Development of Site-Specific CFTR Gene Integration Tools for Testing Gene Editing in Pig Cells

By

Liang Yang

A thesis submitted in conformity with the requirements

for the degree of Master of Science

Department of Laboratory Medicine and Pathobiology

University of Toronto

Development of Site-Specific CFTR Gene Integration Tools for Testing Gene Editing in Pig Cells

Liang Yang

Master of Science

Department of Laboratory Medicine and Pathobiology

University of Toronto

2017

Abstract

Cystic Fibrosis (CF) is a genetic disorder caused by autosomal recessive mutations in the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) gene. In this study, we proposed a novel gene therapy strategy to integrate a transgene expression cassette into *GGTA1* locus utilizing the precise genome cleavage capability of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR associated protein (Cas) and the large packaging capacity of Helper Dependent Adenoviral vector (HD-Ad). Using a *LacZ* reporter gene expression cassette, we determined this novel strategy can achieve stable and sufficient integration in the pig IPEC-J2 cell line. In addition, the *CFTR* transgene mRNA and protein can be successfully detected post HD-Ad delivery. Future experiments include investigating the CFTR functional correction and the impact of enhancing homology directed repair (HDR), which is the major pathway we rely on for transgene integration, on integration efficiency and assess CFTR transgene functional corrections.

Acknowledgements

This Master's project would not have been completed without the help and support from many people. I would like to thank Dr. Jim Hu for agreeing to supervise me and offering me with a fascinating research project. I am very thankful to all the advise and guidance that Dr. Jim Hu provided throughout my Master's project which helped me immensely. Furthermore, I greatly appreciate the suggestions and feedbacks delivered during lab meetings and private discussions from my lab members: Dr. Huibi Cao, Yiqian Zhang, Kyle Siegel, Zhichang Zhou, Jun Li and Emily Xia.

This study was funded by Canadian Institute of Health Research (supervisor's grant) as well as the Hospital for Sick Children (Research Training Competition Scholarship).

I am grateful of Dr. Theo Moraes and Dr. Hartmut Grasemann for agreeing to be my committee members. Not only do they listen carefully to my committee presentations, but they also provided me numerous valuable suggestions and future directions. These constructive criticisms aided me to complete my project.

In addition, I want to express my gratitude toward the Department of Laboratory Medicine and Pathobiology for providing me a remarkable opportunity to study in this top ranked university. Many aspects of my academic abilities were greatly strengthened through the courses offered by this department.

Finally, I am very thankful to my family members for their continuous encouragements and supports throughout my entire life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

α-1,3Gal:	galactose- α-1,3-galactose
Ψ:	Packaging signal of Adenovirus
ΔF508:	Deletion of phenylalanine residue at position 508 of CFTR
A455E:	Substitution of alanine to glutamate at position 455 of CFTR
AAV:	Adeno-associated virus
Ad:	Adenovirus
ASL:	Airway Surface Liquid
ATP:	Adenosine triphosphate
bp:	base pairs
BSA:	Bovine Serum Albumin
cAMP:	cyclic adenosine monophosphate
cDNA:	complementary DNA
crRNA:	CRISPR RNA
CAR:	Coxsackie and Adenovirus Receptor
CF:	Cystic Fibrosis
CFTR:	Cystic Fibrosis Transmembrane conductance Regulator
CPE:	Cytopathic effect
CRISPR/Cas9:	Clustered Regularly Interspaced Short Palindromic Repeat/CRISPR associated protein 9
CT:	Cycle threshold
DSB:	Double Stranded break
eGFP:	Enhanced green fluorescent protein
FBS:	Fetal bovine serum
FG-Ad:	First-Generation Adenoviral vector
G2 phase:	Gap 2 phase of cell cycle
G551D:	substitution of glycine to aspartic acid at position 551 of CFTR
GAPDH:	Glyceraldehyde-3-Phosphate Dehydrogenase
GGTA1:	α-1,3-galactosyltransferase
hCFTR:	Human CFTR transgene
HA:	Homology Arms
HD-Ad:	Helper-Dependent Adenoviral vector

HDR:	Homology Directed Repair
IPEC-J2:	Intestinal Porcine Enterocyte cell line
ITR:	Inverted Terminal Repeats
kb:	kilo base pairs
K18:	Keratin 18
LacZ:	β galactosidase
mRNA:	Messenger RNA
MEM:	Minimum essential medium
MOI:	Multiplicity of infection
NHEJ:	Non-Homologous End Joining
PAM:	Proto-spacer adjacent motif
PBS:	Phosphate buffered saline
PCR:	Polymerase Chain Reaction
Q1412X:	Nonsense mutation at position 1412 of CFTR
rtTA:	Reverse tetracycline-inducible transcription activator
R334W:	Substitution of arginine to tryptophan at position 334 of CFTR
Rpm:	Rotation per minute
RT-qPCR:	Quantitative Reverse Transcription Polymerase Chain Reaction
sgRNA:	single guide RNA
S phase:	Synthesis phase of cell cycle
SDS:	Sodium dodecyl sulfate
SG-Ad:	Second-Generation Adenoviral vector
tracrRNA:	trans-acting CRISPR RNA
TAE:	Tris, acetic acid EDTA buffer
TE:	10 mM Tris-HCl, 1 mM EDTA, pH 8.0
W1282X:	Nonsense mutation at position 1282 of CFTR
X-Gal:	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
UbC:	Ubiquitin C promoter

1. Introduction

1.1 Cystic Fibrosis

Cystic Fibrosis (CF) is the most common monogenic fatal disorder found in the Caucasian population (1). In Canada, it was estimated that 1 in every 3600 newborns is affected by CF (2). This disease was first characterized by Dorothy Andersen nearly 80 years ago (3). In her original paper, she identified this disease in the pancreas of malnourished infants with subsequent lung infections and salt loss during heat waves (3-5). CF is a multi-organ disease and the symptoms associated with this genetic disorder include respiratory failure, meconium ileus (thickened meconium leading to intestinal blockage in newborns), pancreatic insufficiency, malnutrition, male infertility, etc. (6, 7). When CF was first described, the average life expectancy of the patients was only a few months due to digestive tract complications (4). As this disease has been progressively studied and treatment methods were developed, the mean life expectancy of CF patients has significantly increased to an estimated median survival of 50 years for the newborn CF patients of year 2000 (8). The leading cause of morbidity and mortality for CF has shifted from gastrointestinal tract problems to pulmonary complications (9).

1.1.1 Cystic Fibrosis Transmembrane conductance Regulator and CF lung pathology

In 1989, the cause of CF was successfully identified to be a mutation in a gene encoding for Cystic Fibrosis Transmembrane conductance Regulator (CFTR) (10). Not long after the identification of *CFTR* gene, the function of wild type CFTR protein was discovered (11). These monumental steps significantly contributed to the current understanding of the pathophysiology of CF. CFTR is an ATP Binding Cassette-class transporter and it is responsible for the proper chloride/ bicarbonate conductance across the epithelial cell membranes (12). CFTR is found on the apical membrane

of epithelial cells in various organs (12). Gated by cycles of ATP binding and hydrolysis and regulated by cAMP dependent protein kinase A, CFTR allows the flow of selective anions (such as chloride and bicarbonate) down their electrochemical gradient (13). The chloride ion enters the epithelial cell through the Na^+-K^+-$2Cl^-$ cotransporter and exit through the CFTR (14). However, in the absorptive epithelia such as the sweat gland, CFTR is responsible for salt absorption instead of secretion (14). This regulated movement of chloride is essential for the water balance across the epithelial cells through osmosis (15). Besides its role in anion conductance, it was shown that CFTR possesses additional regulatory functions for other epithelial ion channels (16, 17). When mutations occur in the CFTR gene, the CFTR activity or expression will be reduced. In lungs, reduced CFTR function leads to abnormal hydration of airway surface liquid (ASL) (18). Since proper chloride conductance is impaired, water molecules have difficulty flowing into the ASL, which results in thickened mucus build-up in the airway lumen. This is very damaging to the lung because not only it can clog up the airway, but the dehydrated mucus also can decrease mucociliary clearance and provide a place for bacteria to colonize (19). In addition, reduced bicarbonate secretion by CFTR mutation also acidifies the ASL that can lower the capability of host antibacterial mechanisms (20). Opportunistic bacteria such as *Pseudomonas aeruginosa* can establish chronic infections that are difficult to eradicate. Chronic inflammation is also associated with the persistent lung infection and it is characterized by neutrophil infiltration and many other pro-inflammatory reactions (21, 22). These factors all contribute to the gradual decrease in lung function observed in CF patients. In fact, chronic infection and inflammation established in the thickened mucus is the leading cause of CF related morbidity and mortality with approximately 80 – 95% of CF patients eventually succumbing to respiratory failure (21).

1.1.2 Types of CFTR mutation

There are more than 2000 types of mutations identified for CFTR according to the cystic fibrosis mutation database (23). Because of this vast number of mutations recorded, all the mutations were classified into 6 categories (24). In class I mutations, defects exist in CFTR protein production, usually due to non-sense or frameshift mutations, so no full length CFTR mRNA can be synthesized (25). In contrast, class II mutations are usually caused by missense mutations or amino acid deletions, and the full-length mRNA can still be produced (4). However, defects exist in protein structure that lead to the formation of misfolded CFTR protein. As a result, no functional CFTR will reach the plasma membrane and they are targeted for ubiquitin-proteasome and lysosome degradation instead (13, 26). For class III to VI mutations, certain level of CFTR protein is expressed on the cell membrane. These classes retain residue CFTR functions and are associated with milder disease phenotypes (27). More specifically, the class III mutation is categorized by defective channel regulation. Even though the CFTR protein is correctly processed and expressed, disruption in channel gating can significantly affect channel function (24). The most well-known mutation in this class is called G551D mutation (substitution of glycine to aspartic acid at position 551) and it is the third most common CFTR mutation identified (23). Class IV mutations are characterized by decreased channel conductance compared to wild type CFTR. For example, R334W (substitution of arginine to tryptophan at position 334) greatly reduces chloride conductance since this residue was shown to play an important role in coordinating ion-ion interactions inside the pore of CFTR (28). Reduced CFTR protein synthesis is observed in class V mutations. There are many potential causes for decreased protein synthesis including promoter mutation that affects transcription, alteration of amino acid residues that lead to inefficient protein maturation, etc. (24) An example of this class of mutation

is A455E (substitution of alanine to glutamate at position 455) in which the channel formed is misprocessed but has conductive properties similar to wild type CFTR (29). Finally, class VI mutations lead to less stable CFTR proteins in comparison with wild type CFTR. The truncation of last 70 amino acid residues in Q1412X (nonsense mutation that altering glutamine at position 1412 into stop codon) mutation has significantly less half life of mature CFTR protein with no effect on CFTR processing and transport (24).

Even though most of the mutations are extremely rare, one of the mutations identified particularly stands out. It was determined that approximately 70% of CF patients carry two copies and 90% of CF patients carry one copy of CFTR that contains a deletion of the phenylalanine residue at position 508 (ΔF508). This type of mutation impairs the proper folding of CFTR protein (24, 30, 31). Because of the vast number of mutations identified, it is very challenging to study each of the mutations in detail. Therefore, many studies developing therapeutic strategies against CF have a great emphasis on the most common ΔF508 mutation.

1.1.3 Strategies to combat CF

Even though we still do not have a curative strategy against CF, better understanding in this lethal disease enabled researchers to develop therapeutic approaches to increase the survival rate in CF patients. As mentioned previously, a very characteristic symptom found in CF patients is the retaining of thickened mucus. Therefore, airway clearance is essential for the management of CF conditions (1). One approach to limit the airway obstruction by thickened mucus is to use mucolytic agents. Clinical studies have shown that the use of recombinant human DNase through inhalation was associated with improved lung functions and a reduction of pulmonary exacerbation (32). Another approach to reduce the symptoms from dehydrated ASL is through

osmotic agents. Inhaled hypertonic saline can induce water to flow into airway mucus which can enhance mucociliary clearance (1). This method had proven to be effective in improving lung function and reducing pulmonary exacerbation rates in CF patients (33, 34).

One of the major concerns in CF patients is chronic infection by *P. aeruginosa*. Even though chronic *P. aeruginosa* infections are hard to eliminate, early antibiotic treatment can be effective before they develop resistance (35). However, once the infection enters into chronic phase, *P. aeruginosa* is capable of forming biofilms and becomes challenging to eradicate due to their increased resistance against antibiotics, phagocytosis and other components of the host immune system (36). Many studies have investigated the efficacy of inhaled antibiotics. For example, inhaled tobramycin has proven to produce sustained improvements in lung functions for CF patients with identified chronic *P. aeruginosa* infection (37).

Airway inflammation is another important contributor to the eventual lung failure observed in many CF patients. Large numbers of infiltrating neutrophils and a high concentration of proinflammatory cytokines can lead to progressive airway damage (38). Therefore, anti-inflammatory agents such as azithromycin and high dose ibuprofen can significantly reduce the rates of pulmonary exacerbation (39, 40).

Extensive efforts were devoted in developing small molecules that target CFTR channel activities (13). A very successful example of this is a CFTR potentiator named Ivacaftor (VX-770) which was shown to be effective in CF patients carrying the class III G551D mutation (41-43). Nevertheless, lung function can only be partially rescued and not all the patients respond to these drugs (13).

There are many more directions being investigated to benefit the conditions of CF patients and all of these strategies have contributed to the gradual improvement of quality of life in CF patients.

1.1.4 CF gene therapy

The treatment approaches mentioned previously can only slow, but not stop, the progression of lung damage (44). In addition, the drugs that target CFTR activity are only effective for a few types of mutation. As a result, there is still a high demand for a curative strategy against CF. Gene therapy, which works by introducing a functional copy of therapeutic gene into affected cells, has the potential to cure genetic disorders such as CF (45). In fact, CF presents an attractive and suitable gene therapy target because it is a monogenic disorder. This makes developing a gene therapy strategy easier than more complex genetic diseases. Another important reason is the relative easy accessibility of the airway and less invasiveness by aerosol delivery (46). Even though CF affects multiple organs, the lungs are the major site of pathology so they are the target for the majority of CF gene therapy trials (47). Gene therapy attempts initiated shortly after the identification and cloning of the CFTR gene with the first attempt being the utilization of replication deficient adenoviral vector (48). The expression of transgene through adenoviral vector administration in human demonstrated to be transient and the expression suffered after repeated vector administration (49). Adeno-associated viral (AAV) vectors were also investigated to avoid the limitations of adenoviral vectors. AAV is relatively safe and can efficiently transduce non-dividing cells (50). However, a large, placebo-controlled AAV CF clinical trial reported no differences in lung function between the control group and the CFTR receiving group 30 days after administration (48, 51). Besides using viral vectors, trials using other type of

vectors such as cationic lipids complexed with plasmid DNA and DNA nanoparticles have also been performed (48).

In spite of the efforts, all the past CF gene therapy trials had limited efficiency and efficacy due to many limitations (52). In CF patients, the thickened mucus adhering to the airway epithelium and the chronic inflammatory status can greatly impede successful CFTR gene transfer (48). Furthermore, the terminally differentiated epithelial cells have a limited life span so repeated administration of gene transfer vector is necessary. The strong adaptive immune response can interfere with the efficiency of repeated vector administrations as antibody responses to viral gene transfer vectors can eliminate the transduced cells (53, 54).

1.2 CRISPR/Cas9 system

With the rapid development of new technologies, new tools available have made genetic studies much more efficient. Recently, a system called Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/ CRISPR associated proteins (Cas) has been widely utilized for performing genome editing (55).

1.2.1 The discovery and arrangement of CRISPR/Cas9 system

The CRISPR/Cas is a two-component system that was originally identified as a bacterial and archaeal immune system. The main role of this system is to combat infections from various phages (56). The first component of this system is the CRISPR sequence. This sequence is characterized by an array of short palindromic repeats, approximately 24 to 48 nucleotides long, separated by variable spacer DNA of similar lengths (57). There can be more than 100 repeats in the CRISPR region (58). This unique sequence was discovered back in 1987 during the sequencing of *iap* gene found in *Escherichia coli* (59). However, the role of the sequence was not known at

that time. In addition to *E. coli*, the CRISPR sequence can be found in many other bacterial and archaeal species as well (60). In 2005, the seemingly random spacer DNA in the CRISPR region demonstrated homology to numerous known genes found in many different phages (61). The second component of the CRISPR/Cas system is a diverse group of genes named *cas* genes. It was suggested that *cas* genes are functionally associated with the CRISPR sequence since for majority of the cases, *cas* genes were present when the CRISPR sequence was present as well (58). In 2007, the protective role of the CRISPR/Cas system in bacteria and archaea was finally revealed (56).

The CRISPR region is regulated by an upstream leader sequence that is approximately 300 to 500 base pairs in length. The leader sequence was found to be rich in nucleotides A and T and acted as the promoter for the CRISPR region (62). The *cas* genes are located in close proximity to the CRIPSR region. There are many different *cas* genes and each encodes for proteins that carry out different functions (60). For instance, *cas3* gene encodes for a helicase and *cas9* gene encodes for an endonuclease (62). Depending on which *cas* genes are expressed near the CRISPR region, the CRISPR/Cas system can be categorized into six types (63). This study will focus on the type II CRISPR/Cas9 system. In this system, the *cas9* gene is always expressed. As a result, the type II CRISPR/Cas system is also known as the CRISPR/Cas9 system (64). The final element of CRISPR/Cas9 is another small RNA molecule called trans-acting CRISPR RNA (tracrRNA) and is located further upstream of the CRISPR/Cas9 system (65).

1.2.2 The protective mechanism of CRISPR/Cas9 system

The protective mechanism of CRISPR/Cas9 system can be divided into three stages. The first stage takes place upon viral infection. A low proportion of cells acquire a short sequence derived

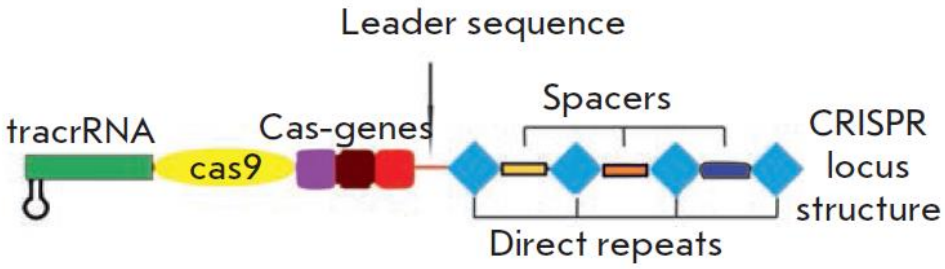


Figure 1. Arrangement of CRISPR repeats and its associated elements in type II CRISPR/Cas system. The palindromic repeats are flanked by the leader sequence which acts as the promoter. *Cas* genes are located upstream of the leader sequence with one of the *cas* gene being the Cas9 endonuclease. TracrRNA is located further upstream and it is responsible for maturation of individual crRNA molecules after transcription of the CRISPR region. Adapted from *Nemudryi et al* (66).

from the viral genetic material and is inserted into the CRISPR locus (67). This phage-derived DNA sequence becomes the variable spacer DNA found within the CRISPR region (68). During the insertion of spacer DNA, a conserved repeat is duplicated and added to the end of the newly integrated spacer DNA (64). The newly formed spacer is always incorporated at the leader end of the CRISPR locus (56). It was shown that Cas9 is required for this immunization step of the CRISPR response (68). The selection of the viral sequence to derive into spacer DNA is not a random process. All integrated spacer DNA contains a short Proto-spacer Adjacent Motif (PAM) that is critical for the function of CRISPR/Cas9 system. In addition, PAM also controls the orientation of the spacer incorporated (69).

In the second stage, the entire CRISPR cassette is transcribed into a long transcript called pre-CRISPR RNA (pre-crRNA). In the subsequent step, tracrRNA directs the maturation of individual

small crRNA molecules from the long pre-crRNA (70, 71). In addition, tracrRNA mediated maturation of crRNA is catalyzed by the enzymes RNase III and Cas9 (64, 65). Within each mature crRNA molecule, there is a different guide sequence that is transcribed from different spacer DNA. The guide sequence is flanked by a conserved repeat at 5' end (72). It was found that the conserved sequence forms a stable stem loop structure which may facilitate the recognition of crRNA by Cas9 (73).

The final stage of the CRISPR/Cas9 system is the interference with foreign DNA or RNA. There are three components involved in this stage: crRNA, tracrRNA, and Cas9 (74). crRNA and tracrRNA hybridize with each other to form a duplex. The 20-nucleotide long guide sequence within the crRNA is responsible for the site-specific DNA cleavage found in the CRISPR/Cas9 system (75). The crRNA and tracrRNA duplex associates with Cas9 and direct it to the target site through complementary base pairing between the guide sequence and the target (75). In addition to recognition by guide sequence, cleavage of the target sequence by Cas9 requires a PAM site to be located immediately downstream of the target sequence (75, 76). Cas9 from different species may require distinct PAM sites. For example, Cas9 from *Streptococcus pyogenes* has the PAM site of 5'-NGG-3' while the *Staphylococcus aureus* PAM site consists of 5'-NNGRRT-3' (77). The PAM sequence is very important for CRISPR/Cas9 mediated resistance and is bound by Cas9 during its recognition (76). It was shown when PAM site is absent, Cas9 is unable to cleave its genomic target even if perfect base pairing exists between the guide sequence and the target (76). Furthermore, if a single point mutation is introduced into PAM, the phage is able to escape the cleavage by this system (78). Once the downstream PAM sequence is recognized and binding of crRNA to genomic target occurs, Cas9 will generate a

double stranded break (DSB) within the target and 3 base pairs upstream of the PAM sequence (79, 80). As a result of this, cellular DNA repair mechanisms will be triggered.

1.2.3 Homology directed repair versus non-homologous end joining

Upon the introduction of DSB by Cas9, the cells primarily utilize two DNA repair pathways: non-homologous end joining (NHEJ) and homology directed repair (HDR) (80, 81). The NHEJ pathway is an efficient and natural mechanism for fixing DNA damages in most organisms (82, 83). Unlike the HDR pathway which will be described later, the NHEJ pathway does not require a template, which makes it simple and effective in most cases (83). In NHEJ, proteins known as Ku will bind to the two ends of DNA with DSB to form a protein-DNA complex (84). Since the broken ends are likely to have overhang base pairs, the Ku-DNA complex can associate with DNA-PKcs to fill-in or trim the ends which is known as the end processing (85). Finally, after the ends became compatible, the two broken ends can be joined together with the assistance of XLF-XRCC4-DNA ligase IV complex through ligation (86). During end processing, some insertion and deletion mutations can be introduced at the DSB site which may lead to uncontrolled mutations (83). For this reason, NHEJ is considered more error-prone than HDR.

In contrast to NHEJ, HDR is a much more sophisticated pathway which allows it to be highly accurate in repairing the DSB due to the usage of a homologous donor DNA template (83). This process involves a complicated pathway of end resection, strand invasion and recombination intermediate resolution (87). If sister chromatid is used as a template, DSB repaired by HDR do not result in altered locus post-repair (88). HDR is a slow process and often outcompeted by the fast rate of NHEJ (81).

1.2.4 Utilization of CRISPR/Cas9 system to achieve precise gene integration

Even though CRISPR/Cas9 system was discovered as a bacterial and archaeal immune system, its powerful potential of generating site-specific cleavage is widely utilized in research studies. For the ease of manipulation by researchers, the two small RNA molecules, crRNA and tracrRNA, can be fused into one chimeric RNA molecule named as single guide RNA (sgRNA) (75, 89). In order to target a specific gene of interest, the guide sequence in sgRNA can be designed to be complementary to a portion of the target gene. As mentioned previously, NHEJ and HDR are two distinct pathways for DSB repair. Even though NHEJ is considered as error prone, it is an effective pathway to be utilized in knockout studies by generating frameshift or nonsense mutations (80). On the other hand, the HDR pathway can be manipulated to achieve precise deletion, substitution, or integration of gene of interest by providing a specifically designed donor sequence as a recombination template (80). In the donor template, it should contain the gene of interest flanked by homology arms (HA), which are sequences homologous to the genomic target (90).

A major problem for utilizing the HDR mediated gene integration is the low efficiency of this pathway. This becomes even more problematic for large inserts (90). Therefore, many strategies were used to boost the HDR efficiency. One study demonstrated that using longer HA can increase the frequency of homologous recombination (91). Furthermore, another idea is to inhibit the major competitor for HDR. A small molecule known as Scr7 can inhibit DNA ligase IV which is a major enzyme involved in the NHEJ pathway (92). In addition, there are also studies aiming for enriching the G2 phase of cell cycle since homologous recombination is restricted to G2 and S phase (93, 94).

1.3 Helper-Dependent Adenoviral vector

In gene therapy, one of the most important factors is the efficiency of gene transfer. Therefore, the nature of the vector used for gene delivery is essential. The ideal vector for a gene therapy study should have the correct tissue tropism while having a decent transgene delivery efficiency and expression in the target cell. Because of the properties of adenovirus, adenovirus derived vector was utilized in the first gene therapy trial for CF (48).

1.3.1 Adenovirus

Adenovirus (Ad) was first discovered in 1953 by isolating the cultures of adenoid tissue (95). It is a non-enveloped, double stranded DNA virus containing a 26 to 40 kb long linear DNA (96). Human Ad is a common pathogen that normally causes mild illness in most patients, yet it is very efficient at spreading the infection (97). It was shown that as little as 5 viral particles are sufficient for propagating the disease (98). Therefore, Ad is very dangerous in the vulnerable population such as people with immunosuppression (97). The identification of Ad's tumorgenicity in rodents triggered many studies on the nature of this virus (99). Out of more than 100 serotypes identified for Ad, Ad serotype 2 (Ad2) and Ad serotype 5 (Ad5) were the most studied (97).

The genome of Ad is characterized by the Inverted Terminal Repeats (ITR) at both ends of the genome which acts as the viral origin of replication. In addition, a packaging signal (ψ) is adjacent to the left ITR (97). The presence of ψ is essential for the encapsidation of viral genome into capsid. The genome of Ad can be categorized into early or late genes depending on when they are expressed in the infection cycle. The early regions are divided into E1A, E1B, E2, E3, and E4 (97). E1A plays an essential role of stimulating host mitogenic activities and the expression of

other viral genes (100). The major role of E1B is to interact with host p53 and prevent p53 induced apoptosis (101). E2, E3, E4 mediates viral replication, altering host immune response and cell signalling respectively (102, 103). After the activation of Major Late Promotor following viral DNA synthesis, late genes, from L1 to L5, are then expressed (97).

The Ad capsid is icosahedral shaped with a size of 70 – 100 nm (97). At each of the vertex of the icosahedron capsid, there is a protruding fiber protein which mediates the initial attachment step of the Ad infection cycle (104). Inside the capsid, the viral DNA is associated with various proteins. These proteins condense the DNA and facilitate the packaging process (105).

The tropism for this virus is the respiratory epithelia and it possesses efficient mechanism of viral gene delivery and expression (52). The entry of Ad5 into the target host cell relies on the Coxsackie and Adenovirus Receptor (CAR) which is expressed on the surface of many tissues (106). In the airway, CAR is expressed at the basolateral membrane of airway epithelial cells (6). Ad5 enters through receptor-mediated endocytosis by clathrin-coated vesicles and enters the cytoplasm by lysing the endosomal membrane of early endosome (107, 108). The virion travels to the nucleus through microtubules and releases the genetic material in the nucleus (109). Subsequently, the early genes are expressed which initiates the production of more virus particles.

1.3.2 Development of Adenoviral vector

Ad based vectors are one of the most commonly utilized vectors for transgene delivery into mammalian cells (110). It possesses numerous advantages as a gene therapy vector. For example, Ad is a non-integrative virus which eliminates the risk of insertional mutagenesis (111). In the First-Generation Adenoviral vector (FG-Ad), the E1 and E3 region are deleted to render the Ad

particle replication incompetent. As a replacement, a transgene expression cassette can be inserted in the E1 and E3 deleted regions to allow the viral particle to carry the gene of interest. Approximately 8 kb can be used for packaging the gene of interest for delivery (112). In some FG-Ads, only the E1 region is deleted and this leads to a smaller packaging capacity comparing to the E1/E3 deleted FG-Ads. Utilization of this type Ad vector achieved targeting both the proliferating and the non-proliferating cells with high efficiency of transgene delivery *in vivo* (113). However, the major disadvantage of this vector system is its high immunogenicity. After vector delivery, the cells receiving the transgene quickly became targeted by host immune reactions and the transgene can only be expressed transiently for approximately 2 to 3 weeks (114).

In order to improve upon the FG-Ad vector system, second-generation adenoviral vectors (SG-Ad) were developed. Based on the FG-Ad vectors, SG-Ad has E2 or E4 deleted in addition to the E1/E3 deletion. The goal is to further increase the packaging capacity of the vector and also limit the expression of viral genes that can trigger undesired immune reactions (115). In animal studies, SG-Ad had demonstrated diminished but not eliminated viral induced inflammation (116). The likely reason for this is low levels of expression of the late genes. The antigenic capsid proteins produced also contribute to triggering the host immune reaction (116).

