiments where pegRNA targeting *HEK3* showed high levels of editing in both the endogenous loci and artificial target template (Figure 5.9). Upon manual inspection of the trimmed NGS reads, detectable levels of +1CTT insertion were found, however, the frequency was below the detection limit of the assay (>0.1%). The exact reason for the lack of editing at the *HEK3* site in this instance remains unclear. Despite prior evidence of high editing efficiency at this locus, the observed discrepancy in the mini library study could be due to a combination of factors, including variability in transgene expression due to positional effects upon integration in the genome, or differences in experimental conditions; therefore further investigation is required to pinpoint the exact underlying cause. In contrast, prime editing at the *RNF2* site increased from day 3 to day 5 and plateaued at day 7 post-transduction, indicating genomic integration of the library construct and prime editing saturation by day 5. Notably, CRISPResso2 analysis confirmed that the majority of edits were the intended TAC insertion at the +1 position of the nick site. To illustrate this point, a sample collected on day 7 post-transduction had over 70% expected TAC edits, with fewer than 5% unintended edits (Figure 5.16). This tendency was consistent across all samples for *RNF2* target, confirming the consistency of the unexpected issue with the other controls.

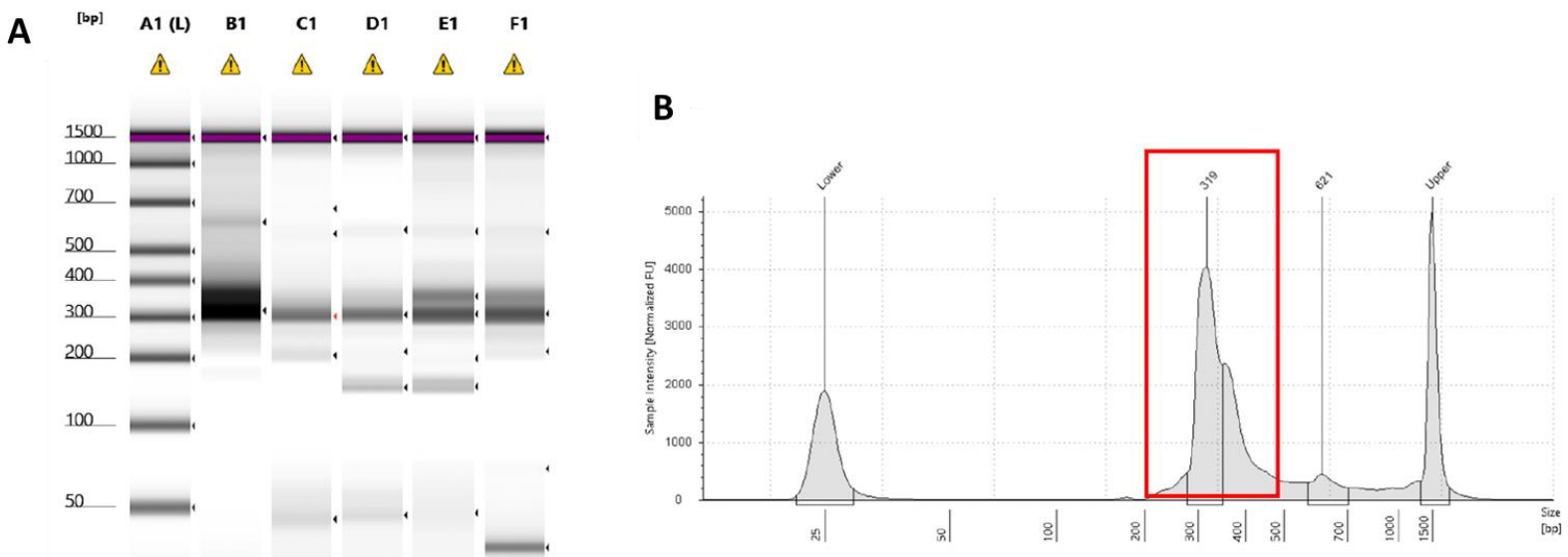

Figure 5.15. Purity of the pooled indexed NGS samples prior to NGS sequencing.

TapeStation was used prior to NGS sequencing to check sample purity. A) Samples from L-R: A1: 100bp ladder; B1: purified, pooled, and indexed mini library samples; C1: unpurified, indexed plasmid library; D1-F1: D3 unpurified, indexed sample replicate 1,2 and 3 respectively. B) A prominent band was observed at 319 bp with low impurities that might affect the NGS reading. From this data I concluded that the NGS sample had a sufficient quality and NGS sequencing could be performed.

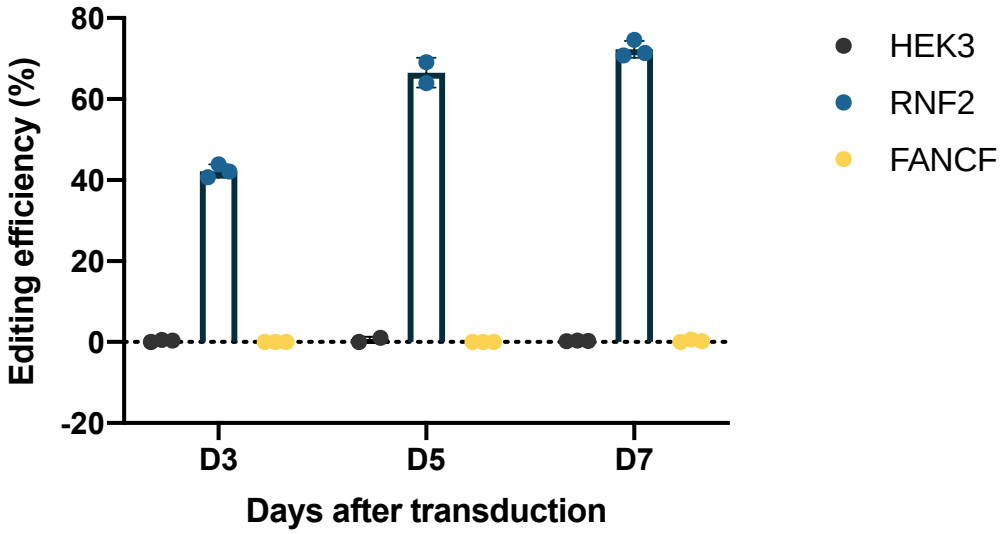

Figure 5.16. Editing efficiency of the positive controls in the library.

Three pegRNAs targeting *HEK3*, *RNF2*, and *FANCF* were used as positive controls in the mini library study. No editing was observed on the *HEK3* and *FANCF* site, but was observed on the *RNF2* site with increasing editing efficiency from D3 to D7. Minimal increases were observed from D5 to D7, indicating that the editing was already saturated on D5.

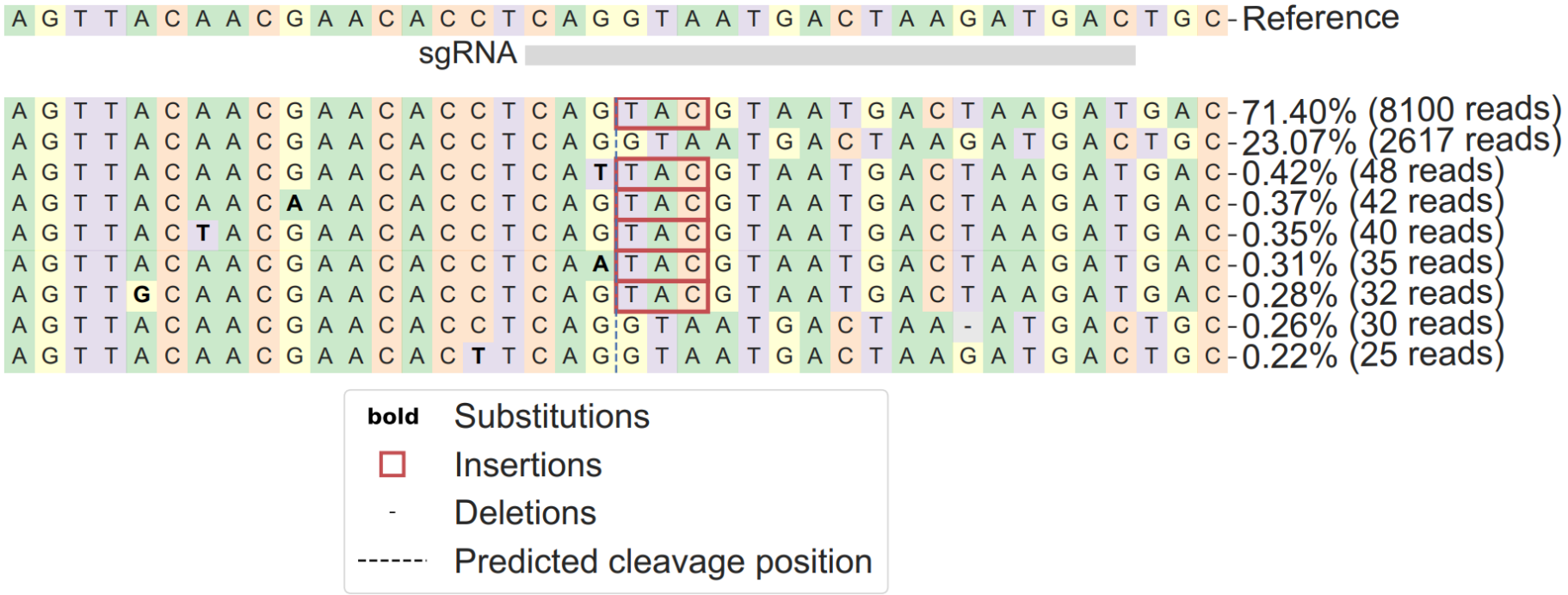

Figure 5.17. Editing efficiency in synthetic *RNF2* sequence on D7 post transduction.

The majority of reads corresponded to the intended edits of TAC insertion in the +1 position of the cut site (71.40%) while unintended edits (row 3-9) occurred in fewer than 3% of the reads.

Following the successful editing observed at the *RNF2* site, pegRNAs targeting the F508del mutation in CFTR were analysed. CRISPResso2 results showed no correct edits for any of the F508del-targeting pegRNAs. Non-specific modifications, including base substitutions, insertions, and deletions, were observed near the F508del region of the synthetic target. However, the overall level of these modifications was low, consistently below 2% across all samples. When analysing the different spacers used in the library, spacer 3 displayed the lowest percentages of modifications, which would be expected given that it is the furthest from the F508del region. Spacer 1 and 2 performed similarly with few pegRNA candidates showing modifications above the detection limit (< 0.1%). This might suggest that with an improved PE configuration, spacer 1 and spacer 2 could perform better in correcting F508del mutation. Due to time constraints, no further troubleshooting could be performed to further optimize the performance of the mini library.

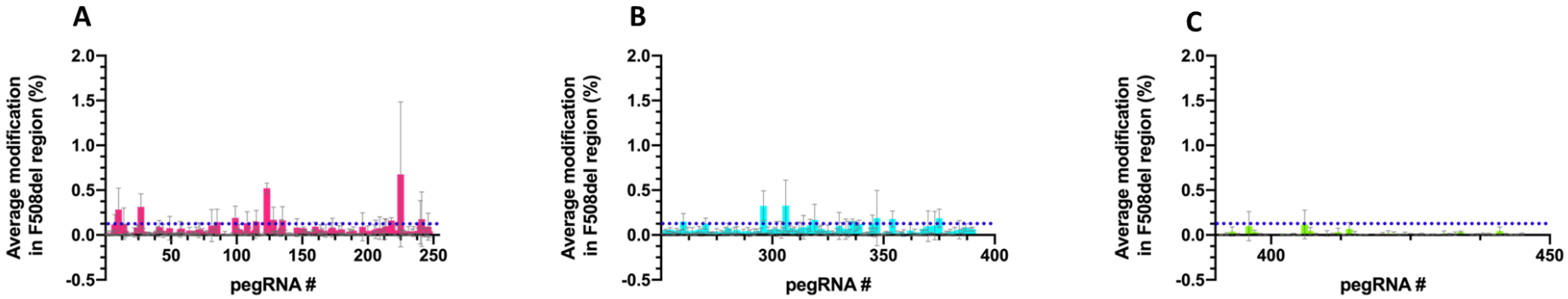

Figure 5.18. Average modification in F508del region of the synthetic target.

Modification in the F508del region of the synthetic target in the library construct from spacer 1 (pegRNA 1 – 250, figure A), spacer 2 (pegRNA 251-390, figure B), and spacer 3 (pegRNA 391-445 , figure C) on D7 post transduction. No intended edit was observed in the F508del region, however, low level of modifications including insertion, deletion, and substitution were observed. Blue dotted line indicates the expected minimum level of modifications above the noise, and therefore considered a true signal.

5.3. Discussion

5.3.1. Utilisation of PEmax-EGFP cell line to ease library screening

Prime editing has a great potential that is unfortunately impeded by the tedious process of selecting the optimal pegRNA sequence to correct specific mutations occurring at a specific target site. In the context of a therapeutic approach, the ideal pegRNA should also have high efficacy and low or no off-target effects. A stable cell line constitutively expressing PEmax-EGFP was generated to adopt a high throughput screening platform for pegRNAs (Figure 5.3, Figure 5.4), and in turn, ease the process of pegRNA selection for a target of interest. In this instance, the *CFTR* F508del target functioned as exemplar.

Several techniques such as western blot, and flow cytometry were used to characterize the cell line. The difference of the Medium and High EGFP expressing group is reflected in the MFI of the cells, where the average MFI of the High GFP group was approximately 2.5 times higher than the MFI of the Medium GFP group. From the western blot (Figure 5.5), it appears that the PEmax protein expression from the HEK293T-PEmax-EGFP is too low to be detected when compared with HEK293T cells transfected with plasmids expressing PEmax-EGFP. However, this does not compromise the ability to perform prime editing as evident in the sequencing result from transfecting the cell line with *HEK3* targeting pegRNA plasmid (Figure 5.6, Figure 5.7, and Figure 5.8). The High 4/B1 clone showed correct CTT insertion with 33% editing efficiency on the first transfection experiment. The second clone that performed consistently was the Med 5/B2 clone, which showed an average of 15% editing in two different replicates (Figure 5.6). However, the editing that was detected by the DECODR algorithm was a CTA insertion, which occurs in over 70% of the total edits, with

the correct editing of CTT insertion detected in only 30% of the total edits. Further experiments must be performed to address this discrepancy including sequencing of the pegRNA expression backbone to conform integrity of the pegRNA. The High 4/B1 clone was selected for subsequent experiments, as it was the only clone that exhibited the correct CTT insertion at the *HEK3* locus.

To validate the feasibility of the mini library strategy, a preliminary experiment was conducted using a single positive control oligo targeting the *HEK3* locus, cloned into the pEF1a.mCherry.hU6.pegRNAacc plasmid, which was then used to generate the lentiviral vector (Table 2.8). Both the endogenous *HEK3* locus and the synthetic target incorporated in the pegRNA construct was PCR amplified and sequenced to determine whether editing had occurred in either region. The sequencing result showed that editing had occurred in both the synthetic target and endogenous locus, and there appears to be a preference for the endogenous locus over synthetic target. This is surprising as the lentivirus tends to integrate in actively transcribed regions of the genome and thus accessible to the prime editor, which may be restricted at endogenous loci. If this is a general trend, it might suggest that the editing efficiency observed from the pegRNA, through this high throughput mini library strategy, is an underestimate, and the pegRNA could perform better in the endogenous context. Previous studies have shown positive correlations between endogenous and integrated target site editing[56,164] in similar experimental setup, where an increased level of editing in the target site was observed with an increased editing in the endogenous target. In Kim et al.[164], generally the editing efficiency observed in the

endogenous sites are lower than in the integrated target sites, however there are some exceptions to this trend.

