Enhancing the Efficiency of CRISPR/Cas9 Precise Gene Editing for Cystic Fibrosis Gene Therapy

by

Kyle E. Seigel

A thesis submitted in conformity with the requirements for the degree of Master of Science
Department of Laboratory Medicine and Pathobiology
University of Toronto

© Copyright by Kyle E. Seigel, 2018
Enhancing the Efficiency of CRISPR/Cas9 Precise Gene Editing for Cystic Fibrosis Gene Therapy

Kyle E. Seigel

Master of Science

Department of Laboratory Medicine and Pathobiology
University of Toronto

2018

Abstract

Cystic fibrosis (CF) is the most common cause of chronic obstructive lung disease in children and young adults, yet there is no cure for this disease. Intimations of curative strategies have emerged from gene therapy, but these approaches have failed in clinical trials due to insufficient safety and efficacy profiles. The molecular technology necessary to overcome preeminent challenges in gene therapy may reside in programmable nucleases – namely CRISPR/Cas9 – which allow for site-specific genomic integration of therapeutic transgenes. However, the clinical utility of these technologies is limited by inherently low efficiencies. Here, we use flow cytometry and PCR assays to demonstrate the efficacy of perturbing DNA repair to overcome such limitations. We demonstrate that overexpressing CtIP and EXO1-4D can enhance the efficiency of CRISPR/Cas9-directed transgene integration by ~3- and ~6-fold, respectively. These discoveries may provide necessary impetus for translating gene therapies into clinical realities for genetic diseases such as CF.
Acknowledgements

My thesis work and development as a scientist would not have been possible without the professional wisdom, scientific expertise and personal support of a number of individuals. I would first like to extend my utmost thanks and gratitude to my supervisor, Dr. Jim Hu. Your expertise were foundational to the development of this project, and your continual encouragement and support helped guide me through the adversities of science. I appreciate your persistent attention to my experiments and challenges in the lab, regardless of how many other things you had going on. This is a true testament to how much you care for your students; for their personal and professional development as scientists. I would also like to thank my committee members: Dr. Martha Brown, Dr. Grant Brown, and Dr. Andras Nagy. The diverse range of scientific expertise that each of you contributed were instrumental to the progression of my project, and I am grateful for your many hours of undivided attention. A special thanks to Dr. Martha Brown, who continually made herself available to discuss my research, as well as to provide technical and moral support. To Dr. Harry Elsholtz, thank you for your professional guidance and support throughout the duration of my studies.

My fellow lab members were also extremely helpful in the development of this project. Dr. Huibi Cao, Linda Li, Cathleen Duan, Fushan Shi, Emily Xia, Amy Zhang, Peter Zhou, Alexandra Georgiou, Leo Yang, and Randolph Kissoon have all contributed to this work in some capacity. Thank you for your helpful contributions in lab meetings, technical support at the bench, and continual friendships. I would also like to extend my gratitude to all the staff in The SickKids-UHN Flow and Mass Cytometry Facility at PGCRL. In particular, thank you Dr. Sherry Zhao and Dr. Emily Reddy for tremendous technical support in designing experiments, operating flow cytometers and analyzing data.

My heartfelt thanks to my family, without whom none of this work would have been possible. To my Mom and Dad, and my brothers Jordan and Mathew: your unconditional love and support form the bedrock upon which my career has thus far been built. Although none of you know precisely what it is that I do, you have continuously supported my endeavours throughout my academic career without question. Science is a fascinating and tremendously rewarding pursuit, but it is not without its hardships. You guys have been there through it all and I am forever
grateful. To my parents in particular, the invaluable sacrifices you have both made for my education, my interests, personal growth and career development are incalculable, without which none of this is possible. Finally, to my late grandfather Dr. Harold O. Seigel: Papa, your voracity for understanding nature and insatiable passion for knowledge reverberate in me to this day; and your contributions to science and technology have been a continual source of inspiration.
CHAPTER 1. INTRODUCTION ................................................................. 1

1.1 Cystic fibrosis .............................................................................. 1
  1.1.1 CF biochemistry ..................................................................... 1
  1.1.2 CF pathophysiology .............................................................. 2
  1.1.3 CF therapies ........................................................................ 3

1.2 Gene therapy ................................................................................ 4
  1.2.1 CF gene therapy ...................................................................... 6

1.3 Gene editing ................................................................................. 7
  1.3.1 Zinc finger nucleases ............................................................. 8
  1.3.2 TALENs ............................................................................... 10
  1.3.3 CRISPR ................................................................................. 11
    1.3.3.1 Discovery and mechanisms ............................................... 11
    1.3.3.2 Genome engineering ....................................................... 12
    1.3.3.3 CRISPR/Cas9 biochemistry and engineered Cas9 variants .... 13
      1.3.3.3.1 dCas9 ........................................................................ 15
      1.3.3.3.2 Base pair editors ......................................................... 16
      1.3.3.3.3 sgRNA engineering ...................................................... 17
    1.3.3.4 CRISPR/Cas9 in genetics and disease modeling ...................... 17
    1.3.3.5 CRISPR/Cas9 in gene and cell therapies .................................. 18

1.4 DNA double-strand breaks ............................................................. 21
  1.4.1 Non-homologous end-joining ................................................ 22
  1.4.2 Homology-directed repair ..................................................... 23
  1.4.3 DSB repair pathway choice .................................................. 24
1.5 Enhancing gene targeting efficiency by perturbing DSB repair ........................................... 29

CHAPTER 2. HYPOTHESIS AND OBJECTIVES ................................................................. 31
2.1 Rationale .................................................................................................................... 31
   2.1.1 CDK1 .................................................................................................................. 31
   2.1.2 Post-translational modification-mimics ............................................................ 31
   2.1.3 CtIP ..................................................................................................................... 32
   2.1.4 EXO1 .................................................................................................................. 33
2.2 Hypothesis .................................................................................................................. 34
2.3 Objectives .................................................................................................................. 34

CHAPTER 3. MATERIALS AND METHODS .................................................................... 35
3.1 Plasmids .................................................................................................................... 35
3.2 Flow cytometry ......................................................................................................... 36
3.3 Junction PCR ............................................................................................................ 37
3.4 Generating the IB3-1-CFTR-EGFP cell line .......................................................... 37
   3.4.1 Transfection ....................................................................................................... 37
   3.4.2 FACS ................................................................................................................ 38
   3.4.3 Puromycin selection and limited dilution ........................................................... 38
3.5 ssDNA donors .......................................................................................................... 39
3.6 T7E1 assay ............................................................................................................... 39
3.7 Data analysis ............................................................................................................. 40

CHAPTER 4. RESULTS .................................................................................................. 41
4.1 Flow cytometry assay for measuring gene targeting efficiency ................................. 41
4.2 CtIP and EXO1-4D significantly enhance gene targeting efficiency ........................... 44
4.3 IB3-1-CFTR-EGFP cell line ..................................................................................... 51

CHAPTER 5. DISCUSSION ............................................................................................. 57
5.1 Future directions ........................................................................................................ 61
5.2 Conclusion ................................................................................................................ 66

REFERENCES ............................................................................................................... 68
List of Tables

Table 1. DSB repair proteins and associated mutations
List of Figures

Figure 1. CRISPR/Cas9-directed genomic DNA cleavage.

Figure 2. NHEJ is used to repair DSBs in G1 when CDK activity is low.

Figure 3. HDR is used to repair DSBs in S/G2 when CDK activity is high.

Figure 4. Schematic diagram of mCherry gene integration into the AAVS1 locus.

Figure 5. Enhancement of precise gene targeting efficiency by CDK1, CtIP and EXO1-4D in HEK293 cells.

Figure 6. Enhancement of precise gene targeting efficiency by CtIP and EXO1-4D in HEK293 cells as measured by flow cytometry.

Figure 7. Enhancement of precise gene targeting efficiency by SCR7 in HEK293 cells.

Figure 8. Junction PCR demonstrating the enhancement of precise gene targeting efficiency by CtIP and EXO1-4D in HEK293 cells.

Figure 9. Schematic representation of the IB3-1-CFTR-EGFP cell line and the associated colour-switch assay.

Figure 10. Investigating the identity and CRISPR cleavage efficiency of IB3-1-CFTR-EGFP cell line clones.

Figure 11. IB3-1 CFTR homology arm sequencing.

Figure 12. Single-stranded mCherry donor production.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>53BP1</td>
<td>p53-binding protein 1</td>
</tr>
<tr>
<td>AAV</td>
<td>Adeno-associated virus</td>
</tr>
<tr>
<td>AAVS1</td>
<td>Adeno-associated virus integration site 1</td>
</tr>
<tr>
<td>ABC</td>
<td>ATP-binding cassette</td>
</tr>
<tr>
<td>AdV</td>
<td>Adenovirus</td>
</tr>
<tr>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>ASL</td>
<td>Airway surface liquid</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>ATM</td>
<td>Ataxia telangiectasia mutated</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BG</td>
<td>Benzylguanine</td>
</tr>
<tr>
<td>BIR</td>
<td>Break-induced replication</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin-dependent kinase</td>
</tr>
<tr>
<td>CF</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>CFTR</td>
<td>Cystic fibrosis transmembrane conductance regulator</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>crRNA</td>
<td>CRISPR RNA</td>
</tr>
<tr>
<td>CtIP</td>
<td>Carboxyl-terminal binding protein (CtBP)-interacting protein</td>
</tr>
<tr>
<td>dCas9</td>
<td>dead-Cas9</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA-PKcs</td>
<td>DNA-dependent protein kinase catalytic subunit</td>
</tr>
<tr>
<td>DSB</td>
<td>Double-strand break</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-stranded DNA</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGFP</td>
<td>Enhanced green fluorescence protein</td>
</tr>
<tr>
<td>ENaC</td>
<td>Epithelium sodium channel</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>ESC</td>
<td>Embryonic stem cell</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>EXO1</td>
<td>Exonuclease 1</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>Glu</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>GTE</td>
<td>Gene targeting efficiency</td>
</tr>
<tr>
<td>HD-Ad</td>
<td>Helper-dependent adenovirus</td>
</tr>
<tr>
<td>HDR</td>
<td>Homology-directed repair</td>
</tr>
<tr>
<td>HEK293</td>
<td>Human embryonic kidney 293</td>
</tr>
<tr>
<td>HPV</td>
<td>Human papillomavirus</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous recombination</td>
</tr>
<tr>
<td>HSPC</td>
<td>Hematopoietic stem and progenitor cells</td>
</tr>
<tr>
<td>Indel</td>
<td>Insertion or deletion</td>
</tr>
<tr>
<td>iPSC</td>
<td>Induced pluripotent stem cell</td>
</tr>
<tr>
<td>LDL</td>
<td>Low-density lipoprotein</td>
</tr>
<tr>
<td>LNP</td>
<td>Lipid nanoparticle</td>
</tr>
<tr>
<td>MEM</td>
<td>Minimal essential media</td>
</tr>
<tr>
<td>MMEJ</td>
<td>Microhomology-mediated end-joining</td>
</tr>
<tr>
<td>MRN</td>
<td>MRE11:RAD50:NBS1</td>
</tr>
<tr>
<td>MSD</td>
<td>Membrane-spanning domain</td>
</tr>
<tr>
<td>NBD</td>
<td>Nucleotide binding domain</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous end-joining</td>
</tr>
<tr>
<td>NPG</td>
<td>Non-obese diabetic (NOD)/Prkd&lt;sup&gt;scid&lt;/sup&gt;/IL-2Rγ&lt;sup&gt;null&lt;/sup&gt;</td>
</tr>
<tr>
<td>NSC</td>
<td>Neural stem cell</td>
</tr>
<tr>
<td>NSG</td>
<td>Immunodeficient, non-obese diabetic (NOD), severe combined immunodeficient (SCID) γ&lt;sup&gt;-/-&lt;/sup&gt;</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>p(A)</td>
<td>Polyadenylation</td>
</tr>
<tr>
<td>PAM</td>
<td>Protospacer-adjacent motif</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PKA</td>
<td>Protein kinase A</td>
</tr>
<tr>
<td>PSC</td>
<td>Pluripotent stem cell</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>RIPA</td>
<td>Radioimmunoprecipitation assay</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>RPA</td>
<td>Replication protein A</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Reverse transcription quantitative real-time polymerase chain reaction</td>
</tr>
<tr>
<td>RVD</td>
<td>Repeat-variable di-residue</td>
</tr>
<tr>
<td>SCID</td>
<td>Severe combined immunodeficient</td>
</tr>
<tr>
<td>SDSA</td>
<td>Synthesis-dependent strand annealing</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>sgRNA</td>
<td>Single-guide RNA</td>
</tr>
<tr>
<td>SIN</td>
<td>Self-inactivating</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>SNCA</td>
<td>Synuclein alpha</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>TALE</td>
<td>Transcription activator-like effector</td>
</tr>
<tr>
<td>TALEN</td>
<td>Transcription activator-like effector nuclease</td>
</tr>
<tr>
<td>tracrRNA</td>
<td><em>trans</em>-activating CRISPR RNA</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
<tr>
<td>ZF</td>
<td>Zinc finger</td>
</tr>
<tr>
<td>ZFN</td>
<td>Zinc finger nuclease</td>
</tr>
<tr>
<td>γ-RV</td>
<td>Gamma-retrovirus</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

1.1 Cystic fibrosis

Cystic fibrosis (CF) is a life-threatening autosomal recessive disease caused by mutations in the CF transmembrane conductance regulator (CFTR) gene. CF constitutes the most common cause of severe chronic obstructive lung disease and exocrine pancreatic dysfunction in children and young adults. The estimated worldwide incidence of CF is 1 in 2,500-3,000 newborns, and the disease affects more than 70,000 individuals worldwide\(^1,2\). Furthermore, although CF is classified as a rare disease, it is the most common lethal genetic disease in the Caucasian population\(^1,2\). Disease onset occurs within the first year of life, and approximately two thirds of patients are diagnosed within this time\(^3\). CF-related complications can persist throughout the duration of life, for which life expectancy is approximately 50 years in the developed world\(^3,4\).

1.1.1 CF biochemistry

CFTR is a polytopic transmembrane protein that functions as a chloride (Cl\(^-\)) and bicarbonate (HCO\(_3^-\)) channel on the apical membranes of secretory epithelia, such as the lungs, sweat glands, pancreas, and other exocrine glands\(^5\). CFTR is a member of the ATP-Binding Cassette (ABC) transporter family, which utilize ATP hydrolysis to transport substrates such as ions, toxins, peptides, drugs, and vitamins across biological membranes\(^6\). CFTR has a minimal architecture common to all ABCs. It is made up of two cytosolic nucleotide binding domains (NBDs) that undergo dimerization upon ATP hydrolysis, and two membrane-spanning domains (MSDs) that come together to form a single transmembrane ion transport channel\(^7\). Most ABC proteins are active transporters, and therefore use this modular structure in order to couple ATP hydrolysis to transport substrates against a concentration gradient\(^6\). In contrast, ATP hydrolysis powers the opening and closing of the CFTR ion channel, which in turn mediates passive electrochemical diffusion of Cl\(^-\) and HCO\(_3^-\) across cell membranes\(^8,9\). CFTR is also unique for its intrinsically
disordered regulatory domain (R domain), which controls channel activity in response to phosphorylation by cAMP-dependent protein kinases, such as protein kinase A (PKA)\textsuperscript{8,9}.

CF disease pathology is a consequence of structural, functional and/or regulatory defects in CFTR due to genetic mutation. More than 2000 CFTR sequence variants have been identified to date, including missense and nonsense mutations, frameshifts, and indels. There are various mechanisms by which these mutations can result in CFTR dysfunction, the most notable of which is defective protein processing. This is exemplified by the most common disease-causing CFTR mutation, the F508 deletion in the NBD1 domain\textsuperscript{10–12}. Almost 50% of CF patients are homozygous for this mutation, and ~85% carry one ΔF508 allele\textsuperscript{4}. ΔF508 disrupts normal topography on the NBD1 surface, such that stabilizing interactions at the interface of NBD1 and MSDs are prevented\textsuperscript{13}. This local structural defect therefore interferes with domain-domain interactions and prevents ΔF508 CFTR from attaining a native conformation. As a consequence, ΔF508 CFTR forms aggregate-prone structures that are recognized by quality control mechanisms in the ER during CFTR maturation, leading to ubiquitination and subsequent proteasome-mediated degradation\textsuperscript{14,15}. This ultimately prevents functional CFTR trafficking to the plasma membrane. Other mutations are rare, but can cause CFTR dysfunction via defective channel regulation or conduction, insufficient protein production, or reduced protein stability\textsuperscript{16}.

1.1.2 CF pathophysiology

In the disease state, CFTR dysfunction results in Cl\textsuperscript{−} and HCO\textsubscript{3}\textsuperscript{−} impermeability across epithelial apical membranes\textsuperscript{5}. Such defective ion conductivity alters the quantity and composition of epithelial fluids, in turn precipitating organ dysfunction. The widespread nature of this abnormality in various tissues is concordant with the diverse clinical manifestations of CF disease pathology, namely respiratory disease, intestinal obstruction, pancreatic dysfunction, salty sweat and male infertility\textsuperscript{5}. However, the main cause of morbidity and mortality in CF is progressive lung disease\textsuperscript{17}.

