Gene Editing in *Prkdc* Severe Combined Immunodeficiency and Ataxia Telangiectasia

Versha Prakash

School of Biological Sciences

Royal Holloway University of London

Research thesis submitted for the degree of Doctor of Philosophy

DECLARATION OF AUTHORSHIP

I, Versha Prakash, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed: 

Dated: 27-07-2017

ABSTRACT

Primary immunodeficiencies (PIDs) are inherited diseases of the immune system. A variety of molecular defects can cause PIDs, including mutations in genes *PRKDC* and *ATM* (ataxia telangiectasia-mutated), both of which in addition lead to DNA repair defects. Due to a variety of factors including gene size and sensitivity to DNA damage, these genetic defects are not easily amenable to gene addition strategies, and alternative gene therapy strategies are actively being sought. One such strategy is genome editing, the topic of the current work.

PRKDC severe combined immunodeficiency (*PRKDC*-SCID) is a rare inherited disease caused by mutations affecting DNA-PKcs (DNA-dependent protein kinase catalytic subunit). DNA-PKcs is a key component of non-homologous end joining (NHEJ), a major defence mechanism against double-strand break DNA damage. NHEJ is also involved in V(D)J recombination, a process necessary to generate functional immunoglobulins and T cells receptors, and hence for the development of B and T lymphocytes. Consequently, defects in DNA-PKcs lead to enhanced sensitivity to DNA-damaging agents, such as ionising radiation, and severe immune defects.

DNA-PKcs deficiency causes SCID in both animals (mice, dogs, horses) and humans; however, the condition is ultra-rare in humans with only two patients identified so far. Like with many other PIDs, transplantation of matched bone marrow remains the only choice of effective treatment, but it is complicated by sensitivity to conditioning agents that cause DNA damage. Gene transfer to haematopoietic stem/progenitor (HSC/HSPCs) cells first shown to have major therapeutic effects in SCID-X1, continues to show promise. Advances in genome editing have further opened possible approaches to treatment by correction of genetic defects.

Previous work in the laboratory using engineered zinc-finger nucleases (ZFNs) and *Prkdc* repair matrixes demonstrated homology-directed gene correction in *SCID* mouse fibroblasts and HSC/HSPCs, leading to *ex vivo* rescue of the *SCID* phenotype with the latter. The current project broadly focuses on two major goals: (i) characterisation and phenotypic analyses of ZFN-targeted *SCID* mouse fibroblasts, including molecular analyses of genomic correction and rescue of DNA-PK activity, and (ii) *in vitro* gene editing of *Prkdc* mutation using CRISPR-Cas9 system.

In this study, we used *Streptococcus pyogenes* derived Cas9 system to repair the *Prkdc* mutation. Co-delivery of guide RNAs (gRNA) on a Cas9 plasmid along with a corrective

donor template demonstrated efficient *Prkdc* modification in mouse fibroblasts; however, this did not restore DNA-PK activity owing to apparently error-prone DNA repair. Nonetheless, this study indicated the potential of CRISPR-Cas9 system for targeted modification of genes involved in DNA repair disorders.

Based on this observation, we explored targeting of *ATM*, which like *Prkdc*, is involved in DNA repair. *ATM* codes for the protein ATM kinase, deficiency of which leads to Ataxia Telangiectasia (AT), a progressive neurodegenerative disease with associated immunodeficiency, radiation sensitivity and susceptibility to malignancies. Correction of mutations on *ATM* could rescue residual ATM kinase activity – a therapeutic approach for treating AT-related immunodeficiency. We selected and designed gRNAs for disease-causing mutations – 5762ins137, 103C>T, 2T>C and 7638del9 – and demonstrated proof-of-concept editing of *ATM* locus in human cells. Given the multitude of mutations on *ATM*, we further designed global approaches for partial *ATM* cDNA knock-in at endogenous N/C-terminals or full *ATM* cDNA knock-in at *hAAVS1* locus for application in the human haematopoietic stem cells.

These studies on *Prkdc* and *ATM* have shown the feasibility of CRISPR/Cas genome editing on both genes. They have also highlighted the possible enhanced risk of homology-dependent DNA repair events being associated with novel mutations at the repair site, which prevented the recovery of fully repaired fibroblast clones in the case of ZFN-treated *Prkdc*. This work, therefore, points to the need for extra caution in therapeutic strategies involving genome editing of DNA repair gene defects.

ACKNOWLEDGEMENTS

This Thesis would not have reached its completion without the support, guidance, mentorship, insight, and friendships of several people. Firstly, I am thankful to my supervisor Rafael Yáñez for his constant guidance and endless discussions over the years. I am grateful for his patience when I was a beginner and for his continued faith in my efforts regardless of my personal setbacks. Under his mentorship, I have had some great opportunities. Although challenging at times, equating to *extremely* long hours of work, I have thoroughly enjoyed and grown from these experiences. Thank you for always pushing and motivating me.

I also thank my advisor Professor George Dickson for providing critical and constructive feedback throughout my project.

I am grateful for the scholarship from Royal Holloway and for the financial support from my family over several years, without which this PhD would not have been possible.

Thanks to Lisa Woodbine for sharing her wealth of knowledge in DNA repair and for elegantly performing several laborious experiments at the Genome Damage and Stability Clinic facility despite her hectic schedule. Thanks to Eric Bennett for taking interest in my work and performing IDAA in Copenhagen. Many thanks to my undergraduate project students – Pramila Gurung and Daniyal Asghar – for bearing with me and for their excellent research associated with my project. I am appreciative of Taeyoung Koo at the Institute of Basic Sciences in South Korea for so aptly performing NGS for me very recently.

Thank you to Neda and members of George's lab for training me during the first year of my PhD. I am also thankful to Simona and Rob for their constant technical support. Thank you to everyone in Office 4-04 for continually keeping up the humour and for some brilliant pranks! Thanks to Dev, Chris and Grant for being excellent sources of brainstorming.

I am incredibly thankful to Helena, Anna, Eva and Kate for being the ideal role models. Special thanks to Daniela for her utmost affection, care, and Italian food at times of most need. Thank you to Claudio for keeping me sane over the last few months and for reminding me of a fun life outside of work. You all have been my family away from home and I would not have made it this far without any of you. Lastly, I thank my brother Devendu for his ever-present faith in my abilities.

TABLE OF CONTENTS

ABBREVIATIONS

AAV	adeno-associated virus
Alt-NHEJ	alternative non-homologous end-joining
A-T	ataxia telangiectasia
ATM	ataxia telangiectasia mutated
BSA	bovine serum albumin
Cas	CRISPR associated systems
CMV	cytomegalovirus
Cre	cyclization recombinase enzyme
CRISPR	clustered regularly interspaced palindromic repeats
crRNA	CRISPR RNA
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DNA PK	DNA protein kinase
DNA PKcs	DNA protein kinase catalytic subunit
DSB	double strand break
eGFP	enhanced green fluorescent protein
FBS/FCS	foetal bovine/calf serum
G418r	G418 resistant
G418s	G418 sensitive
gRNA	guide RNA
HDR	homology-mediated repair
HEK-293T	human embryonic kidney cell line, containing SV40 large T-antigen
HeLa	cervical cancer cell line derived from Henrietta Lacks
HIV	human immunodeficiency virus
HR	homologous recombination
HSC/HSPC	haematopoietic stem cell
HSCT	haematopoietic stem cell transplantation
IDAA	indel detection using amplicon analysis
IDLV	integration deficient lentiviral vector
Indels	insertions/deletions
IPLV	integration proficient lentiviral vector
MMEJ	microhomology mediated end-joining
mTert	mouse telomerase reverse transcriptase
Neo	neomycin gene
Neo-R	neomycin resistance marker selection cassette
NHEJ	non-homologous end joining
PIDs	primary immune disorders
RS-SCID	radiosensitive severe combined immunodeficiency
SCID	severe combined immunodeficiency
TALENS	transcription activator like effector nucleases
TIDE	tracking Indels by decomposition
tracrRNA	trans CRISPR RNA
V(D)J	variable (diverse) joining recombination
WPRE	woodchuck hepatitis post-transcriptional regulatory unit
ZFNs	zinc finger nucleases

TABLE OF FIGURES

TABLE OF TABLES

1 Literature Review

1.1 Gene Therapy

Gene therapy is a paradigm of medicine with enormous potential that relates to treatment of diseases by the modification and regulation of genes (Wolff and Lederberg 1994). Ideally, effective gene therapy should correct the faulty gene to restore its normal function and not cause undesirable effects in the rest of the genome (Yanez and Porter 1998). Gene therapy can be broadly administered using two approaches namely, gene addition and gene editing. Gene addition strategies aim at supplementing the genome with a functional copy of the faulty gene. This is particularly useful for recessive disorders that arise due to lack of or defects in associated genes and their products. Genome editing on the contrary allows permanent modification of the existing genome and can be used for introduction, disruption or repair of genetic loci of interest. Modifications can be carried out *in vivo* through localized or systemic injections directly into cells or tissues of interest, or through *ex vivo* procedures, where the cellular modification is carried out *in vitro* outside the body followed by re-transplantation of these modified cells back into the body (Wolff and Lederberg 1994). The applicability of either gene therapy strategy depends on the genetic defect, the kind of cells affected by the diseases and hence to be treated, and the appropriate use of gene delivery vectors (Verma and Weitzman 2005).