5.3.2. Production of mini library plasmid dependent on quality of oligo pool

The mini library strategy aims to provide a high throughput method for selecting the best pegRNA for a specific target. By designing a construct that is composed of the pegRNA with different 3' extension length, a unique barcode, and a synthetic target, the editing efficiency of each pegRNA could be assessed through NGS, and in turn, the most appropriate pegRNA for the target sequence could be identified and further validated in more relevant models and especially targeting the endogenous locus. This approach, however, relies on the quality of the commercial oligonucleotide pools that was used to generate the elements of the mini library. In this study, it was found that the control pegRNAs were heavily overrepresented in the pool and therefore affecting the representation of the mini library elements. Over 80% of the pegRNAs targeting the F508del mutation in *CFTR* was only represented less than 100 times in the final pool. In comparison, the control pegRNAs were represented on average 16000 times in the final pool. This is likely due to technical aspects of the oligonucleotide synthesis as the same constant primer region was used to PCR amplify each library element to minimize bias. Nevertheless, due to time and fund constraints, it was not possible to reorder the oligonucleotide pool and redo the mini library production, and the results presented in this chapter were obtained from the skewed library. Regardless, a coverage of less than 100 for *CFTR* F508del-targeting pegRNAs should be adequate to detect some editing, but it may not guarantee the identification of the best-performing pegRNA due to under sampling.

In this study MOI of 0.3 was used to ensure that no more than one element of the mini library integrated into genome of the HEK293T-PEmax-EGFP cell line post-transduction. The positive controls, which target *HEK3*, *RNF2*, and *FANCF*, did not show consistent editing. The pegRNA targeting the *HEK3* locus, which was previously shown to edit both on the synthetic and endogenous target (Figure 5.9), was unable to edit the *HEK3* synthetic target when delivered as a component of the mini library. Similarly, the pegRNA targeting *FANCF* also showed no editing. Meanwhile the pegRNA targeting *RNF2* showed a consistent increase in editing efficiency on days 3, 5, and 7 post-transduction. The remaining pegRNAs targeting the F508del mutation on *CFTR* also failed to show any editing. The inconsistent results from the positive controls regrettably render the experiment's outcomes unreliable, making it difficult to draw a definitive conclusion about whether all pegRNAs targeting the F508del mutation would be ineffective if the experiment were to be repeated. One future study to help explain the issue would be to design an experiment which incorporates five different pegRNAs in a plasmid pool, containing the three positive controls from this study, and two negative controls to check whether it is possible to obtain consistent editing from all the positive controls in a smaller plasmid pool. In the section below, several improvements to enhance the mini library strategy are highlighted.

5.3.3. Potential improvements for mini library strategy

The mini library strategy offers a potential solution for a high throughput pegRNA screening. However, some important improvements need to be made to further establish the platform. The utilisation of the HEK293T-PEmax-EGFP cell line generated in this study for the mini library strategy simplified the process of pegRNA selection as it removed the need

for a transfection step after the library transduction i.e. to introduce the PEmax effector component. The EF1α promoter that drives the PEmax-EGFP expression is also stable for a long period of time when compared to other promoter such as CMV[194], which is ideal for this application. The low level of PEmax protein expression that could not be detected through western blot suggested that the initial MOI of the lentivirus used to make the cell line (Table 2.8) could be increased to obtain more copies of the PEmax-EGFP integrated into the genome. Kim et al.[164] used PE2 transfection after the introduction of the library, hence it is possible that a more effective cell line could be generated using much higher MOI than that used in this study. Employing plasmid DNA transfection to achieve high levels of PE protein may increase the prime editing efficiency in a pooled library setting, however, this method can lead to a highly heterogeneous cell population with varying levels of prime editor protein, potentially introducing bias into the results. Nevertheless, this study has demonstrated that the PEmax-GFP cell line is capable of achieving editing at the *HEK3* site in a non-pooled environment.

The design of the mini library construct incorporates the spacer, scaffold, and different length of 3' extension of the pegRNA. Since the commencement of this study, improvements in the field of pegRNA have recently been published. One such improvement is the inclusion of the stable 42 nt pseudoknot at the 3' end of the pegRNA, which is the modified prequeosine1-1 riboswitch aptamer (evopreQ1) motif[66], which improves the stability of the pegRNA and prevents the degradation of the 3' extension region, termed as the enhanced pegRNA (epegRNA). Since the degradation of the 3' extension region is reduced, the editing efficiency of the pegRNA is also increased and therefore the addition

of the evopreQ1 motif could highly improve the pegRNA library construct. As the currently available commercial suppliers could not synthesise oligo pools with oligos longer than 350 nt, it was not possible to directly add the evopreQ1 motif to the design. Rather, the design could be improved by adding another Type II S restriction site on the 3' end of the RT template, such as BbsI, so that the motif could be added through a two-step cloning strategy (Figure 5.19). However, this increases the complexity of the cloning strategy and introduces additional unwanted risks to the final library coverage. Such changes to pegRNA design might enhance future screens.

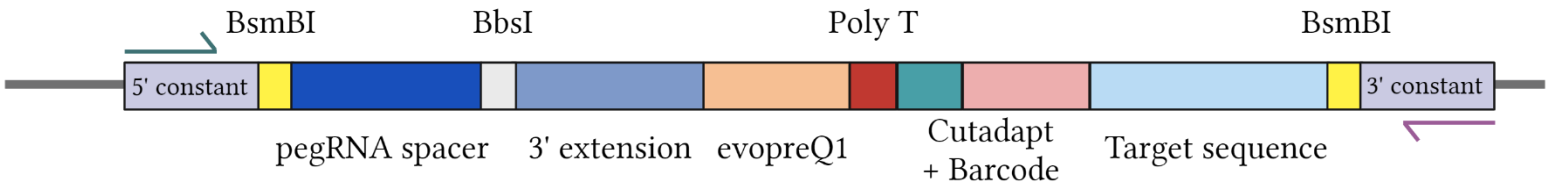

Figure 5.19. Potential improvement to the mini library construct by addition of evopreQ1 motif.

The evopreQ1 motif could be added on the 3' end of the pegRNA 3' extension and substituting the scaffold sequence for a BbsI recognition sequence. This configuration would require a two-step cloning process, which adds to the complexity of the system and might affect the final library coverage.

5.3.4. Challenges associated with correcting the F508del mutation in CFTR

A recent publication from David Liu's group[158] has described in detail the challenges in correcting the F508del mutation in *CFTR.* Using a HEK293T cell line model harbouring the F508del mutation, they found that the F508del mutation is particularly difficult to correct, especially when using the earlier version of PE such as PE2. This was also the case even after optimising the length of RT template and PBS. The addition of the nicking guide RNA (PE3) does not help, with either approach only producing a negligible level of correction (less than 0.5%) and producing higher level of indels (up to 1.5%).

In this study, using the most optimal combination of the equivalent of spacer 2, with primer binding site length of 13bp and RT template length of 41bp, and using a nicking guide RNA in position +104; the system was further re-optimized by combining various PE enhancements such as the most recent version of PE (PE6c), using epegRNA, co-expression of MLH1dn, and incorporation of silent edits improved the editing efficiency to 11%. When the system was adapted to a relevant cell line (16HBE14o-) harbouring the F508del mutation, another improvement was pursued by using a catalytically inactive or "dead" single gRNA (dsgRNA) to increase the accessibility of the F508del locus to the prime editor. When all the enhancements were combined, this further increased the editing efficiency to 51% with 13% indel formation.

These findings highlight why correction of the F508del mutation in *CFTR* is an arduous task, which is not easily accomplished using simple PEmax approach, but rather requires combining all the possible improvements to the PE system. Due to practical time

constraints, and the speed at which the field of prime editing was developing throughout this DPhil, I was not able to incorporate all the improvement iterations into my experiments. In future studies, an effort should be made to incorporate as many improvements as possible when assessing how prime editing could be used to correct a specific disease mutation.

Chapter 6: Discussion

Delivery of gene editing tools such as PE poses a challenge due to the large size of the editor protein. In this study, lentiviral delivery approaches were explored, including 'all-in-one' PE-LV, (Chapter 3) and 'two-vector' delivery for PE and also HITI (Chapter 4). A mini library was also generated for screening optimal pegRNA candidates to target F50del mutation in CFTR. There are a number of benefits, challenges, and potential future developments that I will discuss.

6.1. Viral Delivery of Prime Editing and HITI

In translational research, one aspect of consideration for developing successful gene therapeutics is the delivery method of therapeutic agents. Both viral and non-viral delivery methods have been explored for delivery of CRISPR-Cas9-based gene editing technology, such as PE and HITI. In this thesis, I have focused on LV-based delivery and explored the all-in-one PE-LV (Chapter 3) and two-vector approach for delivering PE and HITI (Chapter 4).

Viral delivery, particularly recombinant LVs, have several advantages due to their large packaging size (up to 9.6 kb[84]), the ability to transduce dividing and non-dividing cells[81,82], and the potential to target specific types of cells in the body through pseudotyping[89]. Third generation LVs, which were used in this thesis, have an improved safety profile due to removal of accessory genes, the use of SIN LTRs, and the division of viral components: four different plasmids contain gag and pol, rev, the pseudotyping glycoprotein, and the transfer genome plasmid, which minimises the possibility of homologous recombination that could

result in replication competent LVs[195]. These advantages make LVs an appealing method of gene delivery.

Due to the ability of the LV to integrate into the host genome, lentiviral delivery is frequently employed for proof-of-principle applications, such as *in vitro* or *in vivo* studies for Chimeric Antigen Receptors T-cell (CAR-T)[196], sickle cell disease[197], and hereditary tyrosinemia type 1[198]. It is also widely used for *ex vivo* gene therapy in hematopoietic stem cells[199–202], and *ex vivo* CAR-T studies[203,204]. Additionally, it has been applied in preclinical gene replacement studies for CF[162], haemophilia B[205], and macular degeneration[206]. Therefore, there is already a strong foundation for the use of LVs for the treatment of genetic diseases, although, LVs have been used less frequently for *in vivo* engineering of CAR-T cells[207].

In preclinical and clinical contexts, both AAVs and LVs have been used to deliver Cas9-based editing tools[208]. Recombinant AAV vectors have been used to deliver the CRISPR-Cas9 components for BE[209–211], PE[111,112,115,160], and HITI[19,78,79,212] for corrections of various mutations. Meanwhile, integrase-competent LVs (ICLVs) are more commonly used for CRISPR-Cas9 knockout library screening[213,214], mouse model generation for acute myeloid leukaemia[215]. Lentiviral delivery has also been used for PE, in particular for establishing pegRNA design rules[164] and more recently epegRNAs[216], prediction of insertion-type editing efficiency[217], and screening of genetic variants in human genome[218,219].

Currently, there are several FDA approved viral-based gene therapy such as Luxturna (AAV based) for treating biallelic RPE65-mutation-associated retinal dystrophy[220]; Zynteglo (LV based) for *ex vivo* delivery of a functional copy of the beta-globin gene to the patient's hematopoietic stem cells for treatment of β-thalassemia[221]; Skysona (LV based) for *ex vivo* delivery of ABCD1 cDNA to haematopoietic stem cells for treatment cerebral adrenoleukodystrophy[222]; Kymriah (CAR-T) for autologous treatment of relapsed/refractory paediatric and young adult acute lymphoblastic leukaemia[223]; and Zolgensma (AAV9 based) for treatment of spinal muscular atrophy[224].

Due to the integrating nature of ICLVs, these vectors are not ideal for in vivo delivery of CRISPR-Cas9 gene editing technologies. This approach carries the risk of insertional mutagenesis due to viral integration[225], particularly when combined with Cas9, which can induce genome instability through nicks or double-strand breaks (DSBs)[195,226]. It is also not desirable to have the cells constantly expressing gene editing machinery, as it increases the risk for off-target editing and immunogenicity. While this may not be a significant issue in proof-of-principle studies performed in this thesis, for translational and therapeutic applications, alternative delivery methods such as IDLVs or lentiviral-derived nanoparticles should be explored to improve safety.

The IDLVs offer similar advantages to ICLVs, including a large packaging capacity and the opportunity for pseudotyping to target various cell types. After IDLV cell entry, the viral RNA is reverse transcribed into cDNA[103]. However, because the integrase enzyme is mutated to render its function redundant (e.g. the D64V mutation[227]), the risk of DNA integration into

the host genome is significantly reduced; although, low levels of genome integration (up to 1 in 200) have still been reported using IDLV[228]. Instead of integrating, the viral DNA forms episomal circles, primarily 1-LTR and 2-LTR circles[103,229]. These episomal circles are still able to mediate transgene expression, but at a lower level than achieved by integrated viral DNA. Over time, the episomes are diluted through successive cell division, or slowly degraded[103]. Due to this transient nature of the episomal circles, IDLV is more suitable for delivery of gene editing technology such as Cas9 or PE, where long term expression is unwanted. There has been reported use of IDLV for Cas9 delivery[101,230,231], but so far there has been no report delivery of PE.

Another approach involves the use of viral-derived particles[232], such as lentiviral-derived nanoparticles that have been employed to deliver Cas9, BE, and PE[233]. Similar, engineered viral-like particles have been used for the delivery of BE and PE[121], and enveloped delivery vehicles have been utilized for Cas9 delivery[207], all in the form of ribonucleoprotein (RNP) complexes. Since the RNPs persist only transiently to carry out the intended genome edit, and are then degraded by normal cellular processes[234], the risk of using the viral derived particles as a delivery vehicle are minimal.

6.1.1. Challenges for the all-in-one delivery strategy

The all-in-one vector approach aimed to provide a simple delivery strategy for PE by only using a single vector to deliver all the necessary components of PE. The construct was developed using rSIV, and also the synthetic hCEF promoter driving expression of the PE2 construct, as the goal was to use this vector for correction of F508del mutation in *CFTR* in

relevant cell models (section 3.2.2). At first glance, the all-in-one approach appears to have a lot of potential, however much optimisation was required to further improve this method.

The total genome size was almost 10 kb (Figure 3.2), and in line with a published study[167], the increase in genome size resulted in decreased titre (Figure 3.4). Although the titre obtained was sufficient for *in vitro* experiments, this would undoubtedly present challenges in scale-up manufacture, particularly for animal studies where higher titres and doses are often required to be administered to elicit measurable effects.

The hU6 promoter was cloned, driving the pegRNA in both forward and reverse orientations. Previous studies in AAV have shown the effect of hU6 directionality in reduction of Cas9 editing efficiency[235]. However, the same effect was not observed using transfection of lentiviral genome plasmid constructs described in this thesis, where the difference in the editing level of the *HEK3* locus was not statistically significant (Figure 3.3). Other studies using the multiple pol III promoter in LVs also showed that there was no difference in sgRNA expression and genome editing potency from differently oriented pol III promoters[236]. This might serve as an early indication that directionality of pol III promoters in LVs does not affect the editing efficiency of PE, although, more studies using rLVs are needed to confirm this.

As the ultimate goal was to use the all-in-one PE-LV for correction of the F508del mutation in *CFTR*, this design was based on the LVs used in the first-in-human clinical trial for CF[162].

However, this added a layer to the editing aspect using PE. The efficiency of PE varies greatly depending on the target cell types, types of editing, and the sequence of the target locus itself which also affect the pegRNA design[164] By using a lentiviral vector previously optimized for lung cell types, the cell line model that could be used as a target was already limited. The hCEF promoter, which was optimised for expression in airway cells, does not show adequate activity in HEK293T cells based on the level of GFP expression observed post-transduction with simple rSIV.F/HN hCEF EGFP LVs (Figure 3.5). No editing of the *HEK3* locus was observed in HEK293T cells (Figure 3.6). Further investigation would be required to determine if this lack of editing was due to the promoter limiting the amount of effector editing protein expressed in transduced cells. This could be tested by repeating the experiment using the same LV vector but replacing the hCEF promoter with either the CMV or EF1α promoter, both of which are known to be active in HEK293T cells, or by increasing the MOI for transduction to deliver more PE machinery into the cells.

Alternatively, it has been reported that PE has very limited editing efficiency in cell lines other than HEK293T, due to absence of mismatch repair mechanism in HEK293T cells[60,64]. Therefore, repeating this experiment with the same setup, but using a cell line where the hCEF promoter is more active, may also struggle to produce any editing. This could be mitigated by providing the MLH1dn to impair the mismatch repair function in the cells, although with the vector capacity already stretched from accommodating the original PE construct, the MLH1dn would need to be introduced in a separate plasmid transfection or vector transduction, thus negating the benefit of the all-in-one approach.