CF respiratory pathology is a consequence of defective CFTR-regulated ion homeostasis between lung epithelia and the airway surface liquid (ASL), resulting in epithelial dehydration and
thickened mucus. The ASL is a region of fluid on the apical surface of airway epithelia, which is composed of a periciliary layer adjacent to the epithelium and a superficial mucus layer. The mucus layer contains gel-forming mucins that trap inhaled particles, which are subsequently removed from the lung via mucociliary clearance\textsuperscript{18}. Furthermore, antimicrobial factors within the ASL are involved in innate and adaptive host immunity to protect the airways from inhaled pathogens. The ASL requires a tightly regulated volume, pH, ionic and nutrient content for sufficient antimicrobial activity, proper ciliary function and mucociliary clearance\textsuperscript{18}.

CFTR directly controls anion flow between the airway epithelia and the ASL by mediating HCO\textsubscript{3}\textsuperscript{−} and Cl\textsuperscript{−} movement, and indirectly by controlling Na\textsuperscript{+} diffusion via the epithelium sodium channel (ENaC)\textsuperscript{19,20}. CFTR normally inhibits ENaC, but its loss of function in CF results in excessive Na\textsuperscript{+} flow into epithelia. These electrochemical changes draw water into the epithelium, depleting ASL depth and immersing cilia in the mucus layer, an environment in which their movement is severely impaired. As a consequence, mucus builds up and airways become obstructed due to insufficient mucociliary clearance\textsuperscript{21}. Viscous and nutrient-rich mucus in turn provide a favourable environment for bacterial colonization. Furthermore, the functions of many antimicrobial factors are inhibited in the abnormally acidified ASL, a consequence of defective CFTR-mediated HCO\textsubscript{3}\textsuperscript{−} transport\textsuperscript{18}. Consequently, CF patients suffer from chronic bacterial infections that facilitate characteristic features of CF lung pathology, such as inflammation and progressive lung disease involving tissue remodelling.

1.1.3 CF therapies

CF survival has dramatically improved from 5 years in the 1960s to now up to 50 years due to clinical advances in treating symptoms and slowing disease progression\textsuperscript{3,4}. The significant progress made in extending life expectancy is largely a result of standardized multi-organ therapies. Such treatments include antibiotics for bacterial lung infections – especially \textit{Pseudomonas aeruginosa}, a hallmark of CF – in addition to mucolytics that release thick pulmonary mucus, chest physiotherapy, and high-calorie nutrition\textsuperscript{22}. Lung transplantation programs have also been beneficial\textsuperscript{23}, and collectively these treatments have resulted in half of CF patients being represented by adults in numerous countries\textsuperscript{24}. 

- 3 -
Novel mutation-specific therapies have recently been introduced into the clinic and many others are in development or clinical trials. High-throughput drug screening has led to the identification and development of two categories of mutation-specific therapies: ‘correctors’ that target the CFTR processing defect, and ‘potentiators’ that improve mutant CFTR channel activity. Ivacaftor, a CFTR potentiator, has been clinically successful in improving CFTR ion transport and pulmonary function in patients with the G551D CFTR genotype\textsuperscript{25}. However, these patients still require additional therapies including pancreatic enzyme replacement, inhaled mucolytics, and antibiotics\textsuperscript{16}. Furthermore, this class of CFTR mutations accounts for only 4-5\% of CF patients\textsuperscript{4}. As previously mentioned, \textasciitilde85\% of CF patients have the ΔF508 CFTR defect. The drug lumacaftor, a ‘corrector’, initially showed promise in rescuing the ΔF508 CFTR folding defect in primary cells\textsuperscript{26}. However, this drug proved ineffective in improving lung function in ΔF508 homozygous patients in clinical trials\textsuperscript{27}. Lumacaftor-ivacaftor combinatorial therapy has also been explored (Orkambi) for the ΔF508 defect. Although this therapy has been approved due to success in clinical trials, the overall efficacy was modest at best\textsuperscript{28}. Accordingly, no effective treatments are available for ΔF508 CFTR patients, the most common CF-causing mutation.

Overall, despite clinical advances in treating symptoms and slowing disease progression, there is still no cure for CF. As small molecules have proven inadequate for dramatically improving CF pathology, distinct efforts have been made to correct the underlying source of the disorder, i.e., the CFTR genetic defect. For these purposes, investigators have long been interested in the prospect of gene therapy for CF ever since the CFTR gene was discovered\textsuperscript{29}.

### 1.2 Gene therapy

Scientists, clinicians and the general public have long been captivated by gene therapy, which involves the correction or replacement of malfunctioning genes that give rise to disease pathologies. Considerable interest in this approach is warranted due to the potential to treat diseases at their genetic origins. Classically, gene therapy has been based on non-targeted, retroviral-mediated insertion of therapeutic transgenes into the host genome. The approach was initially based on gamma-retroviruses (γ-RVs), but was evidently accompanied by numerous limitations and salient risk factors, namely, poor gene transfer efficiencies in progenitor cells and
tumorigenesis due to vector insertion near oncogenes. Accordingly, γ-RVs were largely replaced by safer and more efficacious vectors, such as self-inactivating-γ-RVs (SIN-γ-RVs) and lentiviruses, which have found success in a number of phase 1/2 clinical trials in the past few years. Notably, ex vivo gene transfer and subsequent transplantation of autologous hematopoietic stem and progenitor cells (HSPCs) has been clinically useful in treating the genetic immune diseases Wiskott-Aldrich syndrome (WAS) and X-linked severe combined immunodeficiency (SCID-X1); as well as the blood disease β-thalassaemia, and the neurodegenerative storage diseases adrenoleukodystrophy and metachromatic leukodystrophy.

Despite clinical success using improved delivery vectors, integrating viral vectors carry risks of transcriptional silencing, insertional mutagenesis, and oncogenic transformation. Accordingly, lipid nanoparticles (LNPs) have emerged as an alternative to viral-mediated gene delivery. LNPs similarly have the capacity to protect nucleic acid contents from degradation while allowing cellular uptake of their cargo. However, these delivery methods are relatively generic and cell type-nonspecific. Nonetheless, there are a number of clinical trials currently investigating the utility of LNP-delivered siRNAs for the treatment of viral infections, cardiovascular diseases as well as cancer. In the realm of genetic diseases, LNPs have thus far shown preclinical utility for the treatment of ocular, auditory, and hepatic disorders.

Non-integrating viral vectors have also been successfully used for gene therapy applications. These viral delivery platforms are maintained as episomal vectors expressing therapeutic transgenes. Therefore, while avoiding potential insertional mutagenesis, the lack of permanent genetic modifications restricts the utility of these approaches to targeting low-turnover or non-proliferating cells. Various serotypes of adeno-associated viruses (AAVs) have been employed for these purposes. For example, AAVs are able to stably express coagulation factor IX (FIX) – a protein deficient in hemophilia B patients – from non-proliferating hepatocytes in animal models. Improvements in AAV vector design have since resulted in successful clinical treatments of hemophilia B using this gene therapy approach. Furthermore, AAV-based gene therapies for type 2 Leber congenital amaurosis (LCA) and retinitis pigmentosa (both inherited retinal diseases (IRDs)), recently became the first FDA-approved in vivo gene therapies for clinical commercial use in the United States under the brand name LUXTURNATM (voretigene neparvovec, AAV2-hRPE65v2) by Spark Therapeutics (NCT00999609).
1.2.1 CF gene therapy

Although CF affects multiple organs of the body, the primary cause of morbidity and mortality is progressive lung disease. Furthermore, because of the monogenic nature of CF and the accessibility of airway epithelia to gene therapy vectors, corrective gene therapy for CF lung disease has been a major effort for the last few decades. Three years after the discovery of the CFTR gene in 1989, Rosenfeld et al. successfully delivered replication-deficient first generation adenovirus (AdV) carrying CFTR cDNA to the airway epithelium of rats, from which transgenic CFTR mRNA and protein could be expressed. AdVs are double-stranded, non-integrating episomal viruses with a natural tropism for airway epithelial cells. Furthermore, AdVs possess high transduction efficiencies and relatively low virulence, and were therefore an obvious first choice for CF gene therapy. Alternative non-viral liposomal CFTR delivery had also partially corrected Cl⁻ transport defects in the tracheal epithelium of CFTR knockout mice. These promising preclinical studies precipitated the first CF gene therapy trial. Although the study was small and not placebo-controlled, Zabner et al. demonstrated correction of Cl⁻ transport defects in the nasal epithelium of three CF patients using first generation AdV carrying CFTR cDNA. Similar results were also obtained using liposome-mediated gene transfer. Since then, nine additional AdV CF gene therapy trials have been conducted. However, this approach has ultimately failed due to inefficient gene transfer, unsustainable gene expression, and the induction of immune responses that preclude repeated administrations.

AAVs have also been extensively investigated as a delivery vector for CF gene therapy. Similar to AdVs, AAVs are non-pathogenic and remain episomal in transduced cells in vivo. A number of CF gene therapy trials using AAV vectors have been conducted since 1998. Although initial trials demonstrated sufficient safety profiles, they were ultimately discontinued due to lack of efficacy in improving lung function. More recent studies utilizing pseudotyped lentiviral vectors have demonstrated some efficacy in delivering genes to the airway of mice and pigs, however these have yet to be investigated in clinical trials.

Due to the evident limitations associated with the aforementioned vector delivery systems, our group and others have been investigating the utility of helper-dependent adenoviral (HD-Ad) vectors. HD-Ad vectors are based on AdV serotype 5 with all viral coding sequences deleted. Accordingly, these vectors exhibit stable, long-term transgene expression due to attenuated
immunogenicity, while maintaining adenoviral tropism for epithelial cells\textsuperscript{79}. Our group has demonstrated that \textit{CFTR} knockout mice are protected from acute lung infection following HD-Ad delivery of the \textit{CFTR} gene to mouse airways\textsuperscript{80}. Furthermore, this delivery elicited stable transgene expression for up to 15 weeks in the absence of measurable pulmonary inflammation. HD-Ad vectors have also been used to deliver transgenes to human primary epithelial cells\textsuperscript{81,82}, as well as airway epithelia and submucosal glands of pigs \textit{in vivo}\textsuperscript{83}. However, a major challenge here is that the airway epithelium is a dynamic tissue that undergoes slow yet constant turnover, and therefore therapeutic transgene expression will be limited accordingly\textsuperscript{84}. This limitation illuminates a more broadly diffuse challenge in gene therapy applications. That is, the efficacy of safer, non-integrating vectors is limited by the half-life and finitude of their cognate target cells. Furthermore, although stable transgene expression can be attained using integrating viral vectors, untargeted integration of synthetic genetic elements into the host genome carries risks of transcriptional silencing, insertional mutagenesis, and oncogenic transformation\textsuperscript{41–43}. The requisite technology for ultimately surmounting these barriers have recently come to light with the dawn of programmable nucleases. Such technologies endow the capacity to make targeted and permanent site-specific genomic edits, such as knocking down endogenous genes, or precisely knocking in exogenous transgenes. These technologies can therefore mediate permanent transgene expression while avoiding insertional mutagenesis associated with integrating vectors. Furthermore, our group has recently demonstrated the capacity for targeting porcine airway stem/progenitor cells \textit{in vivo}\textsuperscript{85}. Our modern approach to CF gene therapy is thus to correct \textit{CFTR} mutations into the genome of airway stem/progenitor cells using programmable nucleases, in order to permanently correct CF genetic defects and consequent pathophysiology.

### 1.3 Gene editing

Classical genetics in the time of Mendel and Morgan was empirically limited to spontaneous mutations as the objects of study. The first hints of deliberate genetic manipulation came from Muller (1927)\textsuperscript{86}, and Auerbach & Robson (1947)\textsuperscript{87}, who demonstrated that radiation and chemical treatment, respectively, could potentiate the rate of mutagenesis. More advanced molecular technologies subsequently harnessed inducible transposable elements to transfer genetic
information. However, these procedures were all limited to inducing undirected and stochastic mutations. It wasn’t until the late 1970’s that the first targeted genomic manipulations were produced in yeast and mammalian cells by homologous recombination (HR)\textsuperscript{83–87}. HR can be utilized to transfer genetic information between two homologous genetic elements, and will be discussed in more detail in subsequent sections. Utilizing spontaneous HR for genomic manipulation was targeted and precise, yet suffered great practical inefficiencies requiring powerful selection and laborious characterization\textsuperscript{83–87}.

The key to high-efficiency, precise genome editing was discovered serendipitously through the study of DNA damage and repair, specifically DNA double-strand breaks (DSBs). It was found that the generation of DSBs by ionizing radiation led to sister chromatid crossover events\textsuperscript{92}, and that DSBs are intentionally induced by meiotic cells to facilitate genetic recombination between homologous sequences\textsuperscript{93}. These findings precipitated experiments using specific endonucleases to facilitate HR and targeted gene transfer, which ultimately nucleated modern genome engineering\textsuperscript{94–96}. Artificial nucleases can now be deliberately programmed, endowing researchers with the technical capacity to introduce DSBs at predetermined genomic loci. Following a targeted genomic DSB, the requisite DNA repair mechanisms can then be hijacked to introduce specific genetic modifications. Briefly, gene knock-outs can be generated if the break is repaired by error-prone non-homologous end-joining (NHEJ). In contrast, homology-directed repair (HDR) – which encompasses a number of distinct mechanisms including HR – is required for precise gene repair or site-specific integration of transgenes.

1.3.1 Zinc finger nucleases

Zinc fingers (ZFs) are the most common DNA-binding domains found in the human genome. ZFs are comprised of many subclasses, of which Cys\textsubscript{2}His\textsubscript{2} has been repurposed for genome engineering. Cys\textsubscript{2}His\textsubscript{2} ZF domains are comprised of tandemly repeating zinc-binding motifs of 30 amino acids, in which positions -1, 3, and 6 mediate direct contact with three contiguous base pairs of DNA and contribute to the binding specificity of the motif\textsuperscript{97,98}. Of the natural Cys\textsubscript{2}His\textsubscript{2} domains, Zif-268 has proven the most stable and versatile framework upon to which to engineer
ZF-with novel binding specificities. Modular zinc finger domains with unique specificities can then be linked in tandem in order to recognize novel DNA targets 9-18 bp in length99.

Fusion of ZF tandem repeats to an unbiased restriction endonuclease effectively generates a programmable, site-specific endonuclease. This was first successfully done by Kim et al., who fused the cleavage domain of the FokI endonuclease to a ZF array, in turn creating the first zinc finger nuclease (ZFN)100. Because FokI is an obligate dimer, effective DSB generation requires two ZFNs concomitantly targeting adjacent DNA loci. Despite the evident utility of ZFNs, target site overlap and crosstalk between ZF domains significantly complicate the prediction of binding specificities, and therefore the production of ZFNs targeting user-defined sequences101–103. Therefore, as there is no simple universal code for matching ZFN amino acid sequence to target DNA sequence, selection strategies are often required for the production of ZFNs104,105. ZFNs also appear to have a limited targeting range106. Nonetheless, ZFNs were essentially the first generation of programmable site-specific endonucleases, and accordingly were the first with prospective biotechnological and clinical utility.

ZFNs have been successfully utilized for genome engineering in swine107, cattle108, zebrafish109, mice110, as well as human embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs)111. Prospective clinical utility has been demonstrated for correcting haemophilia112 and sickle-cell in mice113; as well as α1-antitrypsin deficiency in iPSCs114; and Parkinson’s disease-associated alpha-synuclein (SNCA) gene in patient fibroblast-derived iPSCs115. These technologies have since been largely replaced by more powerful and robust programmable nucleases. However, there are a number of active and completed clinical trials based on the pioneering ZFNs due to their chronological primacy. For example, Phase I clinical trials (NCT00842634) have investigated the safety and efficacy of transplantation of autologous CD4+ T cells edited by ZFNs ex vivo to disrupt the CCR5 gene for the treatment of HIV-1116. However, treatment elicited severe adverse effects in some patients and variable efficacies, as viremia was undetectable in only a single patient by the endpoint of the trial. Other active Phase I clinical trials include those for the treatment of mucopolysaccharidosis (MPS) (NCT02702115, NCT03041324), and HPV-positive cervical intraepithelial neoplasia (NCT02800369).
1.3.2 TALENs

Transcription activator-like effector (TALE) proteins are critical virulence factors of the phytopathogenic bacterial genus *Xanthomonas*, which mediate host cell reprogramming by mimicking eukaryotic transcription factors\textsuperscript{117–120}. TALEs are DNA-binding proteins characterized by repeating arrays of 33-35 amino acids, where each array recognizes a single base pair. The repeat-variable di-residue (RVD) region harbours two amino acids that determine the specificity of base pair recognition within each repeat\textsuperscript{121–123}. Furthermore, modular TALE repeats can be linked to FokI in order to generate unique site-specific endonucleases that recognize programmed DNA sequences, analogous to ZFNs. In contrast to ZFNs, however, the simplicity of RVD base pair recognition enables straightforward theoretical design of TALENs for targeting user-defined DNA sequences. TALENs have also demonstrated a greater DNA targeting range with improved specificities\textsuperscript{124–126}. The principle impediments of TALEN usage are large coding sizes (~10 kb) and the repeating nature of the RVDs, which impede efficient assembly and packaging of dimeric TALENs into size-constrained vectors for mammalian cell delivery\textsuperscript{127}. However, investigators have devised superior methods by which to assemble TALENs, and have concurrently demonstrated that TALENs have the capacity to target virtually any DNA sequence\textsuperscript{128–130}. Accordingly, TALEN libraries targeting over 18,000 human protein-coding genes have been constructed using high-throughput Golden-Gate cloning\textsuperscript{131}.