The most current and advanced Ad based vector system is called helper-dependent adenoviral vector (HD-Ad). In HD-Ad, all of the viral components are deleted except for the ITR and ψ (117). This gives HD-Ad a very large packaging capacity compared to all of its predecessors. In addition, the deletion of all the viral genes makes this vector less immunogenic. This property was confirmed by animal study in which the researchers noted a significantly longer transgene expression post-HD-Ad delivery (118). One important note for synthesizing HD-Ad vector is to

make sure the genome size of HD-Ad will not be too small. Studies had shown that Ad genomes smaller than 27 kb are unstable and tend to recombine spontaneously to increase their size (119). Therefore, if the transgene inserted cannot meet the minimum size requirement, stuffer DNA sequences are required to meet this standard. Besides this function, the stuffer sequence was also determined to influence the stability, replication efficiency and *in vivo* activity (111, 120). Since HD-Ad is devoid of all the viral genes, it lacks the capability of replicating autonomously and requires the assistance of helper virus during the production phase (97). The downsides of this vector system are the complicated production and purification procedures and the inevitable helper virus contamination during HD-Ad production.

1.3.3 Production of HD-Ad

The most efficient method of HD-Ad production involves the usage of Cre/LoxP system, which was developed by Frank Graham and his co-worker in 1996 (117). Firstly, the HD-Ad genome is synthesized as a large plasmid DNA through series of cloning in bacteria. Within the plasmid DNA, it should contain the ITR, ψ, transgene expression cassette and the stuffer sequence. Subsequently, the HD-Ad plasmid genome is linearized through restriction digestion to expose the ITR at the two ends. The linearized HD-Ad genome will then be transfected into 293 cells expressing Cre recombinase and Ad E1 gene. At the same time, a modified FG-Ad, which acts as helper virus, is co-transfected into the 293Cre cell line as well. Within the helper virus's genome, its ψ is flanked by two loxP sites which are recognized by Cre. As a result, its ψ will be excised out and render the helper virus genome unpackageable. The role of helper virus is to express Ad viral genes for synthesizing all the necessary components to produce a viable Ad particle (except E1, which is expressed by the 293Cre cell line). Since only HD-Ad genome contains ψ, it will be packaged into the capsid produced instead of the helper genome (111, 117).

Over the past decade, this method of HD-Ad production had been improved to allow easier large scale amplification and decrease the helper virus contamination (121). The replacement of 293Cre by the 116 producer cell line allows the cell to be grown in suspension. This dual growth capability is a drastic improvement over the old system since the rescue and the amplification phase is most efficient with adherent cells while the large scale amplification is more efficient in suspension culture (121). In addition, the 116 producer cell line can express more Cre recombinase compared to the old 293Cre cells. This make the excision of helper virus ψ more complete and thereby decrease helper virus contamination (121).

1.4 Models for studying CF

Despite the promising results of preclinical studies in correcting CFTR channel using recombinant vectors, all the CF gene therapy trials had failed in terms of improving lung function of CF patients (6, 122). Gene transfer to the airway is actually more challenging than previously thought and one of the reasons for this is the models used in the preclinical studies are not very reflective of the conditions of CF patients (123). Therefore, better animal models are needed for testing gene therapy strategies against CF.

1.4.1 Pig as CF gene therapy model

Even though the mouse models have been used for studying CF physiology for more than fifteen years, many problems exist in this model (124). One of the major limitations of the mouse model is the lack of spontaneous airway infection and pancreatic disease found in CF patients (125). The reason for such a drastic difference is because the airway composition in mice is very different from human. Research had shown that mice has alternative non-CFTR ion channel that can compensate for the lack of CFTR expression (126). In addition, the submucosal gland that

expresses abundant CFTR in the human cartilaginous airway can only be found in proximal trachea of mice airway (127).

In contrast, pigs share a very similar airway composition as human (128). However, due to the technological limitations, CF pig models were not available until 2008 using somatic cell nucleus transfer of CF knockout primary fibroblasts (129). After the successful generation of CF pig model, it had been shown to display many similarities in terms of disease pathology in many organs such as the intestine, pancreas, liver and airway comparing with CF patients (6).

1.4.2 IPEC-J2 cell line

In this study, we will be using a cell line derived from pigs to test our gene therapy strategy. Since the strategy we developed is a novel approach, it is important to test it in *in vitro* settings before translating into *in vivo* models. The name of the cell line we will be using is called IPEC-J2. IPEC-J2 cells are intestinal columnar epithelial cells isolated from neonatal pig mid-jejunum in 1989 (130). It is most often been used to characterize the interaction between epithelial cells and enteric bacteria or viruses (130). Even though this is not a CF pig cell line, it is still valid to test our strategy in terms of transgene integration and expression using the IPEC-J2 cells because it contains the target site where we want to insert the transgenes and expresses the CAR receptor. The target site used in this study will be introduced in the next section.

1.5 *GGTA1* as CRISPR/Cas9 target site

An important factor for using the CRISPR/Cas9 system is to choose the target site for transgene insertion. It is very important to integrate the gene of interest in a safe harbour so no regular genes will be disrupted. Alternatively, it is also possible to insert the therapeutic gene at the

location of the mutated gene that we want to correct. In this study, the target site for transgene insertion was chosen to be a locus called *GGTA1*.

1.5.1 *GGTA1* as a safe harbour for transgene integration

The gene *GGTA1* encodes for α-1,3-galactosyltransferase (GGTA1) which functions as producing the carbohydrate galactose- α-1,3-galactose (α-1,3Gal) (131). In pigs, the vascular endothelium of all blood vessels expresses this residue (132). On the contrary, humans do not have this molecule expressed because the *GGTA1* gene was inactivated throughout the course of evolution. This gene is very problematic in the field of xenotransplantation from pigs to human. In fact, the carbohydrate produced by *GGTA1* is responsible for the hyperacute rejection of the graft since humans would produce a high level of anti-α-1,3Gal antibody upon exposure (133). Therefore, studies in the field of xenotransplantation aimed to knockout *GGTA1* to generate *GGTA1* deficient pigs (134, 135). This demonstrates that knocking out this gene would not have any detrimental effects on the pig model. In one study, the researcher successfully knocked-in the *DAF* gene into the *GGTA1* locus and detected DAF expression afterwards. This proves *GGTA1* to be a viable target for transgene integration (133).

1.6 Objectives and hypothesis for this study

Combining the information above, we developed a novel gene therapy strategy for CF. It involves targeting the *GGTA1* locus, a genome safe harbour, through the engineered CRISPR/Cas9 system and integrating a donor *CFTR* gene into the target site by utilizing the cellular homologous recombination pathway (Figure 1). This type of novel approach has to be examined pre-clinically *in vitro* and *in vivo* in a physiologically relevant model. In this project, we will determine the transgene integration efficiency and transgene *CFTR* expression in cultured IPEC-J2 pig cell line.

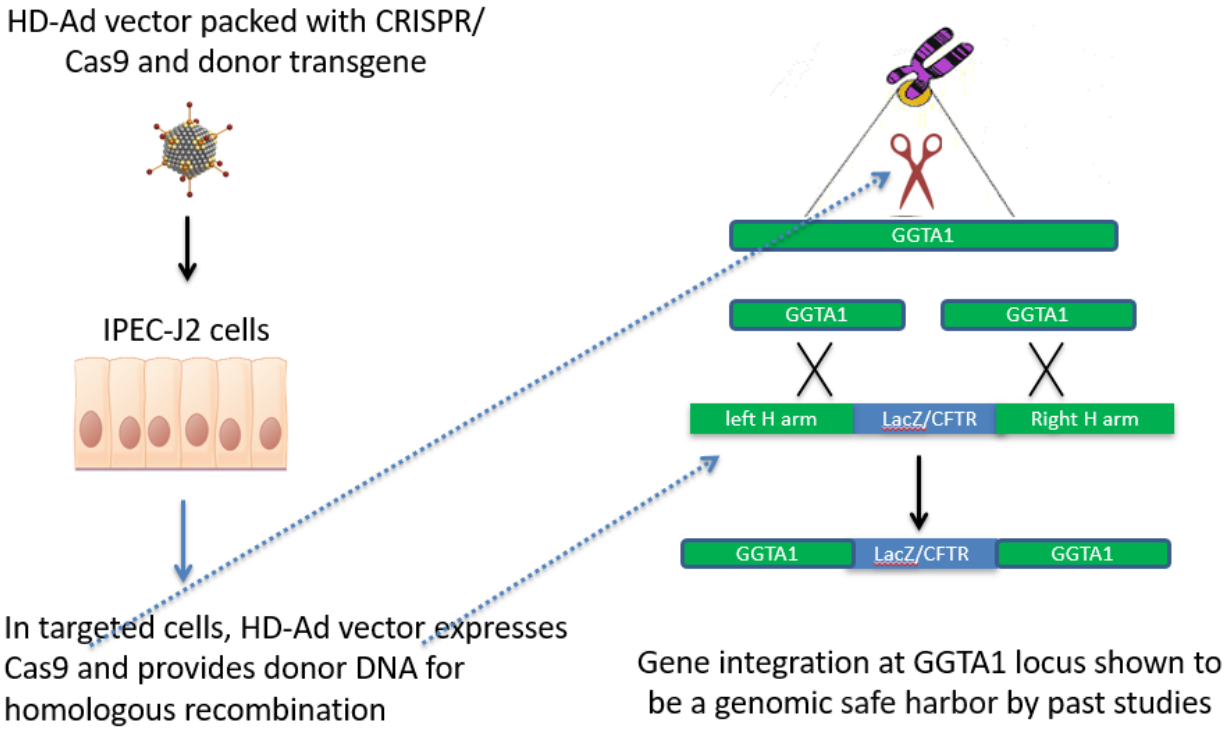


Figure 2. Overview of the gene therapy strategy used in this study.

This gene therapy approach consists of first synthesizing the HD-Ad vector that carries both the CRISPR/Cas9 system as well as the transgene to be delivered. Subsequently, this vector will be transduced into the target IPEC-J2 cells. Successful transduction will lead to expression of both the Cas9 endonuclease and the engineered sgRNA which will generate a double stranded break in the *GGTA1* locus. HDR mediated homologous recombination will allow the precise integration of the transgene expression cassette at the target site.

Hypothesis

We hypothesize that efficient transgene integration into a genome safe harbor of porcine epithelial cells can be achieved by delivering a CRISPR/Cas9 system and a donor expression cassette using HD-Ad vector. If this study can be successfully completed and proven to be

effective, it will be possible to proceed to *in vivo* experiments by testing in the CF pig which is known to be the best animal model for representing human CF symptoms.

The specific aims of this study are:

Aim 1. To examine integration efficiency of a reporter gene in cultured IPEC-J2 pig cell line delivered by HD-Ad vector.

Aim 2. To assess donor CFTR integration and expression in cultured IPEC-J2 pig cell line delivered by HD-Ad vector.

As a novel strategy, the essential objective is to determine transgene integration efficiency before proceeding with future studies. Therefore, an important objective of this study is to determine the transgene integration efficiency. To do this, a reporter gene that can be easily visualized inside the cells after integration is a suitable choice. The reporter *LacZ* was chosen for this purpose because its expression can be detected through a simple staining method. The *LacZ* gene encodes for the protein β-galactosidase which mediates the cleavage of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (136). The product formed by cleavage and dimerization is blue in color which leads to the characteristic blue color of *LacZ* positive cells. One drawback with this method is that it is hard to distinguish between the cells achieving stable integration versus cells only retaining the viral vector genome. To resolve this issue, the transduced cells will be passaged for multiple times before analyzing the integration efficiency. The viral vector cannot replicate so for each passage, the number of cells retaining the initial viral vector will be diluted out and this will not become a problem after multiple passages.

2. Materials and Methods

2.1 Basic molecular biology techniques

2.1.1 Restriction digestion

Restriction enzymes recognize different short DNA sequences known as the restriction site. Upon recognition, the enzymes will cleave the double stranded DNA within or at a constant distance away from the recognition site. The ends produced by restriction digestion may either be blunt ends or sticky ends with complementary overhang. This allows the fragments cleaved by the same enzyme to be readily re-annealed. Restriction digestion was accomplished by incubating the enzyme with target DNA under condition specified by the enzyme provider. In order to stop the reaction, usually heating the reaction to 65°C for 10 minutes is sufficient. However, this was subject to change depending on the nature of the enzyme. The inactivation procedure was specified by the provider as well. All the restriction enzymes used in this study were ordered from New England Biolabs.

2.1.2 Electrophoresis and DNA fragment purification

Agarose gel was prepared by adding 0.7-1% of agarose into Tris Acetic acid EDTA buffer (TAE) with RedSafe Nucleic Acid Staining Solution from FroggaBio. The amount of RedSafe added was total volume of the TAE buffer divided by 20000 since the stock RedSafe was 20000X. DNA samples were mixed with 10X DNA loading dye and compared with ThermoFisher Scientific 1Kb plus DNA ladder.

The electrophoresis apparatus was connected to a power supply and filled to the fill marker with 1x TAE buffer. The gel was placed in the apparatus and the DNA samples were loaded into the

wells of pre-made agarose gel. The electrodes were connected so that the negative electrode corresponded to the top of agarose gel while the positive electrode correspond to the bottom of agarose gel. The voltage was usually set to 100 – 120 V for DNA band resolution.

After the desired band can be clearly resolved, the desired fragment was cut out under long wavelength ultraviolet light. The gel slice was placed in Eppendorf tube and then isolated with QIAquick Gel Extraction Kit from QIAGEN. Steps for isolating the DNA fragment was followed with the protocol (QIAquick Spin Handbook) provided which was also available on www.qiagen.com.

2.1.3 Ligation

The purpose of ligation is to ligate DNA fragments into one piece through the formation of a phosphodiester bond between the 5' phosphate group and the 3' hydroxyl group. For cloning purposes, the ligase used in all the cloning experiments were from ThermoFisher Rapid DNA Ligation Kit. The vector and the insert were usually followed by a 1:3 ratio in the presence of ligation buffer and the reaction mixture was incubated at room temperature for 5 minutes. Upon completion, the reaction mixture was stored at 4°C until use.

2.1.4 Transformation

Stellar competent cells were obtained from Clontech and aliquoted to 100 µl per tube. They were stored at -80°C until use. The first step of transformation was to thaw the competent cells on ice. DNA to be transformed (usually a ligation mixture) was then added to the tube with competent cells. The mixture was incubated on ice for 20 minutes. Subsequently, the mixture was heat shocked at 42°C for 40 seconds and immediately placed on ice. 1 ml of SOC medium was added to the mixture and the cells were incubated at 37°C in a shaker for 45 minutes. Finally,

the cells were plated on 2x YT (16 g tryptone, 10 g yeast extract, 5 g sodium chloride and 15 g bacteriological agar in 1 L distilled water) plates with 100 µg/ml ampicillin and incubated at 37°C overnight without light exposure.

2.1.5 Plasmid isolation through Mini-prep

After transformation, colonies of competent cells transformed with the ligation mixture should be visible after overnight incubation. For mini-prep, colonies were picked from the plate and inoculated aseptically into 1.5 ml of 2x YT medium (same recipe as in 2.1.4 but without 15 g bacteriological agar). Enough aeration was an important factor for the growth of bacteria. The mixtures were then incubated at 37°C overnight to allow the bacteria to amplify. On the next day, the bacterial cells were harvested by spinning at 8000 rpm for 2 minutes. The supernatant after centrifugation was discarded and 150 µl of P1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA), P2 (0.2 M NaOH, 1% SDS) and P3 (3 M potassium acetate, pH 5.5) solutions were added in a stepwise manner. Between each step, the mixtures were inverted 10 times to mix properly. After the addition of P3, the mixtures were incubated on ice for 10 minutes and then centrifuged at 14000 rpm for 10 minutes at 4°C. After centrifugation, the supernatant of each mini-prep (~450 µl) was collected and 900 µl of 100% ethanol was added to each mini-prep supernatant. The mini-prep DNA can then be precipitated through centrifugation at 14000 rpm for 5 minutes at room temperature. Subsequently, the supernatant of each mini-prep was removed and the DNA pellets were washed with 1 ml 70% ethanol. Each mini-prep was centrifuged at 14000 rpm for 1 minute at room temperature. The ethanol solution was removed and the pellet was allowed to air dry at room temperature. The final step was to dissolve the DNA pellet with 40 µl TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) buffer. To verify if the isolated mini-prep DNA is the desired

construct, proper restriction enzyme digestion with addition of 1 µl 10 mg/ml RNaseA per reaction was performed.

2.1.6 Plasmid isolation through Midi-prep

The purpose of midi-prep is to isolate plasmid with higher yield and purity compared to mini-prep. After transformation, colonies were isolated and inoculated aseptically into 1.5 ml 2x YT media. These bacterial cultures were incubated at 37°C for 7 – 8 hours. Then, 300 µl of each of the bacterial culture was transferred into 100 ml 2x YT media containing ampicillin and incubated at 37°C overnight. On the next day, the bacterial cultures were poured into centrifuge bottles and spun at 4800 rpm for 10 minutes. The supernatant from each bottle was removed and 7 ml of P1, P2, and P3 were added sequentially. Between the addition of each solution, proper mixing was ensured. The mixtures were incubated on ice for 10 minutes similar to mini-prep. Next, each of the mixtures contained in the centrifuge bottle was spun at 4800 rpm for 10 minutes. The resulting supernatant was collected, and 13.5 ml isopropanol per reaction mixture was added. Following this step, the midi-prep DNA would be crudely precipitated by centrifugation again at 4800 rpm for 10 minutes. The pellets were resuspended in 400 µl TE buffer and transferred into Eppendorf tubes. 200 µl of 7.5 M ammonium acetate was added to the Eppendorf tubes. The tubes with midi-prep plasmids were incubated on ice for 5 minutes and centrifuged at 4°C for 5 minutes. The resulting supernatant was collected, then 800 µl 100% ethanol was added. After this step, the same procedure was followed as the mini-prep protocol except the final pellet would be dissolved in 400 µl TE instead of 40 µl.

Since RNA was also obtained and mixed with desired DNA construct, 10 µl of 10 mg/ml RNase A was added to the dissolved DNA + RNA mixture and incubated at 37°C for 1 hour to complete

the removal of RNA. To obtain the final purified DNA construct, the midi-prep mixtures post RNase A digestion were purified through phenol chloroform extraction by adding equal volume of phenol: chloroform: isoamyl alcohol (25:24:1). The mixture was vortexed to ensure proper mixing and two layers were observed. The mixtures were subjected to centrifugation at 14000 rpm for 2 minutes. The top aqueous layer was transferred into new Eppendorf tubes and equal volume of chloroform: isopropanol (29:1) was added to each tube. Each tube was centrifuged at 14000 rpm for 5 minutes at room temperature and the top aqueous layer was extracted into new Eppendorf tubes. 2x volume of 100% ethanol and 0.5x volume of 7.5 M ammonium acetate were added to the extracted aqueous layer and the same steps of ethanol precipitation and purification of plasmid DNA was followed as described earlier.

2.1.7 Genomic DNA extraction from adherent cell culture

The cells were first washed with PBS buffer, pH 7.4 2 times and then trypsinized. The cells were collected in 15 ml Falcon tubes and centrifuged at 800 rpm for 5 minutes. The supernatant was discarded and the cell pellet was washed 2 more times with PBS buffer, pH 7.4. To lyse the cells, the cell pellet was then resuspended in digestion buffer (100 mM NaCl, 10mM Tris-HCl pH 8, 25 mM EDTA, pH 8, 0.5% SDS and 0.1 mg/ml proteinase K) and digested overnight in 50°C with adequate shaking. On the next day, the cell lysate was phenol chloroform extracted as mentioned in 2.1.6, and subsequently treated with 10 µl 10 mg/ml RNase A at 37°C for 1 hour. To remove the RNase A, phenol chloroform extraction was performed again to obtain the final purified genomic DNA.

2.1.8 DNA sequencing

All the sequencing reactions were submitted to The Center for Applied Genomics facility at SickKids Hospital. For prepping the DNA to sequence, 200 – 300 ng of DNA sample dissolved in 7 µl of water was prepared. In addition, 50 ng of sequencing primer in 0.7 µl was added. The resulting mixtures were then submitted for sequencing.

2.1.9 Polymerase Chain Reaction (PCR)

The polymerase used for all PCR reactions were Phusion High-Fidelity DNA polymerase from New England Biolabs unless otherwise specified. The protocol for performing PCR using phusion can be found online (137). All the primers were synthesized by ACGT corporation through order submission.

2.2 Molecular cloning

2.2.1 Design the CRISPR/Cas9 plasmid

The sgRNA guide sequence was selected to be 5'-GAGAAAATAATGAATGTCAA-3' in the GGTA1 locus and was intended to be inserted into the pSpCas9(BB)-2A-Puro construct ordered from Addgene. The location of designed guide sequence insertion was immediately upstream of the sgRNA scaffold in the construct. A set of oligonucleotides were ordered which contains the sgRNA guide sequence. The nucleotide sequence of the two oligonucleotides were: 5'-AAACTTGACATT CATTATTTTCTC-3' and 5'-CACCGAGAAAATAATGAATGTCAA-3' (sgRNA-GGTA1 Fwd and Rev). These two oligonucleotides were annealed so it would be compatible with the overhangs created by BbsI digestion. The construct was digested with BbsI and the annealed oligonucleotide was inserted through ligation. After transformation of the ligation mixture, 18 colonies were picked and the mini-prep was performed to isolate the potential constructs. The

potential constructs were then sent for sequencing using the U6 primer with the sequence 5'-GAGGGCCTATTTCCCATGATTCC-3' (U6 Fwd). Constructs that had sgRNA guide sequence correctly inserted were retained. To ease the manipulation of later cloning steps, the entire CRISPR/Cas9 region containing the inserted sgRNA guide sequence was transferred into a commercially available plasmid named as pBSII-SK(+) (Addgene) which contains many restriction sites that can be used in cloning. This construct was named as pBSII-SK(+)-GGTA1 CRISPR.

2.2.2 Construction of pUbCLacZ-Neo GGTA1 plasmid

The first step of synthesizing this vector was to use PCR amplification to obtain the left and right HA required by HDR mediated repair. Genomic DNA from IPEC-J2 cells were extracted and used as the template for HA PCR amplification. For amplifying the left HA, primers 5'- CCAG CGGTGGCGGCCAGGTCGACGGTATCG-3' and 5'-GGAGGCCGAGCGGCCCCTAGTTATCAGCCAAGTC-3' (Infusion-L arm LacZ Fwd and Rev) were used. In the case of right HA, 5'-TCAGCTTTGCA CAAGGAATAGTCAACGAGC-3' and 5'-ATCCACTAGTTCTAGCTCTAAAGCTTCAGCCCAG-3' (Infusion-R arm LacZ Fwd and Rev) were used as primer for PCR. These unusually long PCR primers were designed to introduce a 15-base pair overlap that match with certain restriction digestion ends while amplifying the desired fragment. The 15-base pair overlap was used in the subsequent In-Fusion cloning technique. The resulting correct right HA PCR fragment was purified and inserted into pBSII-SK(+)-UbCLacZ-Neo downstream of the UbCLacZ-Neo gene using the In-Fusion HD Cloning Kit (Clontech Laboratories). The pBSII-SK(+)-UbCLacZ-Neo construct was a modified pBSII-SK(+) plasmid that contains the cDNA of *LacZ* reporter gene fused with neomycin resistant gene. The promoter for regulating the expression of this fused gene was the Ubiquitin C (UbC) promoter. To clone a desired piece into the vector, the steps followed was to first digest the vector and insert with restriction enzyme to create compatible ends, followed by ligation,

transformation, mini-prep isolation and verification. Between each successful cloning steps, the desired construct was purified through midi-prep isolation. The In-Fusion cloning method was used to fuse DNA fragments with linearized vector through recognition of 15 base pairs overlap at the ends of both insert and the vector. The pBSII-SK(+)-UbCLacZ-Neo vector was linearized using XbaI and ApaI, and the right HA was inserted following the protocol provided by In-Fusion HD Cloning Kit User Manual. After the incorporation of right HA, the left HA was inserted using the same In-Fusion cloning procedure but with NotI digestion instead of XbaI and ApaI. Subsequently, the *LacZ* expression cassette, flanked by the left and right HA, was transferred into pC4HSU-PN (modified from (138)) construct using SalI digestion. The pC4HSU-PN vector contains the ITR, ψ, and stuffer DNA required for HD-Ad synthesis. The final step of producing the pUbCLacZ-Neo GGTA1 vector was to insert the CRISPR/Cas9 component from the pBSII-SK(+)-GGTA1 CRISPR into the modified pC4HSU-PN plasmid with AscI digestion. After the successful completion of the pUbCLacZ-Neo GGTA1, restriction digestion analysis and sequencing of the key components were used to verify the completion of the construct.

2.2.3 Construction of pK18CFTR GGTA1 plasmid

The first step of synthesizing this construct was to insert a few useful restriction sites into the vector that would be utilized later. The backbone plasmid chosen for assembling the CFTR expression cassette was the pBSII-SK(+). A set of oligonucleotides with the sequences 5'- TAATG GTTTAAACGTGAGAGCGCTGTAC-3' and 5'-AGCGCTCTCACGTTTAAACCATTAGTA-3' (PmeI-AfeI-Insert Fwd and Rev) was used to insert a PmeI and AfeI restriction sites into pBSII-SK(+). These two oligonucleotides would anneal with each other and form compatible ends with KpnI restriction digestion overhangs. The annealed oligonucleotide pair was inserted into the KpnI digested pBSII-SK(+). With a similar approach, the pC4HSU-PNE (modified from (138)) plasmid

had an AscI site inserted through AgeI digestion using the annealed oligonucleotides 5'-CCGGCGTAGTAGGCGCGCCCA CAGG-3' and 5'- CCGGCCTGTGGGCGCGCCTACTACG-3' (AscI-Insert Fwd and Rev). A construct previously available in the lab was used for producing the CFTR expression cassette. The construct contains a 2.5 kb genomic sequence upstream of the human keratin 18 (*K18*) gene, K18 promoter plus the first intron, hCFTR cDNA (containing a silent mutation to eliminate cryptic splicing with the K18 intron), 3' untranslated region from *K18,* and a polyadenylation signal (139, 140). This sequence will be denoted as K18CFTR afterwards. The K18CFTR fragment was treated with MfeI digestion, removal of overhang by T4 DNA polymerase (New England Biolabs) and SalI digestion. For the modified pBSII-SK(+), it was digested with ApaI, blunted by T4 DNA polymerase, and digested with SalI. The two desired fragments were then ligated together. In order to insert the left HA and right HA, In-Fusion cloning was used. In the case of left HA, primers 5'-TTTAAACGTGAGAGCAGGTCGACGGTATCG-3' and 5'-ACAATTCGGTACAGCCCTAGTTATCAGCCAA GTC-3' (Infusion-L arm CFTR Fwd and Rev) were utilized. PCR amplification. The PCR product was fused with the K18CFTR containing construct at the AfeI restriction site inserted earlier. Similarly, the right HA was incorporated at the EcoRV restriction site using In-Fusion cloning with the primers 5'-ATCGATAAGCTTGATGCTCTAGAACTAGTGGATCCC-3' and 5'- CTGCAGGAATTCGATCTC CTCTATCCTACCTCTAAAGC-3' (Infusion-R arm CFTR Fwd and Rev). The assembled K18CFTR expression cassette and homology arms were cleaved out with PmeI and NotI digestion and the modified pC4HSU-PNE vector was digested with BamHI, blunted by T4 DNA polymerase, and digested with NotI. The fragment of K18CFTR expression cassette was inserted into the digested vector through ligation. The final step to complete the pK18CFTR-GGTA1 construct was to transfer the CRISPR/Cas9 system from pBSII-SK(+)-GGTA1 CRISPR to the K18CFTR inserted

pC4HSU-PNE using AscI digestion and ligation. Restriction digestion analysis and sequencing of key regions in the final construct were carried out to verify if the desired construct was made correctly.

2.3 Cell culture

2.3.1 Maintaining IPEC-J2 cells

The IPEC-J2 cells were grown in Dulbecco's Modified Eagle's Medium/F12 1:1 mixture (Gibco) with 5% Fetal Bovine Serum (FBS, from Wisent), 1% Insulin-Transferrin-Selenium (ThermoFisher Scientific), 1% 10 mg/ml penicillin streptomycin (Wisent), and 5 ng/ml human recombinant epidermal growth factor (ThermoFisher Scientific). This will be referred to as the IPEC-J2 medium afterwards. Cells were usually maintained in 10 cm tissue culture plates (Sarstedt) and passaged when the cells reach 90% confluency.