In future studies, several advancements in both PE and pegRNA design could be incorporated to optimize the all-in-one LV approach. This includes using a mutated Cas9n to enhance editing efficiency[62], employing truncated versions of RT[115,170], or utilizing a smaller PE system altogether[237] to reduce vector size. Furthermore, adding a reporter protein could facilitate troubleshooting, while incorporating epegRNAs[66] could further boost editing efficiency. Additionally, replacing the RT domain entirely could enhance the performance of the PE construct. Viral RT enzymes are more prone to errors compared to DNA polymerases because they lack proofreading mechanisms. Substituting the viral RT with a DNA polymerase and replacing the RNA in the RT template region with DNA may improve the fidelity of the edits introduced into the genome, leading to more accurate gene modifications[238].

6.1.2. Challenges in two-vector delivery strategy

The two-vector approach was explored when the all-in-one delivery was not feasible to be pursued further. This strategy was used to deliver either PE2 or HITI. For delivery of PE2, the Cas9n component was divided into two parts, each fused to an intein fragment to facilitate the protein trans-splicing process. In addition to this, because the bulky and large PE2 sequence was split, any single vector could now package more transgene, so we were able to include a P2A-EGFP reporter linked to the N-terminus fragment, whilst the C-terminus fragment was linked to MMLV-RT using the original flexible linker. These two components were then inserted into two separate lentiviral genome transfer plasmids, with one also carrying the hU6 promoter and pegRNA sequence (Figure 4.7).

For HITI, the Cas9 component with T2A-GFP sequence driven under CMV promoter was encoded by one lentiviral genome plasmid, whilst the sgRNA targeting the intron 10 of CFTR and donor template consisting of CFTR superexon 11-27 was encoded into separate lentiviral genome plasmid (Figure 4.10 and Figure 4.11).

Initially, the editing efficiency of the split PE (section 4.2.4.1) and HITI (section 4.2.5.1) approaches were assessed using simple plasmid transfections to start with. The trans-splicing process of the N-terminus and C-terminus of PE protein was not very efficient (Figure 4.8). As the antibody that was used in the western blotting was raised against the N-terminus of Cas9, the C-terminus could not be detected and therefore it was unknown if the low trans-splicing efficiency was due to low co-transfection level of the two plasmid components. However, editing efficiency of the split PE still averaged around (11.33%±4.04%), which is comparable to the editing efficiency of PE2-GFP and HEK3 +CTT pegRNA plasmid transfection at (17.67%±5.86%) (Figure 4.9B), this indicates that there is a potential for this method to be applied. Due to time constraints, the production of split PE lentiviral vectors and the evaluation of editing efficiency using a two-vector delivery system could not be completed. Lentiviral delivery would be an ideal approach, as it allows for precise control of the MOI and, consequently, the number of split-PE copies delivered to each cell; this method should be explored further in future studies.

While the *CFTR* superexon integration theoretically offers a mutation-agnostic approach for the correction of *CFTR*, the integration efficiency of the HITI superexon donor was very low as evidenced by gel electrophoresis and sequencing results (Figure 4.12). This finding aligns

with other studies reporting similarly low integration rates using HITI[171,183,184], suggesting that this was not the most ideal method for correcting *CFTR* mutations and was therefore not pursued further. An alternative that could be explored for integrating a large genome such as the superexon donor is PASSIGE[239] or PASTE[240], both of which combine PE with large serine recombinases for integration of multi-kilobases of genes. The strategy using PASSIGE is also being pursued by Prime Medicine for correction of *CFTR* using single superexon integration[76] suggesting that this could be a promising course of action for *CFTR* correction.

While providing more room for packaging, allowing the incorporation of reporter gene such as GFP, the two-vector approach also increased the associated challenges of delivery. Firstly, co-infection of two vectors carrying both components of the split PE or HITI is necessary for any editing to occur. The chances of this are lower compared with the all-in-one approach which only required one plasmid or LV to transduce cells to deliver all the necessary components for targeted genome editing, thus lowering the likelihood of successful editing using a split vector approach. In the case of split PE, this was further complicated by the intein-mediated protein trans-splicing process, although the use of Npu intein, known for its efficiency and widespread use[241–243] helped mitigate this issue. Interestingly, studies have shown that untethered PE2, where Cas9n and MMLV-RT are delivered in two separate AAVs without a flexible linker, achieves editing of comparable efficiency to the standard, tethered PE2 system[182]. This could serve as an alternative strategy for split PE, eliminating the intein-mediated trans-splicing process in the two-vector approach. Secondly, from a translational perspective, more vectors and potentially higher doses are needed to deliver to patients to compensate for the reduced efficiency. This both increases the complexity of

delivery, and the risk associated with higher dose of LV delivery such as toxicity and immunogenicity [225,244,245], which has been similarly observed in AAV[246–249].

6.2. Use of HEK3 site as a reference for editing efficiency

The experiments in this study primarily used the *HEK3* site in HEK293T cells as a reference for evaluating editing efficiency (Figure 3.2, Figure 3.6, Figure 4.9, Figure 5.6, and Figure 5.9). This locus was originally described in the GUIDE-seq paper for profiling off-target cleavage by CRISPR-Cas nucleases[250] and has since been adopted as a benchmark in the development of base editing[31,34] and prime editing[56].

The use of the *HEK3* site, especially when targeted in HEK293T cells, is a good start point for benchmarking and as a positive control for editing. However, PE efficiency varies greatly in different cell types even for *HEK3* site. Using the PE3 construct, it was shown that editing efficiency ranges from ~80% in HEK293T cells to ~20% in U2OS cells and 12%~ in HeLa cells for +1 CTT insertion at *HEK3* locus[56]. Recently published studies have shown that the high activity of PE in HEK293T cells is partly due to the lack of mismatch repair[60,64] and the high level of intracellular dNTPs[251]. These conditions are not always reflective of other cell types, particularly primary cells. Another study investigated the effect of positioning within the chromatin context by using a piggyBac reporter, which included a T7 promoter, a *HEK3* target sequence, and a short barcode. This reporter was randomly integrated into the genome of PE2-expressing K562 cells to explore how different chromatin environments influence prime editing efficiency. The results showed that even with consistent *HEK3* target sequence, the editing efficiency varied greatly depending on the genomic location[252].

Therefore, it is likely that the editing efficiency at the *HEK3* site would be higher than at actual target sites in more clinically relevant cells, limiting its translational applicability.

6.3. Using a mini library approach for high throughput pegRNA screening

In Chapter 5, a method was devised for high throughput pegRNA screening for correction of the F508del mutation in *CFTR* using a mini LV library. This strategy was adapted from a study by Kim et al.[164] where they used a library for predicting pegRNA efficiency in different targets in human cells. However, I opted to generate a HEK293T cell line stably expressing PEmax-P2A-EGFP that were isolated after FACS sorting, rather than introducing the PE component via transient transfection. A recently published study for screening human genetic variants also used a similar approach[218] by generating stable PEmax-expressing A549 cells via puromycin selection. My choice and use of selecting FACS sorted single-cell stable cell clones helped address the limitations of not having a cell line expressing a consistent level of the PE protein in each cell. As a result, this approach aimed to improve control over protein expression compared with transient plasmid transfections, where the amount of plasmid introduced into each cell can vary significantly depending on how well complexed the plasmid DNA is with the transfection reagent in any given study. Ultimately, the latter approach was thought to potentially impact the editing readout and would not be useful to help select optimal and potent pegRNAs for correcting the *CFTR* F508del mutation.

The MFI of EGFP fluorescence in various HEK293T-PEmax-EGFP clones (Figure 5.4) and western blot analysis (Figure 5.5) were methods used to evaluate the PEmax expression. Flow cytometry analysis showed significant difference between the MFI level of the

Medium- and High-expressing clones, indicating different level of PEmax protein expression in both groups. Western blotting could not detect the PEmax expression, however the evidence of successful editing suggested that the integration of PEmax-P2A-EGFP construct was achieved, and that the construct was functioning correctly (Figure 5.6, Figure 5.8, and Figure 5.9). Further characterisation, such as using droplet digital PCR (ddPCR), could be performed to determine the number of copies integrated into the genome of these cells. These cell lines were, however, successful in generating the intended edits at the *HEK3* locus after plasmid transfection of the mini library component, mainly comprising the *HEK3* targeting pegRNA and relevant synthetic target (Figure 5.8), and also at the *RNF2* site after transducing the stable cell line with a mini pegRNA-synthetic target library (Figure 5.17), proving the functionality of the integrated PEmax construct that is co-expressed with an in-frame EGFP reporter.

The mini-library approach enables the screening of tens of thousands of pegRNA candidates and can be easily adapted to target other clinically relevant mutations. However, as the library was commercially synthesised, the cost associated with the library increases proportionally with the number of elements included in the library. Predictions based on software such as pegIT[174], pegFinder[253], PrimeDesign[254], or PRIDICT [255], along with the updated version PRIDICT2.0 version; trained using mismatch repair-proficient K562 cells, and ePRIDICT, which accounts for epigenetic features in different chromatin environments[256], can be used for initial screening of potential pegRNA designs for a small-scale preliminary testing. If any of the recommended pegRNAs from the software-based prediction work, then using a library for high throughput screening may not be necessary.

6.4. Prime editing for correction of *CFTR* mutations

Despite all the challenges associated with delivery and pegRNA optimisation, PE remains an attractive option for the correction of disease-causing mutations, such as in *CFTR*. Prime editing has been used successfully to correct CF-causing mutations, L227R- and N1303K, in *CFTR* using patient-derived organoid and human nasal epithelial cells as models[257]. In this thesis, F508del was used as an exemplar mutation of *CFTR* for correction using prime editing, as it is the most common mutation, affecting more than 90% of people with CF[137]. Although treatments using small molecule modulators already exist for individuals with the F508del mutation[137,139,142], these therapies do not address the underlying genetic mutation, meaning they must be taken for life. On top of that, not all people with CF can benefit from these therapies depending on their mutations, and adverse effects are also observed in a small cohort[142]. Prime editing offers a more permanent solution by directly correcting the genetic defect at its source and offers a solution for individuals with mutations that are not suitable for treatment with small molecule modulators.

Efforts toward gene addition therapy for treating CF are currently underway, including using F/HN pseudotyped SIV vectors to deliver the *CFTR* transgene[145,162] , and also using AAV[149]. The use of a strong hCEF promoter to drive the expression of the cDNA further ensures high levels of CFTR protein expression in the lungs. Previous studies have reported that restoring 10% to 35% of CFTR protein activity is sufficient to achieve significant improvements in lung function and ion transport[143,144]. However, in the context of endogenous correction using PE, it is important to recognize that the activity of the endogenous *CFTR* promoter may be lower compared with strong promoters such as hCEF, CMV, or EF1-α. Consequently,

achieving a correction level greater than the 10% baseline at the endogenous locus might be required to observe significant clinical improvements.

The next aspect is to target the appropriate cell type for maximum impact of correction. Although ionocytes express the highest levels of *CFTR* mRNA[132,133], these cells are rare in the lungs, resulting in a relatively low overall production of CFTR protein. In contrast, secretory cells produce the most CFTR protein[258], suggesting that targeting gene editing to secretory cells, rather than ionocytes, may be the most effective strategy for achieving high levels of CFTR expression. Focusing gene editing on the right cell types could improve the phenotypical and, consequently, clinical outcomes, making the therapy more beneficial for patients.

The mini library screening for pegRNA candidates aimed at correcting the F508del mutation did not yield an optimal pegRNA (Figure 5.18). However, this particular mutation has proven to be quite challenging to target. Reports published by David Liu's group[158] had to employ several enhancements to prime editing, such as PE6c, epegRNA, incorporating silent edits, and using dsgRNA to effectively correct the mutation. The PE6c has demonstrated increased editing efficiency, showing up to a 1.5-fold improvement compared with PEmax. Additionally, this has proven to perform well on targets that require longer and structurally more complex RT templates[62,158]. Engineered pegRNA incorporates a pseudoknot motif that prevent degradation by endogenous exonuclease[66,158]. Silent edits can be introduced to disrupt PAM or to install additional silent mutations on the target sequence to evade MMR recognition[56,60,64,158], and thereby enhance the editing efficiency. Lastly, use of dsgRNA, a truncated gRNA with only 14 nt – 16 nt complementarity that enables binding

of Cas9 but not cleavage of target DNA, unwinds the chromatin structure and increases the availability of the target site[158,259]. All these efforts combined could significantly boost prime editing and allows correction even in hard-to-target mutations, such as the F508del. While promising, this approach presents challenges for clinical translation, as it requires the introduction of multiple additional components, which could increase the complexity and associated risks of the therapy. Therefore, a more streamlined and clinically viable strategy must be developed to achieve efficient correction of challenging mutations, such as F508del.

Finally, The selection of pegRNA using the mini library to target the F508del mutation was carried out in HEK293T cells stably expressing PEmax and GFP (section 5.2.1). After pegRNA selection, conducting additional experiments to correct the F508del mutation in the endogenous CFTR gene using a more relevant cell line, such as human lung bronchial epithelial 16HBE14o- cells, would provide significant advantages by enabling more physiologically relevant assessments. Functionality testing of CFTR in this model, such as forskolin-induced CFTR-specific swelling assays[155,173,257], can directly evaluate the restored chloride channel function post-correction. Therefore, this approach ensures that the editing strategy is not only effective at the genetic level but also translates into functional recovery at the cellular level, providing critical insights into its therapeutic potential.

6.5. Conclusions

Prime editing has great potential as a genome editing tool, but faces a number of challenges, such as efficient delivery of the large PE protein to cells. In the studies presented

in this thesis, an all-in-one lentiviral vector for efficient delivery of PE into HEK293T cells was developed and investigated. A two-vector approach was also explored for the delivery of PE and HITI, where transfection of the lentiviral genome plasmid demonstrated editing at the HEK3 site for PE, alongside low-level integration of the CFTR superexon donor for HITI. Additionally, a high-throughput pegRNA screening method using a mini-library approach was tested on the F508del mutation in CFTR. While the challenging nature of the target site required further optimization to identify the optimal pegRNA, these results provide a foundation for future refinement of gene editing approaches for inherited diseases.

References

1. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (1979)* **337**, 816–821 (2012).
2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (1979)* **339**, 819–823 (2013).
3. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).
4. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
5. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
6. Chatterjee, P. *et al.* An engineered ScCas9 with broad PAM range and high specificity and activity. *Nat Biotechnol* **38**, 1154–1158 (2020).
7. Müller, M. *et al.* Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Molecular Therapy* **24**, 636–644 (2016).
8. Redman, M., King, A., Watson, C. & King, D. What is CRISPR/Cas9? *Arch Dis Child Educ Pract Ed* **101**, 213–215 (2016).
9. Lieber, M. R. The Mechanism of Human Nonhomologous DNA End Joining. *Journal of Biological Chemistry* **283**, 1–5 (2008).
10. MAO, Z., BOZZELLA, M., SELUANOV, A. & GORBUNOVA, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)* **7**, 1765–1771 (2008).
11. Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y. & Pomerantz, R. T. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat Struct Mol Biol* **22**, 230–237 (2015).
12. Black, S. J. *et al.* Molecular basis of microhomology-mediated end-joining by purified full-length Polθ. *Nat Commun* **10**, 4423 (2019).
13. Wang, H. & Xu, X. Microhomology-mediated end joining: new players join the team. *Cell Biosci* **7**, 6 (2017).
14. Song, B., Yang, S., Hwang, G.-H., Yu, J. & Bae, S. Analysis of NHEJ-Based DNA Repair after CRISPR-Mediated DNA Cleavage. *Int J Mol Sci* **22**, 6397 (2021).
15. Pawelczak, K. S., Gavande, N. S., VanderVere-Carozza, P. S. & Turchi, J. J. Modulating DNA Repair Pathways to Improve Precision Genome Engineering. *ACS Chem Biol* **13**, 389–396 (2018).

16. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proceedings of the National Academy of Sciences* **95**, 5172–5177 (1998).

17. Kakarougkas, A. & Jeggo, P. A. DNA DSB repair pathway choice: an orchestrated handover mechanism. *Br J Radiol* **87**, 20130685 (2014).

18. Suzuki, K. & Izpisua Belmonte, J. C. In vivo genome editing via the HITI method as a tool for gene therapy. *J Hum Genet* **63**, 157–164 (2018).

19. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).

20. Iyama, T. & Wilson, D. M. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* **12**, 620–636 (2013).

21. Liu, M. *et al.* Methodologies for Improving HDR Efficiency. *Front Genet* **9**, (2019).

22. Aird, E. J., Lovendahl, K. N., St. Martin, A., Harris, R. S. & Gordon, W. R. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun Biol* **1**, 54 (2018).