As with ZFNs, TALENs have been broadly applied for biotechnology and biomedicine, as they have been applied for genome editing in roundworm\textsuperscript{132}, silkworm\textsuperscript{133}, fruitfly\textsuperscript{134}, frog\textsuperscript{135}, cricket\textsuperscript{136}, zebrafish\textsuperscript{137}, rat\textsuperscript{138}, cow and swine\textsuperscript{139}. Delivery of two TALEN pairs targeting the same chromosome has been demonstrated to elicit large indels (>6 kb) in pigs\textsuperscript{139}. These authors used TALENs to generate mono- and biallelic *LDLR* knockout pigs as models of familial hypercholesterolemia\textsuperscript{139}. Interestingly, TALENs have also been used to disrupt particular loci in the rice genome to confer resistance to pathogenic TALEs in *Xanthomonas* infection (the source from which this technology was initially discovered)\textsuperscript{140}. After demonstrating utility for genome editing in human somatic\textsuperscript{125} and pluripotent stem cells\textsuperscript{141}, TALENs were subsequently applied as preclinical anti-viral therapies for HIV infection\textsuperscript{142} and hepatitis B\textsuperscript{143}. Furthermore, Phase I clinical trials have been approved for the use of TALENs to treat HPV-related malignant neoplasms (NCT03226470, NCT03057912).
1.3.3 CRISPR

1.3.3.1 Discovery and mechanisms

When investigating chemical mutagenesis in the 1940s, Auerbach and Robson pronounced that, "…it could be hoped that among chemical mutagens there might be some with particular affinities for individual genes. Detection of such substances not only would be of high theoretical interest but would also open up the long-sought-for way to the production of directed mutations."\(^\text{87}\) It turns out that "such substances" had already been discovered eons ago by archaea, and integrated into an adaptive genetic network nested within an evolutionary arms race against infectious plasmids and phages. These prokaryotic immune systems are now known as the clustered regularly interspaced short palindromic repeats (CRISPR).

Although named differently at the time, CRISPR systems were initially discovered serendipitously through investigating the genetic origins of alkaline phosphatase isozyme conversion in *Escherichia coli*\(^\text{144}\). The authors noted a peculiar set of repetitive sequences with high sequence homology interspaced by variable spacers downstream of their gene of interest. This investigation was indeed the first encounter with CRISPR sequences. Advancements in DNA sequencing technologies in the 1990s uncovered these repetitive sequences in myriads of bacteria and archaea. Despite numerous theories, however, their function remained largely enigmatic. It wasn’t until 2000 that Mojica *et al.* recognized that these sequences must be functionally related\(^\text{145}\). Furthermore, accumulating genomic data eventually revealed CRISPR-associated (*cas*) genes adjacent to the repetitive CRISPR loci. These *cas* genes harboured recognizable subunits indicating functions in DNA repair and recombination, transcriptional regulation and chromosome segregation. Accordingly, their proximal association suggested they were involved in generating the CRISPR loci themselves\(^\text{146}\).

The functional role of CRISPR was first glimpsed independently by a number of groups who discovered that the variable spacer regions separating the repeats were homologous to phage DNA\(^\text{147–149}\). Importantly, they showed that phages are unable to infect host strains containing CRISPR elements with homology to their genomes. These phenomena were also observed in the dairy industry, where investigators were selecting industrial bacteria resistant to phage attack. Here, the authors provided the first empirical evidence for the involvement of *cas* genes in the
generation of CRISPR loci, and additionally as the enzymatic effectors of acquired phage resistance. These discoveries demonstrated that CRISPR can generate inheritable genetic libraries of foreign genetic material, which function as part of a cellular immune system akin to eukaryotic RNA interference (RNAi). Empirical evidence validated this model in subsequent years.

The CRISPR-Cas systems have been categorized into two classes (I and II), and further subdivided into six types (I-VI). Although these subcategories exhibit distinct enzymes and nuanced biochemical pathways, they share the same overarching functional architecture: adaption, expression, and interference. The adaption stage is dependent on Cas1 and Cas2, which form a multiprotein complex and mediate integration of short regions foreign DNA in between CRISPR repeats. These sequences are known as ‘spacers’, and are dependent on the presence of a neighbouring protospacer-adjacent motif (PAM) on the cognate foreign DNA. PAM motifs are typically only a few nucleotides long, and they ensure correct target recognition while avoiding host self-cleavage. The expression stage is characterized by transcription of the CRISPR array into a precursor transcript (pre-crRNA), which is subsequently bound by multi-subunit CRISPR RNA (crRNA)-effector complexes, or, in type II systems, by a single multi-domain Cas9 protein. Following complex formation, mature crRNA is generated via RNA endonuclease activity, or a mechanism involving concerted action of RNase III and trans-activating CRISPR RNA (tracrRNA). These RNA-protein complexes are then responsible for mediating the interference module. Specifically, the activated crRNA-effector complexes recognize the cognate target DNA (i.e., a foreign virus or plasmid) through specific base pair interactions between the target DNA and the homologous spacer region of the crRNA. Target DNA cleavage is then facilitated by a variety of nuclease domains embodied within the crRNA-effector complexes, thus degrading the foreign DNA.

1.3.3.2 Genome engineering

CRISPR-Cas systems are evidently comprised of multiple interacting enzymes, auxiliary proteins, RNA and DNA molecules. Amongst the variety of CRISPR-Cas systems, however, the type II CRISPR-Cas9 system is distinctly parsimonious, requiring only tracrRNA, crRNA, Cas9
nuclease, and host factor RNase III as necessary and sufficient for target DNA cleavage in vitro\textsuperscript{163,164} and in vivo\textsuperscript{165,166}. Accordingly, type II CRISPR-Cas9 has emerged as the archetypal system, and has been the most widely repurposed and engineered. In fact, Jinek \textit{et al.} showed that tracrRNA and crRNA can be combined into a chimeric single-guide RNA (sgRNA)\textsuperscript{163}. This effectively reduced the necessary components for Cas9-directed DNA cleavage and has therefore increased the versatility for biotechnological applications. The modern genome engineering toolkit now consists of ectopic Cas9 (typically purified and codon optimized from \textit{Streptococcus pyogenes}), and a sgRNA harbouring 20 nt of complementarity to the target DNA sequence. Selection of target DNA sequences are limited only by the presence of a neighbouring PAM sequence, which in the case of Cas9 is only ‘NGG’. This system is therefore highly versatile, allowing programmable and directed DNA cleavage at virtually any region within the genome of a variety of organisms.

CRISPR/Cas9 has numerous advantages over preceding technologies like ZFNs and TALENs. Namely, both ZFNs and TALENs must be expressed as pairs for effective cleavage, and require design and production of novel proteins for targeting distinct genomic loci. In contrast, a single Cas9 and sgRNA are sufficient for DNA cleavage, and targeting novel loci simply requires the production of new sgRNA sequences (~20 nt). The nature of RNA-guided targeting also allows for robust multiplexing approaches by expressing multiple sgRNAs, whereas multiplexing with ZFNs or TALENs was severely limited by cloning constraints. The diverse targeting capacities and ease of CRISPR/Cas9 design and production thus render this technology superior to preceding nuclease. Accordingly, once the requisite components for DNA cleavage had been understood\textsuperscript{163}, CRISPR/Cas9 was rapidly applied for genomic engineering in bacteria\textsuperscript{167}, yeast\textsuperscript{168}, as well as mouse and human cell lines\textsuperscript{169,170}. These technologies have also profoundly influenced transgenic animal production in the livestock industry, as CRISPR/Cas9 has already been used to generate transgenic swine\textsuperscript{171,172}, cattle\textsuperscript{173}, sheep\textsuperscript{174}, and goats\textsuperscript{175,176}.

1.3.3.3 CRISPR/Cas9 biochemistry and engineered Cas9 variants

Despite the evident utility of CRISPR/Cas9 for genome engineering, a number of salient limitations are hindering its clinical applications, namely, specificity and efficiency. Accordingly,
a number of research groups have sought to engineer more specific and efficient Cas9 variants. To these ends, an understanding of Cas9 structural biochemistry has proven indubitably useful.

Cas9 is a multi-domain structure which can be broadly dissected into its recognition (REC) and nuclease (NUC) lobes. The REC lobe is essential for sgRNA and target DNA binding, and evidently lacks structural homology with other known proteins and is therefore Cas9-specific\textsuperscript{177}. In contrast, the NUC lobe is comprised of RuvC and HNH nuclease domains, which share structural homology with the superfamily of retroviral integrases such as \textit{Escherichia coli} RuvC\textsuperscript{178} and \textit{Thermus thermophilus} RuvC\textsuperscript{179}; and with phage T4 endonuclease\textsuperscript{180} and \textit{Vibrio vulnificus} nuclease\textsuperscript{181}, respectively. The NUC lobe also harbours the PAM-interacting (PI) domain, which recognizes PAM sequences on the non-complementary strand and confers PAM specificity\textsuperscript{177}.

Following sgRNA:Cas9 complex formation, the active ribonucleoprotein (RNP) complex scans dsDNA for PAM sequences. PAM recognition then stimulates destabilization of the adjacent dsDNA and subsequent R-loop formation\textsuperscript{182} (Fig. 1). During this process, the sgRNA:DNA heteroduplex is stabilized within a positively charged groove at the interface between the REC and NUC lobes of Cas9. As the Cas9 complex interrogates DNA for complementarity, sequential extension of the sgRNA:DNA heteroduplex proceeds distally from the PAM site\textsuperscript{183}. Successful target recognition is highly dependent on complementarity with the first 10-12 nucleotides directly adjacent to the PAM sequence, known as the sgRNA ‘seed region’\textsuperscript{163,169,170,184,185}. Accordingly, mismatches detected early in directional melting prematurely destabilize the ternary complex and prevent Cas9 nuclease activation\textsuperscript{183}. Following successful target recognition, the HNH and RuvC nuclease domains cleave the complementary and non-complementary strands, respectively\textsuperscript{163,177}. 

- 14 -
Figure 1. CRISPR/Cas9-directed genomic DNA cleavage. The active sgRNA:Cas9 RNP complex scans dsDNA for PAM sequences (green). PAM recognition stimulates destabilization of the adjacent dsDNA and R-loop formation ensues. The sgRNA:DNA heteroduplex is stabilized within a positively charged groove, and sequential extension of the heteroduplex proceeds distally from the PAM site. sgRNA complementarity with the first 10-12 nucleotides adjacent to the PAM sequence (the seed region) is essential for successful target recognition and nuclease activation. Correct sequence complementarity stimulates the HNH and RuvC nuclease domains to cleave the complementary and non-complementary strands, respectively.

1.3.3.1 dCas9

The conservation of RuvC and HNH Cas9 nuclease domains has greatly assisted directed mutagenesis to generate novel Cas9 functions. Notably, D10A and H840A mutations in the RuvC and HNH domains, respectively, inactivate the catalytic activity of each respective nuclease. This has allowed for the generation of the catalytically-inactive dead-Cas9 (dCas9) variant, as well as Cas9 nickases. dCas9 was initially utilized for endogenous promoter binding and transcriptional repression in *Escherichia coli*\(^{186}\) and human cells\(^{187}\). Since then, dCas9 has been widely repurposed as a fusion protein platform for the recruitment of effector domains to targeted genomic loci. For example, synthetic transcriptional activators constituted by dCas9-VP64 or -p65 (subunit of nuclear factor kappa B (NF-κB)) fusions have been successfully employed to activate genes in human cells\(^{188,189}\). Researchers have recently used a similar approach to simultaneously activate multiple endogenous neurogenic genes in transgenic mice, resulting in the efficient conversion of astrocytes into functional neurons *in vivo*\(^{190}\). dCas9 has also been used
for targeted epigenomic modifications by fusing dCas9 to histone acetyltransferases (HATs)\textsuperscript{191}. Furthermore, fusions to the Krüppel-associated box (KRAB) mediates site-specific recruitment of a heterochromatin-forming complex that facilitates histone methylation and deacetylation\textsuperscript{192}.

dCas9 variants have also been employed to address off-target edits and increase the specificity of Cas9-induced DNA cleavage. For these purposes, dCas9 has been fused to the unbiased FokI nuclease. Because FokI is an obligate dimer, effective DSB induction is dependent on mutual on-target activity of two dCas9-FokI fusions guided by two distinct sgRNAs to adjacent DNA loci\textsuperscript{193,194}. Although this approach reduces off-target effects, it comes at a cost of reduced efficiency. A similar strategy has been demonstrated by multiplexing Cas9 nickases, where each contain a single point mutation (either D10A or H840A), thus relying on mutual on-target activity of both Cas9 variants for cutting\textsuperscript{195,196}. Other efforts to increase Cas9 specificity have involved directed mutagenesis to generate Cas9 variants with reduced DNA binding energies, which thus require more stringent target interactions for nuclease activation. To this end, eSpCas9 has been engineered by alanine mutagenesis to neutralize positively charged grooves that stabilize DNA during target interrogation\textsuperscript{197}. Furthermore, targeted alanine mutations that disrupt hydrogen bonding with the DNA phosphate backbone have generated the SpCas9-HF variant with improved specificities\textsuperscript{198}. Distinct approaches to improving editing specificity include small molecule- and light-inducible Cas9 variants\textsuperscript{199–201}. These studies demonstrate that controlling the temporal window within which Cas9 is active can substantially alleviate off-target edits.

1.3.3.3.2 Base pair editors

Significant engineering efforts have also been invoked to generate programmable single base-pair editors. Pioneering work by Komor \textit{et al.} generated the first of such enzymes by fusing dCas9 to rAPOBEC1 (a cytidine deaminase), which became the first-generation base editor (BE1)\textsuperscript{202}. This was a significant step forward for genome engineering, as it allowed for directed C→T (or G→A) base pair substitutions, while alleviating the necessity for DSBs. Although BE1 could generate efficient substitutions \textit{in vitro}, \textit{in vivo} editing efficiencies were up to 36-fold lower. Second- and third-generation BEs were engineered by adding a uracil DNA glycosylase inhibitor (UGI), and using a Cas9 nickase instead of dCas9. BE3 has demonstrated superior utility, eliciting permanent mutations in up to 75% of human cells with minimal indel formation (≤1%)\textsuperscript{202}. Further modifications have generated more efficacious fourth-generation base editors\textsuperscript{203}. These
technologies have since been utilized for diverse applications, including targeted base editing in mouse embryos\textsuperscript{204}; wheat and maize\textsuperscript{205}; as well as improving herbicide resistance in rice and generating marker-free gene edits in tomato\textsuperscript{206}.

1.3.3.3 sgRNA engineering

Another effort of Cas9 engineering has been to manipulate the chemical and structural properties of the sgRNA, as means to maximize both cleavage efficiency and target specificity. It was initially found that extending the canonical 20 nucleotide spacer region of the sgRNA does not improve specificity\textsuperscript{196}. In contrast, sgRNA reduction to 17 or 18 nucleotides can generally ameliorate off-target effects while maintaining on-target cutting efficiencies for most sgRNAs tested\textsuperscript{207,208}. However, these truncated sgRNAs have proven incompatible with engineered Cas9 variants like SpCas9-HF due to marginal on-target efficiencies\textsuperscript{198}. The influence of non-canonical sgRNA nucleotide chemistry on Cas9 genome editing has also been investigated. For instance, 2’-O-methyl (M), 2’-O-methyl 3’phosphorothioate (MS), and 2’-O-methyl 3’thioPACE (MSP) modifications of terminal nucleotides improves sgRNA intracellular stability, and significantly increases Cas9 editing efficiencies in CD34+ HSPCs and human primary T cells\textsuperscript{209}. However, additional off-target activity accompanied these improvements. Further modifications to these sgRNAs including interspaced 2’-fluoro groups increased stability and specificity but reduced on-target efficiencies\textsuperscript{210}.