Several gene therapy trials addressing numerous conditions have been performed in the past two decades **(Fig. 1-1)**. As of August 2016, 2,409 gene therapy clinical trials have been completed, are ongoing or have been approved across 37 countries worldwide, with representatives from all five continents (http://www.wiley.com/legacy/wileychi/genmed/clinical/). The clear majority of the gene therapy trials (82.3%) have addressed cancers, monogenic disorders and chronic infection, but only a few have reported clear clinical benefits and in some, individuals experienced severe adverse events related to the gene therapy vectors.

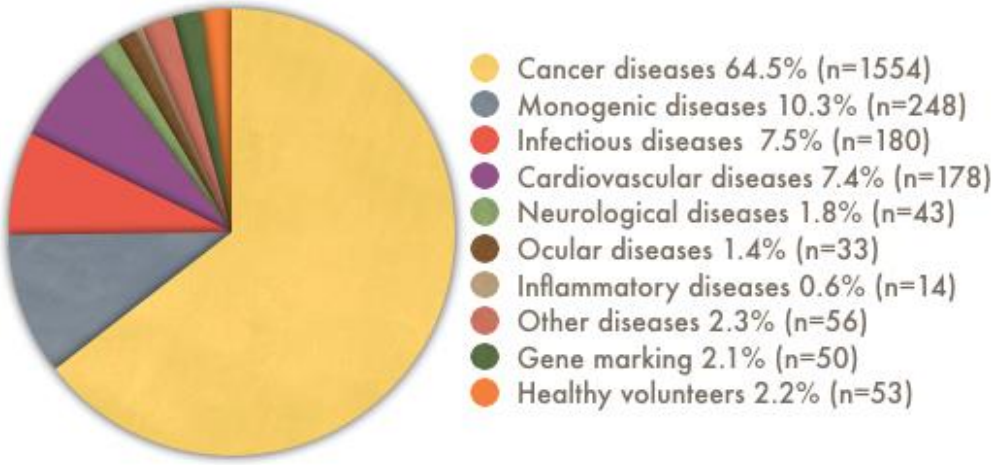


Figure 1-1 Conditions Addressed by Gene Therapy Trials (August 2016)

Source: Journal of Gene Medicine www.wiley.com/legacy/wileychi/genmed/

From the outset, gene therapy aimed to treat inherited disorders, assuming that monogenic diseases would be the easiest to target. The ultimate aim in treating monogenic diseases by gene therapy is the correction of the disorder by the stable transfer of the functioning gene into dividing stem cells to ensure the permanence of the correction, and there has been an array of monogenic disease targeted. After cancers, monogenic diseases are the second most common group of disorders to be targeted by gene therapy **(Fig 1.1)**. Moreover, it also represents the disease group in which the greatest success of gene therapy has been achieved to date (Naldini 2015). Amongst the gene therapy clinical trials for inherited monogenic disorders, cystic fibrosis, the most common inherited genetic disease in Europe and the USA, has been a prime target for gene therapy (Alton, Boyd et al. 2015, Griesenbach, Pytel et al. 2015). The second most common group of inherited diseases targeted are the primary immunodeficiency disorders (PIDs) (Ginn, Alexander et al. 2013, Cicalese and Aiuti 2015, Naldini 2015).

In the first clinical trials of gene therapy for PIDs, retroviral vectors were used in which expression of the normal transgene was driven by the retroviral vector expression. With this approach, successful and durable T cell reconstitution was achieved in patients with X-linked severe combined immunodeficiency (SCID-X1) (Cavazzana-Calvo, Hacein-Bey et al. 2000, Hacein-Bey-Abina, Hauer et al. 2010, Gaspar, Cooray et al. 2011), adenosine deaminase (ADA) deficiency (ADA-SCID) (Blaese, Culver et al. 1995, Kohn, Hershfield et al. 1998, Aiuti, Slavin et al. 2002, Gaspar, Cooray et al. 2011, Cicalese, Ferrua et al. 2016) and Wiskott–Aldrich syndrome (WAS) (Boztug, Schmidt et al. 2010, Moratto, Giliani et al. 2011). Unfortunately, several patients developed leukaemia. These serious adverse events were caused by integration of retroviral vectors in proximity of

transcription initiation sites of genes (including oncogenes) leading to deregulated expression of the targeted oncogenes (Hacein-Bey-Abina, Von Kalle et al. 2003). To counter these adverse effects several improvements have been made in developing safer vectors and are being used in clinical settings.

Overall, concern and scepticism rose over the further deployment of these strategies. But these attitudes are radically changing. Many phase I/II gene therapy clinical trials using improved retro and lentiviral vectors have reported remarkable evidence of efficacy and safety for the treatment of PIDs like WAS (Aiuti, Biasco et al. 2013, Hacein-Bey Abina, Gaspar et al. 2015, Pala, Morbach et al. 2015) and SCID-X1 (Hacein-Bey-Abina, Pai et al. 2014, Sauer, Di Lorenzo et al. 2014). Moreover, these approaches are being translated to blood-borne disorders, such as thalassemia (Cavazzana-Calvo, Payen et al. 2010), sickle cell disease (Hoban, Orkin et al. 2016, Ribeil, Hacein-Bey-Abina et al. 2017), haemophilia (Lheriteau, Davidoff et al. 2015); and metabolic disorders like X-linked adrenoleukodystrophy (Cartier, Hacein-Bey-Abina et al. 2010).

1.2 Gene Therapy Vectors

Gene therapy clinical trials and studies exploit improved vector technologies, ranging from simple naked DNA to live viruses, to deliver therapeutic genes. The unifying objective however of all these clinical projects is to mediate safe and efficient delivery of therapeutic agents into target cells along with stable transgene expression. A variety of vectors and delivery platforms have been investigated in gene therapy studies. Although non-viral approaches are becoming increasingly common, viral vectors remain by far the most popular, having been used in two-thirds of clinical trials performed with adenoviral, retroviral (RV), lentiviral (LV) and adeno-associated viral (AAV) vectors being the most commonly used gene delivery systems in the clinical settings till date **(Fig. 1-2)**.

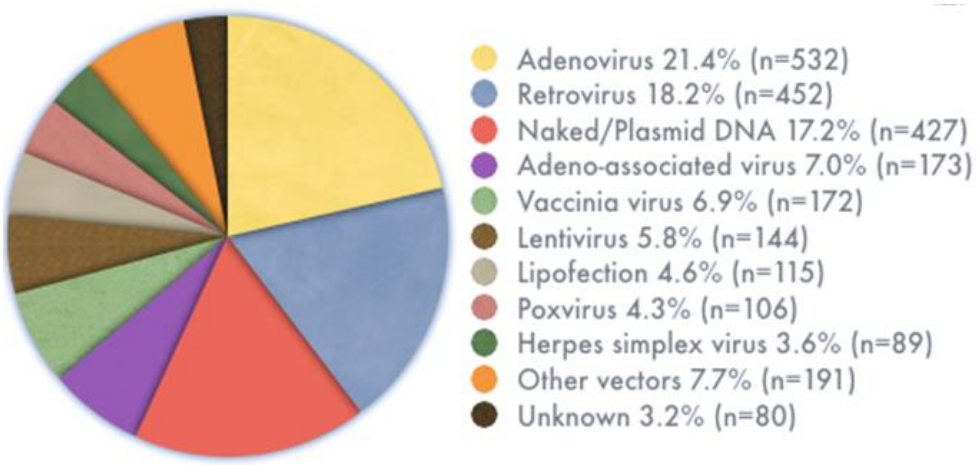


Figure 1-2 Vectors Used in Gene Therapy Clinical Trials (August 2016)

Source: Journal of Gene Medicine *www.wiley.com/legacy/wileychi/genmed/*

The basis of using viruses as vectors lies in harnessing their inherent capability of delivering genetic information into cells and re-populating them to produce large amounts of progeny. The harmful effects of a viral infection mostly occur from the hazardous products of certain viral genes. Therefore, by separating the disease-causing elements, viruses can be used as vectors carrying desired genetic information (Verma and Weitzman 2005).