For each passaging, the cells were first washed with 5 ml Phosphate Buffered Saline (PBS), pH 7.4, two times to remove the residue media and serum. 5 ml 0.25% trypsin-EDTA (ThermoFisher Scientific) was added to the culture and removed by vacuum after the plate was wetted by trypsin. The cells were then incubated at 37°C for 5 minutes for the cells to detach from the bottom of the plate. Then 5 ml of the IPEC-J2 medium was added to resuspend the cells. Finally, 1 ml of suspended cells was added to 9 ml of IPEC-J2 medium in a new plate (1:10 dilution). The passaged cells were incubated at 37°C until the cells were grown to 90% confluency again.

2.3.2 Cell counting

Cell counting was usually required to prepare for transduction and transfection for the next day. The cells were first trypsinized and resuspended in 10 ml growth media. 10 µl of suspended cells were pipetted onto hematocytometer (Hausser Scientific) for cell counting. After the cells were

mounted, the hematocytometer was placed under microscope and number of cells in the five 4x4 squares were counted. During counting, the cells located at the top and right border were included while the cells located at the bottom and left border were excluded. The cell counts were summed up and divided by 5 to take the average. The number of cells in the original suspension is the average cell count multiplied by 10^4.

2.4 HD-Ad production

The protocol for HD-Ad production was modified from Ng *et al* (141).

2.4.1 Rescue

After the successful construction of the pUbCLacZ-Neo GGTA1 and pK18CFTR GGTA1, 10 µg of both plasmid were digested with PmeI to expose the terminal ITR. The digested plasmids were then purified through phenol chloroform extraction and dissolved in sterile TE buffer. The linearized plasmids were then transfected into 116 cells in 6 cm plates at 90% confluency in 3 ml of media. 116 cells were grown in Minimum Essential Medium (MEM, from Gibco) with 10% FBS (Wisent), 1% 10 mg/ml penicillin streptomycin (Wisent) and 100 µg/ml hygromycin B from BioShop. Transfection refers to introducing purified nucleic acids into eukaryotic cells. On the day before transfection, the cells were seeded with numbers according to the recommended seeding density provided by ThermoFisher Scientific website. On the day of transfection, jetPRIME transfection reagent from Polyplus Transfection was used. The 10 µg linearized viral plasmids were diluted in 400 µl of jetPRIME buffer and mixed through vortexing. Next, 20 µl of jetPRIME reagent was added to the DNA buffer mix and left at room temperature for 10 minutes. The transfection mixes were then added to 116 cells. Four hours after transfection, the 116 medium containing the transfection reagent was changed to fresh media. Subsequently, 5

multiplicity of infection (MOI) of NG163 helper virus was added into the transfected cells. 1 MOI equals to 1 active infectious particle per cell. After 48 hours of helper virus addition, some cells became detached and the remaining attached cells became round in shape. This is called cytopathic effect (CPE) which means the infection was successful.

2.4.2 Amplification

After the cells had entered CPE, the cells were viewed under fluorescent microscope to check for GFP expression. The cells were then harvested and collected in 15 ml Falcon tubes (Sarstedt). This was the P0 generation, and the collected cells were frozen in -80°C with the addition of 4% sucrose until use. Similar to the previous step, 116 cells were again seeded in 6 cm plates one night before transduction. Transduction refers to introduction of foreign DNA by using a virus or viral vector. On the next day, the frozen harvest cells were thawed in a 37°C water bath and then frozen in dry ice. The freeze/thaw cycle was repeated 2 more times before use. This step was meant to break open the cells to release the virus particles. After freeze/thaw, media was removed from the seeded cells and 1 ml of the 3 ml cell lysate was added. Furthermore, 2 MOI of helper virus was added to each plate of transduced cells. The cells were incubated at 37°C for 1 hour and gently agitated every 10 minutes after the virus addition. 2 ml of fresh media was added to each plate of transduced cells after 1 hour of virus addition and the cells were incubated at 37°C for 48 hours.

Forty-eight hours post-transduction, the cells were in CPE and GFP fluorescence was examined. The cells were then harvested as P1 generation in 15 ml Falcon tubes. P1 cells were stored in -80°C with the addition of 4% sucrose until use. Similar steps were followed for the generation P2, P3 and P4. However, in P2, the 116 cells were grown in 10 cm plates. In P3, 116 cells were

grown in 15 cm plates. As for P4, 116 cells were grown in eight 15 cm plates. From P1 to P2, 1 ml of cell lysate was added. From P2 to P3, 2 ml of cell lysate were added. Finally, all of P3 lysate were added during P4 transduction.

2.4.3 Large scale amplification

To prepare for large scale amplification, thirteen 15 cm plates of 116 cells at 90% were cultured. Three liters of growth media (MEM + 10% FBS + 1% penicillin streptomycin + 100 µg/ml hygromycin) and 1.5 L of virus media (MEM + 5% FBS + 1% penicillin streptomycin + 100 µg/ml hygromycin) were prepared and filtered. The thirteen 15 cm plates were resuspended and added to a 3 L spinner flask (Bellco biotechnology) containing 1 L growth media. The spinner was set at 75 rpm and the cells were allowed to grow overnight. Over the next 3 days, 500 ml, 500 ml and 1 L of growth media was added to the 3 L spinner flask for the cells to amplify. One day after the addition of all the growth media, the cells were poured in nine 500 ml centrifuge bottles and were centrifuged at 2000 rpm for 10 minutes under 4°C. During centrifugation, the virus media were poured into the spinner flask and warmed up to 37°C. After centrifugation, 1 L of the old growth media was saved and 20 ml was used to resuspend all the cell pellets. The resuspended cells were then transferred into a 250 ml spinner flask. P4 cell lysate after 3 cycles of freeze/thaw was added into the 250 ml spinner flask, together with 1 MOI of helper virus. This mixture was stirred at 75 rpm for 2 hours under 37°C incubation. After incubation, the mixture was transferred to the 3 L spinner flask containing the virus media. Finally, 460 ml of saved old growth media from centrifugation was added to the 3 L spinner flask and the final mixture was incubated at 37°C for 3 days.

2.4.4 Orcein inclusion body staining

Orcein was used to stain the inclusion body of cells which is the location of virus inside an infected cell. The protocol is adapted from Graham *et al* (142). Five mililiters of cells were collected from the 3 L spinner flask and was centrifuged at 1000 rpm for 10 minutes under room temperature. The supernatant was discarded and the cell pellet was resuspended in 0.5 ml of 1% sodium citrate. The resuspended cells were incubated at room temperature for 10 minutes and then fixed using 0.5 ml of Carnoy's Fixative (3:1 methanol to glacial acetic acid) for 10 minutes. After fixing, the cells were resuspended in 1 ml of 1% sodium citrate and then centrifuged at 1000 rpm for 10 minutes at room temperature. The resulting supernatant was removed and the cell pellet was resuspended in 150 µl of 1% sodium citrate. A drop of fixed, resuspended cells was added onto a glass slide (GoldLine microscope slides from VMR) and dried at room temperature for 1 hour. When the drop of cells was thoroughly dried, one drop of Orcein stain (2% Orcein from Sigma Aldrich in 50% glacial acetic acid) was added on top of the dried cells and a cover slip (VMR micro cover glass) was applied. The stained cells were then examined under a light microscope.

2.4.5 Lyse cells

After three days of infection and incubation, the cells were harvested by pouring into six 500 ml centrifuge bottles and were centrifuged at 2000 rpm for 10 minutes under 4°C. The resulting cell pellets were resuspended in 17 ml of 10 mM Tris, pH = 8.0. 5% sodium deoxycholate was prepared by adding 1 g of sodium deoxycholate powder into 2 ml of 10 mM Tris, pH = 8.0. To lyse the cells, 1.5 ml of 5% sodium deoxycholate was added to the harvested cells and was incubated at room temperature for 30 minutes on a tube shaker. Ten microliters of 250 U/µl

benzonase DNase from EMD Millipore and magnesium chloride with the final concentration of 2 mM were added to the cell lysate from previous step. Similar to last step, the mixture was incubated at room temperature for 30 minutes with shaking. The mixture was then centrifuged at 5500 rpm for 15 minutes at 4°C to collect the cellular debris to the bottom of the tube.

2.4.6 Virus purification

For virus purification, 4 tubes of cesium chloride gradient were prepared. For each gradient, 0.5 ml of 1.5 g/ml CsCl, 3 ml of 1.35 g/ml CsCl, 3 ml of 1.25 g/ml CsCl, and approximately 5 ml of cell lysate containing desired virus were slowly and sequentially added to a 14 x 95 mm Beckman ultracentrifuge tube. The interface between 1.25 g/ml CsCl and 1.35 g/ml CsCl was marked using a black marker. The four centrifuge tubes were carefully balanced and placed in the ultracentrifuge for spinning at 35000 rpm for 1 hour at 4°C. Following the completion of centrifugation, the white band located at the marked interface of each centrifuge tubes were drawn using a 3 ml syringe and a 20 guage needle. The virus band were collected into a 15 ml Falcon tube.

Two centrifuge tubes were each loaded with 7 ml of 1.35 g/ml CsCl. The virus band collected were loaded into the 1.35 g/ml CsCl centrifuge tube and topped up to 2 mm from the top of tube with 10 mM Tris, pH 8.0. As for counter balance, 10 mM Tris, pH 8.0 was loaded to the other tube with 1.35 g/mL CsCl. The two centrifuge tubes were spun at 35000 rpm overnight under 4°C.

The virus band after centrifugation was collected in a similar manner and was injected into a 3 ml 10K MW-cutoff Slide-A-Lyzer cassette. The virus was dialyzed in 500 ml 10 mM Tris, pH 8.0 for 24 hours. The 10 mM Tris, pH 8.0 was changed to fresh buffer 2 times during this 24-hour

dialysis. Next, the dialyzed sample was withdrawn from the cassette using a 3 ml syringe and a 18 guage needle. The virus withdrawn were collected in Eppendorf tubes and 2.5 µl was saved for spectrophotometer quantification. Finally, 1/10 volume of 100% glycerol was added to the dialyzed virus and was frozen at -80°C until use.

2.4.7 Spectrophotometer quantification

The saved 2.5 µL virus was diluted 20 times in 47.5 µl distilled water. The diluted sample was read by a spectrophotometer and the OD reading at 260 nm was obtained. Final virus particle number = OD_{260} x dilution factor x 10^{12}.

2.5 Assessing CRISPR cleavage efficiency

2.5.1 T7E1 assay

The purpose of T7E1 assay is to determine the efficiency of the engineered CRISPR/Cas9 is at cleaving the designed target. IPEC-J2 cells were cultured in 6 well plates until 70% confluency. Two wells of cells were transduced with 10 MOI and 50 MOI of HD-Ad UbCLacZ-Neo GGTA1 virus. To compare the cleavage efficiency after virus transduction, another 2 wells of cells received the 2 µg and 3 µg (recommended dosage for transfection of cells grown in 6 well plates) of viral plasmid. As a control for both experiments, the final 2 wells of cells received neither transfection nor transduction. Three days after transfection/transduction, the cells from all the wells were trypsinized and collected by centrifugation at 800 rpm for 5 minutes. The cells were then lysed in 50 µl Cell Lysis Buffer with 2 µl Protein Degrader from the GeneArt Genomic Cleavage Detection Kit (Life Technology). Approximately 350 base pairs in length which incorporated the target site for cleavage. This was amplified through PCR using the follow primer sequences: 5'-ACAACGGCAACTCTCTGGAATGC-3' (T7E1-Fwd) as forward primer and 5'-GCACTCCTTAGCGCT

CGTTG ACTA-3' (T7E1- Rev) as reverse primer. PCR reactions were carried out using the AmpliTaq Gold 360 Master Mix from the same cleavage detection kit. After the PCR products were obtained, 5 µL of each PCR product was added into 1 µl 10X NEBuffer 2.1, 3 µl of distilled water. The PCR products were then denatured at 95°C for 5 minutes and re-annealed by slowly cooling off the mixture to room temperature. 1 µl of T7 endonuclease I (New England Biolabs) was added to re-annealed PCR products and the reaction mixture was incubated at 37°C for 1 hour to complete the cleavage. Finally, all the samples were run on 2% agarose gel at 110 V to separate the bands and a gel image was taken using Alpha Innotech Fluorchem AlphaImager.

2.5.2 ImageJ analysis

The software ImageJ was used to analyze the density of different bands on the gel image. In order to quantify the cleavage efficiency, the density of the two lower bands representing the cleaved fragment were compared with the upper undigested PCR product. Rectangular selection tool was used to outline the area of bands in a single lane and this step was repeated for all the lanes. During this process, it was ensured that each rectangular selection covered the same amount of area in outlining different bands. The outlined areas where then plotted using the plot lane tool and peaks were observed where the bands appeared. Next, the straight line tool was selected to close the bottom of the peak and the wand tool was used to highlight the enclosed peaks. After all the peaks were enclosed and selected, label peaks option was used and the area for all the selected peaks was obtained. To calculate cleavage efficiency for each lane, the formulas for % cleavage is:

$$\% \text{ cleavage} = \frac{\sum Intensities\ of\ cleaved\ bands}{\sum Intensities\ of\ cleaved\ bands\ and\ parental\ band} \times 100\% \quad (143).$$

2.6 Assessing HD-Ad UbCLacZ-Neo GGTA1 integration efficiency

2.6.1 Verifying the activity of HD-Ad UbCLacZ-Neo GGTA1 vector

After the viral titer was quantified, another important factor to verify is if the virus can achieve efficient transgene delivery to the target cell. Therefore, 2 MOI, 10 MOI and 50 MOI of newly synthesized HD-Ad UbCLacZ-Neo GGTA1 vector was transduced into IPEC-J2 cells. 3 days post transduction, the GFP expression (representing the CRISPR expression) was observed under fluorescent microscope. In addition, the *LacZ* expression was examined through a method called β-galactosidase staining. Cells that did not receive any vectors were used as negative control.

2.6.2 β-galactosidase staining

To perform β-galactosidase staining, the cells in 6 well plate to be stained were washed 3 times in 1 ml PBS, pH 8.0. The cells were then fixed in 1 ml 0.5% glutaraldehyde in PBS, pH 8.0 for 15 minutes. After fixing, the cells were again washed for 3 times in 1 ml PBS and 1 ml of β-galactosidase staining solution per well was added (300 µl 40 mg/ml X-Gal, 16 µl $MgCl_2$, 600 µl 100 mM K-ferricynaide, 600 µl 100 mM K-ferrocyanide and 10.5 ml PBS, pH 8.0). The cells were then incubated at 37°C for 3 hours. To stop the reaction, the β-galactosidase staining solution was removed and the cells were washed with PBS, pH 8.0 three times. Finally, the *LacZ* positive blue cells were observed under a light microscope.

2.6.3 Determining *LacZ* integration efficiency

To determine how efficient this novel strategy is at inducing precise transgene integration, IPEC-J2 cells were transduced with 5 MOI, 20 MOI and 50 MOI of HD-Ad UbCLacZ-Neo GGTA1 vector. As a control, IPEC-J2 cells were also transduced with negative control 20 MOI HD-Ad CMVGFP (no *LacZ* gene) and experimental control 20 MOI HD-Ad UbCLacZ-Neo no CRISPR (no HA and no

CRISPR). The transduction was done in duplicates so in total, there were six samples that received transduction. The samples were grown in 6 well plate. The purpose of the duplicate was to divide the experiments into two sets. The first set was used for passaging while the second set was used for β-galactosidase staining to assess the expression of *LacZ* reporter.

All samples were passaged for 8 generations with 1:7 dilution. After the completion of each passage, the second set was stained for *LacZ* expression. The β-galactosidase staining images from the first generation was used to determine the transduction efficiency for each transduction except for the HD-Ad CMVGFP transduction. After 8 passages, the final β-galactosidase staining images were used to assess *LacZ* integration efficiency. For each experimental sample, 9 pictures of β-galactosidase stained cells under 20x objective were taken randomly. For each image, number of *LacZ* positive cells and total number of cells in that image were recorded. The formula used for calculating integration efficiency was:

$$\% \text{ integration} = \frac{\sum LacZ\ positive\ cells\ from\ all\ 9\ images}{(\sum total\ number\ of\ cells\ from\ all\ 9\ images \times transduction\ efficiency)} \times 100\%.$$

2.6.4 Junctional PCR analysis

Junctional PCR analysis was used to verify if the transgene was integrated at the correct locus. A set of primers was designed so that the forward primer will anneal onto the pig genomic DNA and upstream of the target locus. The reverse primer was intended to anneal onto the promoter of transgene integrated. The forward primer sequence used was 5'-AATGTGGACTAACACTGACC TTCC-3' (Junctional PCR-L-LacZ Fwd) which would bind approximately 3500 base pairs upstream of the *GGTA1* sgRNA target sequence. The reverse primer sequence designed was 5'-AAGGCCGAGTCTTATGAGCAG-3' (Junctional PCR-L-LacZ Rev). After 8 passages of transduced IPEC-J2 cells mentioned previously, genomic DNA were extracted and PCR amplification using

the designed primers were carried out. Subsequently, 5 µl of the PCR products were run on a 0.7% agarose gel and visualized under UV.

2.6.5 CRISPR off-target analysis

Potential off-target sites of the sgRNA guide sequence we designed were checked using the online tool crispr.mit.edu. A list of potential off-target sites was predicted and two sites (the top predicted from outside of a gene and the top predicted inside a gene) were picked. The location of the chosen off-target sites were determined using the UCSC genome browser. Similar approach to the junctional PCR analysis were carried out too see if there are undesired integration at incorrect locus. For the top off-target site from outside of any genes, the reverse primer used for detecting integration was the same as the reverse primer used in 2.6.4. The forward primer sequence was redesigned: 5'-GCAGTTCACTCAGGCAATTTC-3' (P1). In addition, a reverse primer, with the sequence of 5'-GTAGAGACCTTTTCTTCCCCATG-3' (P3), was designed to ensure the forward primer used is capable at annealing onto the correct location.

In the case of the top off-target inside a gene, the forward primer used to verify the presence of incorrect transgene integration was 5'-CCCAGCAAATGGATAATAGTATTGG-3' (P1'). The reverse primer used was the same as the reverse primer from 2.6.4. In addition, for a similar purpose mentioned in the previous paragraph, a reverse primer was designed with the sequence of 5'-GCCGTTGAATTTTAGACCTGGC-3' (P3'). PCR amplification was performed for both predicted off-target sites. The template used for amplification was genomic DNA extracted from IPEC-J2 cells transduced with 20 MOI HD-Ad UbCLacZ-Neo GGTA1. Genomic DNA from IPEC-J2 cells transduced with HD-Ad UbCLacZ-Neo without CRISPR element and HA was used as negative

control. The resulting PCR products after amplification were resolved on a 0.7% agarose gel and visualized under UV.

2.6.6 Single cell culturing

An alternative method to verify the integration efficiency obtained is through single cell culturing. After the transduction experiment in 2.6.3 was passed 8 times, a portion of the cells transduced with 20 MOI HD-Ad UbCLacZ-Neo no CRISPR and HD-Ad UbCLacZ-Neo GGTA1 were harvested and diluted to 1 cell per 100 µl using modified IPEC-J2 growth media (20% FBS instead of 5% FBS). The diluted cells were then pipetted into 96 well plates with each well receiving 100 µl of diluted cells. After dilution, the cells were incubated in 37°C for 20 days and all the wells that received a single cell during dilution and plating were marked. After 20 days, all the single cell derived colonies were stained for β-galactosidase activity and the colonies in which all the cells stained blue were counted as successful *LacZ* integration. The % integration efficiency was calculated as

$$\% \text{ integration} = \frac{number\ of\ colonies\ that\ all\ the\ cells\ are\ Lac\ positive}{total\ number\ of\ single\ cell\ colonies} \times 100\%$$

2.6.7 Integration efficiency enhancement

To determine if Scr7 has any effect on increasing the integration efficiency, IPEC-J2 cells were transduced with 20 MOI of HD-Ad UbCLacZ-Neo GGTA1 and treated with 0.1 µM and 1 µM Scr7 20 hours post-transduction. As control, IPEC-J2 cells were transduced with HD-Ad C4HSU empty vector and 20 MOI of HD-Ad UbCLacZ-Neo GGTA1 without Scr7 treatment. The transduced cells were passaged 4 days post-transduction and 3 days for each subsequent generation. In total, the cells were passaged 6 times before integration efficiency of each transduction were analyzed. The method for calculating integration efficiency was the same as mentioned in 2.6.3.

2.7 Determining the expression of hCFTR post HD-Ad delivery

2.7.1 IPEC-J2 transduction and passaging

IPEC-J2 cells were transduced with HD-Ad C4HSU (empty vector), HD-Ad K18CFTR (no CRISPR and HA), 20 MOI and 50 MOI of HD-Ad K18CFTR GGTA1 in 6 well plates. 4 days after integration, the cells were passaged in 1:7 dilution. After the first passage, the cells were allowed to grow for 3 days before the next passage.

2.7.2 Junctional PCR analysis

Genomic DNA samples from 20 MOI of HD-Ad pC4HSU, 20 MOI of HD-Ad UbCLacZ no CRISPR, 20 and 50 MOI of HD-Ad UbCLacZ-Neo GGTA1 were extracted to use as template for junctional PCR analysis. The primer set 5'- AATGTGGACTAACACTGACCTTCC-3' and 5'- GTGGAGTCAACAAGCTA TGTACTGC-3' (Junctional PCR-L-CFTR Fwd and Rev) were used for left arm junctional PCR amplification and the primer set 5'-CCAACAGTCCCATCCCTGATC-3' and 5'-GAAGTGGCTTCACAA AGGCAGTG-3' (Junctional PCR-R-CFTR Fwd and Rev) was used for right arm junctional PCR amplification. Five microliters of the resulting PCR products were loaded onto 0.7% agarose gel and visualized under UV.

2.7.3 Measuring hCFTR mRNA through RT-qPCR

IPEC-J2 cells were transduced with 20 MOI of HD-Ad C4HSU (empty vector), 20 MOI of HD-Ad K18CFTR no CRISPR, 20 MOI of HD-Ad K18CFTR GGTA1 and 50 MOI of HD-Ad K18CFTR GGTA1. The cells were harvested at 3 days, 5 passages and 9 passages post-transduction and the total RNA was extracted using the PureLink RNA Mini Kit (Life Technology) following the corresponding protocol. After RNA extraction, reverse transcription of all the RNA was performed with the SuperScript VILO Master Mix (Thermofisher Scientific) and its corresponding

protocol. The reverse transcribed RNA molecules were used for qPCR analysis to determine CFTR mRNA expression. Primer set 5'-CCTGAGTCCTGTCCTTTCTC-3' and 5'- CGCTGTCTGTATCCTTTCC TC-3' (qPCR-CFTR Fwd and Rev) was used for detecting CFTR transgene mRNA level and the primer set 5'- GTTCGACAGACAGCCGTGTG-3' and 5'- ATGGCGACAATGTCCACTTTGC-3' (qPCR-GAPDH Fwd and Rev) was used for detecting GAPDH mRNA level which acts as internal control. CT values were recorded and relative expression of each sample was calculated as:

Relative expression = $2^{-\Delta\Delta Ct}$ in which $-\Delta\Delta Ct = [(CT_{treatment} - CT_{internal\ control}) - (CT_{standard} - CT_{internal\ control})]$

In this experiment, the standard used in this calculation is the CT value of HD-Ad K18CFTR no CRISPR of the corresponding passage. The resulting relative expression were plotted against the vector and dosage used as treatment.

2.7.4 Measuring hCFTR protein expression through Western blot

In this experiment, a different cell line called IB3 was used. The IB3 cell line is a transformed human CF bronchial cell line and it contains two copies of mutant CFTR gene. One copy carries the ΔF508 mutant *CFTR* gene while the other copy contains W1282X (Nonsense mutation at tryptophan position 1282) mutation (144). The IB3 cells were cultured in DMEM +10% FBS medium until 90% confluency. The cells were then seeded and transduction was performed using 20 MOI of HD-Ad C4HSU, 20 MOI of HD-Ad K18CFTR GGTA1 and 50 MOI of HD-Ad K18CFTR GGTA1 vectors on the next day. The cells were collected 2 days post-transduction for Western blot sample preparation. As for positive control, the 8-3-7 inducible cell line for strong CFTR expression was used (145). The 8-3-7 cells were generated by the work of previous lab members. This modified cell line was derived from IB3 cells described above. Its generation consists of two

steps: transfection of reverse tetracycline-inducible transcription activator (rtTA) expression cassette and transfection of CFTR expression cassette with tetracycline response element. Upon induction with doxycycline (derivative of tetracycline), rtTA will be expressed and turns on CFTR expression by binding to tetracycline response element (145). The 8-3-7 cells were induced with 1 μg/ml of doxycycline and the samples were collected 24 hours post-induction. All the samples were first washed with ice cold PBS to remove the residue culture media and were subsequently scraped off using cell scraper. The cell suspension was centrifuged at 14000 rpm for 20 seconds at 4°C and the supernatant was removed. The cell pellets were resuspended in ice cold RIPA buffer (1% Triton X-100, 0.1% SDS, 150 mM NaCl, 20 mM Tris HCl and 0.5% deoxycholate) containing protease inhibitor (cOmplete Protease Inhibitor Cocktail from Roche) and each sample were lysed on ice for 20 minutes. The next step was to centrifuge each sample at 14000 rpm for 20 minutes at 4°C. The supernatant of each sample was collected and 5X sample buffer was added and thoroughly mixed. The resulting mixture was incubated at room temperature for 10 minutes before storing at -80°C until use.

Polyacrylamide gel with 4% stacking gel and 6% separating gel was prepared and placed in running cassette. One liter of running buffer was prepared and poured into the Western blot running tank so the entire gel was submerged by running buffer (25 mM Tris Base, 0.2 M glycine, 0.1% SDS). Five microliters of PageRuler Plus Prestainied Protein Ladder (ThermoFisher Scientific) and 100 μg of total protein sample from 20 MOI HD-Ad C4HSU, 20 MOI, and 50 MOI of HD-Ad K18CFTR GGTA1 transduced cells were loaded into the wells of polyacrylamide gel. Samples from doxycycline induced 8-3-7 cells and without induction were loaded less. Only 50 μg of total protein samples were loaded. The samples were run at 110 V for 90 minutes until the dye front approaches to the bottom of the polyacrylamide gel.

Transfer buffer (25 mM Tris Base, 0.2 M glycine, 20% methanol, and 0.01% SDS) was prepared and chilled at 4°C before use. Once the sample running was complete, the polyacrylamide gel was removed and submerged in transfer buffer for 10 minutes. In the meantime, a nitrocellulose membrane was submerged in transfer buffer for 10 minutes as well. Before setting up the transfer cassette, two pieces of sponge and filter papers were wet in transfer buffer and the transfer cassette was assembled with the following order: cathode, sponge, filter paper, polyacrylamide gel, nitrocellulose membrane, filter paper, sponge, and anode. The transfer cassette was placed in a running tank and was fully submerged by transfer buffer. The entire running tank was placed on ice to reduce the temperature during transferring. Protein samples were transferred onto nitrocellulose membrane at 100 V for 90 minutes.

After protein transfer, the membrane was cut into two portions. Tris-Buffered Saline with 0.1% Tween 20 (TBST) was prepared from diluting 10 X TBS stock solution. The upper portion contains the CFTR protein was blocked in 5% milk (0.1 g of skim milk powder in 10 ml TBST) at room temperature for 2 hours. The lower portion containing GAPDH was blocked in 5% Bovine Serum Albumin (BSA) (0.1 g Bovine Serum Albumin powder in 10 ml TBST) at room temperature for 2 hours. The blocking solutions were removed and primary mouse anti-CFTR antibody (MAB1660 from R&D Systems) was diluted 1000 times in 5% milk. The anti-CFTR primary antibody was added to the upper portion of the membrane. Similarly, primary mouse anti-GAPDH antibody (mAbcam 9484 from abcam) was diluted 10000 times in 5% BSA and added to the lower portion of the membrane. The upper and lower portions of membrane was incubated in the corresponding primary antibody solution overnight at 4°C.

On the next day, the primary antibody solution was removed and both portions of the membrane were washed in 10 ml TBST 10 minutes for 5 times. Then, goat anti-mouse IgG

secondary antibody conjugated to horseradish peroxidase was diluted in 5% milk (for upper membrane) and 5% BSA (lower membrane) 3000 times. The secondary antibody solutions were added to the corresponding portion of the membrane and incubated at room temperature for 2 hours. After secondary antibody incubation, the upper and the lower portions of the membrane were washed with 10 ml TBST 10 minutes for 5 times. The upper and the lower membrane were then dipped in ECL (Western Lightning Plus-ECL from PerkinElmer) reagents (equal volume of solution A and solution B) and placed in film cassette (Kodak BioMax Cassette). A piece of film (CL-X Posure Film from ThermoFisher Scientific) was placed on top of the membrane. It was developed for 10 minutes with the upper membrane and 1 second with the lower membrane in the dark room. The resulting film was scanned and the region of interest was cropped.