23. Cannan, W. J. & Pederson, D. S. Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin. *J Cell Physiol* **231**, 3–14 (2016).

24. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Mol Cell* **40**, 179–204 (2010).

25. Krenning, L., van den Berg, J. & Medema, R. H. Life or Death after a Break: What Determines the Choice? *Mol Cell* **76**, 346–358 (2019).

26. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765–771 (2018).

27. Boroviak, K., Fu, B., Yang, F., Doe, B. & Bradley, A. Revealing hidden complexities of genomic rearrangements generated with Cas9. *Sci Rep* **7**, 12867 (2017).

28. Kraft, K. *et al.* Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Rep* **10**, 833–839 (2015).

29. Brunner, E. *et al.* CRISPR-induced double-strand breaks trigger recombination between homologous chromosome arms. *Life Sci Alliance* **2**, e201800267 (2019).

30. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109**, (2012).

31. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

32. Yu, Y. *et al.* Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat Commun* **11**, 2052 (2020).

33. Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, (2017).

34. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

35. Jensen, T. I. *et al.* Targeted regulation of transcription in primary cells using CRISPRa and CRISPRi. *Genome Res* **31**, 2120–2130 (2021).

36. Enright, A. L., Heelan, W. J., Ward, R. D. & Peters, J. M. CRISPRi functional genomics in bacteria and its application to medical and industrial research. *Microbiology and Molecular Biology Reviews* **88**, (2024).

37. Clark, T. *et al.* CRISPR activation screens: navigating technologies and applications. *Trends Biotechnol* **42**, 1017–1034 (2024).

38. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat Methods* **10**, 973–976 (2013).

39. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326–328 (2015).

40. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat Methods* **13**, 563–567 (2016).

41. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

42. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2013).

43. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822–826 (2013).

44. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832 (2013).

45. Lopes, R. & Prasad, M. K. Beyond the promise: evaluating and mitigating off-target effects in CRISPR gene editing for safer therapeutics. *Front Bioeng Biotechnol* **11**, (2024).

46. Guo, C., Ma, X., Gao, F. & Guo, Y. Off-target effects in CRISPR/Cas9 gene editing. *Front Bioeng Biotechnol* **11**, (2023).

47. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem* **85**, 227–264 (2016).

48. Pacesa, M. *et al.* Structural basis for Cas9 off-target activity. *Cell* **185**, 4067-4081.e21 (2022).

49. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).

50. Chuai, G. *et al.* DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol* **19**, 80 (2018).

51. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**, W242–W245 (2018).

52. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148 (2016).

53. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187–197 (2015).

54. Wienert, B., Wyman, S. K., Yeh, C. D., Conklin, B. R. & Corn, J. E. CRISPR off-target detection with DISCOVER-seq. *Nat Protoc* **15**, 1775–1799 (2020).

55. Wienert, B. *et al.* Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science (1979)* **364**, 286–289 (2019).

56. Anzalone, A. V *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* (2019) doi:10.1038/s41586-019-1711-4.

57. Zhao, Z., Shang, P., Mohanraju, P. & Geijsen, N. Prime editing: advances and therapeutic applications. *Trends Biotechnol* **41**, 1000–1012 (2023).

58. Liu, Y., Kao, H.-I. & Bambara, R. A. Flap Endonuclease 1: A Central Component of DNA Metabolism. *Annu Rev Biochem* **73**, 589–615 (2004).

59. Keijzers, G., Bohr, V. A. & Rasmussen, L. J. Human exonuclease 1 (EXO1) activity characterization and its function on flap structures. *Biosci Rep* **35**, (2015).

60. Chen, P. F. P. J. P.-F. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635-5652.e29 (2021).

61. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nat Rev Genet* **24**, 161–177 (2023).

62. Doman, J. L. *et al.* Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **186**, 3983-4002.e26 (2023).

63. Yan, J. *et al.* Improving prime editing with an endogenous small RNA-binding protein. *Nature* **628**, 639–647 (2024).

64. Ferreira da Silva, J. *et al.* Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nat Commun* **13**, 760 (2022).

65. Jiang, L. & Yao, S. Enhancing prime editing via inhibition of mismatch repair pathway. *Molecular Biomedicine* **3**, 7 (2022).

66. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat Biotechnol* **40**, 402–410 (2022).

67. Grevet, J. D. *et al.* Domain-focused CRISPR screen identifies HRI as a fetal hemoglobin regulator in human erythroid cells. *Science (1979)* **361**, 285–290 (2018).

68. Haldrup, J. *et al.* Engineered lentivirus-derived nanoparticles (LVNPs) for delivery of CRISPR/Cas ribonucleoprotein complexes supporting base editing, prime editing and in vivo gene modification. *Nucleic Acids Res* **51**, 10059–10074 (2023).

69. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).

70. Suzuki, K. & Izpisua Belmonte, J. C. In vivo genome editing via the HITI method as a tool for gene therapy. *J Hum Genet* **63**, 157–164 (2018).

71. Jasin, M. & Rothstein, R. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harb Perspect Biol* **5**, a012740–a012740 (2013).

72. Rein, L. A. M., Yang, H. & Chao, N. J. Applications of Gene Editing Technologies to Cellular Therapies. *Biology of Blood and Marrow Transplantation* **24**, 1537–1545 (2018).

73. Erbs, V. *et al.* Increased On-Target Rate and Risk of Concatemerization after CRISPR-Enhanced Targeting in ES Cells. *Genes (Basel)* **14**, 401 (2023).

74. Adashi, E. Y., Gruppuso, P. A. & Cohen, I. G. CRISPR Therapy of Sickle Cell Disease: The Dawning of the Gene Editing Era. *Am J Med* **137**, 390–392 (2024).

75. FDA clears prime editors for testing in humans. *Nat Biotechnol* **42**, 691 (2024).

76. Prime Medicine Inc. Prime Medicine Unveils Strategically Focused Pipeline. https://primemedicine.com/pipeline/ (2024).

77. Esposito, F. *et al.* Safe and effective liver-directed AAV-mediated homology-independent targeted integration in mouse models of inherited diseases. *Cell Rep Med* **5**, 101619 (2024).

78. Tornabene, P. *et al.* Therapeutic homology-independent targeted integration in retina and liver. *Nat Commun* **13**, 1963 (2022).

79. Balke-Want, H. *et al.* Homology-independent targeted insertion (HITI) enables guided CAR knock-in and efficient clinical scale CAR-T cell manufacturing. *Mol Cancer* **22**, 100 (2023).

80. Milone, M. C. & O'Doherty, U. Clinical use of lentiviral vectors. *Leukemia* **32**, 1529–1541 (2018).

81. Dufait, I. *et al.* Retroviral and Lentiviral Vectors for the Induction of Immunological Tolerance. *Scientifica (Cairo)* **2012**, 1–14 (2012).

82. Yamashita, M. & Emerman, M. Retroviral infection of non-dividing cells: Old and new perspectives. *Virology* **344**, 88–93 (2006).

83. Argirò, A., Ding, J. & Adler, E. Gene therapy for heart failure and cardiomyopathies. *Revista Española de Cardiología (English Edition)* **76**, 1042–1054 (2023).

84. Seitz, R. Human Immunodeficiency Virus (HIV). *Transfusion Medicine and Hemotherapy* **43**, 203–222 (2016).

85. Dong, W. & Kantor, B. Lentiviral Vectors for Delivery of Gene-Editing Systems Based on CRISPR/Cas: Current State and Perspectives. *Viruses* **13**, 1288 (2021).

86. Lesnik, E. A., Sampath, R. & Ecker, D. J. Rev response elements (RRE) in lentiviruses: An RNAMotif algorithm-based strategy for RRE prediction. *Med Res Rev* **22**, 617–636 (2002).

87. Blesch, A. Lentiviral and MLV based retroviral vectors for ex vivo and in vivo gene transfer. *Methods* **33**, 164–172 (2004).

88. Bose, D., Gagnon, J. & Chebloune, Y. Comparative Analysis of Tat-Dependent and Tat-Deficient Natural Lentiviruses. *Vet Sci* **2**, 293–348 (2015).

89. Durand, S. & Cimarelli, A. The Inside Out of Lentiviral Vectors. *Viruses* **3**, 132–159 (2011).

90. Munis, A. M., Bentley, E. M. & Takeuchi, Y. A tool with many applications: vesicular stomatitis virus in research and medicine. *Expert Opin Biol Ther* **20**, 1187–1201 (2020).

91. Griesenbach, U. *et al.* Assessment of F/HN-pseudotyped lentivirus as a clinically relevant vector for lung gene therapy. *Am J Respir Crit Care Med* **186**, 846–856 (2012).

92. Mitomo, K. *et al.* Toward Gene Therapy for Cystic Fibrosis Using a Lentivirus Pseudotyped With Sendai Virus Envelopes. *Molecular Therapy* **18**, 1173–1182 (2010).

93. Zufferey, R. *et al.* Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery. *J Virol* **72**, 9873–9880 (1998).

94. Schnell, T., Foley, P., Wirth, M., Munch, J. & Uberla, K. Development of a Self-Inactivating, Minimal Lentivirus Vector Based on Simian Immunodeficiency Virus. *Hum Gene Ther* **11**, 439–447 (2000).

95. Koldej, R. M. & Anson, D. S. Refinement of lentiviral vector for improved RNA processing and reduced rates of self inactivation repair. *BMC Biotechnol* **9**, 86 (2009).

96. Suzuki, Y. & Suzuki, Y. Gene Regulatable Lentiviral Vector System. in *Viral Gene Therapy* (InTech, 2011). doi:10.5772/18155.

97. Giquel, B. Quick Guide to All Things Lentivirus. *21 March 2017* https://blog.addgene.org/quick-guide-to-all-things-lentivirus (2017).

98. Alton, E. W. F. W. *et al.* Preparation for a first-in-man lentivirus trial in patients with cystic fibrosis. *Thorax* **72**, 137–147 (2017).

99. Kohn, D. B. *et al.* Autologous Ex Vivo Lentiviral Gene Therapy for Adenosine Deaminase Deficiency. *New England Journal of Medicine* **384**, 2002–2013 (2021).

100. Cavazzana-Calvo, M. *et al.* Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature* **467**, 318–322 (2010).

101. Ortinski, P. I., O'Donovan, B., Dong, X. & Kantor, B. Integrase-Deficient Lentiviral Vector as an All-in-One Platform for Highly Efficient CRISPR/Cas9-Mediated Gene Editing. *Mol Ther Methods Clin Dev* **5**, 153–164 (2017).

102. Saida, H. *et al.* One-year follow-up of transgene expression by integrase-defective lentiviral vectors and their therapeutic potential in spinocerebellar ataxia model mice. *Gene Ther* **21**, 820–827 (2014).

103. Wanisch, K. & Yáñez-Muñoz, R. J. Integration-deficient Lentiviral Vectors: A Slow Coming of Age. *Molecular Therapy* **17**, 1316–1332 (2009).

104. Ling, S. *et al.* Lentiviral delivery of co-packaged Cas9 mRNA and a Vegfa-targeting guide RNA prevents wet age-related macular degeneration in mice. *Nat Biomed Eng* **5**, 144–156 (2021).

105. Patil, S. V. *et al.* Lentiviral mediated delivery of CRISPR/Cas9 reduces intraocular pressure in a mouse model of myocilin glaucoma. *Sci Rep* **14**, 6958 (2024).

106. Pavani, G. *et al.* Ex vivo editing of human hematopoietic stem cells for erythroid expression of therapeutic proteins. *Nat Commun* **11**, 3778 (2020).

107. Naso, M. F., Tomkowicz, B., Perry, W. L. & Strohl, W. R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* **31**, 317–334 (2017).

108. Deyle, D. R. & Russell, D. W. Adeno-associated virus vector integration. *Curr Opin Mol Ther* **11**, 442–7 (2009).

109. Esposito, F. *et al.* Safe and effective liver-directed AAV-mediated homology-independent targeted integration in mouse models of inherited diseases. *Cell Rep Med* **5**, 101619 (2024).

110. Pickar-Oliver, A. *et al.* Full-length dystrophin restoration via targeted exon integration by AAV-CRISPR in a humanized mouse model of Duchenne muscular dystrophy. *Molecular Therapy* **29**, 3243–3257 (2021).

111. Davis, J. R. *et al.* Efficient prime editing in mouse brain, liver and heart with dual AAVs. *Nat Biotechnol* **42**, 253–264 (2024).

112. She, K. *et al.* Dual-AAV split prime editor corrects the mutation and phenotype in mice with inherited retinal degeneration. *Signal Transduct Target Ther* **8**, 57 (2023).

113. Zheng, C. *et al.* A flexible split prime editor using truncated reverse transcriptase improves dual-AAV delivery in mouse liver. *Molecular Therapy* **30**, 1343–1351 (2022).

114. Kumar, N. *et al.* The Development of an AAV-Based CRISPR SaCas9 Genome Editing System That Can Be Delivered to Neurons in vivo and Regulated via Doxycycline and Cre-Recombinase. *Front Mol Neurosci* **11**, (2018).

115. Gao, Z. *et al.* A truncated reverse transcriptase enhances prime editing by split AAV vectors. *Molecular Therapy* **30**, 2942–2951 (2022).

116. Issa, S. S., Shaimardanova, A. A., Solovyeva, V. V. & Rizvanov, A. A. Various AAV Serotypes and Their Applications in Gene Therapy: An Overview. *Cells* **12**, 785 (2023).

117. Kulkarni, J. A. *et al.* Design of lipid nanoparticles for in vitro and in vivo delivery of plasmid DNA. *Nanomedicine* **13**, 1377–1387 (2017).

118. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine* **383**, 2603–2615 (2020).

119. Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine* **384**, 403–416 (2021).

120. Haldrup, J. *et al.* Engineered lentivirus-derived nanoparticles (LVNPs) for delivery of CRISPR/Cas ribonucleoprotein complexes supporting base editing, prime editing and in vivo gene modification. *Nucleic Acids Res* **51**, 10059–10074 (2023).

121. An, M. *et al.* Engineered virus-like particles for transient delivery of prime editor ribonucleoprotein complexes in vivo. *Nat Biotechnol* (2024) doi:10.1038/s41587-023-02078-y.

122. Cheng, Q. *et al.* Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR–Cas gene editing. *Nat Nanotechnol* **15**, 313–320 (2020).

123. Sun, Y. *et al.* In vivo editing of lung stem cells for durable gene correction in mice. *Science* **384**, 1196–1202 (2024).

124. Ratjen, F. *et al.* Cystic fibrosis. *Nat Rev Dis Primers* **1**, 15010 (2015).

125. Cutting, G. R. Cystic fibrosis genetics: From molecular understanding to clinical application. *Nat Rev Genet* **16**, 45–56 (2015).

126. Abubakar Bobbo, K., Ahmad, U., Chau, D.-M., Nordin, N. & Abdullah, S. A comprehensive review of cystic fibrosis in Africa and Asia. *Saudi J Biol Sci* **30**, 103685 (2023).

127. Vaidyanathan, S. *et al.* CFTR genotype analysis of Asians in international registries highlights disparities in the diagnosis and treatment of Asian patients with cystic fibrosis. *Genetics in Medicine* **24**, 2180–2186 (2022).

128. Schrijver, I. *et al.* The Spectrum of CFTR Variants in Nonwhite Cystic Fibrosis Patients: Implications for Molecular Diagnostic Testing. *Journal of Molecular Diagnostics* **18**, 39–50 (2016).

129. Hanssens, L. S., Duchateau, J. & Casimir, G. J. Cftr protein: Not just a chloride channel? *Cells* **10**, (2021).

130. Linsdell, P. Mechanism of chloride permeation in the cystic fibrosis transmembrane conductance regulator chloride channel. *Exp Physiol* **91**, 123–129 (2006).

131. Ramananda, Y., Naren, A. P. & Arora, K. Functional Consequences of CFTR Interactions in Cystic Fibrosis. *Int J Mol Sci* **25**, 3384 (2024).

132. Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).

133. Lei, L. *et al.* CFTR-rich ionocytes mediate chloride absorption across airway epithelia. *Journal of Clinical Investigation* **133**, (2023).