1.2.1 Retroviral Vectors

Retroviral vectors were the first to be developed for transfer of genetic material in gene therapy. Retroviruses belong to the Reteroviridae family of single-stranded RNA spherical viruses (80 to 120 nm diameter) (Vogt and Simon 1999). In the nucleus, the retroviral particle consists of two copies of single stranded RNA genome along with the

enzymes reverse transcriptase, integrase and protease. The viral genome encodes for three proteins for gag, pol and env, which are required for viral replication and packaging. These components are complexed inside a nucleocapsid protein, which is further enclosed inside a second protein shell formed by the capsid protein (Jones and Morikawa 1998). Matrix proteins form a layer outside the core and interact with a cell-derived lipid envelope, which incorporates viral envelope glycoproteins (env), responsible for the interaction with specific cell receptors. Two units form these glycoproteins, a transmembrane that anchors the protein into the lipid bilayer and a surface protein, which binds to the cellular receptor **(Fig. 1-3).**

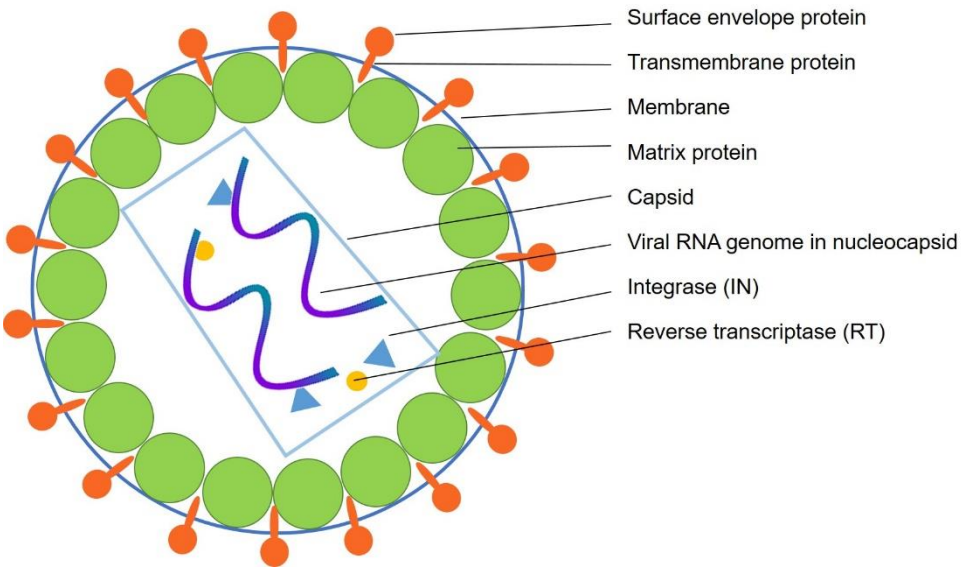


Figure 1-3 Schematic Representation of Retroviral Particle

At the onset of infection, the surface glycoprotein envelope interacts with receptors on the surface of the target cell to gain entry. When inside the cell, the single stranded viral genome is converted into linear double stranded proviral DNA by virus encoded reverse transcriptase. The double stranded DNA is transported to the nucleus of the host cells. For many types of retroviruses, mitosis, is required for the breakdown of the nuclear envelope for the proviral DNA to reach the nucleus. Exceptions to the requirement of cell division for infection are lentiviruses that can actively transport the DNA to the nucleus of non-dividing cells. Once the provirus reaches the host nucleus, it is integrated into the host genome by integrase. The provirus then undergoes transcription and translation along with rest of the genome, resulting in assembly of new viral particles that bud off the surface of the target cell to infect other cells (Robbins and Ghivizzani 1998, Weiss 1998, Kurian, Watson et al. 2000).

Based on the organization of their genomes, the retroviridae are divided into simple and complex retroviruses. Simple retroviruses include oncoretroviruses such as murine leukaemia virus while complex retroviruses comprise of lentiviruses like the human immunodeficiency virus 1 (HIV-1) (Robbins and Ghivizzani 1998). Vectors derived from oncoretroviruses have been attractive for gene therapy as they have a relatively large coding capacity (8-10 kb) and allow stable integration of genetic material into the host genome, leading to persistent long-term expression. Furthermore, they do not transfer viral genes. This avoids the destruction of vector transduced cells by cytotoxic T cells that are specific to viral proteins. In this sense, oncoretroviral vectors are less immunogenic and do not elicit anti-vector immune responses (Robbins and Ghivizzani 1998).

Although numerous clinical studies benefit from the use of oncoretrovirus-based vectors owing to their high efficiency in transducing dividing cells (Robbins and Ghivizzani 1998, Kurian, Watson et al. 2000). There are some important limitations: (i) instability of the viral particle, (ii) low viral titres, (iii) inability to transduce non-dividing cells and (iv) insertional mutagenesis (Robbins and Ghivizzani 1998). These limitations were clear in the first gene therapy trials where retroviral vectors were used for the treatment of X-linked severe combined immunodeficiency (SCID-X1), chronic granulomatous disease (CGD), adenosine deaminase (ADA)-deficient severe combined immunodeficiency (ADA-SCID) and Wiskott-Aldrich syndrome (WAS) (Aiuti, Slavin et al. 2002, Ott, Schmidt et al. 2006, Boztug, Schmidt et al. 2010, Cavazzana-Calvo, Payen et al. 2010). While the treatment was effective for immune restoration, severe complications arose in some of the patients in whom the integrated retroviral vectors led to leukocyte proliferative disorders (Hacein-Bey-Abina, Von Kalle et al. 2003, Ott, Schmidt et al. 2006, Braun, Boztug et al. 2014).

1.2.2 Lentiviral Vectors

Lentivectors resemble retroviral vectors in their ability to integrate into target cell genome resulting in persistent transgene expression. They can transduce both dividing and non-dividing cells, have a larger transgene toleration capacity (~7.5 kb), and a more favourable insertional profile than retroviral vectors (Matrai, Chuah et al. 2010). Moreover, it is possible to minimise the risk of non-specific insertion by modifying the vector design or by employing integration-deficient lentiviral vectors (Yanez-Munoz, Balaggan et al. 2006, Philpott and Thrasher 2007, Wanisch and Yanez-Munoz 2009).

The design of lentivectors is based on the separation of *cis*- and *trans*-acting sequences. This increases the safety of lentivector produced. Components necessary for virus

production are thus divided across packaging, envelope, and transfer plasmid constructs. The lentiviral production systems are divided into three different generations depending upon the packaging plasmid used for production **(Fig. 1-4)** (Vigna and Naldini 2000).

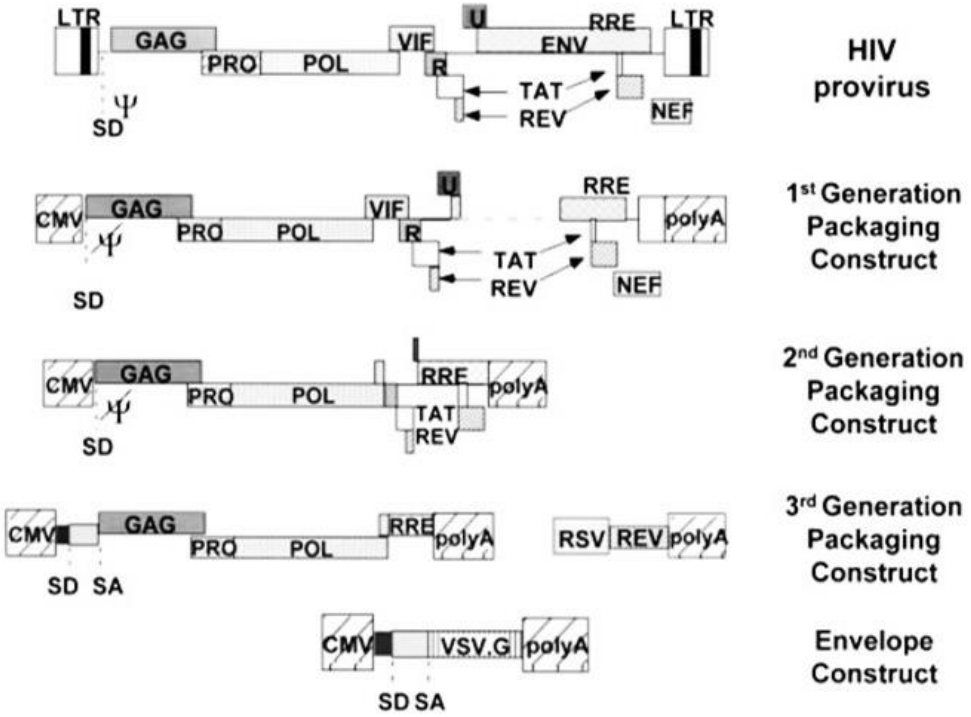


Figure 1-4 Schematic Representation Lentiviral Packaging Constructs

In the packaging vectors, viral long-term repeats (LTRs) are replaced by strong promoter (CMV) and poly-adenylation (poly-A) signal. 1st-generation packaging construct contains all genes encoding for structural and accessory proteins of HIV-1 except for the envelope. In the 2nd-generation packaging construct, all the accessory genes are deleted. In the 3rd-generation packaging construct, the sequences encoding the *Tat* and *Rev* protein are also eliminated. *Rev* is expressed on a separate construct under the control of the RSV promoter. The envelope construct expresses the VSV-G protein under the control of the CMV promoter. PRO, promoter; SD, splice donor; SA, splice acceptor; RSV, rous sarcoma virus; CMV, cytomegalo virus; RRE, Rev-responsive element; VSV-G, vesicular stomatitis virus glycoprotein; ψ, packaging signal. Taken from (Vigna and Naldini 2000).