Table 1. Plasmid constructs used in this study

Plasmid	Details
pSpCas9(BB)-2A-Puro	9289 bp, contains sgRNA scaffold and Cas9 gene linked to eGFP through T2A sequence
pBSII-SK(+)	2961 bp, contains Amp resistance gene and multiple cloning site
pBSII-SK(+)-GGTA1 CRISPR	9621 bp, modified pBSII-SK(+) with CRISPR/Cas9 system inserted into multiple cloning sites
pBSII-SK(+)-UbCLacZ-Neo	9366 bp, modified pBSII-SK(+) with UbCLacZ-Neo transgene inserted into multiple cloning sites
pC4HSU-PN	16278 bp, backbone plasmid for HD-Ad synthesis, contains ITR and ψ
pC4HSU-PNE	9544 bp, backbone plasmid for HD-Ad synthesis, contains ITR and ψ
pUbCLacZ-Neo GGTA1	33165 bp, contains CRISPR/Cas9 system, LacZ-Neo reporter gene flanked by left and right homology arms under the control of ubiquitin c promotor, Amp resistance gene, ITR and ψ
pK18CFTR GGTA1	30289 bp, contains CRISPR/Cas9 system, hCFTR transgene flanked by left and right homology arms under the control of keratin 18 promotor, Amp resistance gene, ITR and ψ

Table 2. Oligonucleotide primers used in this study.

Primer	Sequence (5'-3')	Purpose
sgRNA-GGTA1 Fwd	AAACTTGACATTCATTATTTTCTC	Forward primer of inserting guide sequence into sgRNA scaffold
sgRNA-GGTA1 Rev	CACCGAGAAAATAATGAATGTCAA	Reverse primer of inserting guide sequence into sgRNA scaffold
U6 Fwd	GAGGGCCTATTTCCCATGATTCC	sgRNA guide sequence sequencing
Infusion-L arm LacZ Fwd	CCAGCGGTGGCGGCCAGGTCGACGGTATCG	Forward primer for inserting left homology arm flanking LacZ-Neo reporter through In-Fusion
Infusion-L arm LacZ Rev	GGAGGCCGAGCGGCCCCTAGTTATCAGCCAAGTC	Reverse primer for inserting left homology arm flanking LacZ-Neo reporter through In-Fusion
Infusion-R arm LacZ Fwd	TCAGCTTTGCACAAGGAATAGTCAACGAGC	Forward primer for inserting right homology arm flanking LacZ-Neo reporter through In-Fusion
Infusion-R arm LacZ Rev	ATCCACTAGTTCTAGCTCTAAAGCTTCAGCCCAG	Reverse primer for inserting right homology arm flanking LacZ-Neo reporter through In-Fusion
PmeI-AfeI-Insert Fwd	TAATGGTTTAAACGTGAGAGCGCTGTAC	Forward primer for inserting PmeI and AfeI restriction site
PmeI-AfeI-Insert Rev	AGCGCTCTCACGTTTAAACCATTAGTA	Reverse primer for inserting PmeI and AfeI restriction site
AscI-Insert Fwd	CCGGCGTAGTAGGCGCGCCCACAGG	Forward primer for inserting AscI site
AscI-Insert Rev	CCGGCCTGTGGGCGCGCCTACTACG	Reverse primer for inserting AscI site
Infusion-L arm CFTR Fwd	TTTAAACGTGAGAGCAGGTCGACGGTATCG	Forward primer for inserting left homology arm flanking CFTR transgene through In-Fusion
Infusion-L arm CFTR Rev	ACAATTCGGTACAGCCCTAGTTATCAGCCAAGTC	Reverse primer for inserting left homology arm flanking CFTR transgene through In-Fusion

Infusion-R arm CFTR Fwd	ATCGATAAGCTTGATGCTCTAGAACTAGTGGATCCC	Forward primer for inserting right homology arm flanking CFTR transgene through In-Fusion
Infusion-R arm CFTR Rev	CTGCAGGAATTCGATCTCCTCTATCCTACCTCTAAAGC	Reverse primer for inserting right homology arm flanking CFTR transgene through In-Fusion
T7E1 Fwd	ACAACGGCAACTCTCTGGAATGC	Forward primer for amplifying sequence near GGTA1 target site
T7E1 Rev	GCACTCCTTAGCGCTCGTTGACTA	Reverse primer for amplifying sequence near GGTA1 target site
Junctional PCR-L-LacZ Fwd	AATGTGGACTAACACTGACCTTCC	Forward primer for left arm junctional PCR analysis of LacZ reporter gene
Junctional PCR-L-LacZ Rev/P2/P2’	AAGGCCGAGTCTTATGAGCAG	Reverse primer for left arm junctional PCR analysis of LacZ reporter gene; Reverse primer for verifying off target site 1 and 2
P1	GCAGTTCACTCAGGCAATTTC	Forward primer for verifying off target site 1
P3	GTAGAGACCTTTTCTTCCCCATG	Control reverse primer for verifying effectiveness of P1
P1’	CCCAGCAAATGGATAATAGTATTGG	Forward primer for verifying off target site 2
P3’	GCCGTTGAATTTTAGACCTGGC	Control reverse primer for verifying effectiveness of P1’
Junctional PCR-L-CFTR Fwd	AATGTGGACTAACACTGACCTTCC	Forward primer for left arm junctional PCR analysis of CFTR transgene
Junctional PCR-L-CFTR Rev	GTGGAGTCAACAAGCTATGTACTGC	Reverse primer for left arm junctional PCR analysis of CFTR transgene
Junctional PCR-R-CFTR Fwd	CCAACAGTCCCATCCCTGATC	Forward primer for right arm junctional PCR analysis of CFTR transgene
Junctional PCR-R-CFTR Rev	GAAGTGGCTTCACAAAGGCAGTG	Reverse primer for right arm junctional PCR analysis of CFTR transgene
qPCR-CFTR Fwd	CCTGAGTCCTGTCCTTTCTC	Forward primer for amplifying hCFTR cDNA
qPCR-CFTR Rev	CGCTGTCTGTATCCTTTCCTC	Reverse primer for amplifying hCFTR cDNA
qPCR-GAPDH Fwd	GTTCGACAGACAGCCGTGTG	Forward primer for amplifying pig GAPDH cDNA

qPCR-GAPDH Rev	ATGGCGACAATGTCCACTTTGC	Reverse primer for amplifying pig GAPDH cDNA

Table 3. HD-Ad constructs used in this study.

HD-Ad construct	Details
HD-Ad UbCLacZ-Neo GGTA1	HD-Ad vector that contains CRISPR/Cas9 system, LacZ-Neo reporter gene flanked by left and right homology arms regulated by ubiquitin c promoter
HD-Ad UbCLacZ-Neo no CRISPR	HD-Ad vector that contains LacZ-Neo reporter gene regulated by ubiquitin c promoter
HD-Ad CMVGFP	HD-Ad vector that contains GFP reporter gene regulated by cytomegalovirus promoter
HD-Ad K18CFTR GGTA1	HD-Ad vector that contains CRISPR/Cas9 system, hCFTR transgene flanked by left and right homology arms regulated by keratin 18 promoter
HD-Ad K18CFTR no CRISPR	HD-Ad vector that contains hCFTR transgene regulated by keratin 18 promoter
HD-Ad C4HSU	HD-Ad empty vector

3. Results: construction and verification of tools used in this study

3.1 Design of the two HD-Ad constructs used in this study

The HD-Ad UbCLacZ-Neo GGTA1 construct (Figure 2a) consists of a *Cas9* endonuclease gene linked to an enhanced green fluorescent protein (eGFP) through a sequence called T2A. The T2A sequence can give rise to a self cleaving peptide which allows the expression of both the Cas9 and the eGFP using only one promoter (146). The sgRNA, containing the GGTA1 guide sequence, is driven by the U6 promoter. In addition, the *LacZ* reporter gene, regulated by the UbC promoter, was flanked by the left and right HA. At the ends of the linear construct, ITR and Ψ (only adjacent to the left ITR) can be found. The rest of the construct is occupied by the stuffer sequence from the pC4HSU backbone vector. The HD-Ad K18CFTR GGTA1 construct (Figure 2b) has a similar design compared to the HD-Ad UbCLacZ-Neo GGTA1. It contains the same CRISPR/Cas9 system (with a slight difference in organization), HAs, ITRs, and Ψ. The major difference is the transgene to be integrated is K18CFTR instead of UbCLacZ-Neo. In addition, the total size of the construct is slightly smaller than the previous construct.

3.2 Verification of the major HD-Ad plasmid constructs

3.2.1 pUbCLacZ-Neo GGTA1

Synthesizing HD-Ad vector is a complicated and time-consuming process. Therefore, verification of the construct is necessary before committed to HD-Ad production. From the map generated by SnapGene (Figure 3a), the locations of restriction sites used for the verification step are displayed. Based on the restriction site locations, the length of each set of restriction digestion

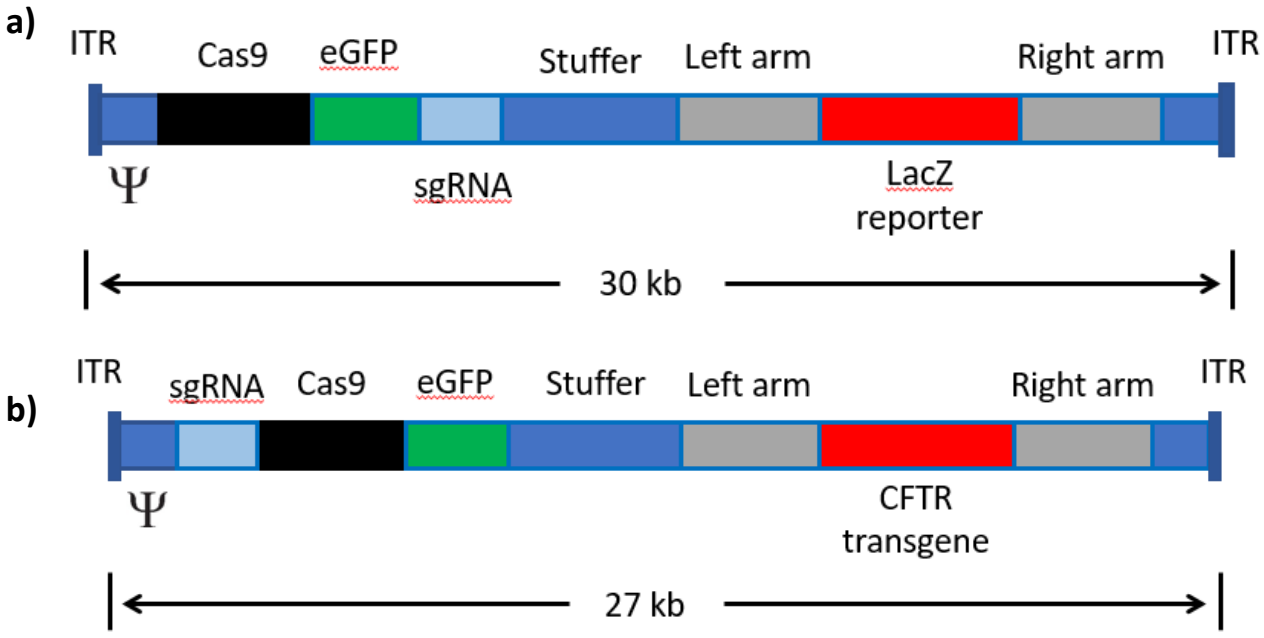


Figure 3. Schematics of the HD-Ad constructs used in this study

a) The 30 kilo base pairs (kb) long HD-Ad UbCLacZ-Neo GGTA1 construct. ITR = Inverted Terminal Repeats; Cas9 = cDNA sequence for Cas9 endonuclease controlled by chicken β-actin promoter; eGFP= enhanced Green Fluorescent Protein; sgRNA = single guide RNA with GGTA1 guide sequence; Stuffer = DNA sequences from human to meet the minimum genome size requirement for adenoviral vector production; Left and Right arm = homology arms required by the process of homologous recombination; LacZ reporter = cDNA sequence for β galactosidase, regulated by UbC promoter. The eGFP sequence is linked to Cas9 cDNA through the T2A linker so they are regulated by the same promoter. b) The 27-kb long HD-Ad K18CFTR GGTA1 construct. Abbreviations are the same as (a). CFTR transgene = modified CFTR cDNA that is regulated by the K18 promoter.

can be predicted. For PmeI and NotI digestion, the predicted fragment lengths will be approximately 14.2 kb, 11.2 kb, 4.9 kb, and 2.9 kb. Furthermore, NotI and SbfI digestion will yield 22 kb, 8.6 kb, and 2.6 kb bands. With similar calculations, NotI, XhoI and MfeI digestion will give rise to 11.8 kb, 10.2 kb, 6.7 kb, and 4.5 kb fragments. Finally, MfeI and PmeI can lead to the formation of 26.3 kb, 4 kb, and 2.9 kb bands. These sets of enzymes were purposely chosen to cover the entire construct. Compared to the predicted sizes of fragments with the gel image (Figure 3b), it is possible to see that the band patterns visualized on the gel correctly matches with the predictions from the information provided by the SnapGene program.

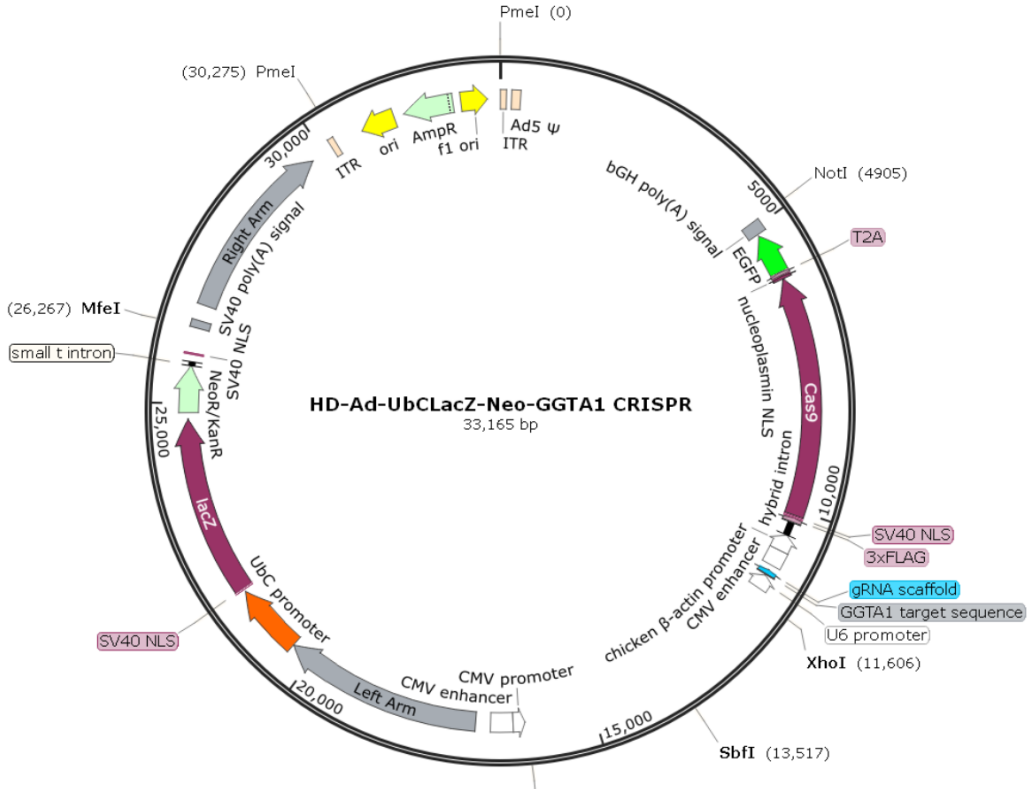
In addition, the sgRNA guide sequence plays a crucial role in this study. As a result, the area encoding for the sgRNA was sequenced (Figure 3c). By comparing the sequencing result with the designed guide sequence, it demonstrates the guide sequence inserted is error free.

3.2.2 pK18CFTR GGTA1

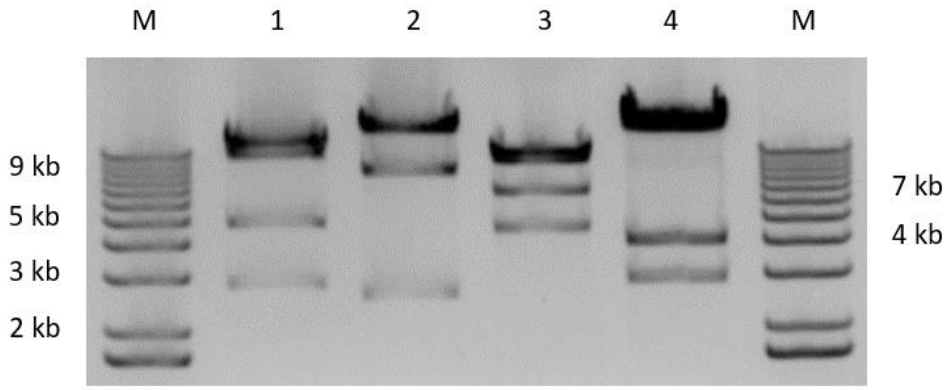
With a similar purpose of 3.2.1, the pK18CFTR GGTA1 plasmid construct was verified as well. From the map generated by SnapGene (Figure 4a), the restriction sites used for verification were displayed. Using similar approach as the previous section, the predicted length of restriction fragments can be calculated. From PmeI and PacI digestion, 27.4 kb, 6 kb, and 2.9 kb fragments will be generated. The second set of restriction cleavage (PacI and BstZ17I) will give rise to 26.6 kb and 3.6 kb bands. BstZ17I and SbfI digestion will lead to the formation of 27.4 kb and 3kb fragments. Furthermore, SbfI and NotI cleavage will yield 23.6 kb, 3.9 kb, and 2.7 kb bands. Finally, the last set of digestions by NotI and PmeI will form 16.6 kb, 8.1 kb, 2.9 kb and 2.7 kb bands. The reason these enzymes were chosen is because the short restriction fragments (<10 kb) resulted from the selected enzymes' cleavage can cover the entire length of the plasmid

construct so each of the important component will be verified. By comparing with the gel image obtained after restriction digestion (Figure 4b), the bands obtained from the gel picture correctly match with the predicted band size by SnapGene. Even though it is hard to determine the size of the large band in each set of enzyme cleavage, the size of small fragments (<10 kb) can be clearly observed and compared with SnapGene predictions. One exception is in set 5 digestion. It is hard to resolve the 2.9 kb and the 2.7 kb bands unless they are separated on a higher concentration agarose gel. However, the problem with running on a higher concentration gel is that the larger molecular size fragments will become unresolvable. Despite of this, it is still possible to tell that the intensity and thickness of the 2.7/2.9 kb band in set 5 is more than the bands with comparable sizes from other sets. Therefore, it is possible to deduce that this band in set 5 is actually consisted of two bands with similar sizes. The sgRNA guide sequence was verified similarly to 3.2.1. The sequencing result demonstrates the sgRNA GGTA1 sequence is exactly the same as the engineered guide sequence.

a)


PmeI (0)
(30,275) PmeI
30,000
ITR
ori
AmpR
f1 ori
Ad5 ψ
ITR
bGH poly(A) signal
5000
NotI (4905)
T2A
EGFP
nucleoplasmin NLS
Cas9
hybrid intron
10,000
SV40 NLS
3xFLAG
gRNA scaffold
GGTA1 target sequence
U6 promoter
XhoI (11,606)
chicken β-actin promoter
CMV enhancer
CMV promoter
15,000
SbfI (13,517)
CMV enhancer
Left Arm
20,000
UbC promoter
SV40 NLS
lacZ
25,000
NeoR/KanR
SV40 NLS
SV40 poly(A) signal
Right Arm
(26,267) MfeI
small t intron
HD-Ad-UbCLacZ-Neo-GGTA1 CRISPR
33,165 bp

b)


M
1
2
3
4
M
9 kb
5 kb
3 kb
2 kb
7 kb
4 kb

c)

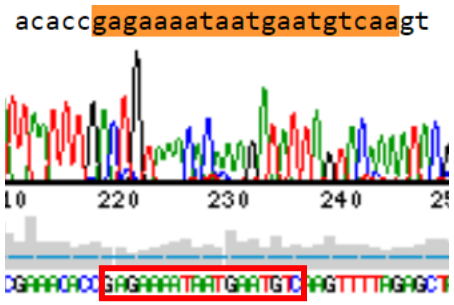
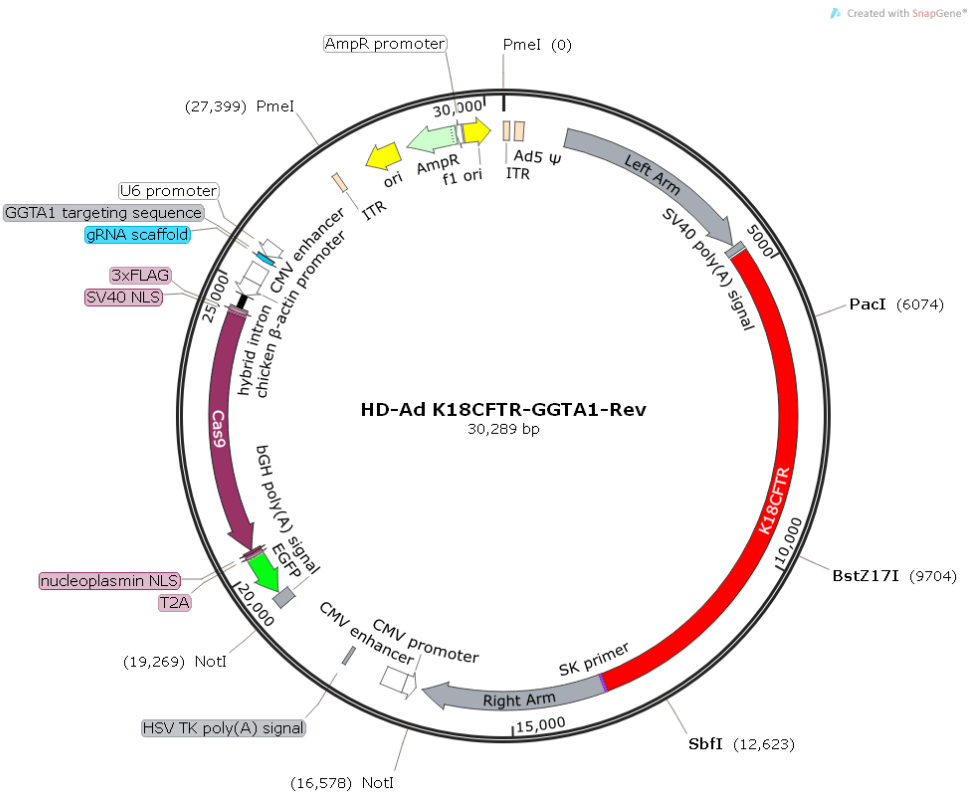

acaccgagaaaataatgaatgtcaagt
220
230
240

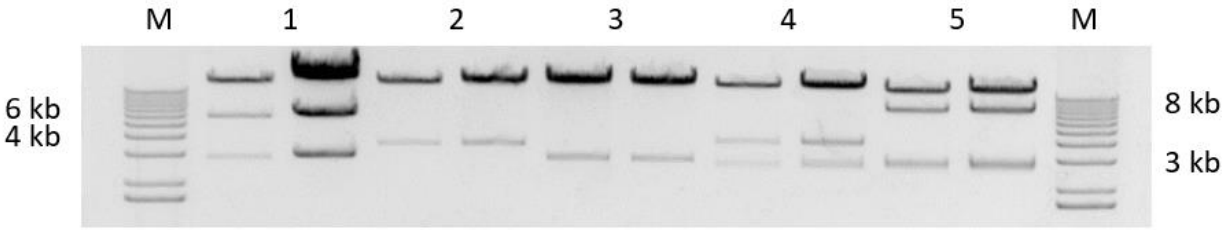
Figure 4. Verification of the pUbCLacZ-Neo GGTA1 plasmid construct

a) Plasmid map of HD-Ad UbCLacZ-Neo GGTA1 before linearization was visualized by the SnapGene program. Important components and enzyme sites are displayed on the map. b) The plasmid was digested with 4 sets of restriction enzymes. M = molecular ladder; 1 = PmeI + NotI; 2 = NotI +SbfI; 3 = NotI + XhoI + MfeI; 4 = MfeI + PmeI. The band sizes of the molecular ladder used are 1.6 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb and 12 kb (from bottom to top respectively). c) the sgRNA guide sequence region was verified through Sanger sequencing. The sequence enclosed by the red box is the GGTA1 guide sequence designed.

a)


Created with SnapGene®
AmpR promoter
PmeI (0)
(27,399) PmeI
30,000
ori
AmpR
f1 ori
Ad5 ψ
ITR
Left Arm
SV40 poly(A) signal
5000
PacI (6074)
U6 promoter
GGTA1 targeting sequence
gRNA scaffold
ITR
CMV enhancer
chicken β-actin promoter
3xFLAG
SV40 NLS
25,000
hybrid intron
Cas9
HD-Ad K18CFTR-GGTA1-Rev
30,289 bp
bGH poly(A) signal
EGFP
K18CFTR
10,000
BstZ17I (9704)
nucleoplasmin NLS
T2A
20,000
(19,269) NotI
CMV enhancer
CMV promoter
SK primer
Right Arm
15,000
SbfI (12,623)
HSV TK poly(A) signal
(16,578) NotI

b)


M
1
2
3
4
5
M
6 kb
4 kb
8 kb
3 kb

c)

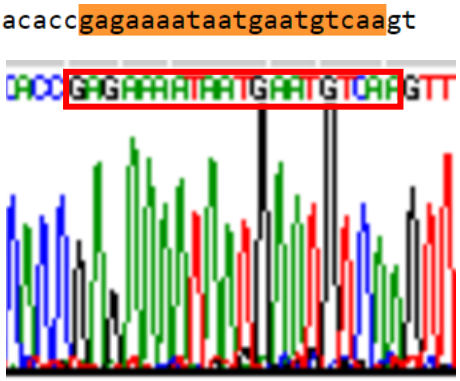

acaccgagaaaataatgaatgtcaagt

Figure 5. Verification of the pK18CFTR GGTA1 plasmid construct.

a) The plasmid map of HD-Ad K18CFTR GGTA1 before linearization is shown by the SnapGene program. Restriction sites used for verification and important elements are displayed on the map. b) The plasmid was digested with 5 sets of restriction enzymes. M = molecular ladder; 1 = PmeI + PacI; 2 = PacI + BstZ17I; 3 = BstZ17I + SbfI; 4 = SbfI + NotI; 5 = NotI + PmeI. The band sizes of the molecular ladder used are 1.6 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb and 12 kb (from bottom to top respectively. The two lanes in each set come from two potential HD-Ad K18CFTR GGTA1 constructs. c) The GGTA1 guide sequence region was verified through Sanger sequencing. The guide region is the 20-nucleotide sequence enclosed by the red box.

3.3 HD-Ad production

3.3.1 HD-Ad UbCLacZ-Neo GGTA1

During HD-Ad production, the GFP fluorescence can be monitored. GFP positive cells represent cells that received our viral genome. The amount of GFP fluorescence also informs us whether our virus particles are being amplified, with the exception of P0 generation. In P0, the green signals observed is actually a mix of both the HD-Ad linear construct used for transfection and the HD-Ad viral vector being synthesized (Figure 5a). Therefore, even though the green signal drastically decreases in P1 (Figure 5b) compared to P0, this does not mean less virus is being produced. As for P2, P3 and P4 (Figure 5c – e), it is possible to observe that the percentage of GFP positive cells increases as the generation progresses. The increase in GFP signal represents an amplification of the virus titer being produced. The left and right image are under the same view with the difference of light source. So, a rough estimation of how much cells are GFP positive can be obtained by comparison of these two images. In P2, there are still a lot of cells

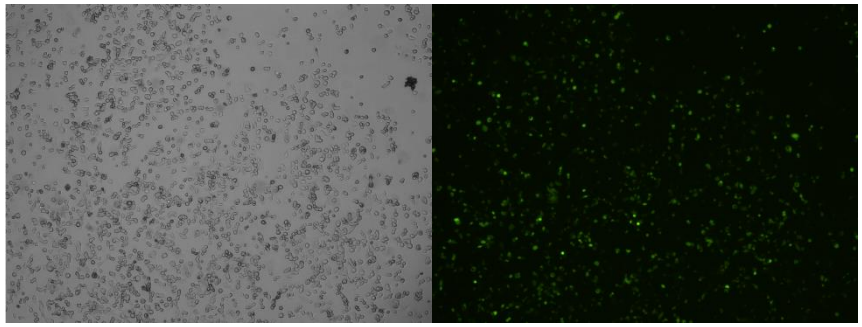
that lack any GFP signals comparing with the left image. As the amplification stage continues to P4, not only the proportion of GFP positive cells significantly increased, but the intensity of GFP signals in each cell also improved. This trend of increasing GFP demonstrates the desired viral vector are actively being amplified for each generation. The time chosen to harvest each generation was 48 hours after transduction because the proportion of lysed cells versus number of cells still in CPE is optimal. If the infected cells are collected earlier, a large proportion of cells are still in the process of virus production and the yield would decrease. On the other hand, if the cells are harvested at a later time point, many cells will be lysed to release the matured viral vector. One problem with this is that the vector is not stable in the media compared to staying inside of a host cell. Therefore, even though the released viral vector in the media can be collected, we have no methods of storing them and preserving their activity. However, the optimal time of harvesting the cells is subject to change depending on the variation of HD-Ad to helper virus ratio.