134. Boucher, R. C. Evidence for airway surface dehydration as the initiating event in CF airway disease. *J Intern Med* **261**, 5–16 (2007).

135. De Boeck, K. Cystic fibrosis in the year 2020: A disease with a new face. *Acta Paediatrica, International Journal of Paediatrics* **109**, 893–899 (2020).

136. De Boeck, K. Cystic fibrosis in the year 2020: A disease with a new face. *Acta Paediatrica, International Journal of Paediatrics* **109**, 893–899 (2020).

137. Taylor-Cousar, J. L. *et al.* Clinical development of triple-combination CFTR modulators for cystic fibrosis patients with one or two F508del alleles. *ERJ Open Res* **5**, (2019).

138. Lukacs, G. L. & Verkman, A. S. CFTR: folding, misfolding and correcting the ΔF508 conformational defect. *Trends Mol Med* **18**, 81–91 (2012).

139. Middleton, P. G. *et al.* Elexacaftor–Tezacaftor–Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele. *New England Journal of Medicine* **381**, 1809–1819 (2019).

140. Lopez, A., Daly, C., Vega-Hernandez, G., MacGregor, G. & Rubin, J. L. Elexacaftor/tezacaftor/ivacaftor projected survival and long-term health outcomes in people with cystic fibrosis homozygous for F508del. *Journal of Cystic Fibrosis* **22**, 607–614 (2023).

141. Volkova, N. *et al.* Disease progression in patients with cystic fibrosis treated with ivacaftor: Data from national US and UK registries. *Journal of Cystic Fibrosis* **19**, 68–79 (2020).

142. Kramer-Golinkoff, E., Camacho, A., Kramer, L. & Taylor-Cousar, J. L. A survey: Understanding the health and perspectives of people with CF not benefiting from CFTR modulators. *Pediatr Pulmonol* **57**, 1253–1261 (2022).

143. Kerem, E. Pharmacologic therapy for stop mutations: how much CFTR activity is enough? *Curr Opin Pulm Med* **10**, 547–552 (2004).

144. Cutting, G. R. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet* **16**, 45–56 (2015).

145. Davies, J. C. *et al.* Lentiviral Gene Therapy for Cystic Fibrosis: A Promising Approach and First-In-Human Trial. *Am J Respir Crit Care Med* (2024) doi:10.1164/rccm.202402-0389CI.

146. Miah, K. M., Hyde, S. C. & Gill, D. R. Emerging gene therapies for cystic fibrosis. *Expert Rev Respir Med* **13**, 709–725 (2019).

147. Moss, R. B. *et al.* Repeated Adeno-Associated Virus Serotype 2 Aerosol-Mediated Cystic Fibrosis Transmembrane Regulator Gene Transfer to the Lungs of Patients With Cystic Fibrosis. *Chest* **125**, 509–521 (2004).

148. Moss, R. B. *et al.* Repeated Aerosolized AAV-CFTR for Treatment of Cystic Fibrosis: A Randomized Placebo-Controlled Phase 2B Trial. *Hum Gene Ther* **18**, 726–732 (2007).

149. Smith, K. (Inizio E. C. 4DMT Presents Positive Interim Data from Phase 1/2 AEROW Clinical Trial of Aerosolized 4D-710 for Modulator-Ineligible/-Intolerant Cystic Fibrosis at 47th European Cystic Fibrosis Conference. https://4dmt.gcs-web.com/news-releases/news-release-details/4dmt-presents-positive-interim-data-phase-12-aerow-clinical-0#top (2024).

150. Steines, B. *et al.* CFTR gene transfer with AAV improves early cystic fibrosis pig phenotypes. *JCI Insight* **1**, (2016).

151. Rowe, S. M. *et al.* Inhaled mRNA therapy for treatment of cystic fibrosis: Interim results of a randomized, double-blind, placebo-controlled phase 1/2 clinical study. *Journal of Cystic Fibrosis* **22**, 656–664 (2023).

152. A Phase 1/2, Randomized, Double-Blinded, Placebo-Controlled, Combined Single and Multiple Ascending Dose Study Evaluating the Safety, Tolerability, and Biological Activity of MRT5005 Administered by Nebulization to Adult Subjects With Cystic Fibrosis. Preprint at https://clinicaltrials.gov/study/NCT03375047 (2017).

153. A Phase 1/2, Multicenter Study Evaluating the Safety, Tolerability, and Biodistribution of RCT2100 with Single-Ascending Doses in Healthy Participants and Multiple-Ascending Doses and Proof-of-Concept in Participants with Cystic Fibrosis. Preprint at https://clinicaltrials.gov/study/NCT06237335 (2024).

154. Kim, H. K. *et al.* Predicting the efficiency of prime editing guide RNAs in human cells. *Nat Biotechnol* **39**, 198–206 (2021).

155. Geurts, M. H. *et al.* Evaluating CRISPR-based prime editing for cancer modeling and CFTR repair in organoids. *Life Sci Alliance* **4**, e202000940 (2021).

156. Liu, G., Yin, K., Zhang, Q., Gao, C. & Qiu, J.-L. Modulating chromatin accessibility by transactivation and targeting proximal dsgRNAs enhances Cas9 editing efficiency in vivo. *Genome Biol* **20**, 145 (2019).

157. Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol* **33**, 1159–1161 (2015).

158. Sousa, A. A. *et al.* Systematic optimization of prime editing for the efficient functional correction of CFTR F508del in human airway epithelial cells. *Nat Biomed Eng* (2024) doi:10.1038/s41551-024-01233-3.

159. Bednarski, C., Tomczak, K., Hövel, B. Vom, Weber, W. M. & Cathomen, T. Targeted integration of a super-exon into the CFTR locus leads to functional correction of a cystic fibrosis cell line model. *PLoS One* **11**, 1–15 (2016).

160. Liu, P. *et al.* Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat Commun* **12**, 2121 (2021).

161. Guschin, D. Y. *et al.* A Rapid and General Assay for Monitoring Endogenous Gene Modification. in 247–256 (2010). doi:10.1007/978-1-60761-753-2_15.

162. Alton, E. W. F. W. F. W. *et al.* Preparation for a first-in-man lentivirus trial in patients with cystic fibrosis. *Thorax* **72**, 137–147 (2017).

163. Gutierrez-Guerrero, A., Cosset, F.-L. & Verhoeyen, E. Lentiviral Vector Pseudotypes: Precious Tools to Improve Gene Modification of Hematopoietic Cells for Research and Gene Therapy. *Viruses* **12**, 1016 (2020).

164. Kim, H. K. *et al.* Predicting the efficiency of prime editing guide RNAs in human cells. *Nat Biotechnol* **39**, 198–206 (2021).

165. Ahuka-Mundeke, S. *et al.* Full-length genome sequence of a simian immunodeficiency virus (SIV) infecting a captive agile mangabey (Cercocebus agilis) is closely related to SIVrcm infecting wild red-capped mangabeys (Cercocebus torquatus) in Cameroon. *Journal of General Virology* **91**, 2959–2964 (2010).

166. Kumar, M., Keller, B., Makalou, N. & Sutton, R. E. Systematic Determination of the Packaging Limit of Lentiviral Vectors. *Hum Gene Ther* **12**, 1893–1905 (2001).

167. Sweeney, N. P. & Vink, C. A. The impact of lentiviral vector genome size and producer cell genomic to gag-pol mRNA ratios on packaging efficiency and titre. *Mol Ther Methods Clin Dev* **21**, 574–584 (2021).

168. Counsell, J. R. *et al.* Lentiviral vectors can be used for full-length dystrophin gene therapy. *Sci Rep* **7**, 79 (2017).

169. Maunder, H. E. *et al.* Enhancing titres of therapeutic viral vectors using the transgene repression in vector production (TRiP) system. *Nat Commun* **8**, 14834 (2017).

170. Zheng, C. *et al.* A flexible split prime editor using truncated reverse transcriptase improves dual-AAV delivery in mouse liver. *Molecular Therapy* **30**, 1343–1351 (2022).

171. Bednarski, C., Tomczak, K., Hövel, B. Vom, Weber, W. M. & Cathomen, T. Targeted integration of a super-exon into the CFTR locus leads to functional correction of a cystic fibrosis cell line model. *PLoS One* **11**, 1–15 (2016).

172. Smirnikhina, S. A. *et al.* P.F508del editing in cells from cystic fibrosis patients. *PLoS One* **15**, 1–11 (2020).

173. Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**, 653–658 (2013).

174. Anderson, M. V., Haldrup, J., Thomsen, E. A., Wolff, J. H. & Mikkelsen, J. G. pegIT - a web-based design tool for prime editing. *Nucleic Acids Res* **49**, W505–W509 (2021).

175. Kawasaki, S. *et al.* Programmable mammalian translational modulators by CRISPR-associated proteins. *Nat Commun* **14**, 2243 (2023).

176. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans -splicing reaction. *FEBS Lett* **583**, 909–914 (2009).

177. De Boeck, K., Zolin, A., Cuppens, H., Olesen, H. V. & Viviani, L. The relative frequency of CFTR mutation classes in European patients with cystic fibrosis. *Journal of Cystic Fibrosis* **13**, 403–409 (2014).

178. Cortijo-Gutiérrez, M. *et al.* Improved Functionality of Integration-Deficient Lentiviral Vectors (IDLVs) by the Inclusion of IS2 Protein Docks. *Pharmaceutics* **13**, 1217 (2021).

179. Cornu, T. I. & Cathomen, T. Targeted Genome Modifications Using Integrase-deficient Lentiviral Vectors. *Molecular Therapy* **15**, 2107–2113 (2007).

180. Wilbie, D., Walther, J. & Mastrobattista, E. Delivery Aspects of CRISPR/Cas for in Vivo Genome Editing. *Acc Chem Res* **52**, 1555–1564 (2019).

181. Horodecka, K. & Düchler, M. CRISPR/Cas9: Principle, Applications, and Delivery through Extracellular Vesicles. *Int J Mol Sci* **22**, 6072 (2021).

182. Grünewald, J. *et al.* Engineered CRISPR prime editors with compact, untethered reverse transcriptases. *Nat Biotechnol* **41**, 337–343 (2023).

183. Aldossary, A. M. Correction of the ΔF508 mutation in the CFTR Gene by CRISPR/Cas9 system. (University College London, 2018).

184. Mention, K. *et al.* Use of adenine base editing and homology-independent targeted integration strategies to correct the cystic fibrosis causing variant, W1282X. *Hum Mol Genet* **32**, 3237–3248 (2023).

185. Shapiro, J. *et al.* Increasing CRISPR Efficiency and Measuring Its Specificity in HSPCs Using a Clinically Relevant System. *Mol Ther Methods Clin Dev* **17**, 1097–1107 (2020).

186. Seki, A. & Rutz, S. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. *Journal of Experimental Medicine* **215**, 985–997 (2018).

187. Tyumentseva, M. A., Tyumentsev, A. I. & Akimkin, V. G. Protocol for assessment of the efficiency of CRISPR/Cas RNP delivery to different types of target cells. *PLoS One* **16**, e0259812 (2021).

188. Min, Y.-L. *et al.* CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. *Sci Adv* **5**, (2019).

189. Domingue, J. C. *et al.* HEK-293 cells expressing the cystic fibrosis transmembrane conductance regulator (CFTR): a model for studying regulation of Cl − transport. *Physiol Rep* **2**, e12158 (2014).

190. Marini, B. *et al.* Nuclear architecture dictates HIV-1 integration site selection. *Nature* **521**, 227–231 (2015).

191. Schröder, A. R. W. *et al.* HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell* **110**, 521–529 (2002).

192. Choi, J.-H. *et al.* Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. *Mol Brain* **7**, 17 (2014).

193. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep* **7**, 2193 (2017).

194. Škalamera, D. *et al.* Generation of a Genome Scale Lentiviral Vector Library for EF1α Promoter-Driven Expression of Human ORFs and Identification of Human Genes Affecting Viral Titer. *PLoS One* **7**, e51733 (2012).

195. Dong, W. & Kantor, B. Lentiviral Vectors for Delivery of Gene-Editing Systems Based on CRISPR/Cas: Current State and Perspectives. *Viruses* **13**, 1288 (2021).

196. Tristán-Manzano, M. *et al.* Physiological lentiviral vectors for the generation of improved CAR-T cells. *Mol Ther Oncolytics* **25**, 335–349 (2022).

197. Hart, K. L. *et al.* A novel high-titer, bifunctional lentiviral vector for autologous hematopoietic stem cell gene therapy of sickle cell disease. *Mol Ther Methods Clin Dev* **32**, 101254 (2024).

198. Nicolas, C. T. *et al.* In vivo lentiviral vector gene therapy to cure hereditary tyrosinemia type 1 and prevent development of precancerous and cancerous lesions. *Nat Commun* **13**, 5012 (2022).

199. Aiuti, A. *et al.* Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. *Science (1979)* **341**, (2013).

200. Fumagalli, F. *et al.* Lentiviral haematopoietic stem-cell gene therapy for early-onset metachromatic leukodystrophy: long-term results from a non-randomised, open-label, phase 1/2 trial and expanded access. *The Lancet* **399**, 372–383 (2022).

201. Biffi, A. *et al.* Lentiviral Hematopoietic Stem Cell Gene Therapy Benefits Metachromatic Leukodystrophy. *Science (1979)* **341**, (2013).

202. Magnani, A. *et al.* Long-term safety and efficacy of lentiviral hematopoietic stem/progenitor cell gene therapy for Wiskott–Aldrich syndrome. *Nat Med* **28**, 71–80 (2022).

203. Labbé, R. P., Vessillier, S. & Rafiq, Q. A. Lentiviral Vectors for T Cell Engineering: Clinical Applications, Bioprocessing and Future Perspectives. *Viruses* **13**, 1528 (2021).

204. Ren, J. *et al.* Multiplex Genome Editing to Generate Universal CAR T Cells Resistant to PD1 Inhibition. *Clinical Cancer Research* **23**, 2255–2266 (2017).

205. Cantore, A. *et al.* Liver-directed lentiviral gene therapy in a dog model of hemophilia B. *Sci Transl Med* **7**, (2015).

206. Campochiaro, P. A. *et al.* Lentiviral Vector Gene Transfer of Endostatin/Angiostatin for Macular Degeneration (GEM) Study. *Hum Gene Ther* **28**, 99–111 (2017).

207. Hamilton, J. R. *et al.* In vivo human T cell engineering with enveloped delivery vehicles. *Nat Biotechnol* (2024) doi:10.1038/s41587-023-02085-z.

208. Chuang, Y.-F. *et al.* Approach for in vivo delivery of CRISPR/Cas system: a recent update and future prospect. *Cellular and Molecular Life Sciences* **78**, 2683–2708 (2021).

209. Davis, J. R. *et al.* Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base editors. *Nat Biomed Eng* **6**, 1272–1283 (2022).

210. Ryu, S.-M. *et al.* Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat Biotechnol* **36**, 536–539 (2018).

211. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* **19**, 770–788 (2018).

212. Esposito, F. *et al.* Safe and effective liver-directed AAV-mediated homology-independent targeted integration in mouse models of inherited diseases. *Cell Rep Med* **5**, 101619 (2024).

213. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science (1979)* **343**, 84–87 (2014).

214. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783–784 (2014).

215. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat Biotechnol* **32**, 941–946 (2014).

216. Yu, G. *et al.* Prediction of efficiencies for diverse prime editing systems in multiple cell types. *Cell* **186**, 2256-2272.e23 (2023).

217. Koeppel, J. *et al.* Prediction of prime editing insertion efficiencies using sequence features and DNA repair determinants. *Nat Biotechnol* **41**, 1446–1456 (2023).

218. Gould, S. I. *et al.* High-throughput evaluation of genetic variants with prime editing sensor libraries. *Nat Biotechnol* (2024) doi:10.1038/s41587-024-02172-9.

219. Ren, X. *et al.* High-throughput PRIME-editing screens identify functional DNA variants in the human genome. *Mol Cell* **83**, 4633-4645.e9 (2023).

220. FDA approves hereditary blindness gene therapy. *Nat Biotechnol* **36**, 6–6 (2018).

221. Asghar, A. A., Khabir, Y. & Hashmi, M. R. Zynteglo: Betibeglogene autotemcel – An innovative therapy for β- thalassemia patients. *Annals of Medicine & Surgery* **82**, (2022).