1.3.3.4 CRISPR/Cas9 in genetics and disease modeling

In its brief history, CRISPR/Cas9 has already significantly impacted the way in which we study genetic networks and the role of genetics in disease pathogenesis. For example, the explosion of massive genomic data sets that has occurred in this century require powerful high-throughput methods for annotating genomes. To this end, CRISPR has proven immensely successful for conducting genome-scale knockout screens in mammalian cell cultures, eliciting substantial phenotypic effects and high validation rates\textsuperscript{211–216}. dCas9-derived base editors have also been repurposed for generating diverse libraries of point mutations within targeted windows of genomic loci, enabling high-throughput screening of genetically-diverse functional variants\textsuperscript{217,218}. The authors demonstrated the utility of such technologies by applying them to identify novel
mutations conferring resistance to the cancer therapeutics imatinib\textsuperscript{217} and bortezomib\textsuperscript{218}. For these high-throughput genetic screens, CRISPR supersedes chemical mutagenesis and RNAi screens. As chemical mutagenesis is undirected, the genetic origins of measurable phenotypes is unknown, and can be costly and laborious to investigate. Furthermore, although RNAi is high-throughput and directed, this approach suffers from low signal-to-noise ratios due to high off-target activity and incomplete gene knockout\textsuperscript{219–221}.

CRISPR systems have also made significant contributions to stem cell research, endowing researchers with the capacity to effectively knockout and knock-in genes in ESCs and iPSCs\textsuperscript{222–225}, as well as transcriptionally regulate genes in ESCs\textsuperscript{226,227}, NSCs\textsuperscript{228}, and iPSCs\textsuperscript{229}. Notably, due to the simplicity of multiplexing sgRNAs, many of these studies demonstrate the feasibility of simultaneously modulating multiple genetic pathways. CRISPR has also been extensively applied for disease modeling and drug screening in ESCs\textsuperscript{230,231}, iPSCs\textsuperscript{232,233}, and a number of organoids. For example, investigators have utilized CRISPR to generate mutant kidney organoids\textsuperscript{234}, as well as intestinal organoids as disease models to study congenital disorders\textsuperscript{235} and colorectal cancer\textsuperscript{236–239}.

1.3.3.5 CRISPR/Cas9 in gene and cell therapies

The first applications of these tools in human cells were hastily welcomed as the dawn of biomedical utopias in which human disease will be eradicated under the auspices of CRISPR/Cas9. Although many of these claims were facile and hyperbolic, these rapidly evolving technologies will undoubtedly have a profound influence on biology and medicine. The low cost and technological simplicity of CRISPR/Cas9 endows researchers with unprecedented capacities for manipulating the genetic profiles of clinically-relevant PSCs, autologous primary cells, and diseased tissues.

CRISPR/Cas9 has already begun to show promise as powerful gene editing tool for gene and cell therapy applications, as evidenced by several preclinical studies and recently emerging clinical trials. For example, LNPs and AAV vectors have been used to deliver Cas9 mRNA and sgRNAs to the liver of a mouse model of tyrosinemia, achieving up to 6% of genetically corrected hepatocytes and full rescue of liver damage\textsuperscript{48}. Furthermore, LNPs have also been used to deliver
CRISPR to the murine inner ear, with the capacity to edit the genome of 20% of hair cells. Interestingly, the power of CRISPR-induced indels has been demonstrated by the correction of a splicing defect in the LAMA2 gene. This gene encodes the α2 chain (Lama2) of the laminin-211 complex expressed in muscle and Schwann cells. Splicing defects in Lama2 result from exon skipping, which in turn produce truncated proteins and ultimately give rise to congenital muscular dystrophy type 1A (MDC1A). Using CRISPR, the authors were able to excise the intronic region containing the mutation, in turn creating a functional donor splice site to create full-length Lama2 protein. Furthermore, AAV-mediated delivery of this CRISPR system in a mouse model of MDC1A resulted in substantial improvements in neuromuscular pathology.

These technologies have also made their way into therapies for cardiovascular diseases. Most notably, investigators have applied the power of CRISPR-induced indels to disrupt the PCSK9 gene. Naturally occurring loss of function mutations in PCSK9 reduce circulating LDL levels and attenuate the risk of coronary heart disease by up to 88%. For these reasons, Ding et al. injected AdV carrying the requisite CRISPR components targeting exon 1 of hepatic PCSK9 into mice. Significant hepatic PCSK9 disruption resulted in a ~90% reduction in circulating PCSK9, and a consequent ~40% reduction in total blood cholesterol. These studies have since been successfully reproduced using AAV vectors, as well as in humanized mouse liver in vivo.

CRISPR has also made significant contributions to cancer therapeutics, namely chimeric antigen receptor (CAR) T cell immunotherapies, which have been transformative for the treatment of refractory blood cancers. Directed gene manipulations facilitated by CRISPR technology have generated synthetic CAR T cell variants with enhanced anti-tumour efficacy in vivo, as well as engineered iPSC-derived natural killer (NK) cells for the treatment of solid tumours. Advancements in CRISPR-edited CAR T cell therapies have since translated into clinical trials for the treatment of relapsed or refractory leukemia and lymphoma (NCT03398967). CRISPR has additionally been used to disrupt the PDCD1 gene encoding programmed death-1 (PD-1) receptor, an immune checkpoint receptor which downregulates cytotoxic T cell proliferation and activity. The activating ligand for this receptor can be overexpressed in tumor cells to subvert the immune system. Accordingly, CRISPR-induced disruption of PDCD1 in cancer patient-derived primary T cells enhanced cytotoxic anti-tumor activity, and in vivo injection of these modified T cells improved the survival of tumor-bearing mice. The efficacy of these treatments has since led
to clinical trials for the treatment of Epstein-Barr virus (EBV)-positive advanced stage malignancies (NCT03044743).

As previously discussed (section 1.2), editing HSPCs for inherited blood disorders are a major effort of gene and cell therapy applications. Autologous HSPCs can be extracted from patients, edited \textit{ex vivo}, and subsequently transplanted back into patients. Patient-derived autologous HSPCs can now be electroporated \textit{ex vivo} with CRISPR-expressing plasmids\textsuperscript{251}, mRNA\textsuperscript{209}, or Cas9:sgRNA RNPs\textsuperscript{252}. The use of RNPs presents a particularly attractive strategy, as it manifests a ‘hit and run’ approach, effectively minimizing nuclease exposure and therefore toxicity and off-target edits\textsuperscript{209}. De Ravin \textit{et al.} utilized this approach to correct mutations in \textit{CYBB}, the gene encoding the catalytic centre of NADPH oxidase, which gives rise to X-linked chronic granulomatous disease (X-CGD)\textsuperscript{253}. Specifically, this group edited X-CGD patient-derived CD34\textsuperscript{+} HSPCs with the requisite CRISPR components, in turn restoring NADPH oxidase function in myeloid cells differentiated from the extracted progenitors \textit{in vitro}. Subsequent transplantation of corrected HSPCs into NSG mice resulted in effective engraftment and sustainable production of functional mature human myeloid and lymphoid cells. These experiments represent a significant step forward from similar studies using ZFNs, which failed to attain appreciable correction rates following transplantation into NSG mice\textsuperscript{113,254}.

These applications have also extended into anti-viral therapy, in particular for disrupting the \textit{CCR5} gene to generate HIV-resistant human primary T-cells\textsuperscript{255,256} and iPSCs\textsuperscript{142,257}. Notably, Xu \textit{et al.} have shown that CRISPR can be used to disrupt the \textit{CCR5} gene in human CD34\textsuperscript{+} HSPCs, and that subsequent transplantation into NPG mice can generate multi-lineage progeny and confer resistance to HIV-1 infection\textsuperscript{258}. These experiments have precipitated the approval of Phase I clinical trials (NCT03164135).

Substantial empirical evidence has also emerged in support of our current model of CRISPR-based CF gene therapy. Namely, CRISPR/Cas9 has been used to correct the \textit{ΔF508} mutation in CF patient-derived iPSCs\textsuperscript{259}. Corrected iPSCs were able to differentiate into mature airway epithelial cells exhibiting physiologically normal CFTR function. Moreover, intestinal stem cells from \textit{ΔF508} CF patients have been corrected \textit{ex vivo}, and subsequently able to form clonally expanded intestinal organoids with restored CFTR function\textsuperscript{260}. These are indeed promising results. However, genome editing of airway stem cells \textit{in vivo} will require significant
improvements in CRISPR-mediated transgene knock-in efficiencies. For these purposes, it is necessary to understand the DSB repair mechanisms used by mammalian cells to repair Cas9-induced genomic breaks.

1.4 DNA double-strand breaks

As with all biological macromolecules, DNA is subject to destabilizing and decomposing forces. However, as the hereditary molecule of life, accurate reconstitution of DNA in response to structural insults is essential to the faithful reproduction of the cells, multicellular networks, organisms, and species in which it encodes. Of the various types of damage DNA can endure, DSBs are indeed the most genotoxic and structurally threatening. DSBs can arise due to extraneous stimuli like ionizing radiation and radiomimetic drugs, or as a consequence of intracellular reactive oxygen species (ROS) or replication fork collapse. Other DSBs are programmed and essential to the reproduction and normal physiology of an organism, such as meiotic crossover and V(D)J recombination, respectively. In an analogous manner, artificial nucleases can be programmed to induce DSBs for deliberate genome editing. However, in contrast to bottom-up, genetically-encoded programmed DSBs, those induced for genome editing are top-down, encoded by scientists and the social institutions thereof. Accordingly, our deliberate genomic manipulations will not have been tested and refined by eons of evolutionary pressures. Understanding the molecular consequences of our programmed DSBs is therefore pertinent to ensuring they are repaired in a manner conducive to our desired outcomes.

Because DSBs are highly genotoxic lesions, employment of the appropriate repair mechanism is tightly regulated and crucial for maintaining genome integrity. In mammalian cells, DSBs are predominantly repaired by one of two pathways, broadly classified as NHEJ and HDR. NHEJ-related pathways commonly repair DSBs by direct re-ligation of the broken ends following some or no processing on the ends themselves. Although NHEJ pathways are ostensibly error-prone due to end processing and template-independent repair, they are probably accurate in many cases. On the contrary, HDR encompasses various distinct subpathways, which share the common feature of requiring a homologous DNA sequence as a template for repair. Accordingly, it is largely accepted that HDR pathways are exclusively error-free.
1.4.1 Non-homologous end-joining

Following a genomic DSB, the kinases ataxia telangiectasia mutated (ATM), ATM and Rad3-related (ATR), and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) rapidly phosphorylate chromatin resident histone H2AX surrounding the break site\textsuperscript{263–266}. This generates γH2AX, which acts as a DNA damage signal. In resting or quiescent vertebrate cells (G0/G1), the Ku70/Ku80 heterodimer (Ku) is first to bind to free DNA ends generated by a DSB, likely due to its shear abundance (~400,000 molecules/cell) and strong dissociation constant (~10\textsuperscript{−9} M)\textsuperscript{267,268}. However, Ku competes for DNA binding with the less abundant MRE11:RAD50:NBS1 (MRN) complex, which promotes alternative end-joining pathways in G1\textsuperscript{269–271}. In canonical NHEJ, the Ku:DNA complex serves as a scaffold for the recruitment of additional repair proteins. The Artemis:DNA-PKcs complex rapidly binds to Ku:DNA ends. Subsequent autophosphorylation of Artemis:DNA-PKcs regulates access of other NHEJ proteins and activates the nucleolytic activity of Artemis, thus initiating DNA end processing\textsuperscript{272–274}. End processing functions to generate short regions of microhomology (≤4 nucleotides) to serve as a template for ligation. Polymerases pol μ and pol λ, which are capable of flexible template-independent and -dependent polymerase activity, respectively, are then recruited to extend the processed DNA ends\textsuperscript{275,276}. The XLF:XRCC4:DNA ligase IV complex subsequently mediates ligation of the two DNA ends (Fig. 2). This enzymatic complex is a flexible ligase with the capacity to ligate complex and incompatible DNA ends across gaps\textsuperscript{277}. Depending on the chemistry of the break, additional auxiliary proteins may be required for repair, including polynucleotide kinase (PNK), aprataxin, and tyrosyl DNA phosphodiesterase 1 (TDP1)\textsuperscript{278}.

Although described here as constrained and sequential mechanistic steps, NHEJ is more accurately depicted as a semi-haphazard process involving malleable iterations of nuclease, polymerase, and ligase activity\textsuperscript{278}. It is evident that these enzymatic pathways and the proteins involved are necessarily flexible. That is, it has proven evolutionarily adaptive to employ robust biochemical pathways capable of repairing a diverse range of broken DNA substrates, as these pathways are largely conserved from yeast to vertebrates.
1.4.2 Homology-directed repair

HDR encompasses a number of distinct repair subpathways, namely HR, synthesis-dependent strand annealing (SDSA), and break-induced replication (BIR). These pathways are commonly initiated by the generation of γH2AX surrounding the DSB site, which function to recruit mediator of DNA damage checkpoint 1 (MDC1) and the MRN complex. These complexes in turn scaffold the break site to recruit additional HDR factors. All HDR pathways are similarly dependent on the generation of long regions of single-stranded DNA (ssDNA) surrounding the break site. Specifically, ssDNA generation is catalyzed by the endonucleolytic activity of the MRN complex, which itself is dependent on the activity of carboxyl-terminal binding protein (CtBP)-interacting protein (CtIP). Following MRN endonucleolytic cleavage, exonuclease resection proceeds bi-directionally via MRN 3’–5’ exonuclease activity, and long range (1-3 kb) 5’–3’ resection mediated by the redundant activities of EXO1 and BLM/DNA2 nucleases. The resulting ssDNA is then rapidly bound by replication protein A (RPA). The RPA coating ssDNA is eventually replaced by RAD51 nucleoprotein filaments, aided by the concerted action of PALB2, BRCA1, and the recombination mediator BRCA2. RAD51 nucleofilaments are the necessary substrates to search for and invade homologous DNA. Following homologous DNA strand invasion, the subset of HDR pathways begin to diverge (Fig. 3).

Canonical HR is a consequence of both strands of the DNA break invading the homologous template. Subsequent polymerization generates joint DNA molecules bearing a double-Holliday junction (HJ), which can undergo considerable branch migration before dissolution. Double-HJ dissolution and consequent separation of the conjoined DNA molecules is dependent on the mutual activity of topoisomerase IIIα and the BLM:RMI1:RMI2 (BTR) complex. Processing by BTR gives rise to non-crossover events that avoid loss of heterozygosity by inter-homologue exchanges. In contrast, if the double-HJs are resolved by the endonucleases MUS81-EME1, SLX1-SLX4 and GEN1, crossovers are more likely to occur. SDSA can occur if one or both strands of the DNA break invades the homologous template, from which extended D-loops are formed instead of double-HJs. Following homology-dependent polymerization of the invading strand, D-loops are processed by helicases and the invading strand is subsequently re-ligated to the other end of the DSB break. In contrast, BIR results exclusively from one-end invasion, and replication can proceed for hundreds of kilobases until the end of the
chromosome\textsuperscript{295,296}. Another DSB repair pathway distinct from both HDR and NHEJ is microhomology-mediated end-joining (MMEJ), which is highly error-prone. Although categorically distinct, MMEJ shares similar features with HDR in that it requires CtIP-dependent MRN nucleolytic action. In contrast to HDR, however, MMEJ ensues following limited resection by MRN and subsequent alignment and ligation of short ssDNA ends, in the absence of any homologous strand invasion\textsuperscript{297} (Fig. 2).

1.4.3 DSB repair pathway choice

The choice of DSB repair mechanism is largely dependent on cell cycle phase\textsuperscript{298}. More specifically, due to the dependence on a homologous sequence, HDR pathways are restricted to S/G2 when a sister chromatid template is available for repair. On the contrary, NHEJ pathways are active throughout the cell cycle. The cell cycle-regulated control of these repair pathways is largely modulated by cyclin-dependent kinases (CDKs), which themselves are master regulators of cell cycle progression. CDKs have extensive regulatory control over DNA end resection, which is generally viewed as the committed step for HDR and functions as the primary regulatory node for DSB repair pathway choice\textsuperscript{299}.

p53-binding protein 1 (53BP1) is a key proximal regulator of DSB repair pathway choice. 53BP1 effectively inhibits end resection, and therefore HDR, principally through its concerted action with Rap1 interacting factor 1 homolog (RIF1). In response to DSBs in G0/G1, ATM kinase-mediated phosphorylation of 53BP1 induces recruitment of RIF1 to DSB sites. RIF1 DSB foci in turn inhibit the accumulation of BRCA1 at DSBs, which ultimately inhibits end resection and promotes repair by NHEJ\textsuperscript{300–304} (Fig. 2). The HDR-inhibitory action of 53BP1 is counteracted by upstream CDK activity as cells enter S phase. More specifically, CDK-mediated phosphorylation of CtIP in S/G2 promotes CtIP:BRCA1 complex formation, in turn relieving the 53BP1-dependent, RIF1-mediated inhibition of end resection\textsuperscript{305}. This regulation is also dependent on CDK-mediated phosphorylation of BRCA1, which promotes UHRF1-mediated ubiquitylation and consequent degradation of RIF1\textsuperscript{306}. Chromatin remodeling has also recently been implicated in this regulatory axis. More specifically, CDK- and ATM-mediated phosphorylation of the chromatin remodeler Cockayne syndrome group B protein (CSB) promotes nucleosome
disassembly, which consequently inhibits and promotes RIF1 and BRCA1 accumulation, respectively\(^{307}\) (Fig. 3).