3.3.2 HD-Ad K18CFTR GGTA1

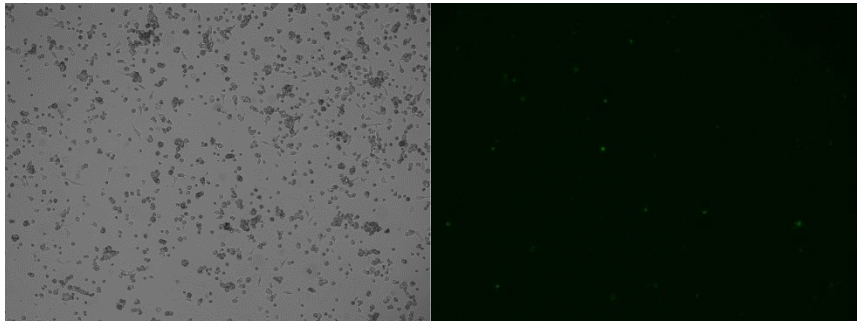
Similar to the previous section, the GFP fluorescence of HD-Ad K18CFTR GGTA1 viral vector production was also monitored. Compared to the HD-Ad UbCLacZ-Neo GGTA1, the proportion of GFP positive cells and the GFP intensity are higher during the production of HD-Ad K18CFTR GGTA1 in every generation. It is interesting to note that in the P4 generation, the amount of vector does not seem to be amplified since the amount and intensity of GFP signals are at a comparable level. However, the image for P4 is only a representative image. In fact, P4 generation is consisted of eight 15 cm plates of 116 producer cells whereas the P3 generation only has one 15 cm plate of 116 cells. Another factor to be considered is that the GFP

fluorescence is already very saturated at P3, so an increase of GFP expression in P4 may not be as noticeable in terms of fluorescence intensity as the amplification from P1 to P3.

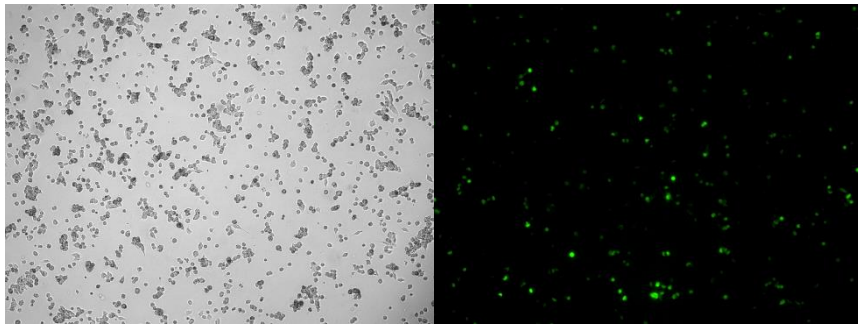
a)



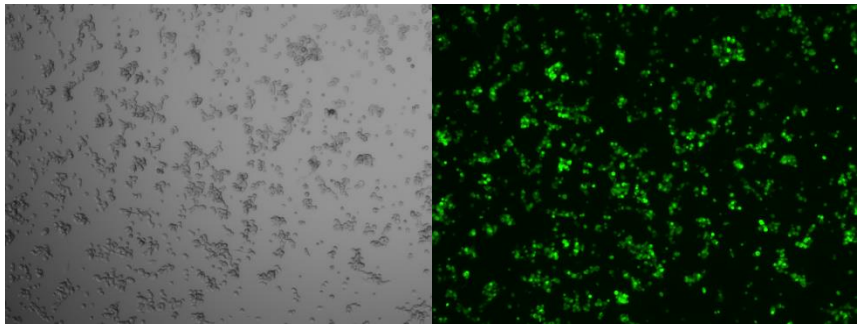
b)



c)



d)



e)

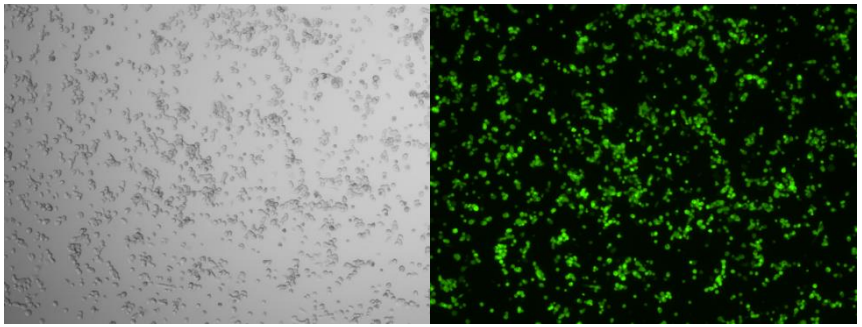
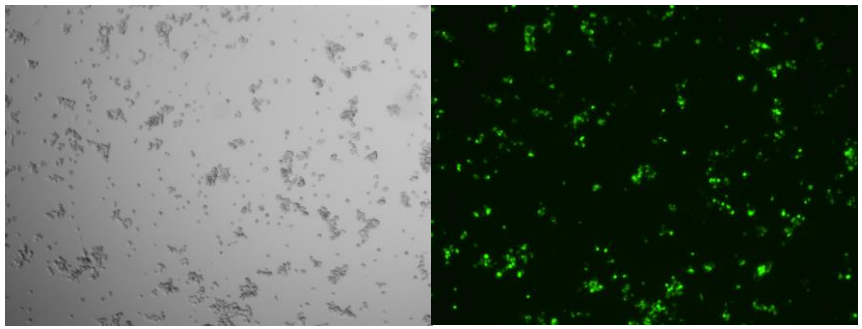


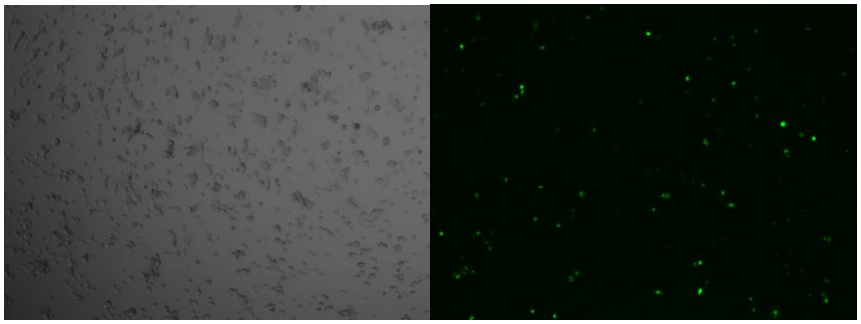
Figure 6. The rescue and amplification stages for HD-Ad UbCLacZ-Neo GGTA1 production.

(a) Image taken of P0 generation by fluorescence microscope 48 hours post-transfection HD-Ad UbCLacZ-Neo GGTA1 linear vector and transduction of NG163 helper virus. The left image represents the infected cells under normal light without fluorescence while the right image is detected at 512 nm of emission wavelength. (b) P1 images taken by fluorescence microscope 48 hours post co-transduction of P0 lysate and NG163 helper virus into 116 producer cells. (c) Fluorescence microscope images of P2 generation 48 hours post co-transduction of P1 lysate and NG163 helper virus into 116 producer cells. (d) P3 images taken by fluorescence microscope 48 hours post co-transduction of P2 lysate and NG163 helper virus into 116 producer cells. (e) Representative fluorescent images of P4 (one of eight 15 cm plates) with same conditions as previous generations except P3 lysate was used.

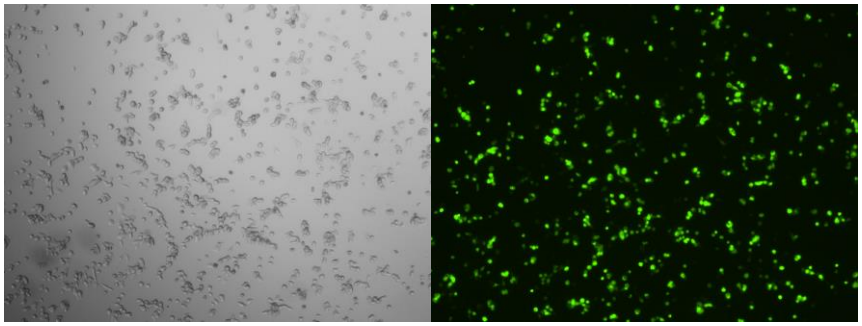
a)



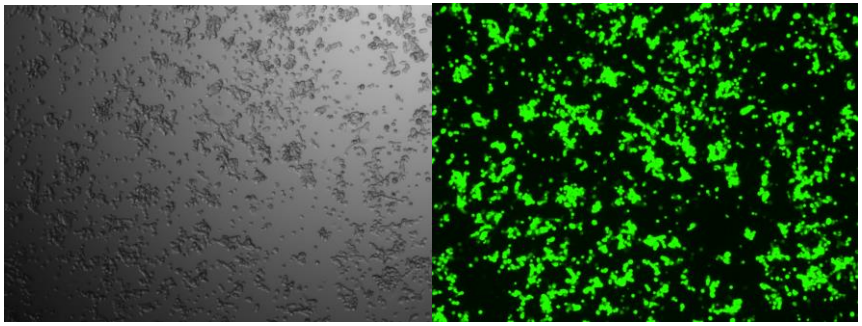
b)



c)



d)



e)

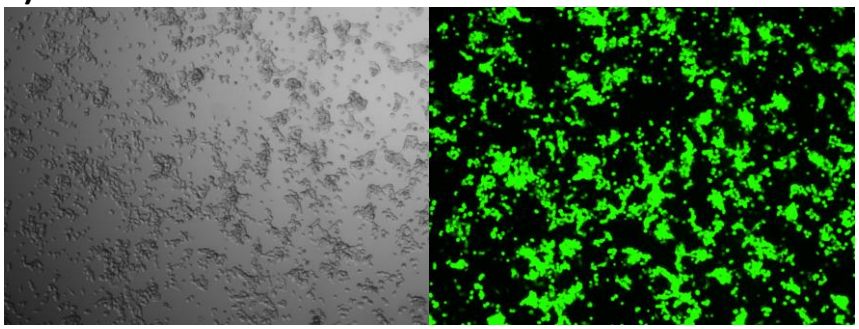


Figure 7. The rescue and amplification stages for HD-Ad K18CFTR GGTA1 production.

a) Images taken by fluorescence microscope at 48 hours post-transfection of HD-Ad K18CFTR GGTA1 linear vector and transduction of NG163 helper virus into 116 producer cells. The left image shows the transfected/transduced 116 cells under normal light microscopy and the right image is the transfected/transduced 116 cells detected at 512 nm of emission wavelength. This generation is denoted as P0. (b) P1 images taken by fluorescence microscope 48 hours post co-transduction of P0 lysate and NG163 helper virus into 116 producer cells. (c) Fluorescence microscope images of P2 generation 48 hours post co-transduction of P1 lysate and NG163 helper virus into 116 producer cells. (d) P3 images with conditions same as (c) except P2 lysate was used. (e) Representative fluorescent images of P4 (one of eight 15 cm plates) with same conditions as previous generations except P3 lysate was used.

3.3.3 Large scale amplification

To compare the difference between infected cells versus uninfected cells, uninfected control 116 producer cells and HD-Ad + helper infected cells 3 days post co-transduction were stained with orcein. In the control population, all of the stained cells are round in shape with approximately the same size as each other (Figure 7a). On the other hand, the infected 116 producer cells have many different characteristics. Firstly, there are many visible lysed cells due to viral infection which are not observed in the control population. Secondly, many enlarged cells with irregular shaped inclusion bodies are visible under the infected population.

After the HD-Ad large scale amplification in the 3L spinner flask, the cells were lysed and ultracentrifuged in cesium chloride gradient. This centrifugation step led to the formation of a milky white band precisely at the interface between 1.25 g/ml cesium chloride and 1.35 g/ml cesium chloride (Figure 7b). There were also other white bands accumulated above the band at interface and those were cellular debris resulted from the cell lysis step. The second centrifugation eliminated the junk bands since only the viral band at the interface was extracted. Therefore, only one milky white band was observed (Figure 7c). Following the method described in 2.4.7, the titer of HD-Ad UbCLacZ-Neo GGTA1 and HD-Ad K18CFTR GGTA1 was determined as 7.2×10^{11} particles/ml and 1.1×10^{12} particles/ml (data not shown).

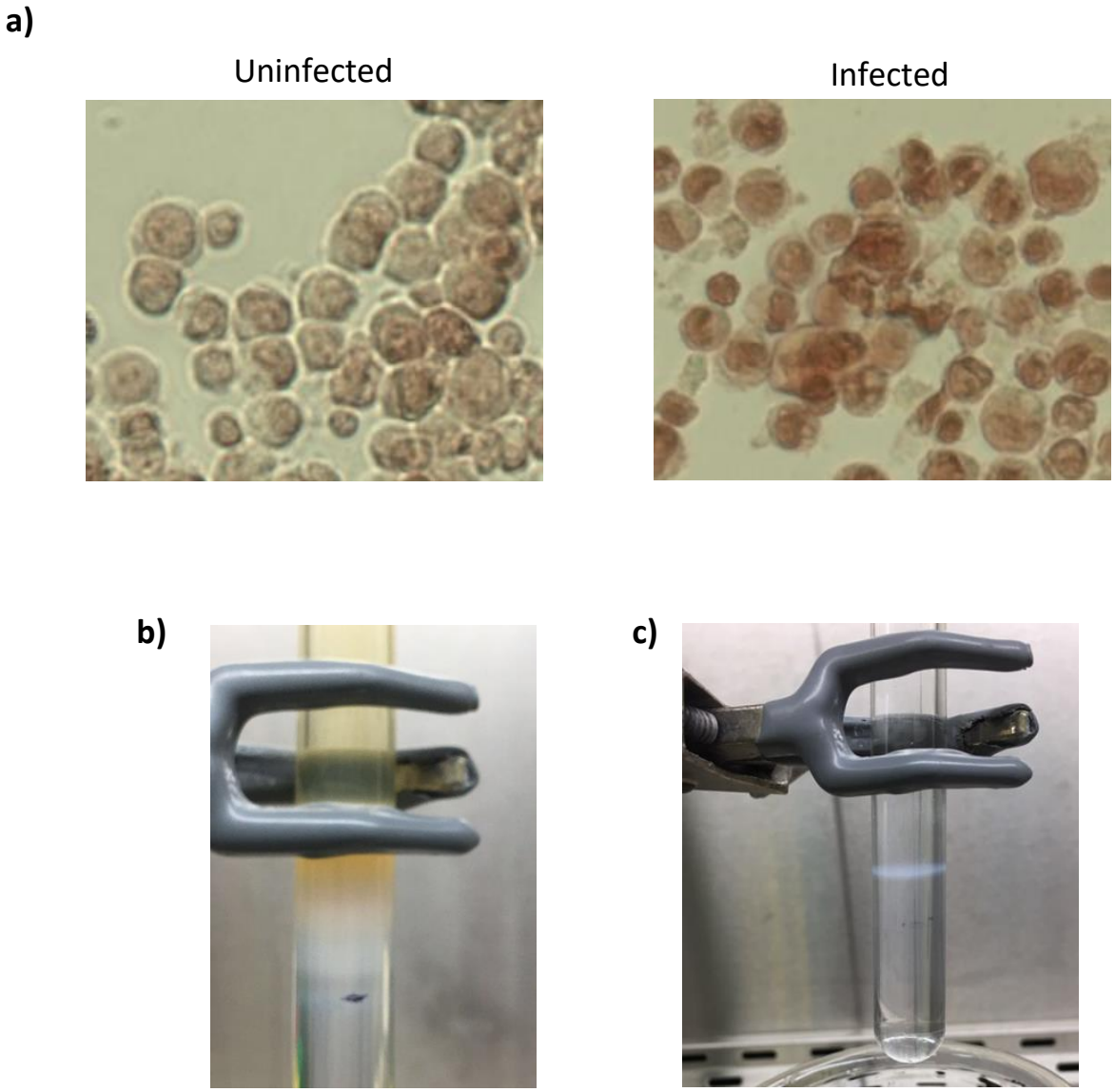


Figure 8. Large scale amplification of HD-Ad vector.

Representative images of the HD-Ad K18CFTR GGTA1 large scale amplification process are shown. (a) Orcein inclusion body staining of HD-Ad infected 116 producer cells and uninfected 116 producer cells from the 3L spinner flask culture. (b) Multiple milky white bands were obtained from ultracentrifugation of cell lysate in cesium chloride gradient. The band located at the marked black line was extracted which represents the true viral band. (c) Second round of ultracentrifugation with the extracted band from previous step was shown.

3.4 Assessing the activity of HD-Ad UbCLacZ-Neo GGTA1 vector

After the successful production of desired vectors, it is important to verify if the vectors produced are active enough to be used in later experiments. Unfortunately for the HD-Ad K18CFTR GGTA1 vector, there are no reporter genes available to test its activity except by monitoring the GFP fluorescence after transduction. The GFP fluorescence of the CFTR vector was examined and it demonstrated a desirable level of expression (data not shown). On the other hand, activity of the HD-Ad UbCLacZ-Neo GGTA1 vector can be determined through both GFP fluorescence as well as β galactosidase staining. As shown in Figure 8a, as little as 2 MOI (on average two active particles per cell) of HD-Ad UbCLacZ-Neo GGTA1 vector can lead to detectable level of GFP fluorescence and a good proportion of cells that are *LacZ* positive. Even though by definition, 2 MOI represents two active viral particles per cell, but the actual distribution of the vector will not be as simple and as straightforward. It is very possible that some cells receive no vectors while other cells receive more than two particles. Judging by the result of β galactosidase staining, the approximate transduction efficiency of 2 MOI vector is approaching 50% (estimation of *LacZ* positive cells divided by total number of cells). When moving to higher vector dosage such as 10 MOI and 50 MOI (Figure 8b and c), the proportion of *LacZ* positive cells significantly increases. With 10 MOI, the approximate transduction efficiency is estimated to be 70%. In the 50 MOI image, it is possible to see that all the cells are *LacZ* positive, indicating a 100% transduction efficiency. Looking at the GFP fluorescence, the amount of GFP signals also increases as the vector dosage increases.

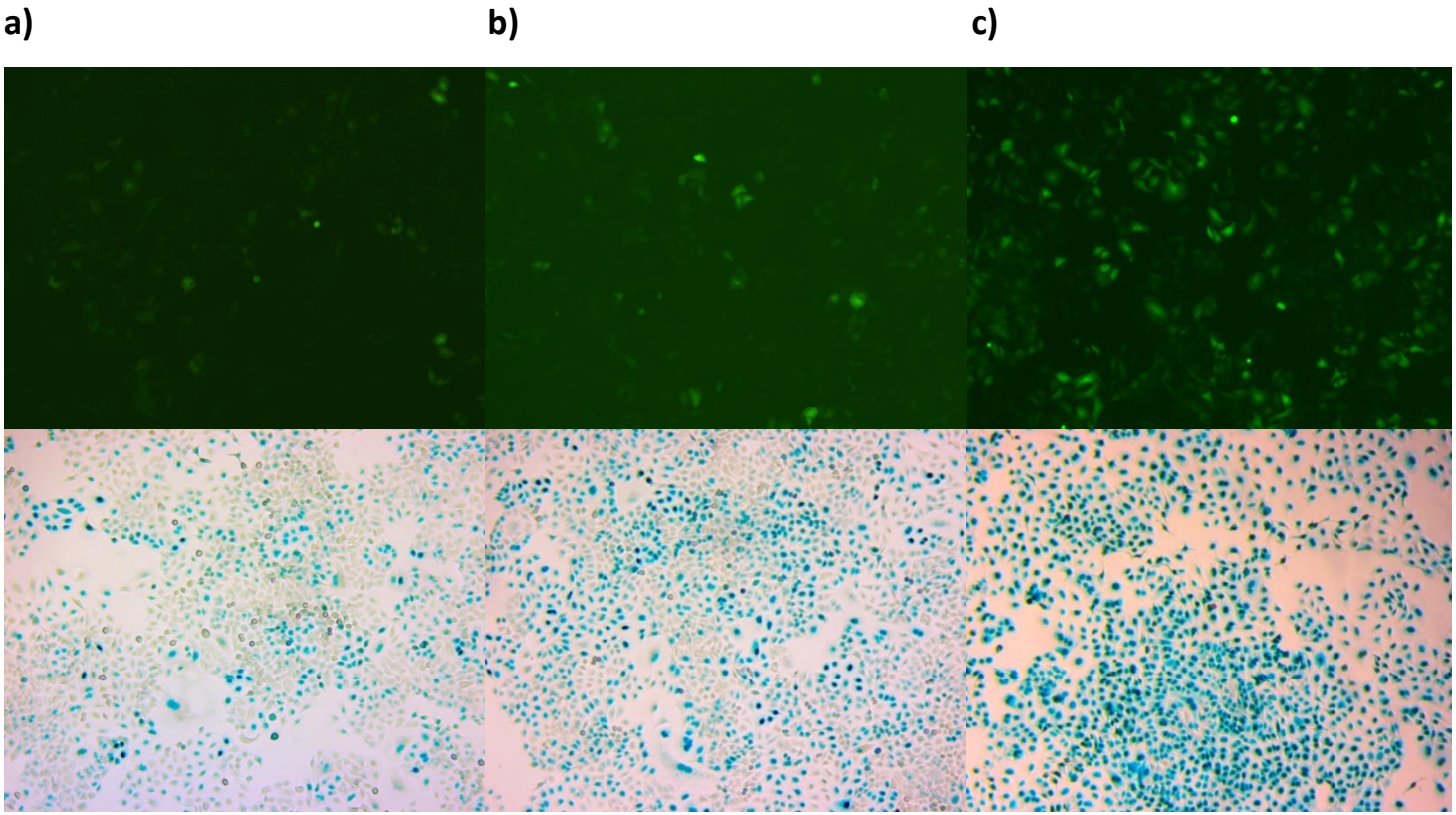


Figure 9. GFP and LacZ expression post HD-Ad UbCLacZ-Neo GGTA1 vector transduction.

IPEC-J2 cells were transduced with the HD-Ad UbCLacZ-Neo GGTA1 and the images were taken 3 days post-transduction. In (a), 2 MOI of the vector was used. The GFP expression (top) was observed under 512 nm of green light fluorescence and the *LacZ* expression (bottom) was visualized under normal light microscope post β galactosidase staining. (b) Same as (a) except 10 MOI of vector was used. (c) Same as (a) except 50 MOI of vector was used.

4. Results: Utilization of the HD-Ad UbCLacZ-Neo GGTA1 vector

4.1 CRISPR/Cas9 cleavage efficiency

The efficiency of the designed sgRNA at cleaving its target is an important factor that can greatly affect the transgene integration efficiency. From lane 2 to lane 6 of the gel image after T7E1 cleavage, 3 bands can be observed (Figure 9). The top parental bands are stronger in the control and transfection group while the opposite is observed in the transduction group. Another major difference between the transfection and the transduction population is the two lower cleaved bands are broader and more smeared in the transfection group than the transduction group, which has more distinct and sharper cleaved bands. In addition, the location of the cleaved bands from transfection have slight difference in sizes compared to the cleaved bands from transduction. Using the procedure described in section 2.5, the cleavage efficiency from lane 3 to lane 6 after normalization with control is: 16.77%, 8.32%, 49.81%, and 66.01% respectively.

4.2 *LacZ* integration efficiency

As previously stated, assessing integration efficiency of a reporter gene give us an opportunity to verify if our developed strategy can lead to stable and sufficient level of integration. The integration efficiency obtained (Figure 10a) from transducing IPEC-J2 cells with 50 MOI of HD-Ad UbCLacZ no CRISPR vector achieved 0.6% of integration efficiency according to the calculation method described in 2.6.3. This low integration efficiency is expected since the vector does not have the potential to induce double stranded break in the host cell's genome. Therefore, the integration of *LacZ* from this virus relies on the intrinsic low frequency of recombination. On the other hand, the utilization of HD-Ad UbCLacZ-Neo GGTA1 leads to a significant increase in

integration efficiency. Transduction of 5 MOI of HD-Ad UbCLacZ-Neo GGTA1 yielded a 2.43% integration efficiency. Similarly, transduction of 20 MOI and 50 MOI of HD-Ad UbCLacZ-Neo GGTA1 achieved 6.84% and 7.58% respectively. From these data, it is interesting to note that the increase of viral vector dosage from 5 MOI to 20 MOI lead to a much more increase in integration efficiency comparing to from 20 MOI to 50 MOI. By observing the images from β galactosidase staining (Figure 10b), the true LacZ signal is a strong blue signal localized in the nucleus as the nuclear localization signal is incorporated on the *LacZ* reporter gene. As a comparison, the background signal that can be observed in the control image has a very faint, disseminated blue signal in the cytoplasm of the stained cells.

4.3 Junctional PCR analysis for *LacZ* integration

To verify if the integration achieved at the designed locus, junctional PCR analysis was performed. The primers used in this experiment were designed according to the plan from Figure 11a. Both the left arm junctional PCR (Figure 11b) and right arm junctional PCR (Figure 11c) demonstrated integration occurred precisely at the designed GGTA1 target site when using the HD-Ad UbCLacZ-Neo GGTA1 vector. Both 5 MOI and 20 MOI transduction revealed visible bands with the correct length of PCR product. However, the bands obtained from 5 MOI transductions are much fainter comparing to the 20 MOI and this observation is consistent in both the left and the right junctional PCR amplification. As a comparison, both HD-Ad C4HSU empty vector and HD-Ad UbCLacZ no CRISPR achieved no detectable integration at the target site since no PCR products can be detected after 35 PCR cycles.

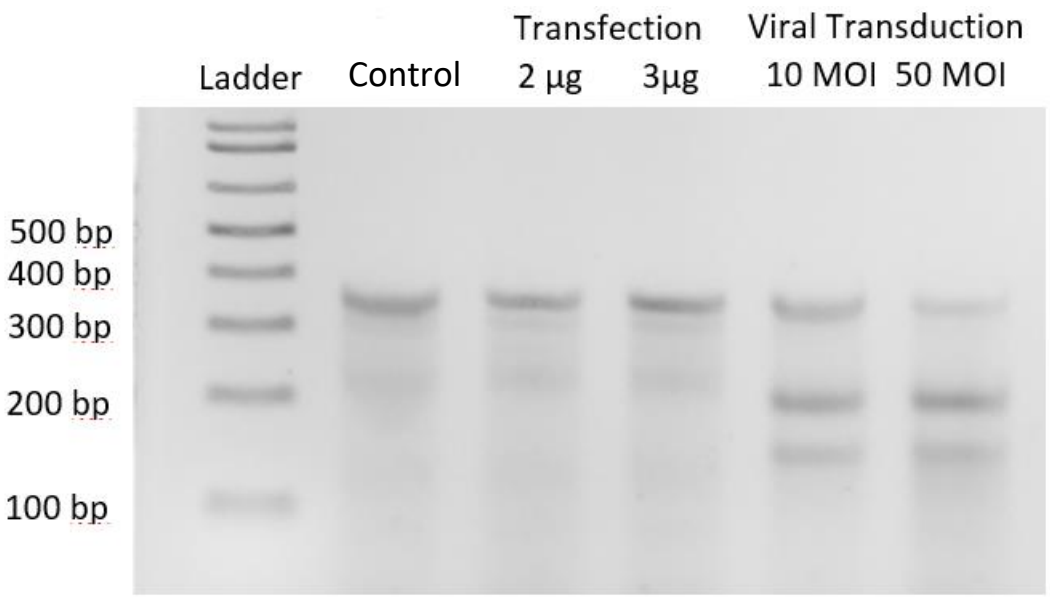
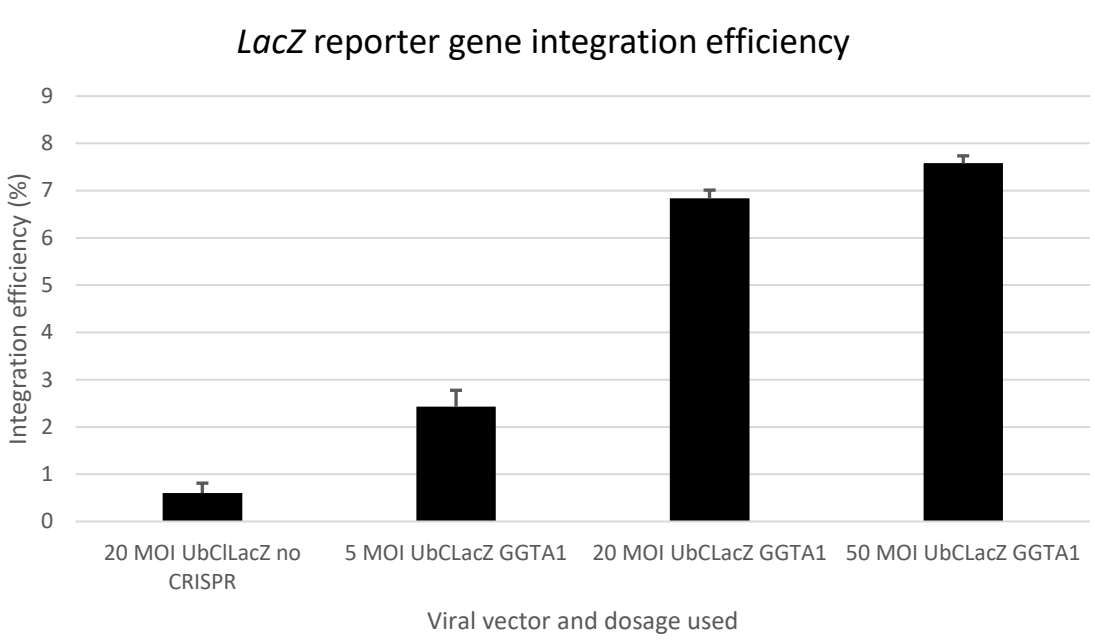


Figure 10. Comparison of CRISPR/Cas9 cleavage efficiency between plasmid vector transfection and HD-Ad transduction.