222. Keam, S. J. Elivaldogene Autotemcel: First Approval. *Mol Diagn Ther* **25**, 803–809 (2021).

223. Awasthi, R., Maier, H. J., Zhang, J. & Lim, S. Kymriah® (tisagenlecleucel) – An overview of the clinical development journey of the first approved CAR-T therapy. *Hum Vaccin Immunother* **19**, (2023).

224. Hoy, S. M. Onasemnogene Abeparvovec: First Global Approval. *Drugs* **79**, 1255–1262 (2019).

225. Schlimgen, R. *et al.* Risks Associated With Lentiviral Vector Exposures and Prevention Strategies. *J Occup Environ Med* **58**, 1159–1166 (2016).

226. Taha, E. A., Lee, J. & Hotta, A. Delivery of CRISPR-Cas tools for in vivo genome editing therapy: Trends and challenges. *Journal of Controlled Release* **342**, 345–361 (2022).

227. Leavitt, A. D., Robles, G., Alesandro, N. & Varmus, H. E. Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. *J Virol* **70**, 721–728 (1996).

228. Suwanmanee, T. *et al.* Integration-deficient Lentiviral Vectors Expressing Codon-optimized R338L Human FIX Restore Normal Hemostasis in Hemophilia B Mice. *Molecular Therapy* **22**, 567–574 (2014).

229. Staunstrup, N. H. *et al.* Hybrid Lentivirus-transposon Vectors With a Random Integration Profile in Human Cells. *Molecular Therapy* **17**, 1205–1214 (2009).

230. Uchida, N. *et al.* Cas9 protein delivery non-integrating lentiviral vectors for gene correction in sickle cell disease. *Mol Ther Methods Clin Dev* **21**, 121–132 (2021).

231. Chavez, M., Rane, D. A., Chen, X. & Qi, L. S. Stable expression of large transgenes via the knock-in of an integrase-deficient lentivirus. *Nat Biomed Eng* **7**, 661–671 (2023).

232. Lu, X. *et al.* Applications and Research Advances in the Delivery of CRISPR/Cas9 Systems for the Treatment of Inherited Diseases. *Int J Mol Sci* **24**, 13202 (2023).

233. Haldrup, J. *et al.* Engineered lentivirus-derived nanoparticles (LVNPs) for delivery of CRISPR/Cas ribonucleoprotein complexes supporting base editing, prime editing and in vivo gene modification. *Nucleic Acids Res* **51**, 10059–10074 (2023).

234. Bloomer, H., Khirallah, J., Li, Y. & Xu, Q. CRISPR/Cas9 ribonucleoprotein-mediated genome and epigenome editing in mammalian cells. *Adv Drug Deliv Rev* **181**, 114087 (2022).

235. Fry, L. E. *et al.* Promoter Orientation within an AAV-CRISPR Vector Affects Cas9 Expression and Gene Editing Efficiency. *CRISPR J* **3**, 276–283 (2020).

236. Kabadi, A. M., Ousterout, D. G., Hilton, I. B. & Gersbach, C. A. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res* **42**, e147–e147 (2014).

237. Lan, T. *et al.* Mini-PE, a prime editor with compact Cas9 and truncated reverse transcriptase. *Mol Ther Nucleic Acids* **33**, 890–897 (2023).

238. Liu, B. *et al.* Targeted genome editing with a DNA-dependent DNA polymerase and exogenous DNA-containing templates. *Nat Biotechnol* **42**, 1039–1045 (2024).

239. Pandey, S. *et al.* Efficient site-specific integration of large genes in mammalian cells via continuously evolved recombinases and prime editing. *Nat Biomed Eng* (2024) doi:10.1038/s41551-024-01227-1.

240. Yarnall, M. T. N. *et al.* Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat Biotechnol* **41**, 500–512 (2023).

241. Iwai, H., Züger, S., Jin, J. & Tam, P.-H. Highly efficient protein trans -splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett* **580**, 1853–1858 (2006).

242. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans -splicing reaction. *FEBS Lett* **583**, 909–914 (2009).

243. Carvajal-Vallejos, P., Pallissé, R., Mootz, H. D. & Schmidt, S. R. Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources. *Journal of Biological Chemistry* **287**, 28686–28696 (2012).

244. Kaiser, R. A. *et al.* Hepatotoxicity and Toxicology of In Vivo Lentiviral Vector Administration in Healthy and Liver-Injury Mouse Models. *Hum Gene Ther Clin Dev* **30**, 57–66 (2019).

245. Carbonaro-Sarracino, D. A. *et al.* Dosing and Re-Administration of Lentiviral Vector for In Vivo Gene Therapy in Rhesus Monkeys and ADA-Deficient Mice. *Mol Ther Methods Clin Dev* **16**, 78–93 (2020).

246. Fang, H., Bygrave, A. M., Roth, R. H., Johnson, R. C. & Huganir, R. L. An optimized CRISPR/Cas9 approach for precise genome editing in neurons. *Elife* **10**, (2021).

247. Hayashi, H., Kubo, Y., Izumida, M. & Matsuyama, T. Efficient viral delivery of Cas9 into human safe harbor. *Sci Rep* **10**, 21474 (2020).

248. Asmamaw Mengstie, M. Viral Vectors for the in Vivo Delivery of CRISPR Components: Advances and Challenges. *Front Bioeng Biotechnol* **10**, (2022).

249. Hanlon, K. S. *et al.* High levels of AAV vector integration into CRISPR-induced DNA breaks. *Nat Commun* **10**, 4439 (2019).

250. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187–198 (2015).

251. Liu, P. *et al.* Increasing intracellular dNTP levels improves prime editing efficiency. *Nat Biotechnol* (2024) doi:10.1038/s41587-024-02405-x.

252. Li, X. *et al.* Chromatin context-dependent regulation and epigenetic manipulation of prime editing. *Cell* **187**, 2411-2427.e25 (2024).

253. Chow, R. D., Chen, J. S., Shen, J. & Chen, S. A web tool for the design of prime-editing guide RNAs. *Nat Biomed Eng* **5**, 190–194 (2020).

254. Hsu, J. Y. *et al.* PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat Commun* **12**, 1034 (2021).

255. Mathis, N. *et al.* Predicting prime editing efficiency and product purity by deep learning. *Nat Biotechnol* **41**, 1151–1159 (2023).

256. Mathis, N. *et al.* Machine learning prediction of prime editing efficiency across diverse chromatin contexts. *Nat Biotechnol* (2024) doi:10.1038/s41587-024-02268-2.

257. Bulcaen, M. *et al.* Prime editing functionally corrects cystic fibrosis-causing CFTR mutations in human organoids and airway epithelial cells. *Cell Rep Med* **5**, 101544 (2024).

258. Okuda, K. *et al.* Secretory Cells Dominate Airway CFTR Expression and Function in Human Airway Superficial Epithelia. *Am J Respir Crit Care Med* **203**, 1275–1289 (2021).

259. Park, S.-J. *et al.* Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. *Genome Biol* **22**, 170 (2021).

Appendix

HITI Donor 1 sequence

CCTAAGCAGAAGTGTATATTCTTCTTAGGATGATAATTGGAGGCAAGTGAATCCTGAGCGTGATTTGATAATG
ACCTAATAATGATGGGTTTTATTTCCAGACTTCACTTCTAATGGTGATTATGGGAGAACTGGAGCCTTCAGAG
GGTAAAATTAAGCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAG
AAAATATCATCTTTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGA
GGACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTGAGTGGAGGTCAA
CGAGCAAGAATTTCTTTAGCAAGAGCAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGAT
ACCTAGATGTTTTAACAGAAAAAGAAATATTTGAAAGCTGTGTCTGTAAACTGATGGCTAACAAAACTAGGAT
TTTGGTCACTTCTAAAATGGAACATTTAAAGAAAGCTGACAAAATATTAATTTTGCATGAAGGTAGCAGCTAT
TTTTATGGGACATTTTCAGAACTCCAAAATCTACAGCCAGACTTTAGCTCAAAACTCATGGGATGTGATTCTT
TCGACCAATTTAGTGCAGAAAGAAGAAATTCAATCCTAACTGAGACCTTACACCGTTTCTCATTAGAAGGAGA
TGCTCCTGTCTCCTGGACAGAAACAAAAAAACAATCTTTTAAACAGACTGGAGAGTTTGGGGAAAAAAGGAAG
AATTCTATTCTCAATCCAATCAACTCTATACGAAAATTTTCCATTGTGCAAAAGACTCCCTTACAAATGAATG
GCATCGAAGAGGATTCTGATGAGCCTTTAGAGAGAAGGCTGTCCTTAGTACCAGATTCTGAGCAGGGAGAGGC
GATACTGCCTCGCATCAGCGTGATCAGCACTGGCCCCACGCTTCAGGCACGAAGGAGGCAGTCTGTCCTGAAC
CTGATGACACACTCAGTTAACCAAGGTCAGAACATTCACCGAAAGACAACAGCATCCACACGAAAAGTGTCAC
TGGCCCCTCAGGCAAACTTGACTGAACTGGATATATATTCAAGAAGGTTATCTCAAGAAACTGGCTTGGAAAT
AAGTGAAGAAATTAACGAAGAAGACTTAAAGGAGTGCTTTTTTGATGATATGGAGAGCATACCAGCAGTGACT
ACATGGAACACATACCTTCGATATATTACTGTCCACAAGAGCTTAATTTTTGTGCTAATTTGGTGCTTAGTAA
TTTTTCTGGCAGAGGTGGCTGCTTCTTTGGTTGTGCTGTGGCTCCTTGGAAACACTCCTCTTCAAGACAAAGG
GAATAGTACTCATAGTAGAAATAACAGCTATGCAGTGATTATCACCAGCACCAGTTCGTATTATGTGTTTTAC
ATTTACGTGGGAGTAGCCGACACTTTGCTTGCTATGGGATTCTTCAGAGGTCTACCACTGGTGCATACTCTAA
TCACAGTGTCGAAAATTTTACACCACAAAATGTTACATTCTGTTCTTCAAGCACCTATGTCAACCCTCAACAC
GTTGAAAGCAGGTGGGATTCTTAATAGATTCTCCAAAGATATAGCAATTTTGGATGACCTTCTGCCTCTTACC
ATATTTGACTTCATCCAGTTGTTATTAATTGTGATTGGAGCTATAGCAGTTGTCGCAGTTTTACAACCCTACA
TCTTTGTTGCAACAGTGCCAGTGATAGTGGCTTTTATTATGTTGAGAGCATATTTCCTCCAAACCTCACAGCA
ACTCAAACAACTGGAATCTGAAGGCAGGAGTCCAATTTTCACTCATCTTGTTACAAGCTTAAAAGGACTATGG
ACACTTCGTGCCTTCGGACGGCAGCCTTACTTTGAAACTCTGTTCCACAAAGCTCTGAATTTACATACTGCCA
ACTGGTTCTTGTACCTGTCAACACTGCGCTGGTTCCAAATGAGAATAGAAATGATTTTTGTCATCTTCTTCAT
TGCTGTTACCTTCATTTCCATTTTAACAACAGGAGAAGGAGAAGGAAGAGTTGGTATTATCCTGACTTTAGCC
ATGAATATCATGAGTACATTGCAGTGGGCTGTAAACTCCAGCATAGATGTGGATAGCTTGATGCGATCTGTGA
GCCGAGTCTTTAAGTTCATTGACATGCCAACAGAAGGTAAACCTACCAAGTCAACCAAACCATACAAGAATGG
CCAACTCTCGAAAGTTATGATTATTGAGAATTCACACGTGAAGAAAGATGACATCTGGCCCTCAGGGGGCCAA
ATGACTGTCAAAGATCTCACAGCAAAATACACAGAAGGTGGAAATGCCATATTAGAGAACATTTCCTTCTCAA
TAAGTCCTGGCCAGAGGGTGGGCCTCTTGGGAAGAACTGGATCAGGGAAGAGTACTTTGTTATCAGCTTTTTT
GAGACTACTGAACACTGAAGGAGAAATCCAGATCGATGGTGTGTCTTGGGATTCAATAACTTTGCAACAGTGG
AGGAAAGCCTTTGGAGTGATACCACAGAAAGTATTTATTTTTTCTGGAACATTTAGAAAAAACTTGGATCCCT
ATGAACAGTGGAGTGATCAAGAAATATGGAAAGTTGCAGATGAGGTTGGGCTCAGATCTGTGATAGAACAGTT
TCCTGGGAAGCTTGACTTTGTCCTTGTGGATGGGGGCTGTGTCCTAAGCCATGGCCACAAGCAGTTGATGTGC
TTGGCTAGATCTGTTCTCAGTAAGGCGAAGATCTTGCTGCTTGATGAACCCAGTGCTCATTTGGATCCAGTAA
CATACCAAATAATTAGAAGAACTCTAAAACAAGCATTTGCTGATTGCACAGTAATTCTCTGTGAACACAGGAT
AGAAGCAATGCTGGAATGCCAACAATTTTTGGTCATAGAAGAGAACAAAGTGCGGCAGTACGATTCCATCCAG
AAACTGCTGAACGAGAGGAGCCTCTTCCGGCAAGCCATCAGCCCCTCCGACAGGGTGAAGCTCTTTCCCCACC
GGAACTCAAGCAAGTGCAAGTCTAAGCCCCAGATTGCTGCTCTGAAAGAGGAGACAGAAGAAGAGGTGCAAGA
TACAAGGCTTTAGCCTAAGCAGAAGTGTATATTCTT

Note: HITI gRNA 1,PAM,partial CFTR intron 10, intron splicing site, CFTR superexon 11-27.