In addition to the relief of 53BP1:RIF1 inhibition, positive stimulatory regulation is also required to promote end resection. In this respect, CtIP has been recognized as a critical regulatory hub for DSB repair pathway choice in mammalian cells, whose activity is a necessary prelude for end resection and therefore HDR\(^{298,299,308–310}\). The involvement of CtIP in HDR regulation is exemplified by low steady-state levels in G1 maintained by ubiquitin-dependent, proteasome-mediated degradation\(^{311–313}\). CtIP degradation is subsequently suppressed during the G1/S transition, in concordance with the upregulation of HDR upon S-phase entry. Its localization at DSBs is also highly regulated. In particular, RNF138 ubiquitin ligase activity displaces Ku from DSB sites while recruiting CtIP, and this process is necessary for the inhibition of NHEJ and promotion of HDR\(^{314,315}\).

CtIP is also heavily phosphorylated by upstream kinases in S/G2, in particular by CDK1. For example, CDK1 phosphorylates CtIP at Thr847 and Ser327 in S/G2, in turn facilitating its interaction with BRCA1, promoting end resection, and ultimately stimulating HDR\(^{282,316–318}\). CDKs also phosphorylates CtIP at 5 other residues, termed 5mCDK. These modifications do not directly promote HDR, but instead facilitate subsequent CtIP phosphorylation at T859 by the ATM kinase\(^ {283}\). T859 phosphorylation promotes CtIP to stimulate end resection and recruit BLM and EXO1, which in turn catalyze the extended end resection necessary for HDR\(^ {283}\). In addition to phosphorylation, CtIP is also constitutively acetylated at Lys432, Lys526, and Lys604\(^ {319}\). Following DNA damage, CtIP is deacetylated by sirtuin 6 (SIRT6), in turn promoting the ability of CtIP to mediate end resection and therefore HDR\(^ {319}\). Furthermore, loss of SIRT6 activity results in severe HDR defects in human cells, exemplifying its role in stimulating CtIP-mediated DSB end resection\(^ {319,320}\). In fact, artificially expressing SIRT6 has been shown to rescue the decline of HDR associated with replicative senescence\(^ {320}\).

Limited resection by MRN:CtIP is sufficient for promoting the error-prone MMEJ pathway. Extended end resection, however, mediated principally via the redundant activities of EXO1 and BLM/DNA2 nucleases, is necessary for commitment to HDR pathways\(^ {284–286}\). Accordingly, EXO1 exonuclease activity is stimulated by CDK-mediated phosphorylation in S/G2, specifically at residues Ser639, Thr732, Ser815, and Thr824\(^ {321}\). The regulation of DSB repair proteins also
appears to be mutually interdependent, in that the activity of EXO1 is also stimulated by BLM, MRN, and RPA\textsuperscript{286,322}. These regulatory pathways effectively restrict long-range end resection to S/G2, and therefore serve as additional levels of HDR regulation.
Figure 2. NHEJ is used to repair DSBs in G1 when CDK activity is low. In G1, ATM-mediated phosphorylation of 53BP1 promotes 53BP1 interactions with RIF1 and PTIP. These interactions in turn inhibit the accumulation of HDR factors such as CtIP, BRCA1, EXO1, BLM and DNA2 at DSB sites. Ku competes for binding with the MRN complex at DSB ends. Stable Ku binding invariably leads to NHEJ with the concerted actions of Artemis:DNA-PKcs, pol µ and λ, and DNA ligase IV. Limited end processing mediated by MRN:CtIP can lead to mutagenic MMEJ.
Figure 3. HDR is used to repair DSBs in S/G2 when CDK activity is high. CDK-mediated phosphorylation of BRCA1 and CtIP promotes their accumulation at DSB sites by relieving the HDR inhibitory effects of 53BP1:RIF1. Chromatin remodelling directed by pCSB further promotes BRCA1 assembly while inhibiting RIF1. Active CtIP and MRN catalyze rate-limiting short-range end resection, and CDK-mediated phosphorylation of EXO1 stimulates long-range end-resection. RAD51 replaces RPA on ssDNA and catalyzes homologous strand invasion, which can subsequently result in SDSA, HR or BIR.

1.5 Enhancing gene targeting efficiency by perturbing DSB repair

The power of programmable nucleases for genome editing lies in their ability to generate directed DSBs. Furthermore, the precise nature of the edit depends on the repair pathway employed to repair the DSB. Gene knockouts can be generated when NHEJ is used to repair the break, as this pathway is prone to generating small mutations or indels. In contrast, due to the dependence of HDR on a homologous template, this pathway can be harnessed to insert exogenous DNA sequences into precise genomic locations. By flanking transgene cassettes – such as CFTR, for example – with loci-specific sequence homology, exogenous DNA can be used as the template for HDR in order to knock-in desired genetic elements. This approach is broadly referred to as gene targeting.

The principle impediment to gene targeting is the inherently low frequency of HDR events. That is, due to the fact that NHEJ is the predominant DSB repair mechanism, the efficiency of precise gene targeting is exceedingly low. In mammalian cell lines, primary cells, PSCs and zygotes, precise gene targeting efficiencies can range anywhere from 0.1-20% (with few exceptions beyond this range), depending on the cell type, the genomic locus being targeted, method of delivery, as well as the size and chemistry of the repair template. This represents a considerable barrier for effective in vivo gene therapy, and more specifically, our CF gene therapy approach to knock-in the CFTR gene into airway stem cells. Here, the low efficiency of gene targeting is further compounded by practical barriers to gene delivery and the host immune response. Accordingly, our approach requires significantly higher gene targeting efficiencies (GTE) to be clinically useful.

It is worth recalling here the broadly applicable preclinical efficacies of CRISPR/Cas9 outlined previously. It should be noted that the vast majority of those cases were using CRISPR/Cas9 to generate knockouts by NHEJ. In the cases where HDR was necessary, small insertions or low efficiencies sufficed for the application. Due to the evident limitations associated with CRISPR/Cas9 gene targeting by HDR, a number of groups have sought to improve current methods and enhance targeting efficiencies. The most widely used approach has been to
administer the small molecule SCR7, a DNA ligase IV inhibitor, in combination with the requisite factors for CRISPR-mediated gene targeting (Cas9, sgRNA, and a HDR template). As previously discussed, DNA ligase IV is responsible for ligating broken ends in DSB repair by NHEJ. Accordingly, by inhibiting this enzyme, investigators have been able to upregulate compensatory HDR and therefore GTE anywhere from ~2 to 20-fold in mammalian cell lines and zygotes. However, these results are often not reproducible. Analogous approaches to small molecule inhibition of NHEJ have been demonstrated using NU7441 and Ku-0060648 to inhibit DNA-PKcs, resulting in ~2 to 4-fold enhancements in human cell lines and mouse embryonic fibroblasts (MEFs). The RAD51-stimulating small molecule RS-1 can also elicit up to 5-fold enhancements in human cell lines and rabbit embryos. Furthermore, the small molecules Brefeldin A and L755507, an ER-Golgi transport inhibitor and a β3-adrenergic receptor agonist, respectively, enhance gene targeting efficiency ~2 to 9-fold in mammalian cell lines and hPSCs by unknown mechanisms.

The cell cycle-regulated control of DSB repair pathway choice also presents an opportunity to potentiate the use of HDR in response to CRISPR/Cas9-generated DSBs. There are several well-established mechanisms for synchronizing proliferating cells in specific cell cycle phases. Yang et al. have shown that the use of nocodazole to synchronize cells in G2/M can enhance gene targeting efficiency ~2 to 4-fold in human cell lines. This approach has also demonstrated utility in hPSCs and neural precursor cells. Another interesting strategy was to fuse Cas9 to a fragment of the human Geminin peptide, which is degraded by the proteosome pathway in G1. This effectively restricts Cas9 expression to S/G2 when HDR is most active, in turn enhancing GTE, albeit modestly. Other approaches to enhancing CRISPR/Cas9 GTE have focused on the nature of the donor template. More specifically, a number of studies have demonstrated greater targeting efficiencies in mice and rats using ssDNA donors as opposed to canonical dsDNA donor templates.
Chapter 2
Hypothesis and Objectives

2.1 Rationale

Researchers have evidently made significant progress in enhancing CRISPR/Cas9 GTE. It should be noted that despite these improvements, large insertions (like those necessary to knock-in CFTR) are more resistant to HDR enhancements. Furthermore, many of these advancements have come with the use of small molecules, which cannot be packaged into viral vectors and are therefore less amenable to in vivo gene therapy. Genetically-encoded factors would be more beneficial for these purposes. In this respect, Orthwein et al. has shown that concurrently overexpressing a CtIP and PALB2 mutant can enhance CRISPR/Cas9 GTE up to 4-fold in non-dividing U2OS 53BP1Δ cells. This study provided proof of principle for genetically-encoded means by which to perturb DSB repair and enhance CRISPR/Cas9 GTE, which could be effectively applied to viral-based gene therapies.

2.1.1 CDK1

Aforementioned studies have evidently demonstrated the utility of perturbing DSB repair pathway choice as effective means by which to enhance CRISPR/Cas9-mediated gene targeting. In line with this reasoning, overexpressing CDK1 is an attractive strategy to increase GTE due its overwhelming control over the balance between HDR and NHEJ.

2.1.2 Post-translational modification (PTM)-mimics

An alternative strategy for increasing GTE would be to specifically harness the influence of CDKs on DSB repair pathway choice, without directly altering CDK expression or activity. This can be accomplished by specifically overexpressing the relevant phospho-products of CDK catalysis that
promote HDR. Many of the CDK DSB repair substrates up-regulate HDR and/or inhibit NHEJ in response to CDK-mediated phosphorylation of specific residues. Furthermore, mutating such residues to Asp or Glu can generate ‘phospho-mimics’ that recapitulate the activity of the phosphorylated proteins. In addition to phosphorylation, DSB repair proteins endure other PTMs such as acetylation and ubiquitylation, which also regulate their capacity to promote HDR in a cell cycle-specific manner. Importantly, mutating acetylated and ubiquitylated residues to Arg can generate DSB repair proteins that mimic their active states, regardless of cell cycle position. These properties make phospho-, as well as acetylated- and ubiquitylated-mimics ideal candidates for effectively potentiating GTE. Crucial HDR proteins that can be ectopically expressed as PTM-mimics are outlined in detail below and summarized in Table 1.

2.1.3 CtIP

As previously outlined, CtIP is recognized as an essential regulatory hub for DSB repair pathway choice in mammalian cells. CtIP activity is negatively regulated in resting cells by ubiquitin-dependent, proteasome-mediated degradation\textsuperscript{311–313}. However, these studies have shown that mutating ubiquitylated residues S276, T315, and Y842 to alanine can block ubiquitylation, in turn generating a hyperactive CtIP variant that constitutively promotes HDR in human cell lines. The relief of G1 ubiquitin-directed, proteasome-mediated degradation of CtIP is endogenously elicited as cells enter S phase. CtIP is then activated by PTMs in S/G2, namely by CDK and ATM-mediated phosphorylation and SIRT6-catalyzed deacetylation. It has been shown that such activity can be effectively mimicked by site-directed mutagenesis to artificially promote HDR throughout the cell cycle. For example, the requirement of Thr847 phosphorylation can be circumvented by a CtIP phospho-mimic in which Thr847 is replaced by Glu (T847E). This mutant demonstrates normal end resection and HDR in the absence of CDK-mediated phosphorylation\textsuperscript{318}, and exogenous expression of CtIP T847E increases aberrant HDR while inhibiting NHEJ\textsuperscript{305,317}. Furthermore, CDK-mediated Ser327 phosphorylation promotes CtIP association with BRCA1, which is necessary to promote end resection\textsuperscript{316,340}. A Ser327 phospho-mimic (S327E) would therefore presumably have similar HDR-promoting activities. In addition, 5mCDK CtIP phosphorylation facilitates subsequent ATM-mediated T859 phosphorylation of CtIP, which in
turn stimulates end resection and the recruitment of additional necessary HDR factors\textsuperscript{283}. Wang \textit{et al.} has demonstrated that a T859E CtIP phospho-mimic bypasses the requirement of 5mCDK phosphorylation, and can alleviate the G1-repression of HDR\textsuperscript{283}.

In addition to phosphorylation, CtIP is also dependent on SIRT6 deacetylase activity to promote end resection\textsuperscript{319}. Importantly, a CtIP deacetylated mimic, in which Lys432, 526, and 604 are mutated to Arg (3KR), relieve the requirement of SIRT6-dependent deacetylation\textsuperscript{319}. This CtIP 3KR mutant would presumably have a greater propensity to promote HDR than its wild-type counterpart.

CtIP evidently endures multiple levels of PTMs that control its ability to promote HDR in accordance with cell cycle phase and DNA damage. It appears that each individual modification contributes to the activation of CtIP, where the collective influence of specific PTMs results in robust HDR-promoting activity. Furthermore, the studies outlined above demonstrate that these modifications can be effectively mimicked by expressing an artificial CtIP harbouring multiple mutations.

\subsection*{2.1.4 EXO1}

Following CtIP/MRN-mediated short-range DSB end resection (\~{}50-100 bp), EXO1 is recruited to the partially-resected ends to catalyze long-range end resection (\~{}1-3 kb), a process that is necessary for HDR\textsuperscript{341}. Similar to CtIP, EXO1 exonuclease activity is stimulated by CDK-mediated phosphorylation in S/G2, specifically at residues Ser639, Thr732, Ser815, and Thr824\textsuperscript{321}. This effectively restricts long-range end resection to S/G2, and therefore serves as an additional level of HDR regulation. Furthermore, Tomimatsu \textit{et al.} demonstrated that an EXO1 phospho-mimic (EXO1-4D) not only rescues the HDR defect associated with CDK inhibition, but also significantly up-regulates HDR in wild-type cells\textsuperscript{321}. 

- 33 -
Table 1. DSB repair proteins and associated mutations

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDK1</td>
<td>-</td>
</tr>
<tr>
<td>CtIP-WT</td>
<td>-</td>
</tr>
<tr>
<td>CtIP-3EA</td>
<td>S327E, T847E, T859E, S276A, T315A, Y842A</td>
</tr>
<tr>
<td>CtIP-M</td>
<td>S327E, T847E, T859E, K432R, K526R, K604R</td>
</tr>
<tr>
<td>CtIP-M2A</td>
<td>S327E, T847E, T859E, K432R, K526R, K604R, S276A, T315A</td>
</tr>
<tr>
<td>EXO1-4D</td>
<td>S639D, T732D, S815D, T824D</td>
</tr>
</tbody>
</table>

2.2 Hypothesis

Overexpressing CDK1, as well as CtIP and EXO1 variants can enhance the efficiency of CRISPR/Cas9-mediated precise gene targeting.

2.3 Objectives

1: Develop an assay to measure gene targeting efficiency.

2: Assess the utility of CDK1, as well as CtIP and EXO1 variants for enhancing CRISPR/Cas9 gene targeting efficiency.

3: Generate an IB3-1 cell line to verify precise gene targeting by a colour-switch assay using ssDNA donors.
Chapter 3
Materials and Methods

3.1 Plasmids

The CRISPR/Cas9 plasmid for the mCherry AAVS1 integration assay was generated by cloning the sgRNA (5'-ACCCCACAGTGGGGCCACTA-3') into pSpCas9(BB)-2A-GFP (PX458) (Addgene #48138) by conventional restriction cloning. The homology arms on the promoterless mCherry donor construct were obtained by PCR from BAC clone: RP11 384-G4. The splice acceptor (SA) and T2A sequence were added upstream of mCherry-poly(A) (from pTRE3G-mCherry) by PCR. The homology arms and SA-T2A-mCherry-p(A) cassette were then cloned into an empty pSEAP backbone342 by In-Fusion cloning (In-Fusion HD Cloning Plus, ClonTech, cat. #638911). The lacZ plasmid was constructed as previously described342. The CDK1 plasmid was obtained from Addgene (Cdc2-HA, #1888). CtIP and EXO1 cDNAs were obtained from Transomic Technologies (cat. #BC030590-seq, and BC007491-seq, respectively). FLAG tags were inserted in frame with each cDNA by PCR. Each cDNA-FLAG cassette was then inserted under the control of a chicken β-actin (CBA) promoter in an empty pSEAP backbone342 by In-Fusion cloning. Mutagenesis was performed using QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies, cat. #210515).

The CRISPR/Cas9 plasmid used for generating the IB3-1 cell line was obtained from OriGene Technologies (cat. #GE100002) containing the sgRNA (5'-GAAGGTGGCCACAACCGAGCTT-3') targeting exon1 of the CFTR gene. The donor plasmid harbouring the EGFP-puromycin cassette flanked by CFTR homology was custom made from OriGene Technologies on a pUC vector. The CRISPR/Cas9 plasmids for CMV-mCherry integration into the genomic EGFP-CFTR locus was generated by cloning the sgRNAs (#1: 5’-AGCAGCTGCACGCCGTAGGTC-3’; #2: 5’-CTCGTGACCACCCCTGACCTA-3’; #3: 5’-GGGCACGGGCGAGCTTGCCCGG-3’) into pSpCas9(BB)-2A-GFP (PX458) (Addgene #48138) by conventional restriction cloning, and the GFP tag was removed by EcoR1 digestion. The donor plasmid designed to integrate CMV-mCherry into the EGFP-CFTR genomic locus was assembled by restriction digestion of CMV-
mCherry-p(A) cassette, while the EGFP homology arms were obtained by PCR. The three fragments were subsequently cloned into an empty pSEAP backbone by In-Fusion cloning. The T7 promoter was inserted in the reverse orientation downstream of the right homology arm by PCR.