After T7E1 treatment, the amplified PCR fragments were loaded on 2% agarose gel. The parental band has a size of ~350 base pairs whereas the size of the cleaved products are approximately 200 base pairs and 150 base pairs in length. Lane 1: 1 kb plus molecular ladder with size of each important band labeled at the left. Lane 2: T7E1 endonuclease I cleavage of PCR product using genomic DNA extracted from cells that received neither transduction nor transfection as template. Lane 3: T7 endonuclease I cleavage of PCR product using genomic DNA extracted from cells that received 2 µg of LacZ plasmid vector as template. Lane 4: T7 endonuclease I cleavage of PCR product using genomic DNA extracted from cells that received 3 µg of LacZ plasmid vector as template. Lane 5: T7 endonuclease I cleavage of PCR product using genomic DNA extracted from cells that received 10 MOI of LacZ viral vector as template. Lane 6: T7 endonuclease I cleavage of PCR product using genomic DNA extracted from cells that received 50 MOI of LacZ viral vector as template.

a)



b)

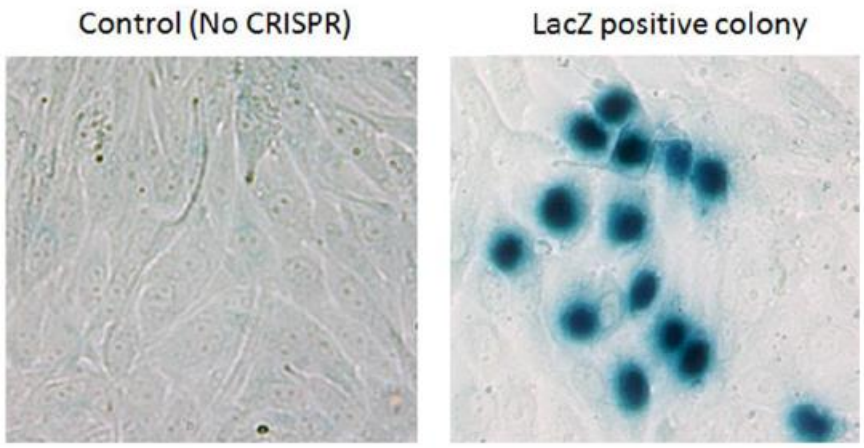


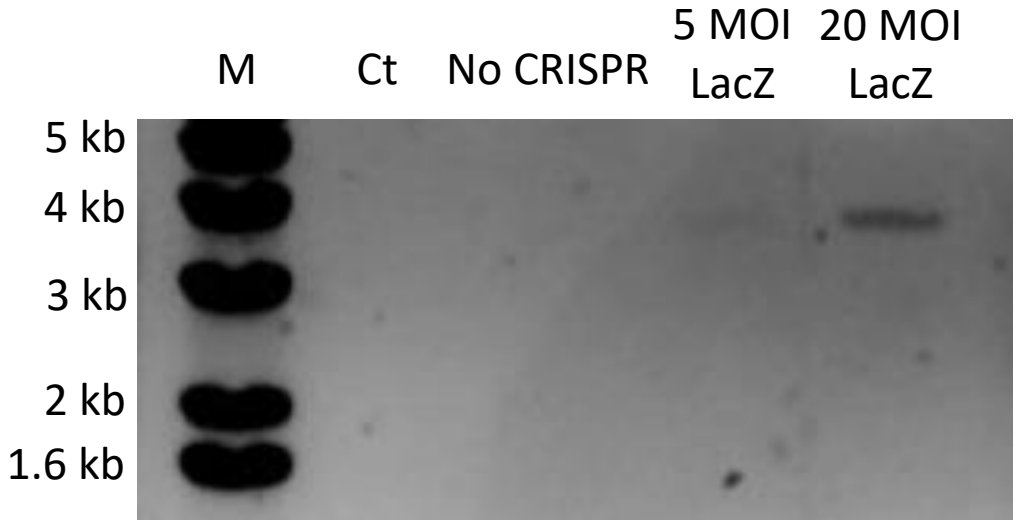
Figure 11. Assessment of *LacZ* report gene integration efficiency.

IPEC-J2 cells were transduced with 20 MOI of HD-Ad UbCLacZ no CRISPR, 5 MOI, 20 MOI and 50 MOI of HD-Ad UbCLacZ-Neo GGTA1 and passaged for 8 generations. In (a), the vector and dosage used was plotted against the % integration efficiency obtained. Data from three independent tests were shown in black solid bar. (b) Representative images from the β galactosidase staining of the 20 MOI HD-Ad UbCLacZ no CRISPR transduced cells and 20 MOI HD-Ad UbCLacZ-Neo GGTA1 transduced cells. The error bars presented are standard error calculated with n = 3.

a)


3.7 kb
4.2 kb
Left Arm
UbCLacZ-Neo
Right Arm

b)


M
Ct
No CRISPR
5 MOI LacZ
20 MOI LacZ
5 kb
4 kb
3 kb
2 kb
1.6 kb

c)

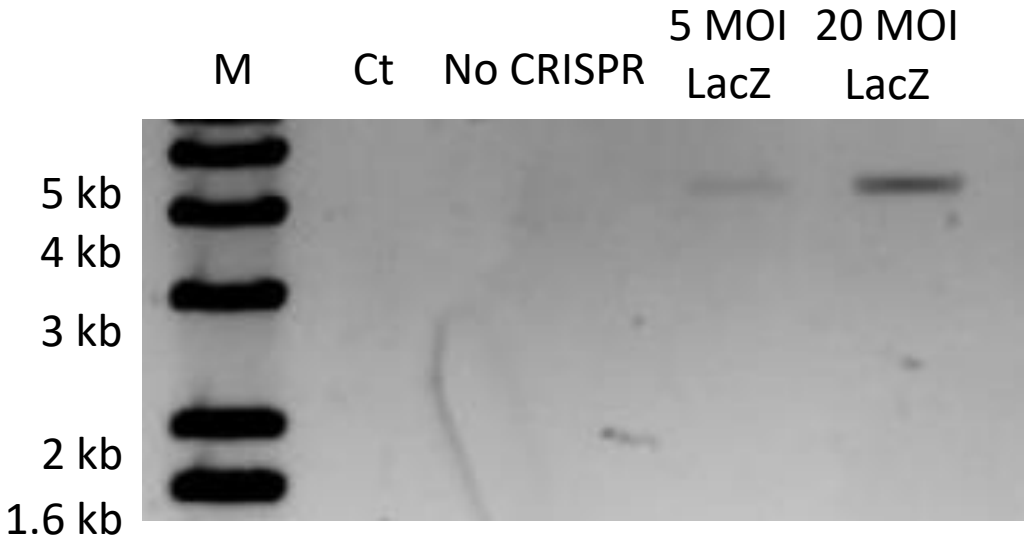

M
Ct
No CRISPR
5 MOI LacZ
20 MOI LacZ
5 kb
4 kb
3 kb
2 kb
1.6 kb

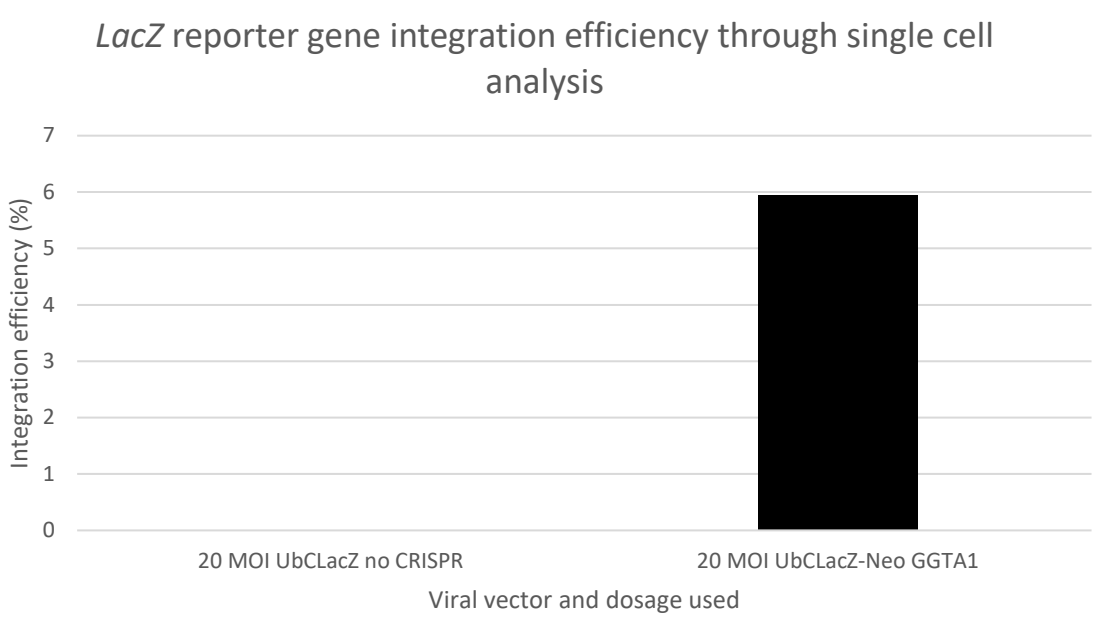
Figure 12. Junctional PCR analysis of HD-Ad UbCLacZ-Neo GGTA1 integration.

(a) Design of junctional PCR primers used in this experiment. The first set of primers spans the whole left arm with the forward primer annealing to upstream of the target site while the reverse primer binding on the promoter of UbCLacZ expression cassette. The second set of primers spans the entire right arm with forward primer binding to the UbCLacZ reporter gene and the reverse primer annealing to downstream of the GGTA1 target site. (b) Gel image of left arm junctional PCR amplification. M = molecular ladder; Ct = genomic DNA from IPEC-J2 cells transduced with HD-Ad pC4HSU empty vector; No CRISPR = genomic DNA from IPEC-J2 cells transduced with HD-Ad UbCLacZ no CRISPR; 5 MOI LacZ = genomic DNA from IPEC-J2 cells transduced with 5 MOI HD-Ad UbCLacZ-Neo GGTA1; 20 MOI LacZ = genomic DNA from IPEC-J2 cells transduced with 20 MOI HD-Ad UbCLacZ-Neo GGTAI. (c) Gel image of right arm junctional PCR amplification. The labels on the gel is same as (b).

4.4 Single cell culturing to confirm integration efficiency

An alternative approach to calculate and verify the integration efficiency obtained is through single cell culturing analysis following the procedures described in 2.6.6. The initial transduction efficiency of 20 MOI HD-Ad UbCLacZ-Neo GGTA1 was 75% and 5 out of 112 single cell colonies have positive LacZ signals for all of the cells in the colony. Therefore, the integration efficiency calculated for 20 MOI HD-Ad UbCLacZ-Neo GGTA1 is 5.95% (Figure 12a), which agrees with the efficiency obtained in 4.2 after normalization with control efficiency. As control, single cell culturing analysis was performed with 20 MOI of HD-Ad UbCLacZ no CRISPR. The integration efficiency for this vector is 0% since 0 out of 132 single cell colonies were successfully cultured that displayed *LacZ* positive signals for every cell in the colony. Representative β galactosidase staining images (Figure 12b) demonstrates that the HD-Ad UbCLacZ no CRISPR transduced colony has no real nuclear LacZ signals detected whereas the colony from the HD-Ad UbCLacZ-Neo GGTA1 transduction has strong nuclear LacZ expression for every single cell from that particular colony.

a)



b)

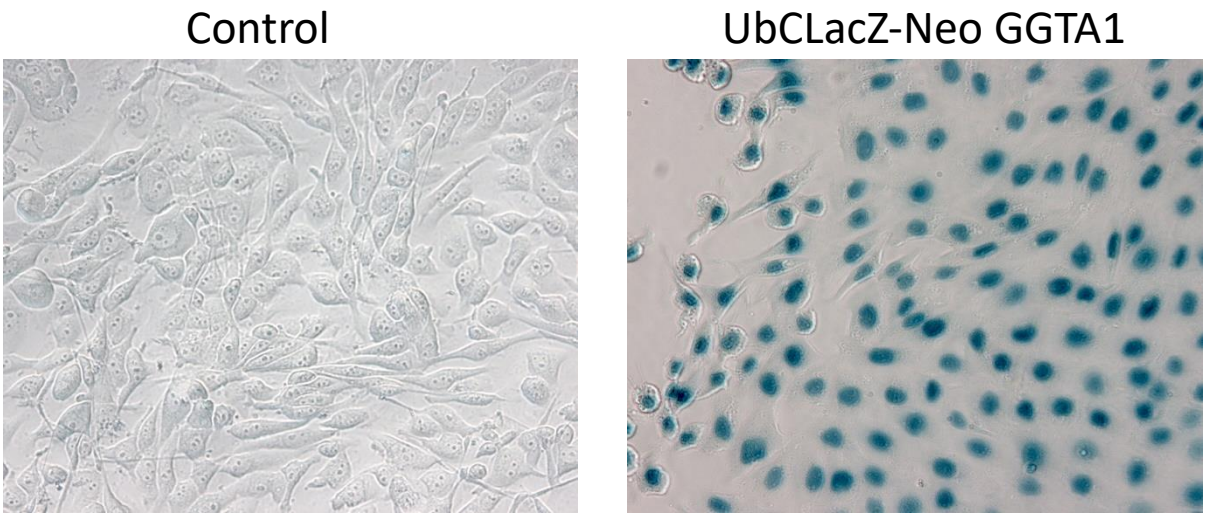


Figure 13. Verifying *LacZ* reporter integration efficiency through single cell culturing analysis. IPEC-J2 cells transduced with 20 MOI of HD-Ad UbCLacZ no CRISPR and HD-Ad UbCLacZ-Neo GGTA1. The cells were passaged 8 times and then plated single cells into wells of 96 well plates through limiting dilution. (a) Calculated integration efficiencies following the formula described in 2.6.6 were plotted against the viral vector and dosage used. 0 out of 132 single cell colonies yielded positive LacZ signal in the control group while 5 out of 112 single cell colonies demonstrated positive LacZ signal in 20 MOI UbCLacZ-Neo GGTA1 transduced group. (b)

Representative images from β galactosidase staining of HD-Ad UbCLacZ no CRISPR and HD-Ad UbCLacZ-Neo GGTA1 transduced IPEC-J2 cells.

4.5 Off-target site analysis

A major concern for utilizing the CRISPR/Cas9 system is whether it will introduce undesired modification at sites that shows high similarities to the designed target sequence. In this study, a PCR based approach was used to analyze undesired genome modifications from the engineered CRISPR/Cas9 system. Primers used in this PCR approach follows the schematics outlined in Figure 13a. When undesired transgene integration occurs at the potential off-target site, primer P1 and P2 (or P1' and P2' from second off-target site) will generate a 1.6 kb long PCR fragment. On the contrary, if no unwanted integration occurs, P1 and P2 (or P1' and P2') will not amplify any fragments. The function of P3 (or P3') is to exclude the possibility of P1 being defective at producing a PCR product even when correct annealing of the primers is achieved. The primer P2 and P2' are actually the same primer and they were used previously in junctional PCR analysis of *LacZ* integration. The amplification product from P1, P3 and P1', P3' are both 1.6 kb in length while no products were detected from P1, P2 and P1', P2' amplifications (Figure 13b).

To further ensure that no incorrect modifications occurred at the potential off-target sites, the PCR products from P1, P3 and P1', P3' were sequenced. The sequencing results were compared with the sequence of potential off-target sites. Only the predicted top off-target site comparison is shown in Figure 13c. There are no random indel mutations detected in the predicted off-target sites.

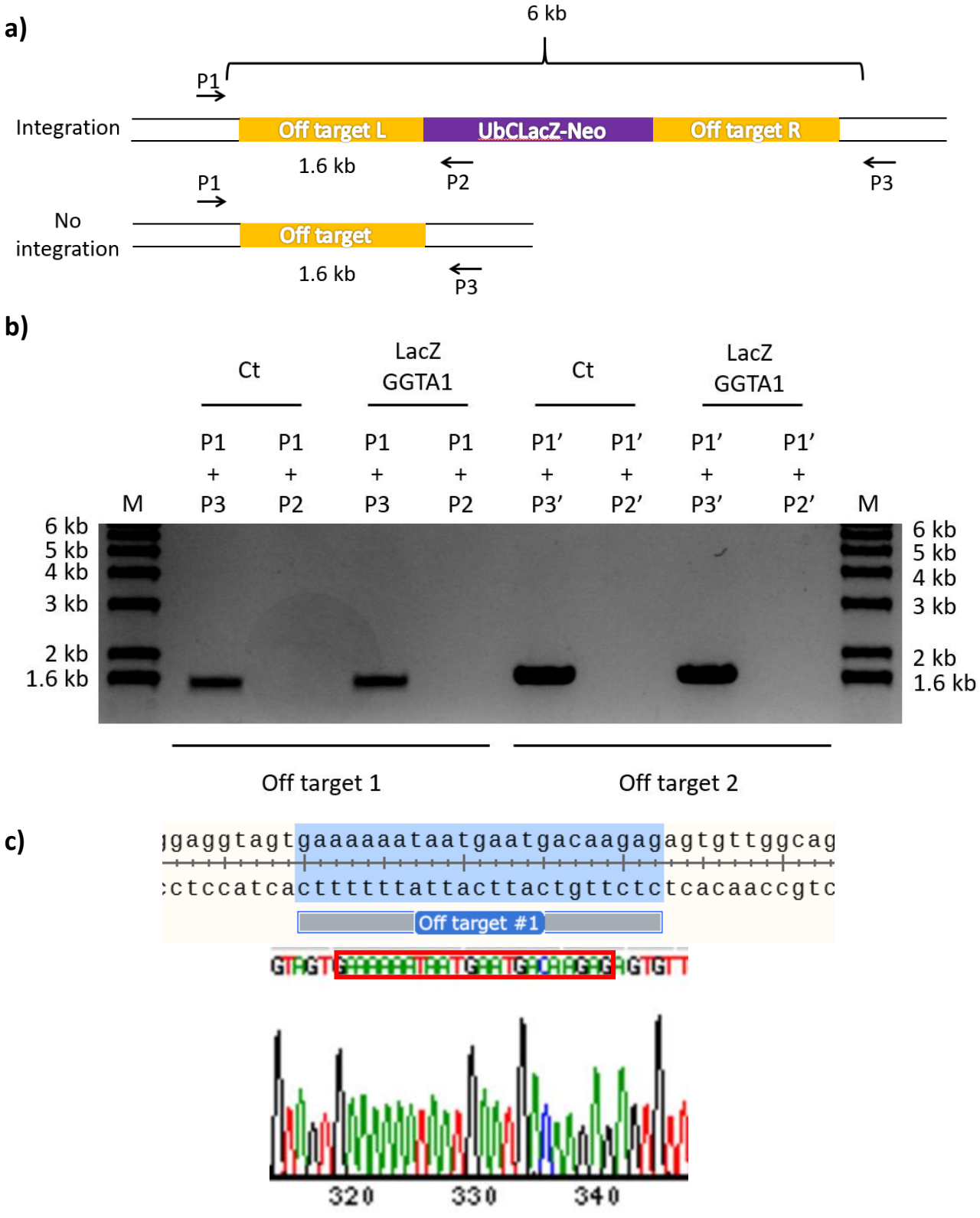

a)
6 kb
P1
Integration
Off target L
UbCLacZ-Neo
Off target R
1.6 kb
P2
P3
P1
No integration
Off target
1.6 kb
P3
b)
Ct
LacZ GGTA1
Ct
LacZ GGTA1
P1 + P3
P1 + P2
P1 + P3
P1 + P2
P1' + P3'
P1' + P2'
P1' + P3'
P1' + P2'
M
6 kb
5 kb
4 kb
3 kb
2 kb
1.6 kb
Off target 1
Off target 2
c)
Off target #1
320
330
340

Figure 14. Verification of undesired genome editing in potential off target sites.

Undesired modification at potential off-target sites was verified through a PCR based approach. (a) For the first off-target site, 3 primers denoted as P1, P2 and P3 were used. P1 anneals approximately 1.6 kb upstream of the predicted off-target site. P2 anneals on the promotor of UbCLacZ expression cassette and P3 anneals slightly downstream of the predicted off-target site. Similarly, 3 primers were designed according to the same strategy for the second off-target site. These primers were denoted as P1', P2', and P3' (not shown on figure). (b) Gel image of the PCR amplification products using the primers labelled on the top. M = molecular ladder; Ct = PCR amplification of HD-Ad UbCLacZ no CRISPR transduced cells (20 MOI); LacZ GGTA1 = PCR amplification of HD-Ad UbCLacZ-Neo GGTA1 transduced cells (20 MOI). (c) Sequencing result of PCR amplification product, obtained using the P1 and P3 primers, from the top predicted off-target site. The red box encloses the predicted off-target site sequence.

4.6 Enhancement of integration efficiency through Scr7

To increase the transgene integration efficiency, the DNA ligase IV inhibitor Scr7 was used. The efficiencies obtained from 6 passages after the transductions of 20 MOI HD-Ad UbCLacZ no CRISPR, 20 MOI HD-Ad UbCLacZ-Neo GGTA1 no Scr7, 20 MOI HD-Ad UbCLacZ-Neo GGTA1 0.1 µM Scr7 and 20 MOI HD-Ad UbCLacZ-Neo GGTA1 1 µM Scr7 were 0.64%, 6.37%, 9.17%, and 12.94% respectively (Figure 14). After normalization of the integration efficiencies obtained with the background (20 MOI UbCLacZ no CRISPR), the integration efficiencies became 5.73%, 8.53%, and 12.3%. Comparing the Scr7 treated groups to the non-Scr7 treated group, a clear increase in integration efficiency can be observed. It is interesting to note that the addition of 1 µM Scr7 lead to a one-fold integration efficiency enhancement.

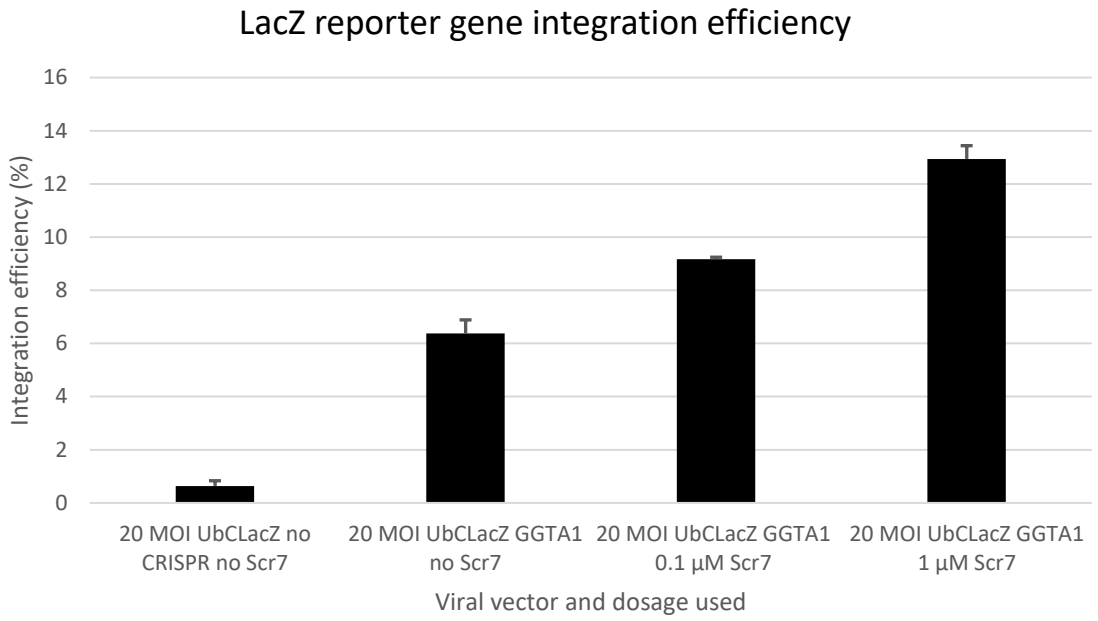


Figure 15. Enhancement of integration efficiency through Scr7 administration.

IPEC-J2 cells were transduced with 20 MOI of HD-Ad C4HSU, HD-Ad UbCLacZ no CRISPR and HD-Ad UbCLacZ-Neo GGTA1. For HD-Ad UbCLacZ-Neo transduction, 0 µM, 0.1 µM and 1 µM of Scr7 were administered. Integration efficiency from passage 6 post-transduction was plotted against the viral vector and dosage using a bar figure. The HD-Ad C4HSU was used as negative β-galactosidase control (not shown in the plot). The HD-Ad UbCLacZ no CRISPR accounted for the background level of *LacZ* expression without the assistance from CRISPR/Cas9 system. The error bars presented are standard error calculated with n = 3.

5. Results: Utilization of the HD-Ad K18CFTR GGTA1 vector

5.1 Junctional PCR analysis for CFTR integration

Similar to the previous aim, junctional PCR analysis was used to verify the position of the CFTR transgene integration. The primers used in this experiment were designed according to the schematic presented in Figure 15a. After the PCR amplification using the designed primers, both left arm junctional PCR (Figure 15b) and right arm junctional PCR (Figure 15c) yielded bands with expected length (3.5 kb and 3.2 kb respectively). The intensities of bands obtained from 20 MOI and 50 MOI HD-Ad K18CFTR GGTA1 transduction are at a comparable level. The negative control used in this experiment is from transduction 20 MOI of HD-Ad C4HSU empty vector. Both the negative control and the transduction of 20 MOI HD-Ad K18CFTR no CRISPR failed to generate the junctional PCR fragment which is demonstrated in the lane 2 and lane 3 of the gel images.

5.2 CFTR mRNA expression

To detect whether the integrated CFTR transgene can be successfully expressed, total RNA was extracted and hCFTR expression were measured using RT-qPCR. The standard for relative expression calculation used was the CT value of cells transduced with 20 MOI of HD-Ad K18CFTR no CRISPR vector in the corresponding passage. In the passage 0, the hCFTR mRNA expression from 20 MOI HD-Ad K18CFTR GGTA1 transduction is 0.925-fold compared to the standard while the 50 MOI HD-Ad K18CFTR GGTA1 transduction induced a 6.23-fold of hCFTR mRNA expression in comparison with the standard (Figure 16a). At passage 5, cells from 20 MOI HD-Ad K18CFTR GGTA1 transduction have 13.04-fold relative expression and cells from 50 MOI HD-Ad K18CFTR GGTA1 transduction have 21.94-fold relative expression (Figure 16b). When moving onto the

passage 9, the relative expressions of hCFTR mRNA for 20 MOI and 50 MOI of HD-Ad K18CFTR GGTA1 are 48.88 and 60.38-fold respectively (Figure 16c).