HITI Donor 3 sequence

CCTCCAATTATCATCCTAAGCAGTGGAGGCAAGTGAATCCTGAGCGTGATTTGATAATGACCTAATAATGATG
GGTTTTATTTCCAGACTTCACTTCTAATGGTGATTATGGGAGAACTGGAGCCTTCAGAGGGTAAAATTAAGCA
CAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTT
GGTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGGACATCTCCAAGT
TTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTGAGTGGAGGTCAACGAGCAAGAATTTC
TTTAGCAAGAGCAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGATACCTAGATGTTTTA
ACAGAAAAAGAAATATTTGAAAGCTGTGTCTGTAAACTGATGGCTAACAAAACTAGGATTTTGGTCACTTCTA
AAATGGAACATTTAAAGAAAGCTGACAAAATATTAATTTTGCATGAAGGTAGCAGCTATTTTTATGGGACATT
TTCAGAACTCCAAAATCTACAGCCAGACTTTAGCTCAAAACTCATGGGATGTGATTCTTTCGACCAATTTAGT
GCAGAAAGAAGAAATTCAATCCTAACTGAGACCTTACACCGTTTCTCATTAGAAGGAGATGCTCCTGTCTCCT
GGACAGAAACAAAAAAACAATCTTTTAAACAGACTGGAGAGTTTGGGGAAAAAAGGAAGAATTCTATTCTCAA
TCCAATCAACTCTATACGAAAATTTTCCATTGTGCAAAAGACTCCCTTACAAATGAATGGCATCGAAGAGGAT
TCTGATGAGCCTTTAGAGAGAAGGCTGTCCTTAGTACCAGATTCTGAGCAGGGAGAGGCGATACTGCCTCGCA
TCAGCGTGATCAGCACTGGCCCCACGCTTCAGGCACGAAGGAGGCAGTCTGTCCTGAACCTGATGACACACTC
AGTTAACCAAGGTCAGAACATTCACCGAAAGACAACAGCATCCACACGAAAAGTGTCACTGGCCCCTCAGGCA
AACTTGACTGAACTGGATATATATTCAAGAAGGTTATCTCAAGAAACTGGCTTGGAAATAAGTGAAGAAATTA
ACGAAGAAGACTTAAAGGAGTGCTTTTTTGATGATATGGAGAGCATACCAGCAGTGACTACATGGAACACATA
CCTTCGATATATTACTGTCCACAAGAGCTTAATTTTTGTGCTAATTTGGTGCTTAGTAATTTTTCTGGCAGAG
GTGGCTGCTTCTTTGGTTGTGCTGTGGCTCCTTGGAAACACTCCTCTTCAAGACAAAGGGAATAGTACTCATA
GTAGAAATAACAGCTATGCAGTGATTATCACCAGCACCAGTTCGTATTATGTGTTTTACATTTACGTGGGAGT
AGCCGACACTTTGCTTGCTATGGGATTCTTCAGAGGTCTACCACTGGTGCATACTCTAATCACAGTGTCGAAA
ATTTTACACCACAAAATGTTACATTCTGTTCTTCAAGCACCTATGTCAACCCTCAACACGTTGAAAGCAGGTG
GGATTCTTAATAGATTCTCCAAAGATATAGCAATTTTGGATGACCTTCTGCCTCTTACCATATTTGACTTCAT
CCAGTTGTTATTAATTGTGATTGGAGCTATAGCAGTTGTCGCAGTTTTACAACCCTACATCTTTGTTGCAACA
GTGCCAGTGATAGTGGCTTTTATTATGTTGAGAGCATATTTCCTCCAAACCTCACAGCAACTCAAACAACTGG
AATCTGAAGGCAGGAGTCCAATTTTCACTCATCTTGTTACAAGCTTAAAAGGACTATGGACACTTCGTGCCTT
CGGACGGCAGCCTTACTTTGAAACTCTGTTCCACAAAGCTCTGAATTTACATACTGCCAACTGGTTCTTGTAC
CTGTCAACACTGCGCTGGTTCCAAATGAGAATAGAAATGATTTTTGTCATCTTCTTCATTGCTGTTACCTTCA
TTTCCATTTTAACAACAGGAGAAGGAGAAGGAAGAGTTGGTATTATCCTGACTTTAGCCATGAATATCATGAG
TACATTGCAGTGGGCTGTAAACTCCAGCATAGATGTGGATAGCTTGATGCGATCTGTGAGCCGAGTCTTTAAG
TTCATTGACATGCCAACAGAAGGTAAACCTACCAAGTCAACCAAACCATACAAGAATGGCCAACTCTCGAAAG
TTATGATTATTGAGAATTCACACGTGAAGAAAGATGACATCTGGCCCTCAGGGGGCCAAATGACTGTCAAAGA
TCTCACAGCAAAATACACAGAAGGTGGAAATGCCATATTAGAGAACATTTCCTTCTCAATAAGTCCTGGCCAG
AGGGTGGGCCTCTTGGGAAGAACTGGATCAGGGAAGAGTACTTTGTTATCAGCTTTTTTGAGACTACTGAACA
CTGAAGGAGAAATCCAGATCGATGGTGTGTCTTGGGATTCAATAACTTTGCAACAGTGGAGGAAAGCCTTTGG
AGTGATACCACAGAAAGTATTTATTTTTTCTGGAACATTTAGAAAAAACTTGGATCCCTATGAACAGTGGAGT
GATCAAGAAATATGGAAAGTTGCAGATGAGGTTGGGCTCAGATCTGTGATAGAACAGTTTCCTGGGAAGCTTG
ACTTTGTCCTTGTGGATGGGGGCTGTGTCCTAAGCCATGGCCACAAGCAGTTGATGTGCTTGGCTAGATCTGT
TCTCAGTAAGGCGAAGATCTTGCTGCTTGATGAACCCAGTGCTCATTTGGATCCAGTAACATACCAAATAATT
AGAAGAACTCTAAAACAAGCATTTGCTGATTGCACAGTAATTCTCTGTGAACACAGGATAGAAGCAATGCTGG
AATGCCAACAATTTTTGGTCATAGAAGAGAACAAAGTGCGGCAGTACGATTCCATCCAGAAACTGCTGAACGA
GAGGAGCCTCTTCCGGCAAGCCATCAGCCCCTCCGACAGGGTGAAGCTCTTTCCCCACCGGAACTCAAGCAAG
TGCAAGTCTAAGCCCCAGATTGCTGCTCTGAAAGAGGAGACAGAAGAAGAGGTGCAAGATACAAGGCTTTAGC
CTCCAATTATCATCCTAAGCAG

Note: HITI gRNA 3, PAM, partial CFTR intron 10, intron splicing site, CFTR superexon 11-27.

Designing pegRNA (R) – code by Dr Jakob Haldrup

```
```{r packages, message=FALSE}
library(readxl)
library(xlsx)
library(ggplot2)
library(vcfR)
library(svMisc)
library(dplyr)
library(stringr)
library(BSgenome)
library(Biostrings)
library(BSgenome.Hsapiens.NCBI.GRCh38)
library(DNABarcodes)
library(gtools)
library(DECIPHER)

```

```{r Barcode generator, eval = FALSE}
site <- "CGT/CTC" #RECOGNITION SEQUENCE OF BsmBI, BUT DIFFERENT CUTSITES. THIS ENABLES THE DigestDNA function to search for the 6 bp recognition sequence isolated, without considering the shiftet cutsites.

cutadapt <- "GTGGACAGCACGATACACCG"

barcodes_20nt <- c()
progress_count <- 0
for (i in 1:1000){
  barc_8 <- paste(c(sample(x = c("A", "T", "G", "C"), size = 8, replace = TRUE)), collapse = "")
  barc <- paste(cutadapt, barc_8, "TT", collapse = "", sep = "")
  digest <- DigestDNA(site, DNAStringSet(barc), strand = "top", type = "position")
  test_1 <- barc %in% barcodes_20nt
  test_2 <- length(digest[[1]]$top) != 0
  test_3 <- length(c(grep("GGGG", barc), grep("AAAA", barc), grep("CCCC", barc), grep("TTTT", barc))) != 0
  test_4 <- substring(barc, first = 21, last = 24) == "GCAT"
  test_final <- c(test_1, test_2, test_3, test_4)
  while (sum(test_final) != 0){
    barc_8 <- paste(c(sample(x = c("A", "T", "G", "C"), size = 8, replace = TRUE)), collapse = "")
    barc <- paste(cutadapt, barc_8, "TT", collapse = "", sep = "")
    digest <- DigestDNA(site, DNAStringSet(barc), strand = "top", type = "position")
    test_1 <- barc %in% barcodes_20nt
    test_2 <- length(digest[[1]]$top) != 0
    test_3 <- length(c(grep("GGGG", barc), grep("AAAA", barc), grep("CCCC", barc), grep("TTTT", barc))) != 0
    test_4 <- substring(barc, first = 21, last = 24) == "GCAT"
    test_final <- c(test_1, test_2, test_3, test_4)
  }

  barcodes_20nt <- append(barcodes_20nt, barc)

  #Progress bar
  progress(value = progress_count/1000*100, progress.bar = TRUE)
  progress_count <- progress_count + 1
#}
```
```

```
#barcodes_15nt <- c()
#progress_count <- 0
#for (i in 1:1000){
 # barc <- paste(c(sample(x = c("A", "T", "G", "C"), size = 15, replace =
TRUE)), collapse = "")
 # digest <- DigestDNA(site, DNAStringSet(barc), strand = "top", type =
"position")
  #test_1 <- barc %in% barcodes_15nt
  #test_2 <- length(grep(barc, barcodes_20nt)) != 0
  #test_3 <- length(digest[[1]]$top) != 0
  #test_4 <- substring(barc, first = 1, last = 5) == "AGACG"
  #test_5 <- substring(barc, first = 1, last = 4) == "TCTC"
  #test_6 <- substring(barc, first = 21, last = 25) == "CGTCT"
  #test_7 <- substring(barc, first = 22, last = 25) == "GAGA"
  #test_8 <- length(c(grep("GGGG", barc), grep("AAAA", barc), grep("CCCC",
barc), grep("TTTT", barc))) != 0
  #test_final <- c(test_1, test_2, test_3, test_4, test_5, test_6, test_7,
test_8)

  #while (sum(test_final) != 0){
    ##barc <- paste(c(sample(x = c("A", "T", "G", "C"), size = 15, replace
= TRUE)), collapse = "")
    #digest <- DigestDNA(site, DNAStringSet(barc), strand = "top", type =
"position")
    #test_1 <- barc %in% barcodes_15nt
    #test_2 <- length(grep(barc, barcodes_20nt)) != 0
    #test_3 <- length(digest[[1]]$top) != 0
    #test_4 <- substring(barc, first = 1, last = 5) == "AGACG"
    #test_5 <- substring(barc, first = 1, last = 4) == "TCTC"
    #test_6 <- substring(barc, first = 21, last = 25) == "CGTCT"
    #test_7 <- substring(barc, first = 22, last = 25) == "GAGA"
    #test_8   <-   length(c(grep("GGGG",   barc),   grep("AAAA",   barc),
grep("CCCC", barc), grep("TTTT", barc))) != 0
    #test_final  <-  c(test_1,  test_2,  test_3,  test_4,  test_5,  test_6,
test_7, test_8)
  #}

  #barcodes_15nt <- append(barcodes_15nt, barc)

  #Progress bar
  #progress(value = progress_count/1000*100, progress.bar = TRUE)
  #progress_count <- progress_count + 1
}

barcodes <- data.frame(barcodes_20nt = barcodes_20nt)
write.csv(x = barcodes, file = "barcodes_CFTR_20221024.csv")

```



```{r Reading in pegRNAs}
file_name    <-    "C:/Users/wians/OneDrive    -    Nexus365/Papers/Mini
Library/RFiles_WS/F508del-new-pegIT-WS.xlsx"
sheets_names <- excel_sheets(file_name)[-1]
n_sheets <- length(sheets_names)

pegRNAs <- c()

for (i in 1:n_sheets+1){
```

```
  # Reading all pegRNAs
  pegs <- as.data.frame(read_excel(path = file_name, sheet = i, range =
"A32:D10000", col_names = TRUE))
  pegs <- na.omit(pegs)

  #adding recommended pegRNA
  rec_pegs <- as.data.frame(read_excel(path = file_name, sheet = i, range
= "A2:L5", col_names = TRUE))
  rec_pegs$sequence <- paste0(rec_pegs$rt_template, rec_pegs$pbs)

  rec_pegs <- rec_pegs[c(1, 9, 10, 13)]
  pegs <- rbind(pegs, rec_pegs)

  pegs       <-       pegs[order(pegs$`#pegRNA`,       pegs$pbs_length,
pegs$rt_template_length),]

  spacers <- as.data.frame(read_excel(path = file_name, sheet = i, range
= "A2:N5", col_names = TRUE))
  spacer <- rep(spacers$spacer, each = nrow(subset(pegs, pegs$`#pegRNA`==
1)))
  gene <- rep(sheets_names[i-1], nrow(pegs))
  gene_id <- rep(i-1, nrow(pegs))
  score <- rep(spacers$score, each = nrow(subset(pegs, pegs$`#pegRNA`==
1)))
  distance   <-   rep(spacers$distance,   each   =   nrow(subset(pegs,
pegs$`#pegRNA`== 1)))
  strand <- rep(spacers$strand, each = nrow(subset(pegs, pegs$`#pegRNA`==
1)))
  pam_disrupted <- rep(spacers$pam_disrupted, each = nrow(subset(pegs,
pegs$`#pegRNA`== 1)))


  pegs$spacer <- spacer
  pegs$gene_id <- gene_id
  pegs$gene <- gene
  pegs$score <- score
  pegs$distance <- distance
  pegs$strand <- strand
  pegs$pam_disrupted <- pam_disrupted


  pegRNAs <- rbind(pegRNAs, pegs)
}

pegRNAs     <-     pegRNAs[order(pegRNAs$gene_id,     pegRNAs$`#pegRNA`,
pegRNAs$pbs_length, pegRNAs$rt_template_length),]
colnames(pegRNAs)[1] <- "pegRNA_id"

write.csv(pegRNAs,"C:/Users/wians/OneDrive                                -
Nexus365/Documents/RFiles/pegRNALibrary/pegRNAs")
```

```{r Subsetting pegRNAs}
CFTR_1 <- subset(pegRNAs, gene == "1 VCV000007105 Substitution" & pegRNA_id
== 1 & pbs_length %in% c(11:20) & rt_template_length %in% c(16:40))
CFTR_2 <- subset(pegRNAs, gene == "1 VCV000007105 Substitution" & pegRNA_id
== 2 & pbs_length %in% c(11:20) & rt_template_length %in% c(27:40))
CFTR_3 <- subset(pegRNAs, gene == "1 VCV000007105 Substitution" & pegRNA_id
== 3 & pbs_length %in% c(11:15) & rt_template_length %in% c(40:50))
```

```
CFTR_1_target                                                                <-
"CCTGGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTC"
CFTR_2_target                                                                <-
"CCTGGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGATACAGAAGCGTC"
CFTR_3_target                                                                <-
"TCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATA"
#HEK3_target  <- "CCTTGGGGCCCAGACTGAGCACGTGATGGCAGAGGAAAGGAAG"

CFTR_1$target <- rep(CFTR_1_target, nrow(CFTR_1))
CFTR_2$target <- rep(CFTR_2_target, nrow(CFTR_2))
CFTR_3$target <- rep(CFTR_3_target, nrow(CFTR_3))

pegRNAs_final <- rbind(CFTR_1, CFTR_2, CFTR_3)

#HEK3 <- subset(pegRNAs_final, pbs_length == 13 & rt_template_length ==
20)
#HEK3$spacer <- "GGCCCAGACTGAGCACGTGA"
#HEK3$gene <- paste(HEK3$gene, "HEK3", sep = "/")

#pegRNAs_final <- rbind(pegRNAs_final, HEK3)


print(pegRNAs_final)
nrow(pegRNAs_final)
```

```{r Oligo Assembly}
barcodes <- read.csv(file = "barcodes_CFTR_20221024.csv")
#pegRNAs_final$barcode_15 <- barcodes_15nt[1:nrow(pegRNAs_final)]
pegRNAs_final$barcode_20 <- barcodes_20nt[1:nrow(pegRNAs_final)]

constant_5_G   <- "CAATCCGCCCTCACTACAACCGGCGTCTCACACCG" #5' constant and
3' constant includes BsmBI site
constant_5_noG <- "CAATCCGCCCTCACTACAACCGGCGTCTCACACC"
Opt_scf                                                                      <-
"GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG
CACCGAGTCGGTGC"
pT <- "TTTTTTT"
constant_3 <- "CGTCTCAAAGCCTCCCTCATCGACGCCAGAGTAG"

library_oligos <- c()
for (i in 1:nrow(pegRNAs_final)){
  spacer_first_base   <-   strsplit(pegRNAs_final[i,]$spacer,   split   =
"")[[1]][1]
  if (spacer_first_base == "G"){
    constant_5 <- constant_5_noG
  } else {
    constant_5 <- constant_5_G
  }

  library_oligo <- paste(constant_5, pegRNAs_final[i,]$spacer, Opt_scf,
pegRNAs_final[i,]$sequence,       pT,       pegRNAs_final[i,]$barcode_20,
pegRNAs_final[i,]$target, constant_3, sep = "")

  library_oligos <- append(library_oligos, library_oligo)
}

pegRNAs_final$library_oligo <- library_oligos
```

```
head(pegRNAs_final)

summary(nchar(pegRNAs_final$library_oligo))

```
```{r Screening for BsmBI sites}

#Filtering based on introduction of BsmBI in the following locations: Target sequence (and spacer), Extension and also filtering for BsmBI sites that are introduced at junction between e.g. Extension and scaffold. The last type of BsmBI sites will not be detected, if just checking the target and/or extension seperately, as the BsmBI site itself is split between different elements.

site <- "CGT/CTC" #RECOGNITION SEQUENCE OF BsmBI, BUT DIFFERENT CUTSITES. THIS ENABLES THE DigestDNA function to search for the 6 bp recognition sequence isolated, without considering the shiftet cutsites.

dna <- DNAStringSet(pegRNAs_final$library_oligo)

#Finding number of BsmBI recognition sites
digest_library <- DigestDNA("CGT/CTC" , dna, strand = "top", type = "position")

#Checking number of recognition sites for each pegRNA library oligo. Only and ONLY 2 sites should be present in each oligo. If not,
#the oligo should be flagged as containing excessive BsmBI targets
bsmbi_check <- c()
for (i in 1:length(digest_library)){
  if (length(digest_library[[i]]$top) != 2){
    bsmbi_check <- append(bsmbi_check, TRUE)
  } else {
    bsmbi_check <- append(bsmbi_check, FALSE)}
}

pegRNAs_final$bsmbi_check <- bsmbi_check

# CUTADADPT CHECK
#Checking number of motifs for the cutadadpt. Only and ONLY 1 site should be present in each oligo. If not,
#the oligo should be flagged as containing excessive cutadapt motifs
cutadapt_library <- DigestDNA("ATGCA/TGCAT" , dna, strand = "top", type = "position")
cutadapt_check <- c()
for (i in 1:length(cutadapt_library)){
  if (length(cutadapt_library[[i]]$top) != 1){
    cutadapt_check <- append(cutadapt_check, TRUE)
  } else {
    cutadapt_check <- append(cutadapt_check, FALSE)}
}

pegRNAs_final$cutadapt_check <- cutadapt_check


summary(pegRNAs_final)


subset(pegRNAs_final, cutadapt_check == TRUE)


write.xlsx2(pegRNAs_final, file = "pegRNAs_library_CFTR_20221024.xlsx")
```

```
```  }

  library_oligo <- paste(constant_5, pegRNAs_final[i,]$spacer, Opt_scf,
pegRNAs_final[i,]$sequence,   pT,   pegRNAs_final[i,]$barcode_20,
pegRNAs_final[i,]$target, constant_3, sep = "")

  library_oligos <- append(library_oligos, library_oligo)
}
pegRNAs_final$library_oligo <- library_oligos
head(pegRNAs_final)

summary(nchar(pegRNAs_final$library_oligo))

```
```{r Screening for BsmBI sites}