The first-generation packaging plasmid provides all *gag* and *pol* sequences, the viral regulatory genes *tat* and *rev* and the accessory genes *vif*, *vpr*, *vpu* and *nef*. Identification of the HIV genes disposable for transfer of the genetic cargo allowed the engineering of the multiple-attenuated second-generation packaging systems (Zufferey, Nagy et al. 1997). In these, the four accessory genes, *vif*, *vpr*, *vpu* and *nef*, are removed without negative effects on vector yield or infection efficiency, and as such improving lentivector safety, since any replication-competent lentivirus would be devoid of all virulence factors. The third-generation lentiviral system further improves on the safety of the second-generation. It consists of a split-genome packaging system in which the *Rev* is expressed

from a separate plasmid under the control of a heterologous promoter, such as RSV **(Fig. 1-4)** (Dull, Zufferey et al. 1998, Vigna and Naldini 2000, Cockrell and Kafri 2007).

The lentiviral transfer plasmid which encodes the transgene is depicted in **Fig. 1-5**. The transgene is flanked by LTR sequences and typically, it is the sequences between and including the LTRs that are integrated into the host genome (Dull, Zufferey et al. 1998, Vigna and Naldini 2000). For safety reasons, lentiviral vectors are all replication incompetent and may contain an additional deletion in the 3'LTR, rendering the virus self-inactivating (SIN) after integration (Zufferey, Dull et al. 1998). Key components of the third-generation lentiviral vector system are listed in **Table 1.1.**

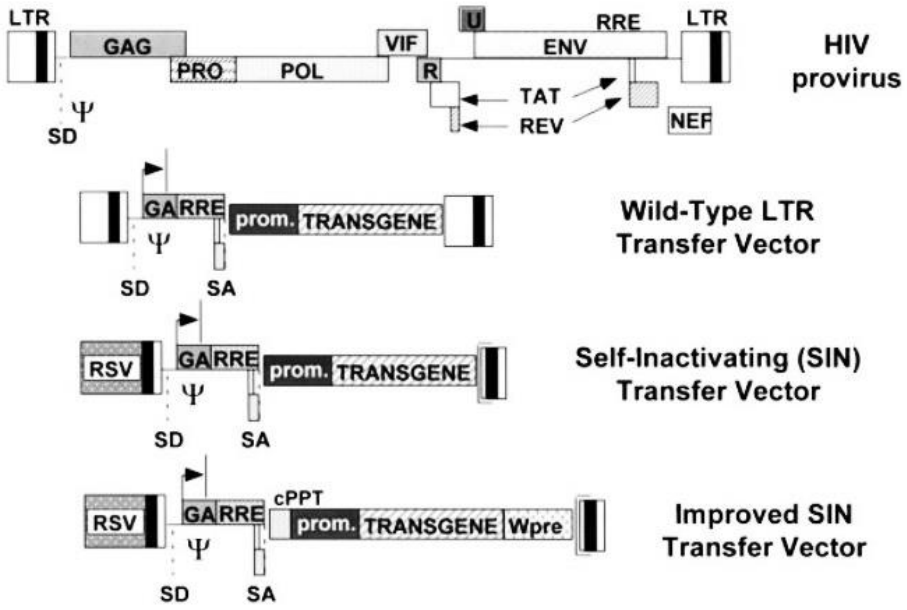


Figure 1-5 Schematic of Lentivector Transfer Plasmids

Transgene is driven by an internal promoter. The wild-type LTR transfer vector maintains the unmodified HIV LTRs. The SIN transfer vector carries a large deletion in the U3 region (represented by [) of the 3'LTR completely inactivating its activity. The 5' LTR is modified by substituting the U3 region with a strong promoter (RSV). The improved SIN vector contains cPPT and WPRE sequences. ψ, packaging signal; prom., internal promoter; SD, splice donor; SA, splice acceptor; GA, a portion of *gag*; cPPT: central polypurine tract; WPRE, woodchuck hepatitis post-transcriptional regulatory element; LTR, long-terminal repeat. Taken from (Vigna and Naldini 2000).

Table 1-1 Glossary of Third-Generation Lentiviral Vector Components

PLASMID	ELEMENT	DESCRIPTION
ENVELOPE	VSV-G	Vesicular stomatitis virus G glycoprotein; broad tropism envelope protein used to pseudo type most lentiviral vectors.
PACKAGING	Gag	Precursor structural protein of the lentiviral particle containing Matrix, Capsid and Nucleocapsid components.
	Pol	Precursor protein containing Reverse Transcriptase and Integrase components.
	Rev	Binds to the Rev Response Element (RRE) within un-spliced and partially spliced transcripts to facilitate nuclear export.
TRANSFER	cPPT/cTS	Central polypurine tract; recognition site for proviral DNA synthesis. Increases transduction efficiency and transgene expression.
	Ψ (*Psi*)	RNA target site for packaging by Nucleocapsid.
	RRE	Rev Response Element; sequence to which the Rev protein binds.
	WPRE	Woodchuck hepatitis virus post-transcriptional regulatory element; sequence that stimulates the expression of transgenes via increased nuclear export.
	LTR	Long terminal repeats; U3-R-U5 regions found on either side of a retroviral provirus.
	U3	*Unique 3'; region at the 3' end of viral genomic RNA. Contains sequences for activation of viral genomic RNA transcription.*
	R	*Repeat region found within both the 5' and 3' LTRs of retro/lentiviral vectors.*
	U5	*Unique 5'; region at the 5' end of the viral genomic RNA.*
	5' LTR	Acts as an RNA pol II promoter. The transcript begins at the beginning of R, is capped, and proceeds through U5 and the rest of the provirus.
	3' LTR	Terminates transcription started by 5' LTR by the addition of a poly A tract just after the R sequence.

1.2.3 Adeno-Associated Viral Vectors

To minimize the risk of insertional mutagenesis, vectors capable of site-specific integration or devoid of integration are deemed suitable. Adeno-associated viruses (AAV) are explored as one such possibility for maintaining prolonged transgene expression in non-replicating cells, where they persist in a non-integrating and non-replicating episomal form (Berns and Giraud 1995, Gao, Alvira et al. 2002, Grimm and Kay 2003, Verma and Weitzman 2005, Perabo, Endell et al. 2006, Daya and Berns 2008). Twelve human serotypes of AAV (AAV-1 to AAV-12) have thus far been identified, with the best characterized and most commonly used being AAV2 (Gao, Alvira et al. 2002, Daya and Berns 2008). These serotypes differ in their tropism, or the types of cells they infect, making AAV a very useful system for preferentially transducing specific cell types. The lack of pathogenicity of the virus, the persistence of the virus, and the many available serotypes have increased AAV's potential as a delivery vehicle for gene therapy applications. AAV vectors are currently among the most frequently used viral vectors for gene therapy (Carter 2005, Perabo, Endell et al. 2006, Daya and Berns 2008).

AAV is a small (25 nm), non-enveloped virus that packages a linear single-stranded DNA genome. It belongs to the family Parvoviridae and is placed in the genus Dependovirus, because productive infection by AAV occurs only in the presence of a helper virus, usually either adenovirus or herpesvirus (Gao, Alvira et al. 2002). In the absence of helper virus, AAV (serotype 2) can set up latency by integrating into chromosome 19q13.4 (Shi, Arnold et al. 2001, Grimm and Kay 2003).

The small 4.8 kb single-stranded DNA AAV genome consists of two open reading frames encoding for viral genes replication (*Rep*) and capsid (*Cap*), flanked by two 145 bp inverted terminal repeats (ITRs) (Daya and Berns 2008, Kwon and Schaffer 2008, Kotterman and Schaffer 2014). These ITRs base pair to allow for synthesis of the complementary DNA strand. *Rep* and *Cap* are translated to produce multiple distinct proteins. These include Rep78, Rep68, Rep52, Rep40, which are required for the AAV life cycle; and VP1, VP2, VP3, proteins associated with the formation of capsid (Kotterman and Schaffer 2014). When constructing an AAV transfer plasmid, the transgene is placed between the two ITRs, and *Rep* and *Cap* are supplied in *trans* **(Fig. 1-6)** (Kotterman and Schaffer 2014). In addition to *Rep* and *Cap*, AAV requires a helper plasmid containing genes from adenovirus. These genes (*E4*, *E2a* and *VA*) mediate AAV replication (Perabo, Endell et al. 2006, Daya and Berns 2008). The transfer plasmid, *Rep*/*Cap*, and the helper plasmid are transfected into HEK293 cells, which contain the

adenovirus gene *E1+*, to produce infectious AAV particles (Kotterman and Schaffer 2014) **(Fig. 1-6)**.