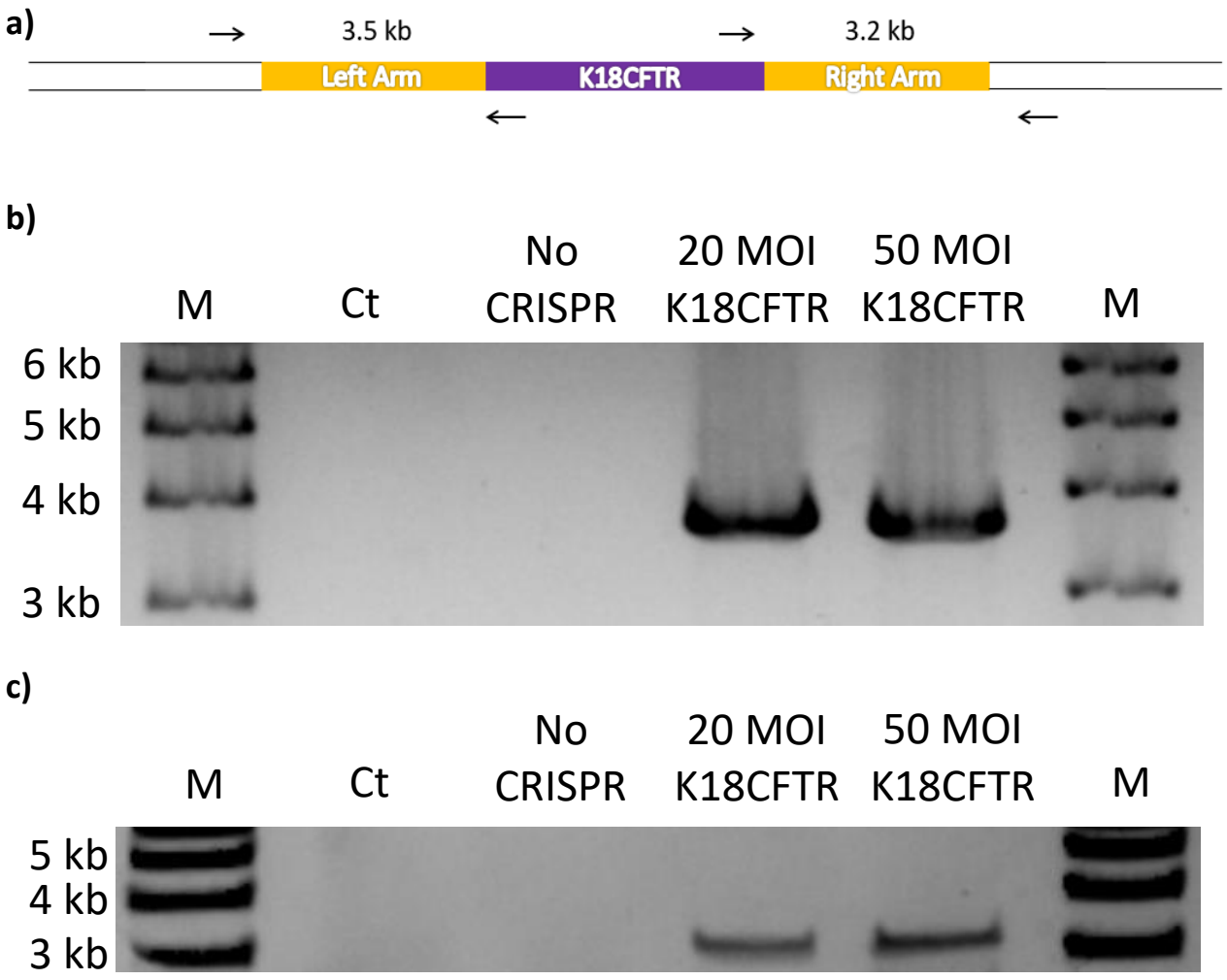
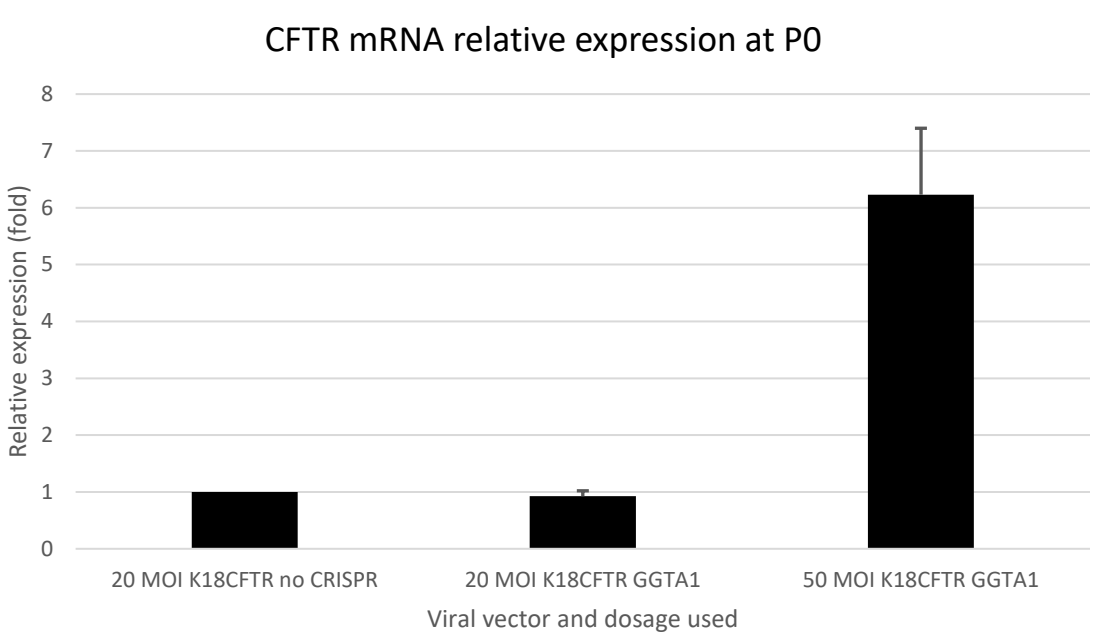


Figure 16. Junctional PCR analysis of HD-Ad K18CFTR GGTA1 integration.

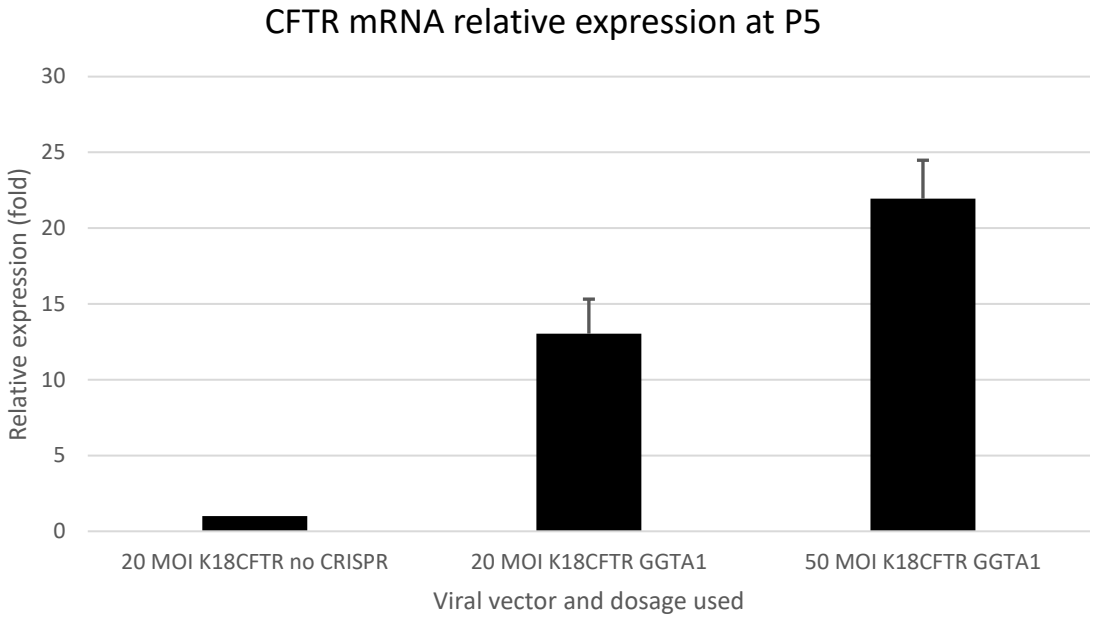
(a) Schematics of the primer design and expected PCR fragment length used in this experiment. The first set of primers covers the whole left arm with the forward primer annealing to upstream of the target site while the reverse primer binding on the promoter of K18CFTR expression cassette. The second set of primers spans the entire right arm with forward primer binding to

the K18CFTR reporter gene and the reverse primer annealing to downstream of the GGTA1 target site. (b) Gel image of left arm junctional PCR amplification. M = molecular ladder; Ct = PCR amplification using genomic DNA from IPEC-J2 cells transduced with HD-Ad C4HSU empty vector; No CRISPR = PCR amplification using genomic DNA from IPEC-J2 cells transduced with HD-Ad K18CFTR no CRISPR; 20 MOI LacZ = PCR amplification using genomic DNA from IPEC-J2 cells transduced with 20 MOI HD-Ad K18CFTR GGTA1; 50 MOI LacZ = PCR amplification using genomic DNA from IPEC-J2 cells transduced with 50 MOI HD-Ad K18CFTR GGTA1. (c) Gel image of right arm junctional PCR amplification. The labels on the gel is same as (b).

a)


CFTR mRNA relative expression at P0
Relative expression (fold)
0
1
2
3
4
5
6
7
8
20 MOI K18CFTR no CRISPR
20 MOI K18CFTR GGTA1
50 MOI K18CFTR GGTA1
Viral vector and dosage used

b)


CFTR mRNA relative expression at P5
Relative expression (fold)
0
5
10
15
20
25
30
20 MOI K18CFTR no CRISPR
20 MOI K18CFTR GGTA1
50 MOI K18CFTR GGTA1
Viral vector and dosage used

c)

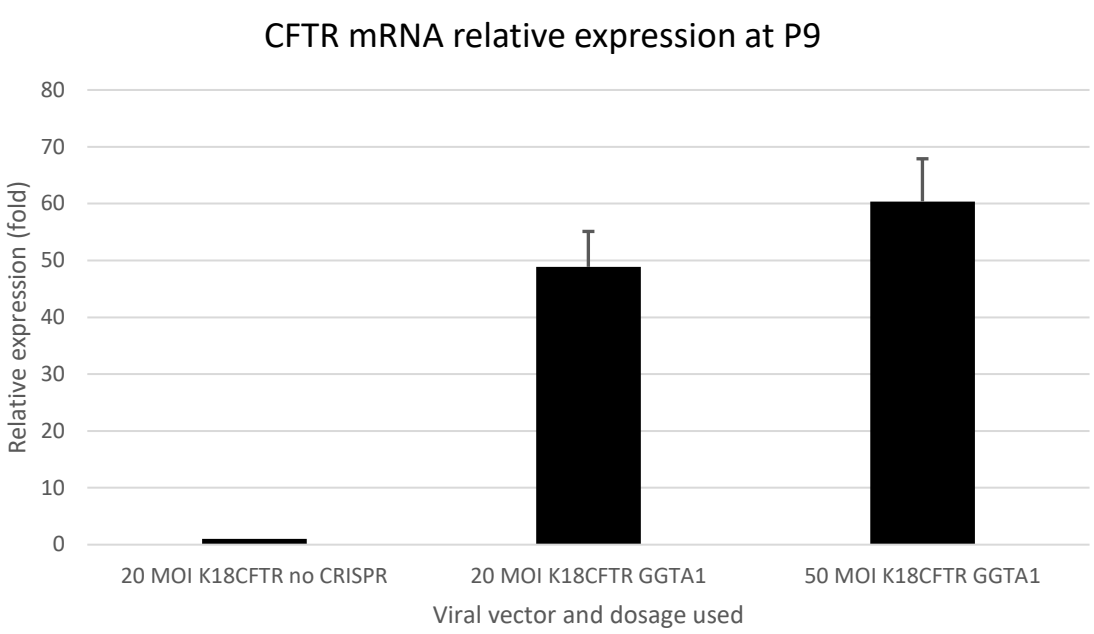


Figure 17. hCFTR transgene mRNA expression at various passages post HD-Ad K18CFTR GGTA1 transduction.

(a) Relative expression of hCFTR mRNA from P0 of 20 MOI and 50 MOI HD-Ad K18CFTR GGTA1 transduction was plotted. The expression of hCFTR mRNA of 20 MOI K18CFTR no CRISPR was set as the standard for calculating relative expression and GAPDH expression was used as internal control. (b) Same as (a) except the mRNA was extracted and measured 5 passages after the initial transduction. (c) Same as (a) except the mRNA was extracted and measured 9 passages after the initial transduction. The error bars presented are standard error calculated with n = 6.

5.3 hCFTR protein expression

To detect whether transgene hCFTR protein can be expressed, protein samples were extracted and Western blot was performed. In the upper panel of Figure 17, a very strong expression of CFTR can be observed in the lane with positive control (8-3-7 cells induced with doxycycline). Besides the very strong band at the top, there is also a faint thin band beneath it. In comparison, 8-3-7 cells without induction lacks detectable CFTR expression which is expected. In the lane with protein sample from 20 MOI and 50 MOI of HD-Ad K18CFTR GGTA1 transduction, a very faint band can be observed at the same location as the strong band in positive control. As for negative control (20 MOI HD-Ad C4HSU transduction), no CFTR signal was detected.

In the lower panel, GAPDH expression can be detected in all of the samples which serves as loading control. The bottom band represents the real GAPDH signal while the upper bands are non-specific detections based on the size comparison with the protein ladder. The intensity of GAPDH band is lower in the -Dox and +Dox lane comparing to the other three samples since only half the amount of protein sample was loaded in these two lanes.

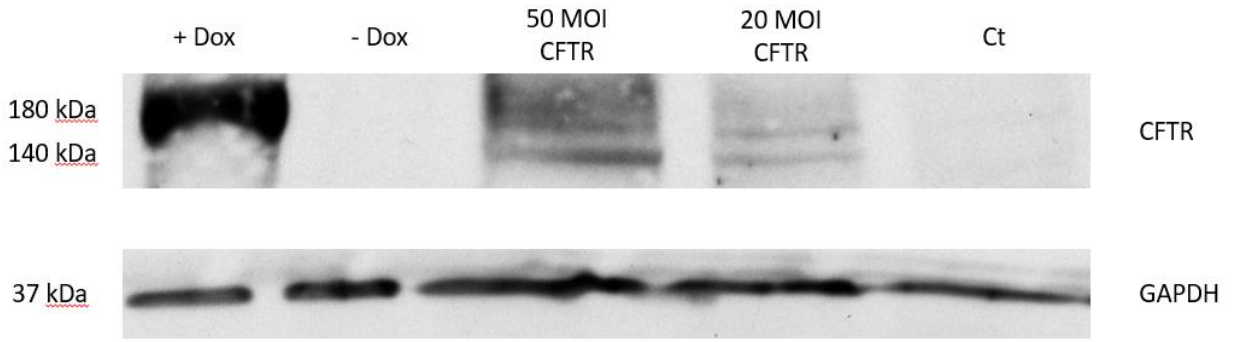


Figure 18. hCFTR transgene protein expression 3 days post HD-Ad K18CFTR GGTA1 transduction.

Ct = sample collected from cells transduced with 20 MOI of HD-Ad C4HSU empty vector. 20 MOI CFTR = sample collected from cells transduced with 20 MOI of HD-Ad K18CFTR GGTA1. 50 MOI CFTR = sample collected from cells transduced with 50 MOI of HD-Ad K18CFTR GGTA1. – Dox = sample collected from 8-3-7 cells without doxycycline induction. + Dox = sample collected from 8-3-7 cells with doxycycline induction. All the samples were collected 3 days post corresponding vector transduction. 100 µg of total protein was loaded for the first 3 lanes while 50 µg of total protein was loaded for the -Dox and +Dox lanes.

6. Discussion and Future directions

Our results demonstrated that our novel strategy was successful in *in vitro* settings. By packaging an engineered CRISPR/Cas9 system and a transgene expression cassette into a HD-Ad vector, stable and precise transgene integration can be achieved. There are several advantages with this system. First of all, the utilization of the popular and readily available CRISPR/Cas9 system can lead to efficient double stranded break at the designed target locus. Furthermore, the deletion of almost all the viral genes in HD-Ad vector minimizes the immunogenicity related to the delivery of this vector. In addition, the large packaging capacity gives us an opportunity to package multiple elements in the same vector and thereby achieving simultaneous delivery of all the necessary components.

In this study, the pig intestinal epithelial cell line IPEC-J2 was used for all the *in vitro* testing due to the lack of access to pig airway epithelial cell lines. Even though it is ideal to use airway cells for all the analysis since the respiratory system is the most affected in CF, the intestine is still a major organ that displays CF related symptoms. IPEC-J2 cell line is a satisfactory replacement cell line for this study because it expresses the CAR receptor required for HD-Ad vector entry and it contains the GGTA1 target locus for our CRISPR/Cas9 system. When cultured in the correct media, this cell line can be grown at a fast rate which reduces the waiting time for each analysis.

HD-Ad UbCLacZ-Neo GGTA1 transduction demonstrated that efficient transgene can be expressed in the IPEC-J2 cell line post-transduction. However, the signals obtained from β galactosidase staining are stronger than GFP fluorescence even though the same dosage of viral vector was used (Figure 8). Since β galactosidase is an enzyme, so it can catalyze the cleavage of numerous X-Gal molecules. The blue color detected actually represents how many products are

being formed resulting from β galactosidase cleavage. On the other hand, GFP is a fluorescence protein whose signal amplification relies on increasing its level of expression. Therefore, β galactosidase staining is a more sensitive detection method since it has an inherited signal amplification process during its detection. However, this does not mean *LacZ* is a superior reporter compared to GFP. The disadvantage of using *LacZ* is that the cells must be fixed before staining while GFP fluorescence can be monitored with live cells.

The differences between vector plasmid transfection versus HD-Ad transduction on CRISPR cleavage efficiency were compared (Figure 9). It is possible to see that the cleavage efficiencies obtained from viral vector transduction are significantly higher comparing to plasmid vector transfection. In addition, the cleaved bands were designed to be approximately 200 base pairs and 150 base pairs in length. However, the background bands and bands from transfection are all slightly above 200 base pairs while the sizes of bands from transduction are more accurate according to the original design. In fact, the location of the bands from transfection is identical to the no DNA control. From this observation, it is unclear if the bands obtained from transfection are actual cleaved products or increased background smear. Nevertheless, this comparison demonstrates that the cleavage efficiency obtained from HD-Ad transduction is far more superior compared to plasmid vector transfection. The potential reason for the superior efficiency from transduction is that the HD-Ad vector has an efficient mechanism that allows the delivery of a very large construct (greater than 30kb in length) to the nucleus of target cells. In comparison, the transfection efficiency can be greatly influenced by the size of plasmid delivered. Even though the cellular uptake of the DNA transfected is relatively independent from the size of the construct, nuclear delivery was reduced with increasing construct sizes (147). The relatively less efficient construct delivery in transfection results in less CRISPR/Cas9 being

expressed in the target cell which leads to less cleavage of the designed target. The most important message from this figure is to show transduction is a much better approach for delivering large constructs compared with transfection.

Using the HD-Ad UbCLacZ-Neo GGTA1 vector, the reporter gene *LacZ* can be inserted into the target cell's genome at the correct location (Figure 11). Integration efficiency was highest observed in cells transduced with 50 MOI of HD-Ad UbCLacZ-Neo GGTA1 comparing to other dosage administered (Figure 10). This is expected since integration efficiency should increase as the amount of vector used increases. However, there was only less than 1% gain in integration efficiency comparing with cells transduced with 20 MOI of vector. As the increase in integration efficiency is not significant and large dosage of vectors introduced may be toxic to the cells, 20 MOI was the dosage preferred in this study.

The single cell culturing analysis was performed with the purpose of providing an alternative approach to verify *LacZ* integration efficiency. As mentioned in introduction, only the cells that have achieved stable integration can replicate the transgene introduced during cell division. If all the cells in a colony originally derived from a single cell expresses the transgene, then it means a stable integration had taken place. On the contrary, if no cells or only a portion of the cells expresses the transgene in the colony derived from a single cell, then the original cell was non-integrated. As shown in the representative image in Figure 12, all the cells have strong nuclear β galactosidase expression which confirms the theory behind this assay. The integration efficiency calculated from this approach is approximately the same as the approach described in 2.6.3 which proves the efficiency obtained was accurate.

Previous study has demonstrated that a mixed culture of 10% wild type non-CF cells with 90% human bronchial epithelial cells has normal ion transport properties (148). Therefore, the goal for our integration efficiency in this study is set as 10%. However, even with 50 MOI of the *LacZ* vector, the integration efficiency obtained did not meet this standard. To improve the efficiency, a NHEJ inhibitor, Scr7, was used together with vector transduction. With 1 μM of Scr7, the *LacZ* integration efficiency was boosted to 12.3% after normalization with control. This is slightly over a 1-fold increase of integration efficiency compared to a lack of Scr7 treatment (Figure 14). Through the enhancement by Scr7, the integration efficiency successfully reached our goal of 10%.

To determine whether our designed CRISPR/Cas9 system can lead to integration at undesired location, online tool was used to analyze the most probable off-target sites for the sgRNA designed. The top two off-target sites were analyzed using a PCR based approach that works in the same way as junctional PCR analysis. If undesired integration occurs, then the primer P1 and P2 will generate a 1.6 kb PCR fragment (Figure 13). However, we cannot confidently conclude the absence of integration if this primer pair did not generate any product. A potential reason for lack of product is that the primer P1 may be defective at amplifying the PCR product even if undesired integration occurs. The primer P2 is the same as the reverse primer used in left arm junctional PCR in Figure 11 so it is effective at amplifying the integrated product. To exclude this possibility, primer P3, which binds downstream of the off-target site, was designed so that if P1 and P3 can generate a PCR product while P1 and P2 cannot, then we can confidently claim no integration occurred in this off-target site. Results demonstrated that both off-target sites did not have any incorrect *LacZ* integration. In addition, mismatches inside the off-target sites were not identified, which suggests no indel from NHEJ took place in the off-target sites as well.

In the analysis on effects of HD-Ad K18CFTR GGTA1 transduction, the junctional PCR analysis proves that the human CFTR transgene successfully achieved stable integration at the designed locus (Figure 15). In addition, the mRNA expression of CFTR transgene was also verified (Figure 16). Comparing with the no CRISPR control, the HD-Ad K18CFTR GGTA1 transduced cells have significantly more folds of CFTR mRNA expression after 9 passages. The relative expression of CFTR increases because during each passage, the number of vectors retaining in the transduced cells gets diluted. Without the assistance of CRISPR, the rate of integration is intrinsically low. Therefore, the relative expression increases dramatically as the retained vectors decreases over the passages. The difference of mRNA expression between 20 MOI and 50 MOI of initial transduction is noticeable in the early passages. This is due to with higher dosage of vectors introduced, the number of retained plasmid without integration is higher. However, the relative difference between these two groups reduces as the cells were passed for more generations. Since in the previous aim, the integration efficiency between the 20 MOI group and 50 MOI group only differ by 1%, this can be used to explain why CFTR mRNA levels are comparable between the 20 and 50 MOI groups after passage 9. Unlike the previous aim which focuses on HD-Ad UbCLacZ-Neo GGTA1, it is very challenging to quantify the integration efficiency of CFTR transgene due to the lack of methods to distinguish between the individual cells that are hCFTR positive versus hCFTR negative. However, a sequencing approach may be used to attempt quantifying integration efficiency for CFTR. The region with predicted integration can be enriched and these enriched regions will all be sequenced. From the sequencing results, it is possible to determine how many of these regions contain the CFTR insert and divide by total number of fragments sequenced.

Transgene hCFTR protein expression was analyzed. There are two bands detected in the 20 MOI and 50 MOI lane and they correspond to the C band and B band during typical CFTR western blot detections. The C bands represents fully matured CFTR protein while B band represents immature CFTR. The heavy C band and relatively lighter B bands in the 20 MOI and 50 MOI suggest most of the CFTR produced are fully matured. On the other hand, B band is not visible under the +Dox lane which means almost all CFTR proteins synthesized in 8-3-7 cells after doxycycline induction are fully matured. The reason why less total proteins are loaded in the -Dox and +Dox lanes is because this cell line is known to induce a very high level of CFTR expression upon doxycycline induction. Loading too much sample may make identifying the real position of the CFTR band and distinguishing between the C and B bands challenging.

A major concern for this strategy is that the mutant CFTR protein is still present even after CFTR transgene integration in the *GGTA1* locus. Therefore, the mutant CFTR protein can still be expressed and may be potentially problematic for the cell even if a correct CFTR transgene is integrated into the genome. Studies have shown that the expression of mutant ΔF508 CFTR mutant protein can form aggregates inside the cell (149, 150). To avoid this issue, an additional sgRNA sequence that designed to knockout the mutant CFTR locus can be added to the design of our HD-Ad vector. This is relatively easy to accomplish because of the large packaging capacity of the HD-Ad vector. In addition, it is possible to change our *GGTA1* target locus for integration into the CFTR locus to ensure the mutant CFTR gene can be inactivated.

In the future, a very important and useful study to be performed is to analyze if CFTR transgene integration can lead to functional correction in the target cells. Even though CFTR mRNA can be detected after multiple passages of transduced IPEC-J2 cells, it is still important to verify whether this expression is enough to achieve a functional benefit. To perform this experiment, a CFTR

knockout pig cell line must be generated. After this cell line becomes available, the cells can be grown in air liquid interface and the CFTR function can be tested through Ussing chamber analysis or iodide efflux assay. In addition, Scr7 treatment can be used post HD-Ad K18CFTR GGTA1 transduction to see if the expression of CFTR can be enhanced after multiple passages similar to the approach used in HD-Ad UbCLacZ-Neo GGTA1 integration efficiency enhancement. Furthermore, additional genes that target the HDR pathway can be investigated and incorporated in the HD-Ad vector to further enhance the integration efficiency. Proteins such as CtIP was shown to promote homologous recombination (151). If this protein can be modified to be constitutively expressed, HDR will be enhanced which can then lead to increase in integration efficiency. Such modified proteins can be included in the HD-Ad vector and tested for its effects.

Another direction worth investigating is the utilization of whole genome sequencing in determining potential off-target sites. Since the CRISPR/Cas9 system can tolerate mismatches at the 5' end of the guide sequence upon target recognition, Cas9 mediated double stranded break may occur in off-target sites that are not perfectly complementary to the guide sequence (152). The strategy used in this study for determining off-target sites relies on computer programs for screening sequences flanked by PAM site that have high similarities with the on-target. Then, the top predicted off-target sites were verified through PCR based approach. In comparison, whole genome sequencing offers an unbiased method to determine unwanted editing across the entire genome. After treatment with the CRISPR/Cas9 system, the edited genomic DNA can be extracted and deep sequenced to determine all the single nucleotide variants and small indels compared with reference genome. To identify whether these changes in genome are resulted from CRISPR/Cas9 editing, the location of all the single nucleotide variants and indels can be

validated to see if they are inside of any putative off-target sites determined by software (153, 154).

The successful completion of this project suggests that our novel strategy can indeed mediate stable precise transgene integration and transgene expression in *in vitro* settings. The next big step for our study after functional analysis is to translate the *in vitro* work into *in vivo*. As mentioned previously, pigs are considered as one of the best animal model for mimicking human CF conditions. Therefore, we plan to utilize our strategy in CF pigs to determine if stable transgene integration and CFTR expression can be achieved in the pig airway. Even though translating into animal study is a big challenge for us, we believe our vector system has many possibilities and potentials that will enable us to achieve our ultimate goals.

7. References

1. **Cohen-Cymberknoh M, Shoseyov D, Kerem E.** 2011. Managing cystic fibrosis: strategies that increase life expectancy and improve quality of life. American journal of respiratory and critical care medicine **183:**1463-1471.
2. **Stenekes SJ, Hughes A, Grégoire M-C, Frager G, Robinson WM, McGrath PJ.** Frequency and Self-Management of Pain, Dyspnea, and Cough in Cystic Fibrosis. Journal of Pain and Symptom Management **38:**837-848.
3. **Andersen DH.** 1938. Cystic fibrosis of the pancreas and its relation to celiac disease: A clinical and pathologic study. American Journal of Diseases of Children **56:**344-399.
4. **Elborn JS.** Cystic fibrosis. The Lancet **388:**2519-2531.
5. **Davis PB.** 2006. Cystic fibrosis since 1938. American journal of respiratory and critical care medicine **173:**475-482.
6. **Yan Z, Stewart ZA, Sinn PL, Olsen JC, Hu J, McCray PB, Jr., Engelhardt JF.** 2015. Ferret and pig models of cystic fibrosis: prospects and promise for gene therapy. Human gene therapy. Clinical development **26:**38-49.
7. **Rowe SM, Miller S, Sorscher EJ.** 2005. Cystic fibrosis. The New England journal of medicine **352:**1992-2001.
8. **Dodge JA, Lewis PA, Stanton M, Wilsher J.** 2007. Cystic fibrosis mortality and survival in the UK: 1947-2003. The European respiratory journal **29:**522-526.
9. **Jat KR, Walia DK, Khairwa A.** 2015. Anti-IgE therapy for allergic bronchopulmonary aspergillosis in people with cystic fibrosis. The Cochrane database of systematic reviews**:**Cd010288.
10. **Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al.** 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science **245:**1066-1073.