3.2 Flow cytometry

3x10⁵ HEK293 cells (cultured in MEM (Thermo Fisher Scientific, cat. #11095080) + 10% FBS (Wisent Bioproducts, cat. #080-450) + 1% Pen/Strep (Thermo Fisher Scientific, cat. #15070063) in a humidified 37 °C, 5% CO₂ incubator) were seeded into tissue culture-treated polystyrene 6-well plates (Corning, cat. #353046), and transfected the following day (~70-80% confluency) using jetPRIME (Polyplus-transfection, cat. #114-07) according to the manufacturer’s protocol (jetPRIME:DNA = 2:1) (CRISPR/Cas9: 500 ng; Donor: 1 µg; lacZ, CtIP, EXO1, CDK1: 2 µg). For SCR7-treated cells, SCR7 (Xcess Biosciences, cat. #M60082-2) was diluted to the indicated concentrations in DMSO and added to the culture media 18 hr post-transfection. 48 hr post-transfection, cells were either harvested for flow cytometry or passed 1/8 into new 6-well plates. Passaged cells were subsequently passed 5 additional times over the course of 3 weeks before harvesting for flow cytometry. When harvesting for flow cytometry, cells were detached by incubating in 0.5 mL trypsin-EDTA (0.05%) (Thermo Fisher Scientific, cat. #25300062) for 1 min at 37 °C (incubator), suspended in serum-containing media, and spun down at 220 g. Cells were then washed with PBS (1X, pH 7.4), and subsequently re-suspended in PBS. Dead cell staining was performed using LIVE/DEAD Fixable Violet Dead Cell Stain Kit (Thermo Fisher Scientific, cat. #L34955) according to the manufacturer’s protocol for unfixed cells. Briefly, cells were washed twice with PBS + 1% FBS, and re-suspended in 1 mL PBS + 1% FBS. Cells were filtered through 70 µM nylon mesh into 5 mL round-bottom polystyrene tubes (Falcon, cat. #352052). Flow cytometry was performed using an LSRII-CFI (Becton Dickinson), and data analysis was done using FlowJo v.10.
3.3 Junction PCR

HEK293 cells were cultured, seeded, transfected and passaged as detailed in flow cytometry methods (section 3.2). Following transfection and subsequent passaging for 3 weeks, cells were trypsinized (as in flow cytometry methods), spun down at 220 g, and washed with PBS. Cells were subsequently lysed in 500 µL of RIPA buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) + 650 µg/mL proteinase K (Thermo Fisher Scientific, cat. #EO0491) + 100 µg/mL RNase A at 55 °C with gentle shaking overnight. Genomic DNA was purified by phenol-chloroform extraction and ethanol precipitation. Specifically, 1X volume of phenol:chloroform:isoamyl alcohol (25:24:1, v/v) (Thermo Fisher Scientific, cat. #15593031) was added to cell lysates, briefly vortexed, and spun down at 14,000 rpm for 2 min. The top aqueous layer was extracted into a fresh microfuge tube. 0.5X volume of 7.5 M ammonium acetate and 2X volume of 100% ethanol were added and incubated at -20 °C for at least 1 hr. The mixture was then spun down at 14,000 rpm at 4 °C for 5 min and subsequently washed with 1 mL 70% ethanol. Purified genomic DNA was re-suspended in Tris-EDTA (TE) buffer (pH 8.0). 1 µg of genomic DNA was used as template in each PCR reaction, and PCR products were run on non-denaturing 1% agarose gels. PCR primers illustrated in Fig. 5 (primer set #1: Fwd: 5'-GCTTTGCCACCCTATGCTGACAC-3', Rev: 5'-TGTGCACCTTGAAAGCGCATGAACCTGTGACACAC-3'); primer set #2: Fwd: 5'-TTAGCCACTCTGTGCTGACCACTCTGCTGACACAC-3', Rev: 5'-AGTCACCCAGAGACAGTGCACCTA - 3'). Band intensities were quantified using ImageJ.

3.4 Generating the IB3-1-CFTR-EGFP cell line

3.4.1 Transfection

2x10⁵ IB3-1 cells (cultured in LHC-8 (Thermo Fisher Scientific, cat. #12678017) + 10% FBS (Wisent Bioproducts, cat. #080-450) + 1% Pen/Strep (Thermo Fisher Scientific, cat. #15070063) in a humidified 37 °C, 5% CO₂ incubator) were seeded into a tissue culture-treated polystyrene
6-well plate (Corning, cat. #353046). The following day, seeded cells were transfected with CRISPR/Cas9 targeting CFTR exon1 and the EGFP-puromycin donor DNA (1 µg each) using FuGENE HD Transfection Reagent (Promega, cat. #E2311) according to the manufacturer’s protocol (FuGENE:DNA=3:1).

3.4.2 FACS

Transfected cells were expanded to 10 cm plates and passaged for 2 weeks. One of the 10 cm plates was trypsinized (as in flow cytometry methods, section 3.2) and spun down at 220 g. Cells were then washed with PBS, and subsequently re-suspended in 1 mL PBS. Single EGFP+ cells were then sorted into 10 collagen-coated (type I bovine collagen, Advanced Biomatrix, cat. #5005), tissue culture-treated, polystyrene 96-well plates (Corning, cat. #C353227) containing LHC-8 + 20% FBS + 1% Pen/Strep culture media.

3.4.3 Puromycin selection and limited dilution

Transfected cells were incubated in LHC-8 + 10% FBS + 1% Pen/Strep supplemented with 1 µg/mL puromycin for 2 weeks. Every 2 days, plates were washed with PBS (1X, pH 7.4) and fresh puromycin-containing media added. Fluorescence microscopy was then used to identify EGFP+, puromycin-resistant colonies. Sterile cylinders were placed around each individual colony to allow for isolated trypsinization, re-suspension in serum-containing media (supplemented with 1 µg/mL puromycin), and subsequent transfer to 48-well plate. These colonies were expanded to 6-well plates. Each colony was then diluted in LHC-8 + 10% FBS + 1% Pen/Strep (supplemented with 1 µg/mL puromycin) to 5 cells/mL, and 100 µL pipetted into 96-well plates. Plates were monitored 12 hr later to mark wells which only contained a single cell. These clones were allowed proliferate and then expanded to larger culture vessels. A total of 6 clones were successfully expanded and healthy at this stage.
3.5 Generating ssDNA donors

T7 in vitro transcription was performed using TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific, cat. #K0441) according to the manufacturer’s protocol using the CMV-mCherry donor (targeting EGFP-CFTR) as a template. Reaction products were then digested with 1 µL DNaseI at RT for 15 min, followed by enzyme inactivation with 1 µL 25 mM EDTA at 65 ºC for 10 min. RNA was then purified using the PureLink RNA Mini Kit (Thermo Fisher Scientific, cat. #12183018A) according to the manufacturer’s protocol for purification from liquid samples. Purified RNA was then used for gene-specific reverse transcription with the SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific, cat. #18090010) according to the manufacturer's protocol. Products of this reaction were then digested with 1 µL RNaseH at 37 ºC for 20 min to remove template RNA. RNaseH was inactivated by incubation at 65 ºC for 10 min. The remaining ssDNA was purified using the PureLink RNA Mini Kit according to the manufacturer's protocol for purifying from liquid samples.

3.6 T7E1 Assay

2x10⁵ IB3-1-CFTR-EGFP cells (from each of the 6 clones) were seeded into tissue culture-treated polystyrene 12-well plates (Corning, cat. #C353225), and the following day transfected with CRISPR/Cas9 (targeting EGFP-puromycin) (650 ng) and ssDNA mCherry donor (350ng) using FuGENE HD Transfection Reagent (Promega, cat. #E2311) according to the manufacturer’s protocol (Fugene:DNA = 3:1). 3 different CRISPR/Cas9 plasmids containing various sgRNAs targeting different EGFP sequence motifs were tested for each clone (see plasmids methods, section 3.1). 2 days post-transfection, transfectants were expanded to 6-well plates. 3 days later, genomic DNA from each condition was harvested and purified using the GeneArt Genomic Cleavage Detection Kit (Thermo Fisher Scientific, cat. #A24327) according to the manufacturer’s protocol. PCR was used to amplify the EGFP sequence from genomic DNA. PCR products were purified using the NucleoSpin Gel and PCR Cleanup Kit (ClonTech, cat. #740609). Purified PCR products were subsequently denatured and randomly reannealed using a thermocycler (95 ºC for 5 min, 95→85 ºC (-2 ºC/sec), 85→25 ºC (-0.1 ºC/sec), 4 ºC hold). Randomly reannealed
amplicons were then digested with 1 µL of T7E1 (New England Biolabs, cat. #M0302) for 45 min at 37 °C, and subsequently run on a non-denaturing 2% agarose gel.

3.7 Data analysis

Graphical presentation and statistical analysis were performed using Prism 6 (GraphPad). All data are presented as mean ± SEM. Statistics were performed using unpaired, two-tailed Student’s t-tests. Differences between means were considered statistically significant when p<0.05 and p<0.01, and were represented graphically as * and **, respectively.
Chapter 4

Results

4.1 Flow cytometry assay for measuring gene targeting efficiency

An assay to measure GTE requires two plasmid constructs: one for expressing CRISPR/Cas9 and a sgRNA, and another harbouring the donor DNA to be integrated into the Cas9 cleavage site. The Cas9 protein is co-expressed with a sgRNA from a single construct, as seen in Fig. 4A. Cas9 is expressed as a single transcript with EGFP, and the intervening T2A sequence facilitates cleavage during protein translation, giving rise to individual Cas9 and EGFP proteins. EGFP fluorescence can therefore be used as a measure of transfection efficiency. The sgRNA is designed to target Cas9 to induce a DSB within the first intron of the human PPP1R12C gene on chromosome 19. This genomic region has been commonly referred to as the AAVS1 locus since the observation that AAV integrates here. AAV infection is not associated with any known pathophysiology, and both hES and hiPS cells with disrupted PPP1R12C genes retain pluripotency. The AAVS1 locus is therefore considered a ‘genomic safe harbour’ for transgene integration. Furthermore, this gene is ubiquitously expressed in all primary human cells studied, as well as in commonly used transformed cell lines. For these reasons, the AAVS1 locus is an ideal genomic locus to study precise gene targeting.

The donor DNA plasmid contains a promoterless mCherry cassette, which is flanked on both sides by ~1 kb sequences that are homologous to the AAVS1 genomic DNA (Fig. 4B). These homology arms facilitate site-specific integration of intervening plasmid sequence via HDR at the target DSB site. A splice acceptor site and T2A signal were inserted upstream of the mCherry ORF, with a poly(A) signal downstream. Since AAVS1 exon 1 contains the translation start site, insertion of this cassette into intron 1 gives rise to a single transcript driven by the endogenous AAVS1 promoter. Because of the T2A signal upstream of mCherry, translation produces two separate polypeptides: exon 1 of AAVS1, and mCherry. The absence of an upstream promoter on the donor plasmid ensures that mCherry is only expressed from the endogenous AAVS1 promoter.
following a successful integration event. This approach is analogous to gene trapping strategies\textsuperscript{347}. mCherry fluorescence is therefore a reliable measure of precise gene targeting, which can be quantified by flow cytometry. This is demonstrated by the fact that transfection of either the donor or CRISPR/Cas9 plasmid alone produces negligible mCherry expression (Fig. 4C). In contrast, appreciable levels of mCherry expression are only attained following co-transfection of both donor and CRISPR/Cas9 plasmids.
Figure 4. Schematic diagram of mCherry gene integration into the AAVS1 locus. a) CRISPR/Cas9 gene expression cassette containing the sgRNA and Cas9 flanked by a FLAG tag, NLS, T2A-EGFP, and a poly(A) signal. b) Promoterless mCherry donor construct containing the mCherry ORF downstream of a splice acceptor site and T2A, and upstream of a poly(A) signal. The entire cassette is flanked on either side by ~1kb of homologous sequence to the AAVS1 locus. Following a successful gene integration event, mCherry expression is driven by the endogenous AAVS1 promoter. c) HEK293 cells were transfected with the indicated plasmids and subsequently harvested 48 hr post-transfection for analysis by flow cytometry.
4.2 CtIP and EXO1-4D significantly enhance GTE

CtIP and EXO1 mutants were generated by site-directed mutagenesis (see Table 1 for a full list of mutants). CtIP and EXO1 mutants, as well as CDK1, were then cloned into expression plasmids under the control of the chicken β-actin promoter, which is frequently used to drive high levels of gene expression in mammalian expression vectors. To assess the ability of CDK1, as well as CtIP and EXO1 variants to enhance GTE, the respective expression plasmids were co-transfected with the AAVS1 CRISPR/Cas9 and donor construct into HEK293 cells. 48 hr post-transfection, cells were harvested for flow cytometry analysis. Representative flow cytometry data is shown in Fig. 5A. The depicted populations are gated on GFP+ cells, and the proportion of mCherry+ is expressed as a percentage of the parent population. Because GFP is constitutively expressed from the CRISPR/Cas9 plasmid, this effectively controls for transfection efficiency. To assess enhancements in GTE, data were normalized to the CRISPR + donor condition. Co-transfection with CtIP-WT, -3EA, -M, -M2A, -M3A as well as EXO1-4D and CDK1 elicited significant enhancements in GTE up to ~2-3 fold (Fig. 5C). Furthermore, absolute targeting efficiencies reached up to ~20% (Fig. 5B).

Subsequent studies investigated the possibility that observed enhancements in GTE were due in part to a greater total amount of transfected DNA in experimental vs control conditions. For these purposes, a lacZ expression plasmid (which theoretically would not influence GTE) was co-transfected with either the donor plasmid alone, or together with the CRISPR + donor condition. From these studies, it was evident that the total amount of DNA was indeed influencing the observed GTE measurements (Fig. 5D). Furthermore, the increase in mCherry+ cells elicited by co-transfection with lacZ (in the absence of CRISPR) indicated that there was a significant amount of background mCherry expression from the donor plasmid. This is puzzling due to the absence of an upstream promoter, but could be the result of weak cryptic promoter activity in the upstream homology arm. Although the background expression may hinder the reliability of these data, these experiments were taken as a sufficient initial screen to demonstrate that CtIP mutants were no more effective than the wild-type CtIP protein. Accordingly, subsequent experiments were limited to the investigation of CtIP-WT, CDK1 and EXO1-4D.
Figure 5. Enhancement of precise gene targeting efficiency by CDK1, CtIP and EXO1-4D in HEK293 cells. HEK293 cells were transfected with the indicated plasmids and subsequently harvested 48 hr post-transfection for analysis by flow cytometry. a) Representative flow cytometry experiment. The populations displayed are gated on GFP+ cells to control for transfection efficiency. b) Average gene targeting efficiency. c) Fold increase in gene targeting efficiency. All values are normalized to Donor+CRISPR condition. d) lacZ control reveals background mCherry expression from the Donor plasmid. Data presented as mean ± SEM (n=3). * P < 0.05, ** P < 0.01.
In order to dilute background mCherry expression from the donor plasmid, HEK293 cells were passed 6 times over the course of 3 weeks post-transfection before flow cytometry analysis. This effectively eliminated residual plasmid in cells and ensured that mCherry expression was only a consequence of successful integration events. This is supported by the fact that in lacZ controls, mCherry expression is an order of magnitude higher in the presence of CRISPR, despite equal amounts of transfected DNA (Fig. 6A, panels 2,4). However, co-transfection of lacZ with CRISPR + donor enhanced GTE values relative to CRISPR + donor controls (panels 3,4). This can be explained due to the greater total amount of DNA transfected in the lacZ condition, which effectively stabilizes CRISPR and donor plasmids, thus allowing greater opportunity for integration events. Accordingly, GTE enhancements in the lacZ condition were subtracted from GTE values in experimental samples in order to control for the amount of transfected DNA. Corrected GTE values were then normalized to the CRISPR + donor condition. These experiments demonstrated that CtIP-WT and EXO1-4D can elicit up to 3- and 6-fold enhancements in GTE, respectively, whereas CDK1 had no significant effect (Fig. 6C). Absolute integration efficiencies measured up to 0.5-1% of total cells (Fig 6B). Furthermore, various combinations of factors were assessed for potential additive or synergistic effects on GTE enhancement, but revealed no significant increases (Fig. 6E).
Figure 6. Enhancement of precise gene targeting by CtIP and EXO1-4D in HEK293 cells as measured by flow cytometry. HEK293 cells were transfected with the indicated plasmids and subsequently passed 6 times over the course of 3 weeks. Cells were then harvested for analysis by flow cytometry. a) Representative flow cytometry experiment. b) Average gene targeting efficiency of controls and individual factors. c) Fold change in gene targeting efficiency induced by individual factors. d) Average gene targeting efficiency of controls and combinations of factors. e) Fold change in gene targeting efficiency induced by combinations of factors. Data presented as mean ± SEM (n=3 or 4). * $P < 0.05$, ** $P < 0.01$. 
SCR7 is a small molecule inhibitor of DNA ligase IV that has been widely utilized for enhancing GTE. Investigators have been able to upregulate HDR and therefore GTE anywhere from ~2 to 20-fold in mammalian cell lines and zygotes using SCR7\textsuperscript{315,316}. Here, we investigated the utility of SCR7 using our flow cytometry assay and compared these results to those obtained by CtIP-WT and EXO1-4D. HEK293 cells treated with various concentrations of SCR7 were similarly passed 6 times over the course of 3 weeks, and subsequently harvested for flow cytometry analysis. These data demonstrate that 0.1\textmu M SCR7 was the most effective concentration used, eliciting up to 3-fold enhancements in GTE (Fig. 7A). Thus, the increase in GTE evoked by CtIP-WT over-expression is comparable to that of SCR7, while EXO1-4D is significantly more effective in this respect (Fig. 7B).