The separation of *Rep* and *Cap* facilitates the viral pseudo-typing, which is the process of mixing capsids and genomes from different viral serotypes (Wu, Asokan et al. 2006). Use of these pseudo-typed viral vectors improves the transduction efficiency and alters the tropism. One common example is AAV2/5, where AAV2 is the genome and AAV5 is the capsid. The vector targets neurons that are not efficiently transduced by AAV2/2. Furthermore, this serotype shows wide spread expression throughout the brain thus indicating improved transduction profile (Georgiadis, Duran et al. 2016).

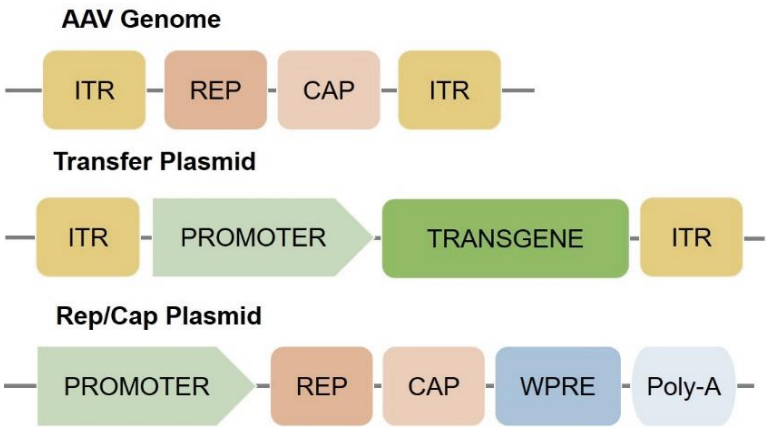


Figure 1-6 Schematic of AAV Plasmid System

To produce an adeno-associated viral (AAV) vector, the AAV genome is split into independent plasmids. A Rep/Cap plasmid encoding viral proteins necessary for replication and formation of AAV particle. The transgene flanked between AAV inverted terminal repeats (ITRs) is supplied on a separate plasmid construct. WPRE, woodchuck hepatitis B posttranscriptional regulatory element; Rep, replication; Cap, capsid, Poly-A, polyadenylation signal.

As part of its lysogenic cycle, wild-type AAV integrates into the host genome at a specific site, AAVS1 on human chromosome 19. This site is favoured due to the presence of a Rep binding element; however, random integrations may occur at a much lower frequency. As a replication-incompetent virus, AAV cannot enter the lytic cycle without help. Another virus, such as adenovirus or herpes simplex virus, or a genotoxic agent such as UV radiation or hydroxyurea, is necessary for lytic cycle activation (Grimm and Kay 2003, Daya and Berns 2008).

When recombinant-AAV (rAAV) is used, the Rep protein is supplied in *trans*, eliminating the ability of rAAV to integrate into its preferred site of genomic integration on human chromosome 19, termed AAVS1 (Burger, Gorbatyuk et al. 2004, Dismuke, Tenenbaum et al. 2013). Instead, the rAAV genome is typically processed into a double stranded,

circular episome through double stranded synthesis. These episomes can concatemerise, producing high molecular weight structures that are maintained extrachromosomally. rAAV are more likely than wild-type AAV to integrate at non-homologous sites in the genome (Burger, Gorbatyuk et al. 2004). Nonetheless, most rAAV particles are thought to be maintained in episomes or concatemers. Episomes differ profoundly from viral particles produced during a lytic cycle. rAAV episomes can develop chromatin-like organization and persist in non-dividing cells for a period of years without damaging the host cell. In contrast, viral particles produced during a lytic cycle are quickly released through cell lysis. Episomal stability enables long-term transgene expression in non-dividing cells and is a key advantage of rAAV (Dismuke, Tenenbaum et al. 2013).

1.3 DNA Repair

1.3.1 Overview

Repair of DNA double strand breaks (DSBs) plays a critical role in the maintenance of the genome. DSB arise frequently because of replication fork stalling and due to the attack of exogenous agents. DSBs can be generated by environmental factors such as ionizing radiation, by cellular metabolic products and as recombination intermediates. In cycling cells, DSBs occur mainly during replication. Repair of broken DNA is essential for cellular survival. Left unrepaired or incorrectly repaired, they can lead to genomic changes that may result in cell death or cancer. DSBs are the most critical type of DNA damage to the cells, as it is believed that a single unrepaired DSB is sufficient for inducing apoptosis (Sonoda, Hochegger et al. 2006, Helleday, Lo et al. 2007, Davis and Chen 2013, Shibata and Jeggo 2014).

Two major pathways, homologous recombination (HR) and non-homologous end-joining (NHEJ) have evolved to deal with these lesions, and are conserved from yeast to vertebrates. Despite the conservation of these pathways, their relative contribution to DSB repair varies greatly between these two species. HR plays a dominant role in any DSB repair in yeast, whereas NHEJ significantly contributes to DSB repair in vertebrates. This active NHEJ requires a regulatory mechanism to choose HR or NHEJ in vertebrate cells (Sonoda, Hochegger et al. 2006, Lieber 2010, Davis and Chen 2013).

1.3.2 Homologous Recombination

Whereas NHEJ is a conceptually simpler cascade of events, homologous recombination is a complex process that exploits undamaged homologous templates of DNA to restore and repair damaged DNA sequences. Homologous recombination is initiated by CtIP/MRE11-dependent resection of double stranded DNA ends (Li and Heyer 2008). The extension of resection is highly regulated and involves the repositioning of 53BP1 via a BRCA1-dependent process (Liu and West 2004, Li and Heyer 2008). The single-stranded tails generated by resection are rapidly bound by Replication protein A (RPA). Nucleoprotein filaments subsequently form after replacement of RPA by RAD51 via a process that involves BRCA2. The nucleoprotein filaments promote invasion of the undamaged homologue, strand displacement and D-loop formation. The displaced strand pairs with the broken and abandoned strand, generating a heteroduplex molecule and a Holliday junction. Subsequently, synthesis using the undamaged strand as a template and ligation repairs the DSB. Finally, resolution of the heteroduplex molecule

generates crossover or non-crossover products depending on the direction of resolution **(Fig. 1.7)** (Paigen and Petkov 2010).

1.3.3 Non-Homologous End-Joining

In mammalian cells, NHEJ DNA repair pathway requires 6 core proteins – Ku70/80 heterodimer, DNA-PKcs, XRCC4 (X-ray complementing Chinese hamster gene) 4), DNA ligase IV, Artemis and XLF or Cernunnos **(Fig. 1-7)** (Lees-Miller and Meek 2003, Meek, Gupta et al. 2004, Mahaney, Meek et al. 2009, Hammel, Yu et al. 2010, Lieber 2010, Davis and Chen 2013). Deletion or inactivation of any of these core NHEJ factors induces marked sensitivity to ionising radiation and other DSB-inducing agents, as well as defects in the V(D)J recombination. In general terms, NHEJ is thought to proceed through the following stages: i) detection of the DSB and tethering or protection of the DNA ends; (ii) DNA-end processing to remove damaged or non-ligatable groups; and (iii) DNA ligation (Lees-Miller and Meek 2003).

DNA ends are first recognised by the NHEJ factor DNA-PK, which is composed of the DNA PKcs and Ku70/80 heterodimer (Blunt, Finnie et al. 1995, Jackson and Jeggo 1995, Secretan, Scuric et al. 2004). After initial loading of the Ku70/80 heterodimer onto DNA ends, DNA PKcs is recruited to form a DNA end synapsis, ensuring protection from exonuclease activities and juxtaposition of DNA ends. The presence of Ku70/80 and DNA-PKcs at DNA ends is not rigid but constitutes a dynamic equilibrium of DNA-bound and DNA-free protein. Trans autophosphorylation of the ABCDE cluster of DNA PKcs (7 phosphorylation sites between residues 2,609 and 2,647) causes conformational changes that facilitate Artermis nuclease activity (Douglas, Sapkota et al. 2002, Secretan, Scuric et al. 2004, Uematsu, Weterings et al. 2007), which is required for the opening of hairpin sealed coding joints. During processing of coding ends, nucleotides can be lost due to exonuclease activity, and non-template nucleotides can be added by terminal deoxynucleotidyl transferase. This contributes tremendously to the diversity of antigen receptor repertoire. Trans autophosphorylation of PQR cluster of DNA-PKcs (6 phosphorylation sites between residues 2,203 and 2,056) functions to limit further end processing and to specifically promote end-joining (Ding, Reddy et al. 2003, Block, Yu et al. 2004, Lieber 2010, Davis and Chen 2013). Finally, the DNA ends are ligated by XRCC4-ligase IV complex, prompted by XLF.