11. **Bear CE, Li CH, Kartner N, Bridges RJ, Jensen TJ, Ramjeesingh M, Riordan JR.** 1992. Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR). Cell **68:**809-818.
12. **Riordan JR.** 2008. CFTR function and prospects for therapy. Annual review of biochemistry **77:**701-726.
13. **Mall MA, Galietta LJ.** 2015. Targeting ion channels in cystic fibrosis. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society **14:**561-570.
14. **Akabas MH.** 2000. Cystic fibrosis transmembrane conductance regulator. Structure and function of an epithelial chloride channel. The Journal of biological chemistry **275:**3729-3732.
15. **Boucher RC.** 2007. Airway surface dehydration in cystic fibrosis: pathogenesis and therapy. Annual review of medicine **58:**157-170.
16. **Gabriel SE, Clarke LL, Boucher RC, Stutts MJ.** 1993. CFTR and outward rectifying chloride channels are distinct proteins with a regulatory relationship. Nature **363:**263-268.
17. **Mall M, Hipper A, Greger R, Kunzelmann K.** 1996. Wild type but not deltaF508 CFTR inhibits Na+ conductance when coexpressed in Xenopus oocytes. FEBS letters **381:**47-52.
18. **Matsui H, Grubb BR, Tarran R, Randell SH, Gatzy JT, Davis CW, Boucher RC.** 1998. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. Cell **95:**1005-1015.
19. **Flume PA, Mogayzel PJ, Jr., Robinson KA, Rosenblatt RL, Quittell L, Marshall BC.** 2010. Cystic fibrosis pulmonary guidelines: pulmonary complications: hemoptysis and pneumothorax. American journal of respiratory and critical care medicine **182:**298-306.
20. **Pezzulo AA, Tang XX, Hoegger MJ, Abou Alaiwa MH, Ramachandran S, Moninger TO, Karp PH, Wohlford-Lenane CL, Haagsman HP, van Eijk M, Banfi B, Horswill AR, Stoltz DA, McCray PB, Jr., Welsh MJ, Zabner J.** 2012. Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. Nature **487:**109-113.
21. **Lyczak JB, Cannon CL, Pier GB.** 2002. Lung Infections Associated with Cystic Fibrosis. Clinical Microbiology Reviews **15:**194-222.
22. **Hartl D, Gaggar A, Bruscia E, Hector A, Marcos V, Jung A, Greene C, McElvaney G, Mall M, Döring G.** Innate immunity in cystic fibrosis lung disease. Journal of Cystic Fibrosis **11:**363-382.
23. April 25, 2011, posting date. Cystic Fibrosis Mutation Database [Online.]
24. **Wang Y, Wrennall JA, Cai Z, Li H, Sheppard DN.** 2014. Understanding how cystic fibrosis mutations disrupt CFTR function: from single molecules to animal models. The international journal of biochemistry & cell biology **52:**47-57.
25. **Rafeeq MM, Murad HAS.** 2017. Cystic fibrosis: current therapeutic targets and future approaches. Journal of translational medicine **15:**84.
26. **Okiyoneda T, Barriere H, Bagdany M, Rabeh WM, Du K, Hohfeld J, Young JC, Lukacs GL.** 2010. Peripheral protein quality control removes unfolded CFTR from the plasma membrane. Science **329:**805-810.
27. **Hirtz S, Gonska T, Seydewitz HH, Thomas J, Greiner P, Kuehr J, Brandis M, Eichler I, Rocha H, Lopes AI, Barreto C, Ramalho A, Amaral MD, Kunzelmann K, Mall M.** 2004. CFTR Cl- channel function in native human colon correlates with the genotype and phenotype in cystic fibrosis. Gastroenterology **127:**1085-1095.
28. **Gong X, Linsdell P.** 2004. Maximization of the rate of chloride conduction in the CFTR channel pore by ion-ion interactions. Archives of biochemistry and biophysics **426:**78-82.
29. **Sheppard DN, Ostedgaard LS, Winter MC, Welsh MJ.** 1995. Mechanism of dysfunction of two nucleotide binding domain mutations in cystic fibrosis transmembrane conductance regulator that are associated with pancreatic sufficiency. The EMBO journal **14:**876-883.
30. **Kim SJ, Skach WR.** 2012. Mechanisms of CFTR Folding at the Endoplasmic Reticulum. Frontiers in pharmacology **3:**201.

31. **Beaudet AL.** 1992. Genetic testing for cystic fibrosis. Pediatric clinics of North America **39:**213-228.
32. **Fuchs HJ, Borowitz DS, Christiansen DH, Morris EM, Nash ML, Ramsey BW, Rosenstein BJ, Smith AL, Wohl ME.** 1994. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. The New England journal of medicine **331:**637-642.
33. **Robinson M, Regnis JA, Bailey DL, King M, Bautovich GJ, Bye PT.** 1996. Effect of hypertonic saline, amiloride, and cough on mucociliary clearance in patients with cystic fibrosis. American journal of respiratory and critical care medicine **153:**1503-1509.
34. **Elkins MR, Robinson M, Rose BR, Harbour C, Moriarty CP, Marks GB, Belousova EG, Xuan W, Bye PT.** 2006. A controlled trial of long-term inhaled hypertonic saline in patients with cystic fibrosis. The New England journal of medicine **354:**229-240.
35. **Ho SA, Lee TW, Denton M, Conway SP, Brownlee KG.** 2009. Regimens for eradicating early Pseudomonas aeruginosa infection in children do not promote antibiotic resistance in this organism. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society **8:**43-46.
36. **Hoiby N, Ciofu O, Bjarnsholt T.** 2010. Pseudomonas aeruginosa biofilms in cystic fibrosis. Future microbiology **5:**1663-1674.
37. **Ramsey BW, Pepe MS, Quan JM, Otto KL, Montgomery AB, Williams-Warren J, Vasiljev KM, Borowitz D, Bowman CM, Marshall BC, Marshall S, Smith AL.** 1999. Intermittent administration of inhaled tobramycin in patients with cystic fibrosis. Cystic Fibrosis Inhaled Tobramycin Study Group. The New England journal of medicine **340:**23-30.
38. **Elizur A, Cannon CL, Ferkol TW.** 2008. Airway inflammation in cystic fibrosis. Chest **133:**489-495.
39. **Saiman L, Marshall BC, Mayer-Hamblett N, Burns JL, Quittner AL, Cibene DA, Coquillette S, Fieberg AY, Accurso FJ, Campbell PW, 3rd.** 2003. Azithromycin in patients with cystic fibrosis chronically infected with Pseudomonas aeruginosa: a randomized controlled trial. Jama **290:**1749-1756.
40. **Konstan MW, Byard PJ, Hoppel CL, Davis PB.** 1995. Effect of high-dose ibuprofen in patients with cystic fibrosis. The New England journal of medicine **332:**848-854.
41. **Kapoor H, Koolwal A, Singh A.** 2014. Ivacaftor: a novel mutation modulating drug. Journal of clinical and diagnostic research : JCDR **8:**Se01-05.
42. **Vauthier V, Housset C, Falguieres T.** 2017. Targeted pharmacotherapies for defective ABC transporters. Biochemical pharmacology.
43. **Accurso FJ, Rowe SM, Clancy JP, Boyle MP, Dunitz JM, Durie PR, Sagel SD, Hornick DB, Konstan MW, Donaldson SH, Moss RB, Pilewski JM, Rubenstein RC, Uluer AZ, Aitken ML, Freedman SD, Rose LM, Mayer-Hamblett N, Dong Q, Zha J, Stone AJ, Olson ER, Ordonez CL, Campbell PW, Ashlock MA, Ramsey BW.** 2010. Effect of VX-770 in persons with cystic fibrosis and the G551D-CFTR mutation. The New England journal of medicine **363:**1991-2003.
44. **Davies JC.** 2006. Gene and cell therapy for cystic fibrosis. Paediatric respiratory reviews **7 Suppl 1:**S163-165.
45. **Pittman JE, Ferkol TW.** 2015. The Evolution of Cystic Fibrosis Care. Chest **148:**533-542.
46. **Griesenbach U, Geddes DM, Alton EW.** 2004. Gene therapy for cystic fibrosis: an example for lung gene therapy. Gene therapy **11 Suppl 1:**S43-50.
47. **Davies JC.** 2006. Gene and cell therapy for cystic fibrosis. Paediatric respiratory reviews **7, Supplement 1:**S163-S165.
48. **Kreindler JL.** 2010. Cystic fibrosis: exploiting its genetic basis in the hunt for new therapies. Pharmacology & therapeutics **125:**219-229.
49. **Harvey BG, Leopold PL, Hackett NR, Grasso TM, Williams PM, Tucker AL, Kaner RJ, Ferris B, Gonda I, Sweeney TD, Ramalingam R, Kovesdi I, Shak S, Crystal RG.** 1999. Airway epithelial CFTR

mRNA expression in cystic fibrosis patients after repetitive administration of a recombinant adenovirus. The Journal of clinical investigation **104:**1245-1255.

50. **Malik AK, Monahan PE, Allen DL, Chen BG, Samulski RJ, Kurachi K.** 2000. Kinetics of recombinant adeno-associated virus-mediated gene transfer. Journal of virology **74:**3555-3565.
51. **Moss RB, Milla C, Colombo J, Accurso F, Zeitlin PL, Clancy JP, Spencer LT, Pilewski J, Waltz DA, Dorkin HL, Ferkol T, Pian M, Ramsey B, Carter BJ, Martin DB, Heald AE.** 2007. Repeated aerosolized AAV-CFTR for treatment of cystic fibrosis: a randomized placebo-controlled phase 2B trial. Human gene therapy **18:**726-732.
52. **Kennedy MJ.** 2002. Current status of gene therapy for cystic fibrosis pulmonary disease. American journal of respiratory medicine : drugs, devices, and other interventions **1:**349-360.
53. **Flotte TR, Zeitlin PL, Reynolds TC, Heald AE, Pedersen P, Beck S, Conrad CK, Brass-Ernst L, Humphries M, Sullivan K, Wetzel R, Taylor G, Carter BJ, Guggino WB.** 2003. Phase I trial of intranasal and endobronchial administration of a recombinant adeno-associated virus serotype 2 (rAAV2)-CFTR vector in adult cystic fibrosis patients: a two-part clinical study. Human gene therapy **14:**1079-1088.
54. **Moss RB, Rodman D, Spencer LT, Aitken ML, Zeitlin PL, Waltz D, Milla C, Brody AS, Clancy JP, Ramsey B, Hamblett N, Heald AE.** 2004. Repeated adeno-associated virus serotype 2 aerosol-mediated cystic fibrosis transmembrane regulator gene transfer to the lungs of patients with cystic fibrosis: a multicenter, double-blind, placebo-controlled trial. Chest **125:**509-521.
55. **Sander JD, Joung JK.** 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature biotechnology **32:**347-355.
56. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.
57. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science **322:**1843-1845.
58. **Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV.** 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology direct **1:**7.
59. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. Journal of bacteriology **169:**5429-5433.
60. **Haft DH, Selengut J, Mongodin EF, Nelson KE.** 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. **1:**e60.
61. **Bolotin A, Quinquis B, Sorokin A, Ehrlich SD.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology (Reading, England) **151:**2551-2561.
62. **Jansen R, Embden JD, Gaastra W, Schouls LM.** 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Molecular microbiology **43:**1565-1575.
63. **Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV.** 2015. An updated evolutionary classification of CRISPR-Cas systems. Nature reviews. Microbiology **13:**722-736.
64. **Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nature reviews. Microbiology **9:**467-477.
65. **Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E.** 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature **471:**602-607.

66. **Nemudryi AA, Valetdinova KR, Medvedev SP, Zakian SM.** 2014. TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. Acta Naturae **6:**19-40.
67. **Heler R, Wright AV, Vucelja M, Bikard D, Doudna JA, Marraffini LA.** 2017. Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. Molecular cell **65:**168-175.
68. **Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA.** 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature **519:**199-202.
69. **Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C.** 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (Reading, England) **155:**733-740.
70. **Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K.** 2010. Transcription, processing and function of CRISPR cassettes in Escherichia coli. Molecular microbiology **77:**1367-1379.
71. **Djordjevic M, Djordjevic M, Severinov K.** 2012. CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs. Biology direct **7:**24.
72. **Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP.** 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell **139:**945-956.
73. **Kunin V, Sorek R, Hugenholtz P.** 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome biology **8:**R61.
74. **Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F.** 2013. Multiplex genome engineering using CRISPR/Cas systems. Science **339:**819-823.
75. **Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E.** 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337:**816-821.
76. **Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA.** 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature **507:**62-67.
77. **Feng Y, Chen C, Han Y, Chen Z, Lu X, Liang F, Li S, Qin W, Lin S.** 2016. Expanding CRISPR/Cas9 Genome Editing Capacity in Zebrafish Using SaCas9. G3 (Bethesda, Md.) **6:**2517-2521.
78. **Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. Journal of bacteriology **190:**1390-1400.
79. **Gasiunas G, Barrangou R, Horvath P, Siksnys V.** 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. **109:**E2579-2586.
80. **Zhang JP, Li XL, Li GH, Chen W, Arakaki C, Botimer GD, Baylink D, Zhang L, Wen W, Fu YW, Xu J, Chun N, Yuan W, Cheng T, Zhang XB.** 2017. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome biology **18:**35.
81. **Iliakis G, Wang H, Perrault AR, Boecker W, Rosidi B, Windhofer F, Wu W, Guan J, Terzoudi G, Pantelias G.** 2004. Mechanisms of DNA double strand break repair and chromosome aberration formation. Cytogenetic and genome research **104:**14-20.
82. **Gu J, Lieber MR.** 2008. Mechanistic flexibility as a conserved theme across 3 billion years of nonhomologous DNA end-joining. Genes & development **22:**411-415.
83. **Shen S, Loh TJ, Shen H, Zheng X, Shen H.** 2017. CRISPR as a strong gene editing tool. BMB reports **50:**20-24.
84. **Aravind L, Koonin EV.** 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome research **11:**1365-1374.
85. **Zhu S, Peng A.** 2016. Non-homologous end joining repair in Xenopus egg extract. Scientific reports **6:**27797.

86. **Riballo E, Woodbine L, Stiff T, Walker SA, Goodarzi AA, Jeggo PA.** 2009. XLF-Cernunnos promotes DNA ligase IV–XRCC4 re-adenylation following ligation. Nucleic Acids Research **37:**482-492.
87. **Cortez C** 2015, posting date. CRISPR 101: Homology Directed Repair. [Online.]
88. **Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM.** 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics **194:**1029-1035.
89. **Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F.** 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology **31:**827-832.
90. **Paix A, Folkmann A, Seydoux G.** 2017. Precision genome editing using CRISPR-Cas9 and linear repair templates in C. elegans. Methods (San Diego, Calif.).
91. **Beumer KJ, Trautman JK, Mukherjee K, Carroll D.** 2013. Donor DNA Utilization During Gene Targeting with Zinc-Finger Nucleases. G3 (Bethesda, Md.) **3:**657-664.
92. **Maruyama T, Dougan SK, Truttmann M, Bilate AM, Ingram JR, Ploegh HL.** 2015. Inhibition of non-homologous end joining increases the efficiency of CRISPR/Cas9-mediated precise [TM: inserted] genome editing. Nature biotechnology **33:**538-542.
93. **Heyer WD, Ehmsen KT, Liu J.** 2010. Regulation of homologous recombination in eukaryotes. Annual review of genetics **44:**113-139.
94. **Yang D, Scavuzzo MA, Chmielowiec J, Sharp R, Bajic A, Borowiak M.** 2016. Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. Scientific reports **6:**21264.
95. **Rowe WP, Huebner RJ, Gilmore LK, Parrott RH, Ward TG.** 1953. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.) **84:**570-573.
96. **Campos SK, Barry MA.** 2007. Current Advances and Future Challenges in Adenoviral Vector Biology and Targeting. Current gene therapy **7:**189-204.
97. **Saha B, Wong CM, Parks RJ.** 2014. The Adenovirus Genome Contributes to the Structural Stability of the Virion. Viruses **6:**3563-3583.
98. **Musher DM.** 2003. How contagious are common respiratory tract infections? The New England journal of medicine **348:**1256-1266.
99. **Yabe Y, Trentin JJ, Taylor G.** 1962. Cancer induction in hamsters by human type 12 adenovirus. Effect of age and of virus dose. Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.) **111:**343-344.
100. **Frisch SM, Mymryk JS.** 2002. Adenovirus-5 E1A: paradox and paradigm. Nat Rev Mol Cell Biol **3:**441-452.
101. **Lowe SW, Ruley HE.** 1993. Stabilization of the p53 tumor suppressor is induced by adenovirus 5 E1A and accompanies apoptosis. Genes & development **7:**535-545.
102. **Horwitz MS.** 2004. Function of adenovirus E3 proteins and their interactions with immunoregulatory cell proteins. The journal of gene medicine **6 Suppl 1:**S172-183.
103. **Weitzman MD.** 2005. Functions of the adenovirus E4 proteins and their impact on viral vectors. Frontiers in bioscience : a journal and virtual library **10:**1106-1117.
104. **Engler JA, Hong JS.** 2007. Characterization of the adenovirus fiber protein. Methods in molecular medicine **131:**281-298.
105. **Mirza MA, Weber J.** 1982. Structure of adenovirus chromatin. Biochimica et biophysica acta **696:**76-86.
106. **Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, Hong JS, Horwitz MS, Crowell RL, Finberg RW.** 1997. Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. Science **275:**1320-1323.

107. **Wickham TJ, Mathias P, Cheresh DA, Nemerow GR.** 1993. Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. Cell **73:**309-319.
108. **Wiethoff CM, Wodrich H, Gerace L, Nemerow GR.** 2005. Adenovirus protein VI mediates membrane disruption following capsid disassembly. Journal of virology **79:**1992-2000.
109. **Leopold PL, Ferris B, Grinberg I, Worgall S, Hackett NR, Crystal RG.** 1998. Fluorescent virions: dynamic tracking of the pathway of adenoviral gene transfer vectors in living cells. Human gene therapy **9:**367-378.
110. **Ginn SL, Alexander IE, Edelstein ML, Abedi MR, Wixon J.** 2013. Gene therapy clinical trials worldwide to 2012 - an update. The journal of gene medicine **15:**65-77.
111. **Vetrini F, Ng P.** 2010. Gene Therapy with Helper-Dependent Adenoviral Vectors: Current Advances and Future Perspectives. Viruses **2:**1886-1917.
112. **Bett AJ, Prevec L, Graham FL.** 1993. Packaging capacity and stability of human adenovirus type 5 vectors. Journal of virology **67:**5911-5921.
113. **Kay MA, Glorioso JC, Naldini L.** 2001. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. Nature medicine **7:**33-40.
114. **Yang Y, Li Q, Ertl HC, Wilson JM.** 1995. Cellular and humoral immune responses to viral antigens create barriers to lung-directed gene therapy with recombinant adenoviruses. Journal of virology **69:**2004-2015.
115. **Sondhi D, Stiles KM, De BP, Crystal RG.** 2017. Genetic Modification of the Lung Directed Toward Treatment of Human Disease. Human gene therapy **28:**3-84.
116. **Chirmule N, Hughes JV, Gao GP, Raper SE, Wilson JM.** 1998. Role of E4 in eliciting CD4 T-cell and B-cell responses to adenovirus vectors delivered to murine and nonhuman primate lungs. Journal of virology **72:**6138-6145.
117. **Parks RJ, Chen L, Anton M, Sankar U, Rudnicki MA, Graham FL.** 1996. A helper-dependent adenovirus vector system: removal of helper virus by Cre-mediated excision of the viral packaging signal. Proc. Natl. Acad. Sci. U. S. A. **93:**13565-13570.
118. **Toietta G, Koehler DR, Finegold MJ, Lee B, Hu J, Beaudet AL.** 2003. Reduced inflammation and improved airway expression using helper-dependent adenoviral vectors with a K18 promoter. Molecular therapy : the journal of the American Society of Gene Therapy **7:**649-658.
119. **Parks RJ, Graham FL.** 1997. A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. Journal of virology **71:**3293-3298.
120. **Parks RJ, Bramson JL, Wan Y, Addison CL, Graham FL.** 1999. Effects of Stuffer DNA on Transgene Expression from Helper-Dependent Adenovirus Vectors. Journal of virology **73:**8027-8034.
121. **Palmer D, Ng P.** 2003. Improved system for helper-dependent adenoviral vector production. Molecular therapy : the journal of the American Society of Gene Therapy **8:**846-852.
122. **Griesenbach U, Alton EW.** 2012. Progress in gene and cell therapy for cystic fibrosis lung disease. Current pharmaceutical design **18:**642-662.
123. **Liu X, Luo M, Trygg C, Yan Z, Lei-Butters DC, Smith CI, Fischer AC, Munson K, Guggino WB, Bunnell BA, Engelhardt JF.** 2007. Biological Differences in rAAV Transduction of Airway Epithelia in Humans and in Old World Non-human Primates. Molecular therapy : the journal of the American Society of Gene Therapy **15:**2114-2123.
124. **Wilke M, Buijs-Offerman RM, Aarbiou J, Colledge WH, Sheppard DN, Touqui L, Bot A, Jorna H, de Jonge HR, Scholte BJ.** 2011. Mouse models of cystic fibrosis: phenotypic analysis and research applications. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society **10 Suppl 2:**S152-171.
125. **Fisher JT, Zhang Y, Engelhardt JF.** 2011. Comparative biology of cystic fibrosis animal models. Methods in molecular biology (Clifton, N.J.) **742:**311-334.
126. **Clarke LL, Grubb BR, Yankaskas JR, Cotton CU, McKenzie A, Boucher RC.** 1994. Relationship of a non-cystic fibrosis transmembrane conductance regulator-mediated chloride conductance to organ-level disease in Cftr(-/-) mice. Proc. Natl. Acad. Sci. U. S. A. **91:**479-483.

127. **Pack RJ, Al-Ugaily LH, Morris G, Widdicombe JG.** 1980. The distribution and structure of cells in the tracheal epithelium of the mouse. Cell and tissue research **208:**65-84.
128. **Liu X, Luo M, Zhang L, Ding W, Yan Z, Engelhardt JF.** 2007. Bioelectric properties of chloride channels in human, pig, ferret, and mouse airway epithelia. American journal of respiratory cell and molecular biology **36:**313-323.
129. **Rogers CS, Hao Y, Rokhlina T, Samuel M, Stoltz DA, Li Y, Petroff E, Vermeer DW, Kabel AC, Yan Z, Spate L, Wax D, Murphy CN, Rieke A, Whitworth K, Linville ML, Korte SW, Engelhardt JF, Welsh MJ, Prather RS.** 2008. Production of CFTR-null and CFTR-DeltaF508 heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. The Journal of clinical investigation **118:**1571-1577.
130. **Brosnahan AJ, Brown DR.** 2012. Porcine IPEC-J2 intestinal epithelial cells in microbiological investigations. Veterinary microbiology **156:**229-237.
131. **Yang YG, Sykes M.** 2007. Xenotransplantation: current status and a perspective on the future. Nature reviews. Immunology **7:**519-531.
132. **Galili U, Tibell A, Samuelsson B, Rydberg L, Groth CG.** 1995. Increased anti-Gal activity in diabetic patients transplanted with fetal porcine islet cell clusters. Transplantation **59:**1549-1556.
133. **Kim JW, Kim HM, Lee SM, Kang MJ.** 2012. Porcine Knock-in Fibroblasts Expressing hDAF on alpha-1,3-Galactosyltransferase (GGTA1) Gene Locus. Asian-Australasian journal of animal sciences **25:**1473-1480.
134. **Dai Y, Vaught TD, Boone J, Chen SH, Phelps CJ, Ball S, Monahan JA, Jobst PM, McCreath KJ, Lamborn AE, Cowell-Lucero JL, Wells KD, Colman A, Polejaeva IA, Ayares DL.** 2002. Targeted disruption of the alpha1,3-galactosyltransferase gene in cloned pigs. Nature biotechnology **20:**251-255.
135. **Lai L, Kolber-Simonds D, Park KW, Cheong HT, Greenstein JL, Im GS, Samuel M, Bonk A, Rieke A, Day BN, Murphy CN, Carter DB, Hawley RJ, Prather RS.** 2002. Production of alpha-1,3-galactosyltransferase knockout pigs by nuclear transfer cloning. Science **295:**1089-1092.
136. **Burn SF.** 2012. Detection of beta-galactosidase activity: X-gal staining. Methods in molecular biology (Clifton, N.J.) **886:**241-250.
137. 2017, posting date. PCR Protocol for Phusion® High-Fidelity DNA Polymerase. [Online.]
138. **Sandig V, Youil R, Bett AJ, Franlin LL, Oshima M, Maione D, Wang F, Metzker ML, Savino R, Caskey CT.** 2000. Optimization of the helper-dependent adenovirus system for production and potency in vivo. Proc. Natl. Acad. Sci. U. S. A. **97:**1002-1007.
139. **Koehler DR, Chow YH, Plumb J, Wen Y, Rafii B, Belcastro R, Haardt M, Lukacs GL, Post M, Tanswell AK, Hu J.** 2000. A human epithelium-specific vector optimized in rat pneumocytes for lung gene therapy. Pediatric research **48:**184-190.
140. **Koehler DR, Sajjan U, Chow YH, Martin B, Kent G, Tanswell AK, McKerlie C, Forstner JF, Hu J.** 2003. Protection of Cftr knockout mice from acute lung infection by a helper-dependent adenoviral vector expressing Cftr in airway epithelia. Proc. Natl. Acad. Sci. U. S. A. **100:**15364-15369.
141. **Ng P, Parks RJ, Graham FL.** 2002. Preparation of helper-dependent adenoviral vectors. Methods in molecular medicine **69:**371-388.
142. **Graham FL, Prevec L.** 1991. Manipulation of adenovirus vectors. Methods in molecular biology (Clifton, N.J.) **7:**109-128.
143. **Wang X, Wang Y, Huang H, Chen B, Chen X, Hu J, Chang T, Lin RJ, Yee JK.** 2014. Precise Gene Modification Mediated by TALEN and Single-Stranded Oligodeoxynucleotides in Human Cells. PLoS One **9**.
144. **Gruenert DC, Willems M, Cassiman JJ, Frizzell RA.** 2004. Established cell lines used in cystic fibrosis research. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society **3 Suppl 2:**191-196.

145. **Ye L, Chan S, Chow YH, Tsui LC, Hu J.** 2001. Regulated expression of the human CFTR gene in epithelial cells. Molecular therapy : the journal of the American Society of Gene Therapy **3:**723-733.
146. **Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, Kim MK, Shin BA, Choi SY.** 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One **6:**e18556.
147. **McLenachan S, Sarsero JP, Ioannou PA.** 2007. Flow-cytometric analysis of mouse embryonic stem cell lipofection using small and large DNA constructs. Genomics **89:**708-720.
148. **Dannhoffer L, Blouquit-Laye S, Regnier A, Chinet T.** 2009. Functional properties of mixed cystic fibrosis and normal bronchial epithelial cell cultures. American journal of respiratory cell and molecular biology **40:**717-723.
149. **Du K, Karp PH, Ackerley C, Zabner J, Keshavjee S, Cutz E, Yeger H.** 2015. Aggregates of mutant CFTR fragments in airway epithelial cells of CF lungs: New pathologic observations. Journal of Cystic Fibrosis **14:**182-193.
150. **Luciani A, Villella VR, Esposito S, Brunetti-Pierri N, Medina D, Settembre C, Gavina M, Pulze L, Giardino I, Pettoello-Mantovani M, D'Apolito M, Guido S, Masliah E, Spencer B, Quaratino S, Raia V, Ballabio A, Maiuri L.** 2010. Defective CFTR induces aggresome formation and lung inflammation in cystic fibrosis through ROS-mediated autophagy inhibition. Nat Cell Biol **12:**863-875.
151. **Sartori AA, Lukas C, Coates J, Mistrik M, Fu S, Bartek J, Baer R, Lukas J, Jackson SP.** 2007. Human CtIP promotes DNA end resection. Nature **450:**509-514.
152. **Pattanayak V.** 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. **31:**839-843.
153. **Smith C, Gore A, Yan W, Abalde-Atristain L, Li Z, He C, Wang Y, Brodsky RA, Zhang K, Cheng L, Ye Z.** 2014. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. Cell stem cell **15:**12-13.
154. **Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, Hwang J, Kim JI, Kim JS.** 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nature methods **12:**237-243, 231 p following 243.