![Figure 7](image_url)

**Figure 7. Enhancement of precise gene targeting efficiency by SCR7 in HEK293 cells.** a) HEK293 cells were transfected with the indicated plasmids. SCR7 was added to the culture media 18 hr post-transfection at the indicated concentrations. Cells were passed 6 times over the course of 3 weeks, and then harvested for analysis by flow cytometry. b) Comparing the respective fold changes associated with SCR7, CtIP-WT, and EXO1-4D as measured by flow cytometry 3 weeks post-transfection. Data presented as mean ± SEM (n=3). * P < 0.05.
In order to confirm the GTE enhancements observed by flow cytometry, junction PCR was performed at the AAVS1 integration site. Briefly, HEK293 cells were transfected with AAVS1 CRISPR/Cas9 and mCherry donor plasmids, as well as either CtIP-WT, EXO1-4D, or CDK1. Transfectants were passed 6 times over the course of 3 weeks, and genomic DNA was subsequently harvested and purified. Limited-cycle PCR was performed to amplify the junction site using the indicated primers (Fig 8A). Band intensities for each condition were quantified and normalized to their respective control amplicons (Fig. 8B). As in the flow cytometry experiments, GTE enhancements in the lacZ condition were subtracted from GTE values in experimental samples in order to control for the amount of transfected DNA. Corrected GTE values were then normalized to the CRISPR + donor condition. These data largely corroborate those obtained from flow cytometry experiments, demonstrating ~3- and 3.5-fold enhancements by CtIP-WT and EXO1-4D, respectively. Furthermore, there was no significant enhancement by CDK1, and no additive or synergistic effects by combining various factors (Fig. 8C).
Figure 8. Junction PCR demonstrating the enhancement of precise gene targeting efficiency by CtIP and EXO1-4D in HEK293 cells. a) Representation of genomic DNA containing a successfully integrated mCherry cassette. Primer set 1 is specific for the inserted cassette, while primer set 2 amplifies the wild-type AAVS1 locus. b) Representative junction PCR experiment. HEK293 cells were transfected with the indicated plasmids and subsequently passed 6 times over the course of 3 weeks. Genomic DNA was then harvested and purified from each treatment condition. Primer set 1 was used to amplify the mCherry insert, and primer set 2 was used as a PCR control. c) Fold change in gene targeting efficiency as measured by junction PCR. Data presented as mean ± SEM (n=3 or 4). * P < 0.05, ** P < 0.01.
4.3 IB3-1-CFTR-EGFP cell line

One of the main limitations of the AAVS1 mCherry assay is that it cannot distinguish between on-target and potential off-target integration events. Although the assay was designed so that mCherry expression is driven by the endogenous AAVS1 promoter, it could also be inserted under the control of an off-target promoter. We sought to generate a cell line that could be used for an assay to distinguish between these distinct integration events. For these purposes, we generated a cell line from IB3-1 cells, which are immortalized bronchial epithelial cells from a CF patient49. To generate the cell line, CRISPR/Cas9 was programmed to cut immediately upstream of CFTR exon 1 and insert an EGFP-puromycin cassette (Fig. 9A). Once the cell line had been obtained, CRISPR/Cas9 was programmed to cut within the EGFP locus and insert an mCherry expression cassette (Fig. 9B). A precise integration event would therefore knockout EGFP expression and induce mCherry expression, whereas an off-target insertion would be double positive. These distinct integration events could be readily measured by flow cytometry.

In order to generate the cell line, CRISPR/Cas9 targeting CFTR exon 1 and the requisite donor DNA (Fig. 9A) were co-transfected into IB3-1 cells. Initially, we tried to isolate clonal integrants by sorting EGFP+ cells using FACS. However, despite sorting close to 1000 single cells, not a single well contained a growing colony after 3 weeks post-sorting. Puromycin selection was thereafter performed on the remaining transfected plates in order to isolate clonal integrated cells. Subsequent limited dilution allowed for single cell pipetting into 96-well plates. Single cell clones were allowed proliferate (Fig. 10A), and were then expanded to larger culture vessels (Fig. 10B). To this stage, 6 colonies were successfully expanded and healthy.
Figure 9. Schematic representation of the IB3-1-CFTR-EGFP cell line and the associated colour-switch assay.  

a) To generate the cell line, Cas9 was programmed to cut immediately upstream of CFTR exon1. The donor DNA to be integrated harbours an EGFP-puromycin expression cassette flanked on either side by ~500 bp of sequence homologous to the CFTR locus.  

b) Cas9 is programmed to cut the IB3-1-CFTR-EGFP cell line within the EGFP sequence, and subsequently insert an mCherry expression cassette.
Figure 10. Investigating the identity and CRISPR cleavage efficiency of IB3-1-CFTR-EGFP cell line clones. 

a) Fluorescence microscopy image of clone #2 isolated from puromycin selection (20X magnification) and b) following 10 passages grown to confluency (5X magnification). 

c) PCR amplification of the wild-type CFTR locus and d) the transgenic EGFP sequence from purified genomic DNA of each clone. 

e) T7E1 measuring CRISPR cleavage efficiency in each clone. Each clone was transfected with CRISPR (targeting EGFP) and single-stranded mCherry donor DNA (Fig. 11), and genomic DNA was harvested and purified one week post-transfection. PCR was used to amplify EGFP from each clone, the products of which were denatured and randomly re-annealed. Re-annealed amplicons were then incubated with T7E1 endonuclease to cleave amplicons with base-pair mismatches, and then run on a 2% agarose gel.
In order to verify that the EGFP cassette had been inserted in the correct locus, we tried to amplify the EGFP-CFTR junction from the 6 clones obtained from selection. However, despite being able to amplify the wild-type CFTR locus (Fig. 10C) and the EGFP sequence (Fig. 10D), we were unable to amplify the EGFP-CFTR junction. A myriad of primer sets and PCR conditions were attempted with no success for any of the 6 clones obtained, indicating off-target integrations. Because all 6 clones appeared to have off-target integrations, we speculated that the homology arms flanking the EGFP-puromycin cassette on the donor DNA may not have perfect sequence homology to the IB3-1 CFTR locus, as this is a transformed cell line. To investigate this possibility, we PCR amplified the CFTR locus from genomic DNA isolated from IB3-1 cells, and Sanger sequencing was used to analyze the PCR products. However, both homology arms used on the donor DNA in fact harboured 100% identity to the IB3-1 CFTR genomic DNA (Fig. 11).

However, off-target insertions should not necessarily prohibit the use of these cell lines for the colour switch assay, as the requisite homology arms on the mCherry donor do not span beyond the EGFP cassette (Fig. 9B). Furthermore, numerous studies have demonstrated the utility of linear single-stranded donor DNA (as opposed to circular double-stranded plasmids), for CRISPR/Cas9-mediated gene targeting. Therefore, we also wanted to assess the utility of CtIP and EXO1-4D for enhancing GTE using ssDNA donors. In order to generate ssDNA from the mCherry donor, a T7 promoter was cloned in the reverse orientation downstream of the homology arms (Fig 9B). T7 in vitro transcription followed by DNaseI digestion and RNA purification generates large quantities of donor transcripts. These transcripts are subsequently used for gene-specific reverse transcription, followed by RNaseH digestion and ssDNA purification. Products of each step were run on an agarose gel and can be seen in Fig. 12. In order to verify the identity of the ssDNA, it was used as a template for PCR, purified, and sequenced.
Figure 11. IB3-1 CFTR homology arm sequencing. Genomic DNA from IB3-1 cells was harvested as described in methods. PCR was used to amplify the region corresponding to the homology arms on the IB3-1 CFTR donor plasmid. PCR products were purified and analyzed by Sanger sequencing.
CRISPR/Cas9 (targeting EGFP) and the ssDNA mCherry donor were co-transfected into each of the 6 IB3-1-CFTR-EGFP clones. 3 different sgRNAs targeting different EGFP sequence motifs were tested for each clone. One week post transfection, genomic DNA from each condition was harvested and purified. The T7E1 genomic cleavage assay was performed to assess CRISPR/Cas9 cutting efficiency. This assay relies on the fact that Cas9-induced DSBs are frequently repaired by NHEJ, resulting in indels at the break site. The sequence surrounding the break site is then amplified from genomic DNA by PCR, the products of which are then denatured and randomly re-annealed. This results in mutated sequence (a consequence of Cas9 cleavage) pairing with wild-type sequence. The randomly annealed amplicons are then incubated with T7E1 endonuclease, which cleaves mismatched dsDNA. The cleaved amplicons can then be visualized as two distinct bands on an agarose gel, and are a measure of CRISPR/Cas9 cutting efficiency (Fig. 10E). Only one of the three sgRNAs tested resulted in appreciable cutting efficiencies. Despite successful cutting, however, when the transfected cells were visualized by fluorescence microscopy prior to harvesting genomic DNA, no mCherry fluorescence was evident.
Chapter 5
Discussion

The technological power of CRISPR/Cas9 gene editing has already made a substantial impact on genetics and genomics; agriculture and livestock; stem cell research and tissue engineering; the development of organoids and disease models; as well as a range of gene and cell therapies. Of particular interest is the applicability of CRISPR/Cas9 gene editing for 

More specifically, the power to generate permanent genomic changes in diseased tissues without the use of integrating viral vectors. This represents a tremendous technological leap for our ability to correct heritable diseases at their genetic origins. Our group has sought to apply these technologies specifically for CF gene therapy. Using our HD-Ad vectors, our goal is to deliver CRISPR/Cas9 and the wild-type CFTR gene to permanently correct the CFTR genetic defect in airway stem/progenitor cells, for which we have recently demonstrated successful gene delivery in vivo. However, the utility of this approach is still limited by the inherently low efficiency of CRISPR/Cas9 gene targeting by HDR. In these studies, we have discovered genetically-encoded means by which to significantly enhance GTE, which may improve the clinical utility of in vivo gene therapies.

Our approach to enhancing GTE was largely based on previous studies demonstrating the feasibility of perturbing DSB repair in favour of HDR. Accordingly, we sought to overexpress CDK1, which has extensive regulatory control over the balance between DSB repair pathways, whose activity promotes HDR while inhibiting NHEJ. There was also strong theoretical and empirical support to investigate more proximal regulators of DNA end resection, namely CtIP and EXO1, and site-specific mutants thereof. Our AAVS1-mCherry integration assay initially appeared to be a reliable measure by which to test these factors (Fig. 4C). Initial screens using this assay suggested that CDK1, as well as the CtIP and EXO1 variants could enhance GTE by ~2 to 3-fold (Fig. 5C). However, further investigation revealed that there was significant background mCherry expression from the donor plasmid itself (Fig. 5D). This is puzzling, as there is no upstream promoter on the mCherry donor cassette (Fig. 4B). The background expression is therefore likely a result of weak cryptic promoter activity in the upstream homology arm.
However, there is no documented evidence for promoter activity in intron 1 of the PPP1R12C gene in the literature.

Due to the background expression from the donor plasmid, we reasoned that harvesting transfectants 48 hr post-transfection for flow cytometry was not the most reliable measure of targeting efficiencies. However, these experiments were taken as a sufficient initial screen to demonstrate that CtIP mutants were no more effective than the wild-type protein for enhancing GTE (Fig. 5C). Transfectants of subsequent experiments were passaged over the course of 3 weeks prior to flow cytometry. This effectively diluted the background mCherry expression, such that mCherry expression was only a consequence of successful integration events (Fig. 6). These experiments revealed that CtIP-WT and EXO1-4D can significantly enhance GTE by ~3 and 6-fold, respectively. CDK1 surprisingly had a negative impact on GTE with respect to controls, but this reduction was not statistically significant.

There are a number of potential reasons why CDK1 did not upregulate GTE in these studies. Firstly, the catalytic activity of CDK1 is dependent on binding to its cognate cyclin proteins, which are themselves transcriptionally restricted to S/G2.348,349 Therefore, ectopic CDK1 expression would presumably only upregulate HDR in the cell cycle phases in which HDR is already most active. Such upregulation is also conditional on the relative concentrations of CDK1 to cyclins. That is, if cyclin proteins are endogenously limiting, ectopically overexpressed CDKs would in fact not potentiate CDK activity, and therefore have no influence on GTE. Furthermore, although the effect was not statistically significant, there was a trend towards reduced GTE in response to CDK1 overexpression (Fig. 6). This can potentially be explained by the fact that CDK1 has, in some cases, been shown to down-regulate HDR. For example, CDK1 activity has been shown to inhibit RAD51 nucleoprotein filament formation in M phase Xenopus extracts350, and forced activation of CDK1 in interphase cells can attenuate HDR in human cell lines351. CDKs also control a multitude of metabolic processes, epigenetic modifications and transcriptional programs352. Therefore, their overexpression may be cytotoxic, in turn reducing GTE in transfected cells.

Based on our initial screen using the AAVS1-mCherry assay, the CtIP mutants tested were no more effective in potentiating GTE than the wild type protein (Fig. 5). Accordingly, they were
not included in subsequent experiments. Retrospectively, however, the reliability of these experiments appears to have been severely hampered by background expression from the donor plasmid. For example, the direction of the CDK1 effect was reversed following donor plasmid dilution (Fig. 5C vs Fig. 6C). Flow cytometry following vector dilution was likely a more accurate measure of GTE, and this effect was confirmed by semi-quantitative junction PCR (Fig. 8). In hindsight, therefore, one cannot utilize the initial screens as a reliable measure of GTE, nor potential enhancements thereof.

With that in mind, it may not be reasonable to speculate on why the CtIP mutants were not more effective than the wild-type protein, as it is evident that they were not reliably tested. Future studies should therefore subject these mutants to the experimental protocol used in Fig 6. However, it is worthwhile to retrospectively speculate on the shortcomings of the CtIP mutants generated in these studies. Firstly, the mutants tested concurrently harboured a number of mutations. This may destabilize the native structure of the protein and therefore preclude its potential utility for perturbing DSB repair. Such a phenomenon could be verified by RT-qPCR and western blot. For these reasons, a more suitable approach would be to create a small library of CtIP mutants, each possessing individual mutations and various combinations thereof. However, assuming these combinatorial mutants did not influence native protein folding, there are still limitations to this approach due to the quaternary structure of CtIP. That is, CtIP functions as a homotetramer\textsuperscript{353}, and therefore ectopically-expressed mutant peptides may readily form quaternary structures with wild type conspecifics. This would presumably subject these proteins to wild type regulatory stimuli, thus reducing potential efficacies of these mutants. A viable solution to this limitation would be to co-express siRNAs targeting endogenous CtIP with siRNA-resistant CtIP mutants.

It is evident that the results obtained here for the potential GTE-enhancing effects of the CtIP mutants studied were inconclusive. Perhaps surprisingly, however, CtIP-WT significantly enhanced GTE up to ~3-fold, as evidenced by flow cytometry (Fig. 6) and junction PCR (Fig. 8). This may be a consequence of overcoming the endogenous down-regulation of CtIP-dependent end resection natively mediated by ubiquitylation\textsuperscript{311–313}. That is, CtIP overexpression may saturate or exceed the capacities of the cognate E3 ubiquitin ligases to ubiquitylate CtIP, thus overcoming the associated suppression of end resection.
The substantial increase in GTE in response to EXO1-4D overexpression is unsurprising, as it has been shown to possess hyper-resective activity and potentiate HDR\(^{321}\). However, it is evident that the magnitude of the ~6-fold increase observed using flow cytometry was not entirely recapitulated by the ~3.5-fold increase observed using junction PCR (Fig. 6C vs Fig. 8C). This discrepancy may be a consequence of off-target insertions. The flow cytometry assay measures loci-independent mCherry expression within cells, whereas a positive junction PCR product is dependent on a precise insertion at the AAVS1 locus. The 6-fold enhancement measured by flow cytometry in response to EXO1-4D may therefore reflect an up-regulation of both on- and off-target insertions. If this is indeed the case, the observed enhancements for CtIP would not have been a consequence of off-target insertions, as the fold increases measured by flow cytometry and junction PCR were virtually identical at ~3-fold (Fig. 3C, Fig. 5C). The discrepancy between junction PCR and flow cytometry data for EXO1-4D could also be explained by the semi-quantitative nature of junction PCR. Due to the exponential amplification of DNA in PCR, saturation of amplicons may reduce the dynamic range of absolute band intensities, and thus the fold enhancement in GTE.