Cells use two mechanistically distinct end-joining pathways to repair DNA DSBs. Classical or c-NHEJ leads to minimal sequence alterations at the repair junctions, whereas alt-NHEJ (also known as microhomology-mediated end joining or MMEJ)

causes extensive deletions (as well as insertions) that scar the break sites following repair (Betermier, Bertrand et al. 2014, Shibata and Jeggo 2014).

Alt-NHEJ or MMEJ is a mutagenic DSB repair mechanisms that use alignment of microhomologous sequences (1-16 nucleotides) flanking the initiating DSB to align the ends of repair (Wang, Perrault et al. 2003). Because of which, the process is associated with deletions and insertions that mark the original break site, as well as chromosome translocations. Whether MMEJ has a physiological role or is simply a back-up repair mechanism is a matter of debate (Wang, Perrault et al. 2003, Wang and Xu 2017)

Classical or c-NHEJ is active throughout the cell cycle and is initiated when the Ku70–Ku80 heterodimer binds to DNA ends with high affinity. Ku then recruits the Ser/Thr kinase DNA-PKcs to phosphorylate several downstream targets, including the terminal end-processing enzyme Artemis that cleaves single-stranded overhangs, and DNA ligase 4 (LIG4) and the scaffold protein XRCC4, which catalyse the ligation of DNA ends. Alt-NHEJ, which is most active in the S and G2 phases of the cell cycle, is dependent on signalling by poly(ADP-ribose) polymerase 1 (PARP1) and relies on 5′–3′ resection of DNA by MRN (MRE11–RAD50–NBS1) and CtBP-interacting protein (CtIp) **(Fig. 1.7)** (Shibata and Jeggo 2014). Base pairing at the resected ends drives their annealing to promote synapsis of opposite ends of a DSB. Annealed ends are subject to fill-in synthesis by the low-fidelity DNA polymerase θ (Pol θ), which stabilizes the annealed intermediates and promotes end joining, primarily by DNA ligase 3 (LIG 3). Alt-NHEJ introduces deletions and insertions that scar the break sites following repair. The deletions are caused by extended nucleolytic processing, whereas the insertions result from the activity of Pol θ (Hammel, Yu et al. 2010, Black, Kashkina et al. 2016).

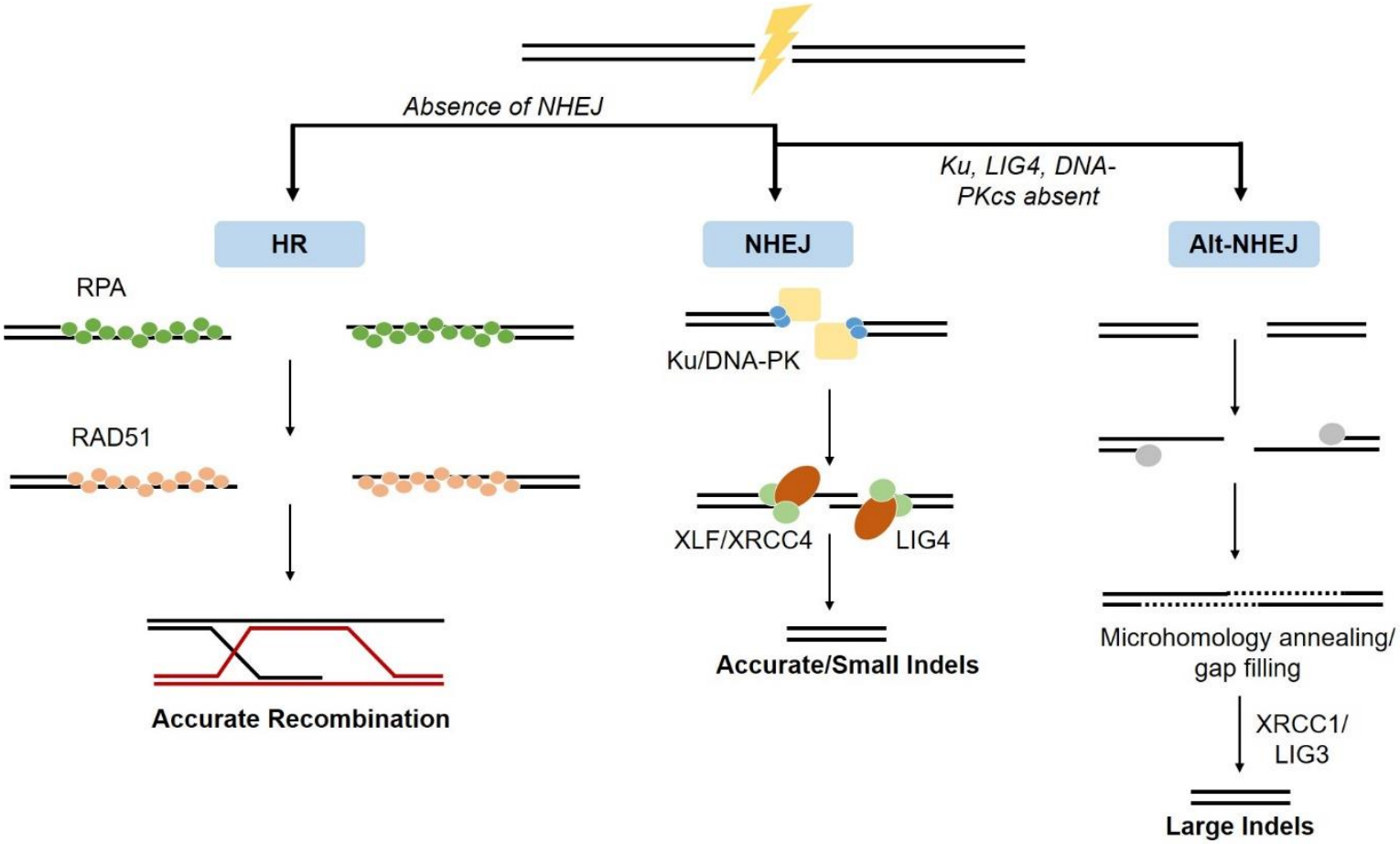


Figure 1-7 Schematic of DNA Repair Pathways

Non-homologous end-joining (NHEJ) is the first choice of double stranded break (DSB) repair pathway in mammalian cells. However, if NHEJ does not ensure in late s or G2 phases of cell cycle, cells exploit the use of homologous recombination (HR). Separately, in absence of key NHEJ proteins such as Ku, DNA ligase IV (LIG4) and DNA protein kinase catalytic sub-unit (DNA-PKcs), the mutagenic alternative NHEJ (alt-NHEJ) predominantly takes place. Indels: Insertions or deletions; DNA-PK: DNA protein kinase; RPA: replication protein A; XRCC1/4: X-ray complementing Chinese hamster gene; XLF: XRCC4-like factor; LIG 3: DNA ligase 3. Adapted from (Shibata and Jeggo 2014).

1.4 Genome Editing

1.4.1 Overview

Gene editing is a gene therapy approach that relies on designer nucleases to recognise and cut specific DNA sequences, and subsequently exploit innate cellular DNA repair pathways, namely non-homologous end joining and homology directed repair, to introduce targeted modifications in the genome **(Fig. 1-8)**.

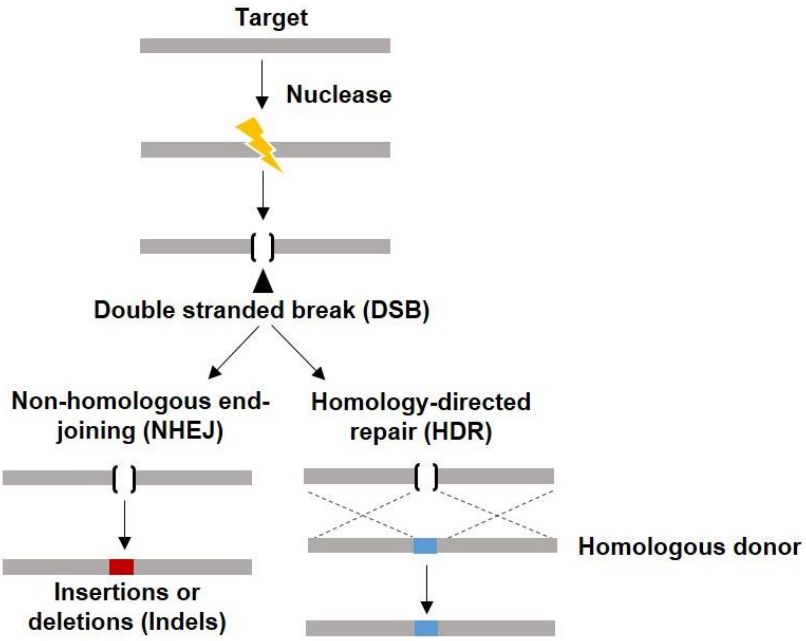


Figure 1-8 Resolution of a Double Stranded Break

A nuclease is targeted towards a defined genomic locus, introducing a DSB. This may undergo one of two major repair pathways known as non-homologous end joining (NHEJ) or homology directed repair (HDR), depending upon cell cycle stage and availability of a DNA donor template. NHEJ is an error prone mechanism, which causes small insertions or deletions (Indels) upon ligating the ends of the DNA break. HDR is a precise mechanism which repairs the break by using a homologous donor template, normally the sister chromatid.