#Filtering based on introduction of BsmBI in the following locations:
Target sequence (and spacer), Extension and also filtering for BsmBI sites
that are introduced at junction between e.g. Extension and scaffold. The
last type of BsmBI sites will not be detected, if just checking the target
and/or extension seperately, as the BsmBI site itself is split between
different elements.

site <- "CGT/CTC" #RECOGNITION SEQUENCE OF BsmBI, BUT DIFFERENT CUTSITES.
THIS ENABLES THE DigestDNA function to search for the 6 bp recognition
sequence isolated, without considering the shiftet cutsites.

dna <- DNAStringSet(pegRNAs_final$library_oligo)

#Finding number of BsmBI recognition sites
digest_library <- DigestDNA("CGT/CTC" , dna, strand = "top", type =
"position")

#Checking number of recognition sites for each pegRNA library oligo. Only
and ONLY 2 sites should be present in each oligo. If not,
#the oligo should be flagged as containing excessive BsmBI targets
bsmbi_check <- c()
for (i in 1:length(digest_library)){
  if (length(digest_library[[i]]$top) != 2){
    bsmbi_check <- append(bsmbi_check, TRUE)
  } else {
    bsmbi_check <- append(bsmbi_check, FALSE)}
}
pegRNAs_final$bsmbi_check <- bsmbi_check

# CUTADADPT CHECK
#Checking number of motifs for the cutadadpt. Only and ONLY 1 site should
be present in each oligo. If not,
#the oligo should be flagged as containing excessive cutadapt motifs
cutadapt_library <- DigestDNA("ATGCA/TGCAT" , dna, strand = "top", type =
"position")
cutadapt_check <- c()
for (i in 1:length(cutadapt_library)){
  if (length(cutadapt_library[[i]]$top) != 1){
    cutadapt_check <- append(cutadapt_check, TRUE)
  } else {
    cutadapt_check <- append(cutadapt_check, FALSE)}
}
pegRNAs_final$cutadapt_check <- cutadapt_check

summary(pegRNAs_final)
```
```

```
subset(pegRNAs_final, cutadapt_check == TRUE)

write.xlsx2(pegRNAs_final, file = "pegRNAs_library_CFTR_20221024.xlsx")
```

NGS Analysis – sample trimming with CutAdapt and demultiplexing (Python) – code by Dr Jakob Haldrup

```
import subprocess

# Define input and output directories
input_directory                                                    =
"/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/G5/"
output_directory                                                   =
"/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G5/"

# Define adapter sequences
adapter_fwd = "GTGGACAGCACGATACACCG"   # Replace with forward adapter
sequence
adapter_rev = "CGGTGTATCGTGCTGTCCAC"   # Replace with reverse adapter
sequence

# Specify trim lengths for forward and reverse reads
trim_length_fwd = 60  # Replace with the desired trim length for forward
reads
trim_length_rev = 100  # Replace with the desired trim length for reverse
reads

# Construct cutadapt command for paired-end reads with individual trim
lengths
cutadapt_command = [
    "cutadapt",
    "-g", adapter_fwd,
    "-A", adapter_rev,
    "--discard-untrimmed",  # Discard reads without adapter sequence
    "--match-read-wildcards",  # Require a perfect match for the adapter
    "--length", str(trim_length_fwd),    # Trim forward reads to the
specified length
    "--length", str(trim_length_rev),    # Trim reverse reads to the
specified length
    "-o", output_directory + "trimmed_forward.fastq.gz",
    "-p", output_directory + "trimmed_reverse.fastq.gz",
    input_directory + "forward_reads.fastq.gz",
    input_directory + "reverse_reads.fastq.gz"
]

# Execute cutadapt command and capture output
try:
    process      =      subprocess.run(cutadapt_command,      check=True,
stdout=subprocess.PIPE, stderr=subprocess.PIPE, text=True)
    print("Cutadapt command executed successfully!")
    # Display detailed output
    print("Cutadapt detailed output:")
    print(process.stdout)
    print("Errors or warnings:")
    print(process.stderr)
except subprocess.CalledProcessError as e:
    print(f"Error executing cutadapt command: {e}")
import pandas as pd
from Bio import SeqIO
from collections import Counter
import gzip

def count_barcodes_and_write_excel(trimmed_fastq_file, excel_input_file,
excel_output_file):
```

```
    # Read the Excel sheet with barcodes and names
    df_names = pd.read_excel(excel_input_file)

    # Make sure the column names match the case in your Excel file
    barcode_column = 'barcode_20'  # Change to the exact name in your Excel
file
    name_column = 'Name'  # Change to the exact name in your Excel file

    # Create a dictionary to map barcodes to names
    barcode_to_name          =          dict(zip(df_names[barcode_column],
df_names[name_column]))

    barcodes = []

    # Open trimmed FastQ file and extract barcodes
    with gzip.open(trimmed_fastq_file, "rt") as fastq_handle:
        fastq_records = SeqIO.parse(fastq_handle, "fastq")
        for record in fastq_records:
            # Assuming the barcode is the first 10 nucleotides
            barcode = str(record.seq[:10])
            barcodes.append(barcode)

    # Count occurrences of each barcode
    barcode_counts = Counter(barcodes)

    # Convert barcode counts to a pandas DataFrame
    barcode_df        =        pd.DataFrame(list(barcode_counts.items()),
columns=[barcode_column, "Count"])

    # Add a 'Name' column based on the mapping of barcodes to names
    barcode_df['Name'] = barcode_df[barcode_column].map(barcode_to_name)

    # Reorder columns to have 'Name' in column C
    result_df = barcode_df[[barcode_column, "Count", "Name"]]

    # Writing the DataFrame to Excel
    result_df.to_excel(excel_output_file, index=False)

if __name__ == "__main__":
    trimmed_fastq_file                                                    =
"/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G2/aligned_rea
ds.fastq.gz"
    excel_input_file                                                      =
"/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G2/pegRNA_CFTR
.xlsx"  # Adjust the path
    excel_output_file                                                     =
"/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G2/barcode_cou
nts_with_names.xlsx"

    count_barcodes_and_write_excel(trimmed_fastq_file,  excel_input_file,
excel_output_file)

    print(f"Barcode counts with names written to {excel_output_file}")
Barcode        counts        with        names        written        to
/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G2/barcode_coun
ts_with_names.xlsx
```

NGS Analysis – CRISPResso2 batch analysis – code by Dr Jakob Haldrup

```
# Loading pegRNA information from pegIT Excel file
import openpyxl

file_path =
'/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G6/pegRNA_CFTR
.xlsx'
workbook = openpyxl.load_workbook(file_path)  # Load the Excel workbook
sheet = workbook.active  # Access the active sheet

fastq_dict = {}  # Dictionary to store FASTQ barcodes and corresponding
values
oligo_name_list = []  # List to hold oligo names (not used in the current
script)
oligo_name_dict = {}  # Dictionary to map unique barcodes to oligo names

# Loop through the first 454 rows of the Excel sheet
for i in range(454):
    oligo_name = sheet.cell(row=(i+2), column=1).value  # Get oligo name
from the first column
    oligo_name = oligo_name.replace(" ", "_")  # Replace spaces with
underscores in oligo name
    oligo_barcode = sheet.cell(row=(i+2), column=15).value  # Get oligo
barcode from the 15th column
    oligo_unique_barcode = oligo_barcode[20:]  # Extract the unique part
of the oligo barcode

    fastq_dict[oligo_unique_barcode] = []  # Initialize an entry in
fastq_dict for the unique barcode
    oligo_name_dict[oligo_unique_barcode] = oligo_name  # Map the unique
barcode to its oligo name

# Loading pegRNA information from pegIT Excel file
import openpyxl

file_path =
'/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G6/pegRNA_CFTR
.xlsx'
workbook = openpyxl.load_workbook(file_path)  # Load the Excel workbook
sheet = workbook.active  # Access the active sheet

fastq_dict = {}  # Dictionary to store FASTQ barcodes and corresponding
values
oligo_name_list = []  # List to hold oligo names (not used in the current
script)
oligo_name_dict = {}  # Dictionary to map unique barcodes to oligo names

# Loop through the first 454 rows of the Excel sheet
for i in range(454):
    oligo_name = sheet.cell(row=(i+2), column=1).value  # Get oligo name
from the first column
    oligo_name = oligo_name.replace(" ", "_")  # Replace spaces with
underscores in oligo name
    oligo_barcode = sheet.cell(row=(i+2), column=15).value  # Get oligo
barcode from the 15th column
    oligo_unique_barcode = oligo_barcode[20:]  # Extract the unique part
of the oligo barcode

    fastq_dict[oligo_unique_barcode] = []  # Initialize an entry in
fastq_dict for the unique barcode
```

```
    oligo_name_dict[oligo_unique_barcode] = oligo_name  # Map the unique
barcode to its oligo name
# Loading cutadapt 5' trimmed FASTQ file
import gzip

file_path                                                                =
"/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G6/aligned_rea
ds.fastq.gz"

trimmed_data = []  # Initialize the list to store trimmed data

try:
    with gzip.open(file_path, "rt") as trimmed:  # Open the gzipped FASTQ
file in text mode
        for line in trimmed:  # Iterate through each line in the file
            trimmed_data.append(line)     #  Append  each  line  to  the
trimmed_data list
except FileNotFoundError:
    print("File not found. Please check the file path.")  # Handle file
not found error
except Exception as e:
    print(f"An error occurred: {e}")  # Handle any other exceptions
# Sorting FASTQ output into separate lists
read_ID_list = []  # List to store read IDs
sequence_list = []  # List to store sequences
quality_list = []  # List to store quality scores

# Extracting data from trimmed_data
for i in range(0, len(trimmed_data) - 1, 4):  # Loop through every fourth
line
    read_ID_list.append(trimmed_data[i])  # Append read ID
    sequence_list.append(trimmed_data[i + 1])  # Append sequence
    quality_list.append(trimmed_data[i + 3])  # Append quality score

# Sorting FASTQ output into a dictionary based on barcode (fastq_dict)
for i in range(len(read_ID_list)):
    barcode = sequence_list[i][0:10]  # Extract the first 10 characters as
barcode
    if barcode in fastq_dict:  # Check if barcode exists in fastq_dict
        read_ID = read_ID_list[i]  # Get corresponding read ID
        sequence = sequence_list[i]  # Get corresponding sequence
        quality = quality_list[i]  # Get corresponding quality score
        assembly = [read_ID, sequence, "+\n", quality]  # Create a FASTQ
entry
        for j in range(len(assembly)):
            fastq_dict[barcode].append(assembly[j])  # Append entry to the
dictionary
# Writing separate FASTQ files for each oligo
import os

# Output folder path
output_folder                                                            =
'/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G6/'

# Initialize a list to keep track of R1 FASTQ file names
fastq_r1_list = []

# Writing separate FASTQ files for each oligo
for i in fastq_dict:
    filename = os.path.join(output_folder, oligo_name_dict[i] + "_r1" +
".fastq")  # Construct the output filename
```

```
    fastq_r1_list.append(oligo_name_dict[i] + "_r1")  # Add the filename
to the list
    reads = fastq_dict[i]  # Retrieve the list of reads for the current
barcode
    if fastq_dict[i] != []:  # Check if there are any reads to write
        with open(filename, "w") as output:  # Open the file for writing
            for line in reads:  # Iterate through the list of reads
                output.write(line)  # Write each line to the output file

    print(oligo_name_dict[i])  # Print the oligo name for confirmation

import os
import pandas as pd

# Writing separate FASTQ files for each oligo
fastq_r1_list = []  # List to store names of R1 FASTQ files
barcode_sequences = []  # List to store barcode sequences
barcode_counts = []  # List to store counts of reads per barcode

for i in fastq_dict:
    filename = os.path.join(output_folder, oligo_name_dict[i] + "_r1" +
".fastq")  # Construct the output filename
    fastq_r1_list.append(oligo_name_dict[i] + "_r1")  # Add the filename
to the list
    reads = fastq_dict[i]  # Retrieve the list of reads for the current
barcode

    # Counting the number of reads for each barcode
    barcode_counts.append(len(reads) // 4)  # Each read consists of 4 lines
in a FASTQ file

    # Extracting the first 10 characters as the barcode sequence
    if len(reads) >= 2:  # Ensure there are enough reads to extract a
barcode
        barcode_sequence = reads[1][0:10]  # Extract the barcode sequence
        barcode_sequences.append(barcode_sequence)  # Append to the list
    else:
        # Handle the case where there aren't enough reads
        barcode_sequences.append("N/A")  # Append "N/A" for insufficient
reads

    if fastq_dict[i] != []:  # Check if there are reads to write
        with open(filename, "w") as output:  # Open the file for writing
            for line in reads:  # Iterate through the list of reads
                output.write(line)  # Write each line to the output file

    print(oligo_name_dict[i])  # Print the oligo name for confirmation

# Creating a DataFrame and writing it to an Excel file
df = pd.DataFrame({
    'Barcode': fastq_r1_list,  # Column for R1 FASTQ file names
    'Barcode_Sequence': barcode_sequences,  # Column for barcode sequences
    'Read_Count': barcode_counts  # Column for counts of reads
})

excel_filename = os.path.join(output_folder, 'read_counts.xlsx')    #
Construct the output Excel filename
df.to_excel(excel_filename, index=False)  # Write the DataFrame to an Excel
file without row indices
print(f"Excel file '{excel_filename}' generated with read counts and
barcode sequences.")  # Confirmation message
```

```
import os

# Specify the output folder for batch files
output_folder                                                    =
"/Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G1/"

# Writing a tab-separated batch file for running CRISPResso2 (batch mode)
spacer_names = ["CFTR_spacer_1", "CFTR_spacer_2", "CFTR_spacer_3"]  # List
of spacer names

for i in range(len(spacer_names)):
    spacer = spacer_names[i]  # Current spacer name
    oligo_names = ["name"]  # Initialize oligo names list
    fastq_r1 = ["fastq_r1"]  # Initialize R1 FASTQ file names list
    fastq_r2 = ["fastq_r2"]  # Initialize R2 FASTQ file names list

    for j in fastq_dict:  # Iterate through each barcode in fastq_dict
        oligo_name = oligo_name_dict[j]   # Get the corresponding oligo
name
        oligo_names.append(oligo_name)  # Add oligo name to the list
        fastq_r1.append(oligo_name + "_r1.fastq")  # Construct and add R1
FASTQ filename
        fastq_r2.append(oligo_name + "_r2.fastq")  # Construct and add R2
FASTQ filename

    merge = [oligo_names, fastq_r1, fastq_r2]  # Merge lists into a single
list
    batch_file_name = os.path.join(output_folder, spacer + "_batch.txt")
# Construct the batch file name

    with open(batch_file_name, "w") as file:  # Open the batch file for
writing
        for x in zip(*merge):  # Zip the lists together for tab-separated
output
            file.write("{0}\t{1}\t{2}\n".format(*x))  # Write each line to
the batch file
cd /Users/jakobhaldrupjensen/Python/NGSlib/Data/MiSeq/Output/G2/
```