We also investigated potential synergistic or additive effects associated with co-expression of CtIP-WT and EXO1-4D (Fig. 6D,E). However, although these combinatorial approaches increased GTE, the effects were not statistically significant, and thus no more effective than the individual factors. Because CtIP and EXO1 function in contiguous biochemical pathways in HDR, their co-overexpression may lead to dramatic hyper-resection phenotypes and severe genomic instability. Concurrently overexpressing these factors may therefore be synthetically lethal and undesirable for our purposes.

The IB3-1-CFTR-EGFP cell line would have been a useful model to test the efficacies of CtIP and EXO1-4D, as it would have been able to distinguish between on- and off-target integration events. Perhaps ironically, it was evident that all 6 clones obtained for the cell line harboured EGFP-puromycin cassettes knocked in at off-target genomic loci. This could not be explained by lack of sequence identity between the donor homology arms and CFTR genomic DNA, as the genomic DNA was sequenced and revealed 100% identity. The results of this cell line generation underscore another significant challenge facing CRISPR technology, that is, specificity. Although not the focus of these studies, the ability to eliminate (or at least minimize) off-target genomic
cleavage events will be especially important for CRISPR therapeutics. This exemplifies the utility of highly specific Cas9 variants such as eSpCas9 and SpCas9-HF$^{197,198}$, which should be used in these studies going forward.

Despite the fact that IB3-1 cell line clones harboured off-target insertions, we reasoned that this would not abolish the utility of the clones. The requisite homology arms on the mCherry donor do not span beyond the EGFP cassette (Fig. 9B), and therefore these clones could still theoretically be used to assess precise insertions. Accordingly, we sought to test the utility of ssDNA donors using this assay, as they have become the preferred donor template for many CRISPR gene targeting applications$^{333-338}$. However, upon co-transfection of the mCherry ssDNA donor and CRISPR/Cas9 into each of these clones, no mCherry expression was evident by fluorescence microscopy. This may have been a result of the integrity of the ssDNA production. As seen in Fig. 12, the purified ssDNA migrates as a smear and faster than the corresponding RNA of similar size. It remains unclear if this is simply the nature of how long (>2kb) ssDNA migrates on a non-denaturing agarose gel, or if the product is a heterogeneous mixture of various species. Future experiments should perform linear PCR (using a single primer) on the plasmid mCherry donor template to generate ssDNA, in order to observe its migratory behavior on an agarose gel. This will help evaluate the identity of the ssDNA. Furthermore, plasmid donor DNA should be co-transfected with the CRISPR/Cas9 plasmid into the IB3-1 cell line clones and monitored for mCherry expression. These experiments will help explain the inability to elicit mCherry expression in the IB3-1 cell line clones using ssDNA donors in these studies.

### 5.1 Future directions

Further studies will undoubtedly need to confirm the efficacy of these approaches in primary human cells to evaluate their applicability for clinical therapies. For such purposes, these factors will need to be cloned into our HD-Ad vectors together with CRISPR/Cas9 and donor DNA. Changing the context of the donor DNA from plasmids to HD-Ad vectors may influence the utility of these factors, as it has been shown that the type of DSB repair pathway employed can be
dependent on the nature of the repair template\textsuperscript{354–356}. Indeed, there are evident differences in the behavior of viral DNA and small circular plasmids within the nucleus\textsuperscript{357,358}.

If the efficacies of these factors are translatable to viral vectors in the context of primary cells and animal models, this will have a tremendous impact on clinical applications for CF gene therapy. Preliminary data from our laboratory suggests that we can achieve permanent transgene integration in \textasciitilde 10% of transformed airway cells using our CRISPR HD-Ad vectors, without CtIP or EXO1-4D (unpublished data). If the addition of CtIP or EXO1-4D to our CRISPR HD-Ad vectors elicits GTE enhancements consistent with the data obtained here, we may be able to achieve therapeutically-significant levels of gene editing in airway stem/progenitor cells. Additional barriers to this goal include \textit{in vivo} gene delivery and the host immune response. To these ends, we have recently shown that our vectors can effectively target airway stem/progenitor cells in primary culture, as well as in mouse and pig lungs \textit{in vivo}\textsuperscript{85}. Furthermore, studies have shown that \textasciitilde 10% of wild type CFTR expression in CF airways is sufficient to correct lung pathology\textsuperscript{359,360}. Therefore, if the GTE enhancing strategies discovered here can be applied to our HD-Ad vectors, in parallel with improvements in subverting the host immune response, we may be able to achieve permanent correction of CF lung pathology.

Additional genetically-encoded approaches to enhancing CRISPR/Cas9 GTE have recently been documented. For example, Canny \textit{et al.} has generated a small ubiquitin peptide variant, i53, that inhibits 53BP1 accumulation at DSB sites, in turn down-regulating NHEJ and promoting compensatory HDR\textsuperscript{361}. The authors demonstrate that i53 can be used to enhance CRISPR/Cas9 GTE by up to 5.6-fold in mouse and human cell lines using ds- and ssDNA donors. Furthermore, the upregulation of HDR using i53 was largely dependent on the presence of CtIP. This suggests that potential synergistic effects on GTE could be elicited by co-expressing i53 and CtIP (or mutant CtIP variants). This may indeed be true for EXO1 as well, although not tested in this study. Due to the large cloning capacity of HD-Ad vectors (\textasciitilde 36 kb), i53 could be readily packaged into our vectors together with either CtIP or EXO1-4D, in order to further enhance CRISPR/Cas9 GTE.

Despite the advancements made here and elsewhere for enhancing CRISPR/Cas9 GTE, there are still unanswered questions and further room for improvement. It is commonly understood in the
research community that the favourability of NHEJ over HDR represents the sole and principle impediment to CRISPR/Cas9 GTE. However, there are likely additional cellular and molecular barriers to GTE that have been unexplored and largely overlooked, yet may be equally relevant. These include the bias for sister chromatids as templates for repair in HDR, and the necessity of exogenous donor DNA to spatially co-localize with the integration site for successful gene targeting.

When HDR is employed to repair DSBs in post-replicative cells, the sister chromatid template is physically linked to the DSB by the cohesin complex and associated factors. Cohesin maintains spatial co-localization of sister chromatids, and is crucial for efficient HDR in yeast and mammalian cells. Accordingly, cohesin is readily recruited to DSBs via the DNA damage response, and genome-wide sister chromatid cohesion is up-regulated in response to DSBs in S/G2. Furthermore, the ability of a particular sequence to function as a template for recombinational repair is primarily limited by its proximity to the site of DNA damage. Unsurprisingly, the frequency at which sister chromatids are used as templates for HDR is orders of magnitude greater than that of ectopic sequences. Efficient HDR is therefore dependent on the spatiotemporal co-localization of the template sequence with the DSB. It logically follows that during the stage of the cell cycle in which HDR is most efficient (S/G2), the sister chromatid physically associated with the DSB will predominate as the template for repair over exogenous DNA that lacks a spatial bias for the genomic integration site. This is supported by the fact that down-regulating essential cohesin subunits dramatically increases aberrant gene conversion between homologous chromosomes. The biased use of sister chromatids as templates for HDR may therefore be a significant barrier suppressing CRISPR/Cas9 GTE.

The challenge of the sister chromatid bias is further compounded by the fact that exogenous DNA does not have the liberty of unrestricted nuclear diffusion. In fact, some reports have shown that plasmid DNA is >100-fold less mobile than soluble proteins of similar dimensions, and that DNA ranging from 21 bp to 6 kb is completely immobile. Similarly, others have reported that plasmids >5 kb accumulate in stagnant clusters at the periphery of the nucleus at sites of nuclear entry. On the contrary, some reports suggest that DNA <400 bp rapidly diffuses throughout the nucleus and cytoplasm, whereas others have demonstrated that plasmid mobility and localization depends on its transcriptional output. It appears that the spatiotemporal
dynamics of exogenous DNA within the nucleus remains relatively unclear. However, one could reasonably conclude that an exogenous donor template would not have the liberty of unrestricted nuclear diffusion, thus further amplifying the limitation of lacking a spatial bias for the genomic integration site.

The impetus of conventional wisdom in this field has restricted investigative efforts on potentiating compensatory HDR by inhibiting NHEJ. This is straightforwardly obvious, as gene targeting depends on HDR. This approach also subverts the sister chromatid bias by artificially upregulating HDR in G1 – i.e., in the absence of a sister chromatid. However, the lack of an active spatial bias of the donor template for the target site raises additional concerns in this context. That is, if HDR is employed at targeted DSBs in G1 in the absence of a spatiotemporally associated donor template, ectopic recombination between non-allelic genomic sequences is a foreseeable consequence. Such recombination events can lead to loss of heterozygosity (LOH), gross chromosomal rearrangements (GCRs) and tumorogenesis\textsuperscript{379,380}. These potential consequences deserve more attention from the research community investigating ways to enhance CRISPR-mediated gene targeting.

A number of potential strategies independent of DSB repair pathway choice could be employed to circumvent the prospective deleterious consequences mentioned above. One idea would be to minimize the use of sister chromatids as templates for repair in HDR in S/G2, thus raising the probability that exogenous sequences would be used as templates for repair. This could be accomplished by targeting critical regulators of cohesin dynamics, such as sororin and WAPL, to reduce the physical association of sister chromatids in S/G2\textsuperscript{381,382}. However, this approach introduces the possibility of adverse cytogenetic effects that are observed in cohesinopathies\textsuperscript{383}. A potential solution to this would be to localize perturbations in cohesin dynamics to the target integration site. This could be accomplished by using dCas9 as a fusion platform to direct cohesin regulators to target sites, analogous to the use of dCas9 for transcriptional and epigenetic regulation. This approach would additionally require the use of an orthogonal programmable nuclease to induce the targeted DSB. The principle drawback here would therefore be the necessity for excessive molecular machinery and attendant cloning capacities.
There may be a more parsimonious and effective way to overcome a number of these challenges simultaneously. That is, active spatiotemporal recruitment of donor DNA molecules to the target integration site. The approach would involve covalently linking ssDNA donors to the Cas9 enzyme inducing the targeted DSB. This could be effectively accomplished by fusing Cas9 to SNAP-tag, a synthetic mutant of the DNA repair protein O^6-alkylguanine-DNA alkyltransferase (AGT) that reacts specifically with benzylguanine (BG) derivatives to generate covalent linkages\(^3\). BG derivatives can themselves be covalently linked to ssDNA donors if the ssDNA is generated using terminal amine-modified oligos. The covalent linkage between Cas9-SNAP and BG-ssDNA can then be generated in vivo following co-transfection of Cas9-SNAP and BG-ssDNA, or by in vitro methods (see online New England Biolabs, SNAP-tag Technologies for details). Interestingly, studies have shown that Cas9 remains strongly associated with DSB cleavage products following nucleolytic catalysis\(^1\). Cas9-SNAP-BG-ssDNA would therefore establish and maintain donor co-localization following DSB induction, thus increasing the probability that the localized donor will be used as a template for subsequent repair. Cas9-SNAP-BG-ssDNA would be a large nucleoprotein complex, theoretically capable of making targeted genomic DSBs while delivering the template donor DNA to the target locus. Using this engineered approached, therefore, not only would Cas9 function as a programmable nuclease, but additionally as a molecular motor that guides transgene cassettes to their cognate target loci.

This approach would simultaneously overcome a number of limitations currently facing efficient CRISPR-mediated gene targeting. Namely, this would impose a spatiotemporal bias on the donor DNA for its cognate target integration site. Accordingly, this would increase the probability that exogenously supplied donor DNA is used as a template for Cas9 DSB repair over the sister chromatid in S/G2. This could also be used in parallel with strategies that enhance the use of HDR in G1. In this context, spatiotemporal co-localization of donor DNA would mitigate the possibility of deleterious recombination events at the target site in the absence of a sister chromatid. This approach would also presumably reduce the frequency of off-target insertions, as the donor DNA would be spatially restricted to its target site, and therefore less available for recombination at distant off-target loci.

The main limitation of this approach would be the requirement for chemical modifications on the donor DNA which cannot be genetically encoded. Therefore, these could not be collectively
packaged into viral vectors. However, Cas9-SNAP-BG-ssDNA nucleoprotein complexes could be readily synthesized in vitro and packaged into LNPs for ex vivo and in vivo delivery. These components could also be packaged separately and assembled in vivo. That is, Cas9-SNAP could be encoded and delivered in viral vectors in parallel with LNP-delivered BG-ssDNA. LNPs are able to effectively deliver proteins and nucleic acids to the brain\textsuperscript{385}, retina\textsuperscript{45,46}, inner ear\textsuperscript{47}, liver\textsuperscript{48}, as well as the nasal epithelium\textsuperscript{56,386} and lung\textsuperscript{54,387}. This approach could therefore be amenable to a variety of gene therapy applications, including lung gene therapy for CF.

5.2 Conclusion

Gene therapies have attracted substantial attention from physicians, scientists and the public alike, and have found broad ranging therapeutic applications. However, many gene therapy clinical trials have failed in the last few decades due to insufficient understanding of the major challenges and a lack of the requisite technologies. Classical approaches involving γ-RTs were initially effective but elicited deleterious genotoxic effects due to random genomic insertions. On the contrary, superior integrating vectors like SIN-γ-RVs and lentiviruses with improved safety profiles have demonstrated utility in a number of clinical trials. However, they still rely on semi-random integration, and therefore carry the risk of insertional mutagenesis and oncogenic transformation. Semi-random integration also gives rise to transgene expression heterogeneity amongst targeted cells, thus generating cells with differential potency for cell therapies. Although non-integrating vectors such as AdVs and AAVs subvert these risks and therefore possess better safety profiles, their efficacies for gene therapy are limited by the half-life and finitude of their cognate target cells.

The advent of programmable nucleases such as CRISPR/Cas9 will greatly attenuate these risks and overcome impeding limitations. The targeting capacities of CRISPR/Cas9 to permanently correct gene code will improve the efficacies of non-integrating vectors, while mitigating the genotoxic risks of insertional mutagenesis associated with integrating viruses. Directed genome edits also have the advantage of preserving endogenous regulatory elements and therefore the native spatiotemporal patterns of gene expression. Furthermore, Cas9 has proven to be a robust
platform for rationally-engineered mutants with novel genomic targeting functions. As a consequence, this technology has opened the door to inducing genetic manipulations that were hitherto unfeasible, such as making single base pair edits and targeted epigenetic modulations.

Despite the power of CRISPR/Cas9, immanent challenges remain for the application of these technologies into the clinic. In particular, hijacking HDR of targeted DSBs to insert therapeutic transgenes has thus far remained inefficient with limited solutions. Here, we present the use of CtIP and EXO1-4D as novel, genetically-encoded strategies to perturb DSB repair in favour of HDR and consequently enhance GTE. These studies have demonstrated that CtIP and EXO1-4D overexpression can upregulate CRISPR/Cas9 GTE up to ~3- and 6-fold, respectively, in human cell lines. These discoveries represent a significant step forward for our CF gene therapy strategy, but future studies will be necessary to address a host of lingering unanswered questions. In particular, assessing the utility of these strategies in the context of viral vectors in primary cells and in vivo animal models will be pertinent to clinical translations.

Our efforts to generate the IB3-1-CFTR-EGFP cell line also illuminate the potential for insidious off-target edits, which remain a preeminent concern for CRISPR therapeutics. Furthermore, the long-term consequences of transiently manipulating DSB repair have yet to be thoroughly investigated. The appropriate clinical implementations of CRISPR/Cas9 technologies will be incumbent on our fastidious attention to these salient risk factors. Notwithstanding these persistent challenges, however, CtIP and EXO1-4D may be clinically significant for viral gene therapy for genetic diseases such as CF. Our high capacity HD-Ad vectors provide a unique opportunity to co-package these factors with CRISPR/Cas9 and donor DNA on a single vector, which can readily target airway stem/progenitor cells. The realization of this approach with the requisite technologies may ultimately provide a cure for CF, the most common genetic disease in the Caucasian population.
References


29. Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti,


39. Cartier, N. et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked


49. High, K. H., Nathwani, A., Spencer, T. & Lillicrap, D. Current status of haemophilia


101. Isalan, M., Choo, Y. & Klug, A. Synergy between adjacent zinc fingers in sequence-


212. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells


230. Zhang, W. *et al.* A Werner syndrome stem cell model unveils heterochromatin alterations


302. Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A. D., Feldhahn, N., Resch, W., Oliveira, T. Y., Chait, B. T., Nussenzweig, A., Casellas, R., Robbiani, D. F. & Nussenzweig, M. C. Rifl prevents resection of DNA breaks and promotes immunoglobulin class switching. Science 339, 711–715 (2013).


376. Shimizu, N., Kamezaki, F. & Shigematsu, S. Tracking of microinjected DNA in live cells