NHEJ repairs the lesion by directly re-joining the DSBs ends together in a process that does not require a repair template. Although NHEJ-repair can be accurate, repeated repair of the same DSB eventually results in insertion and deletion mutations (indels) joining the break site. Indels created can cause frameshift mutations leading to mRNA degradation by nonsense-mediated decay or result in formation of truncated non-functional proteins (Lees-Miller and Meek 2003). In contrast, HDR allows for repair of the targeted genomic region. Since for HR to take place, it is not necessary for the two homologs to be chromosomal, the process can be exploited by supplying a donor or 'template' DNA to the cells and transferring that information into its genome (Thomas

and Capecchi 1986, Capecchi 1989). Upon introduction of a DSB, HDR machinery utilises exogenous single- or double stranded DNA templates that has sequence similarity to the break site, to repair the lesion, incorporating any changes encoded in the template DNA. Homology-directed gene repair offers potential advantages. First, it avoids the potential risks of harmful mutations associated with random non-specific integration of therapeutic DNA. Second, it preserves the original expression and regulation patterns of the gene. Third, it can correct both recessive and dominant disorders, unlike gene addition which are restricted to treating recessive disorders (Yanez and Porter 1998).

In the context of a disease-causing locus, NHEJ can be exploited to excise or disrupt deleterious sequences, or even restore the reading frame of a gene **(Fig. 1-9)**. In contrast, HDR which requires a donor DNA containing sequences homologous to those adjacent to the DSB, can be used to repair a mutation, or to knock-in a block of exons (“superexon”) or a full cDNA at either the endogenous locus (reconstituting the wild-type sequence) or at a genomic ‘safe harbour’ (a region of DNA where transgenes can integrate and express in a predictable manner without insertional mutagenesis or perturbation of gene function) (Sadelain, Papapetrou et al. 2011) **(Fig. 1-9)**.

Gene editing thus opens the possibility of permanently modifying a genomic sequence of interest by enabling targeted disruption, insertion, excision, and correction in both *ex vivo* and *in vivo* settings **(Fig. 1-10).**

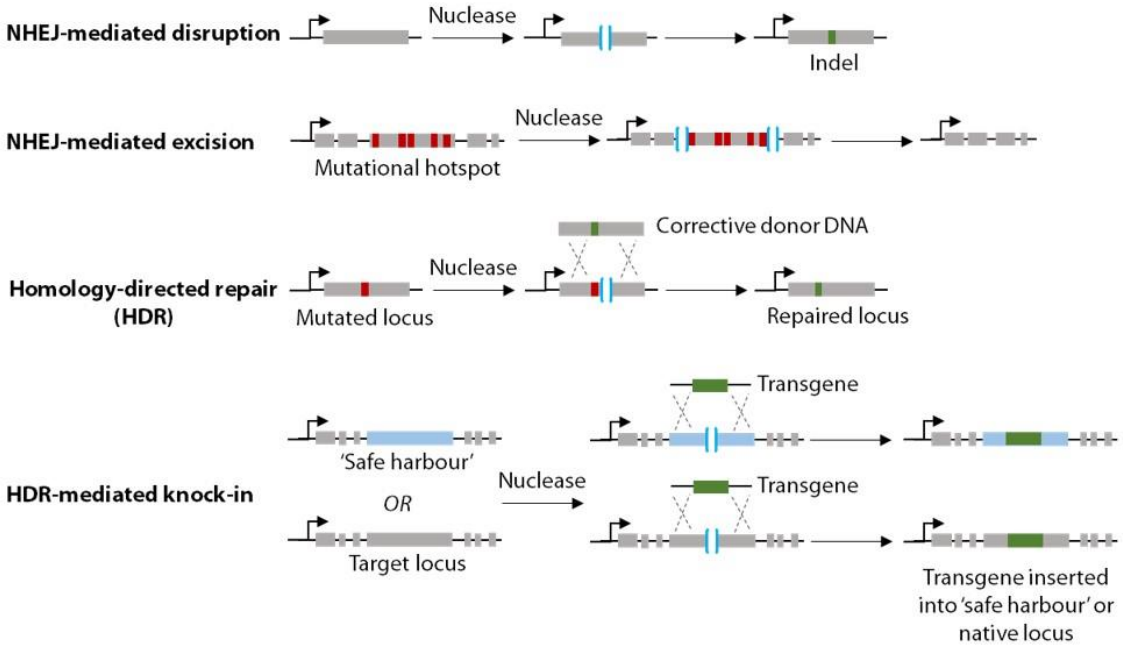


Figure 1-9 Gene Editing Strategies Using DNA Repair Pathways

1) Non-homologous end joining (NHEJ) can be used to disrupt genomic sequences because of insertions or deletions (Indels). This can cause frameshift mutations leading to an early stop codon (or restoration of the reading frame ssby splice site disruption). 2) NHEJ can mediate targeted deletions. This requires generation of double strand breaks (DSBs) on both sides of the target genomic sequence, which then deletes the intervening sequence while NHEJ re-joins the DNA ends. 3) Homology-directed repair (HDR) can be used to correct a specific mutation by introducing a nuclease-mediated DSB (in proximity to the target site) in the presence of a homologous donor DNA containing corrective sequence. Upon recombination, the repair template corrects the mutated locus. 4) Likewise, by supplying exogenous DNA on the donor template flanked between regions of homology, HDR can be used to mediate targeted gene insertion or knock-in.

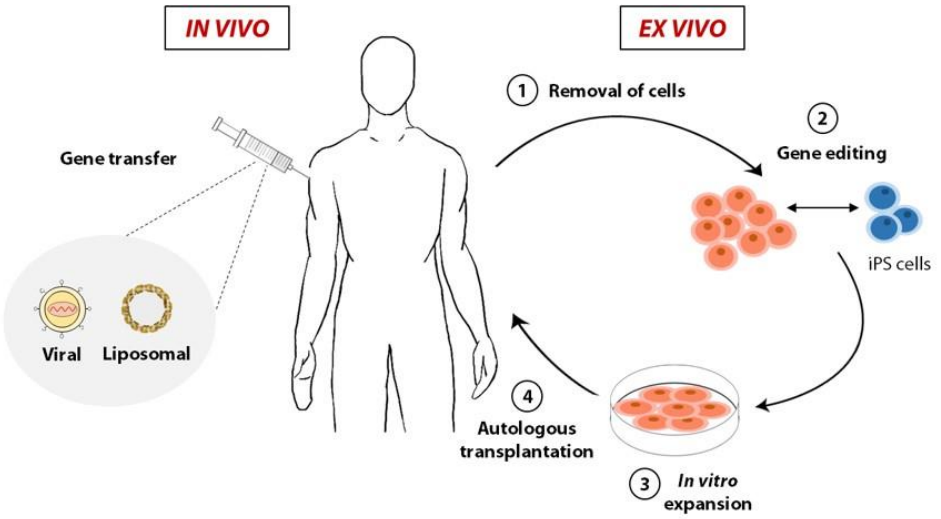


Figure 1-10 *Ex Vivo* and *In Vivo* Gene Editing Approaches

In vivo approaches involve direct transfer (denoted by the syringe) of genome editing reagents such as programmable nucleases and donor templates to the human body. In this instance, two prominent gene transfer agents, viral vectors and liposomes are shown. Ex-vivo is centred on correction of the genetic defect outside of the body. This is a staged-approach whereby: 1) Patient cells are obtained. 2) Gene editing is performed in vitro. This involves delivery of nucleases on their own or concomitantly with repair template. Optionally, the patient cells can be programmed into induced pluripotent stem cells (iPSCs) before or after gene editing. Once corrected, iPSCs may be differentiated into cell types of interest. 3) The genetically corrected cells are characterized and expanded. 4) The corrected cells are then re-grafted back into the patient through autologous transplantation.

To date, four major classes of programmable nucleases have been developed to enable targeted genome editing. These include mega-nucleases and their derivatives, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR-associated Cas9 nuclease. Based on their mode of DNA recognition these nucleases can be classified into two broad categories. First, mega-nucleases, TALENs and ZFNs that recognise and specifically bind DNA via protein-DNA interactions and second, CRISPR-based Cas9 nuclease, which is guided to specific DNA sequence by small RNA molecules forming a RNA-DNA hybrid.

Meganucleases are natural endonucleases with large (>14 bp) recognition sites. Their DNA binding domains are also responsible for cleavage of target sequences (Silva, Poirot et al. 2011). ZFNs and TALENs on the other hand are chimeric enzymes consisting of a DNA binding domain fused to sequence specific *Fok*I nuclease domain (Porteus and Carroll 2005, Urnov, Miller et al. 2005). Targeting using these requires extensive protein engineering and complex molecular cloning. In contrast, Cas9 can be easily targeted to DNA sequences by changing the base pair sequence of the guide RNA that binds directly the target DNA. Cas9 can also be used to create DSBs at multiple

sites, known as multiplexing, within the same cell by expression of distinct guide RNA molecules. All four nucleases have been shown to achieve efficient genome editing in a wide range of organisms and mammalian systems and therapeutic benefit in monogenic disease models (Prakash, Moore et al. 2016).

1.4.2 Zinc Finger Nucleases

Zinc finger nucleases (ZFNs) are powerful tools used for making directed genomic modifications in experimental organisms for functional studies or for creating models of human genetic diseases (Porteus and Carroll 2005, Carroll 2011). A ZFN is a fusion protein and has a modular structure that is composed of two domains: a DNA-binding zinc-finger protein (ZFP) domain and the nuclease domain derived from the FokI restriction enzyme (Porteus and Carroll 2005) **(Fig. 1-11)**.

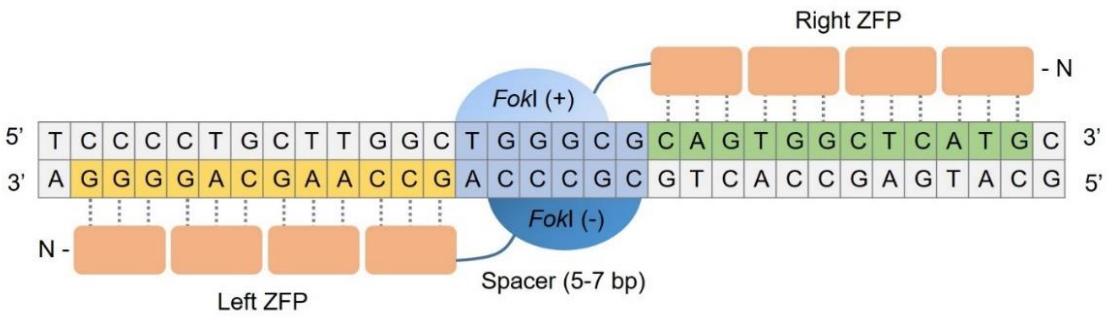


Figure 1-11 Schematic Representation of ZFNs

Each ZFN monomer is made of a zinc finger protein (ZFP) at the amino terminus and the FokI nuclease domain at the carboxyl terminal. Each ZFP binds with 3 base pairs of the sequence. Target sequences of ZFN pairs are typically 18-36 base pairs in length, excluding a 5-7 base pair spacer region.

The structurally separated DNA-binding domain of *Fok*I can be replaced with ZFPs to create ZFNs, and the FokI nuclease domain must dimerize to cleave DNA. Thus, two ZFN monomers are required to form an active nuclease; each monomer must bind to adjacent half-sites that are separated by spacers of 5–7 bp **(Fig. 1-11)**. This requirement for dimerization doubles the length of recognition sites, which substantially increases the specificity of ZFNs. However, the wild-type FokI domain can still form homodimers to cleave DNA when one monomer binds to DNA, which often leads to unwanted off-target effects. The FokI dimeric interface was artificially modified to generate obligate heterodimeric forms, which substantially reduced off-target effects and ZFN cytoxicity (Miller, Holmes et al. 2007, Szczepek, Brondani et al. 2007).

The sequence specificity of ZFNs is determined by ZFPs, which consist of tandem arrays of C2H2 zinc-fingers — the most common DNA-binding motif in higher eukaryotes

(Durai, Mani et al. 2005). Each zinc-finger recognizes a 3-bp DNA sequence (Urnov, Rebar et al. 2010, Carroll 2011), and 3–6 zinc-fingers are used to generate a single ZFN subunit that binds to DNA sequences of 9–18 bp **(Fig. 1-11)**.

The co-crystal structure of a ZFP bound to DNA showed that zinc-finger–DNA interactions are modular in nature; each zinc-finger interacts almost independently with a 3-bp DNA sequence (Miller, Holmes et al. 2007). Indeed, new ZFPs with desired specificities can be constructed by modular assembly of pre-characterized zinc-fingers (Kim, Lee et al. 2009). However, ZFNs created using this fast and convenient method often either lack DNA targeting activity or are cytotoxic owing to off-target effects (Cornu and Cathomen 2010, Ramirez, Certo et al. 2012). Cell-based selection methods and modular assembly methods that account for context dependence between neighbouring zinc-fingers have been developed to yield functional ZFNs (Maeder, Thibodeau-Beganny et al. 2008, Maeder, Thibodeau-Beganny et al. 2009). Nonetheless, it remains challenging to construct ZFNs with high activity and low cytotoxicity using publicly available resources.

1.4.3 CRISPR-Cas9 System

Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated (Cas) are essential components of nucleic-acid-based adaptive immune systems that are widespread in bacteria and Archae. Originally identified in *Escherichia coli* (Ishino, Shinagawa et al. 1987), the CRISPR-mediated immune system relies on small guide RNAs for sequence-specific detection and destruction of invading pathogens, including viruses and plasmids (Barrangou, Fremaux et al. 2007, Wiedenheft, Zhou et al. 2009, Mali, Esvelt et al. 2013).

The CRISPR/Cas immunity occurs in phases **(Fig. 1-12)**. Firstly, in the infection phase, in response to a viral infection, bacteria and archaea acquire and integrate short fragments of invading nucleic acid (known as protospacers) into their chromosome proximal to CRISPR loci. These repetitive loci serve as genetic ‘memory cards’ by maintaining a record of encounters with a pathogen. Once the spacer is acquired, the repeats duplicate maintaining the architecture of repeat-spacer-repeat. In the next phase of immunity which encompasses expression and interference, CRISPR loci with acquired spacers are transcribed into precursor CRISPR RNA (pre-crRNA) which is subsequently processed into a library comprising of short CRISPR-derived RNAs (crRNAs) with each corresponding to a complementary nucleic-acid sequence from invading pathogen. These short crRNAs then bind to large protein surveillance complex - Cas protein

complex (Cascade) - and mediate detection of any foreign pathogen by surveying the intra-cellular space (Brouns, Jore et al. 2008, Wiedenheft, Zhou et al. 2009).

The CRISPR loci are approximately 20-50 base pairs in length separated by unique spacer regions. Within a CRISPR locus the repeats are conserved, but sequences in different CRISPR loci differ both in sequence and length. Even more, the repeat-spacer units in a CRISPR locus varies widely within and across different species. Owing to this variation, the identification of these loci was initially limited but now with use of computational programmes that can carry out widespread comparative analyses, related structures have been revealed – an (A+T)-rich leader sequence designated as a promoter element, and set of four CRISPR-associated (cas 1-4) genes (Jansen, Embden et al. 2002).

Phylogenetic analyses based on Cas1 sequences led to evidence for classification of distinct CRISPR types – Type I, II and III (Kunin, Sorek et al. 2007). Type I and III share some comparable features, like, specialized Cas endonucleases processes the pre-crRNAs, and once mature, each rRNA assembles into a large multi-Cas protein complex capable of recognizing and cleaving nucleic acids complementary to the crRNA. In contrast, type II systems process pre-crRNAs using a ternary complex composed of a trans-activating crRNA (termed tracrRNA), RNA specific ribonuclease RNase III, and a type-II CRISPR-associated endonuclease Cas9. The Cas9 nuclease with crRNA and tracrRNA binds to and cleaves dsDNA protospacer sequences homologous to crRNA spacer and lying adjacent to a short protospacer adjacent motif (PAM) recognised by Cas9 (Jinek, Chylinski et al. 2012). Cas9 is thought to be the sole protein responsible for crRNA-guided silencing of foreign DNA. The specificity of the CRISPR-Cas system relies on tightly bound crRNA, which efficiently guides Cas9 to its target—the complementary DNA fragment (Wiedenheft, Zhou et al. 2009).

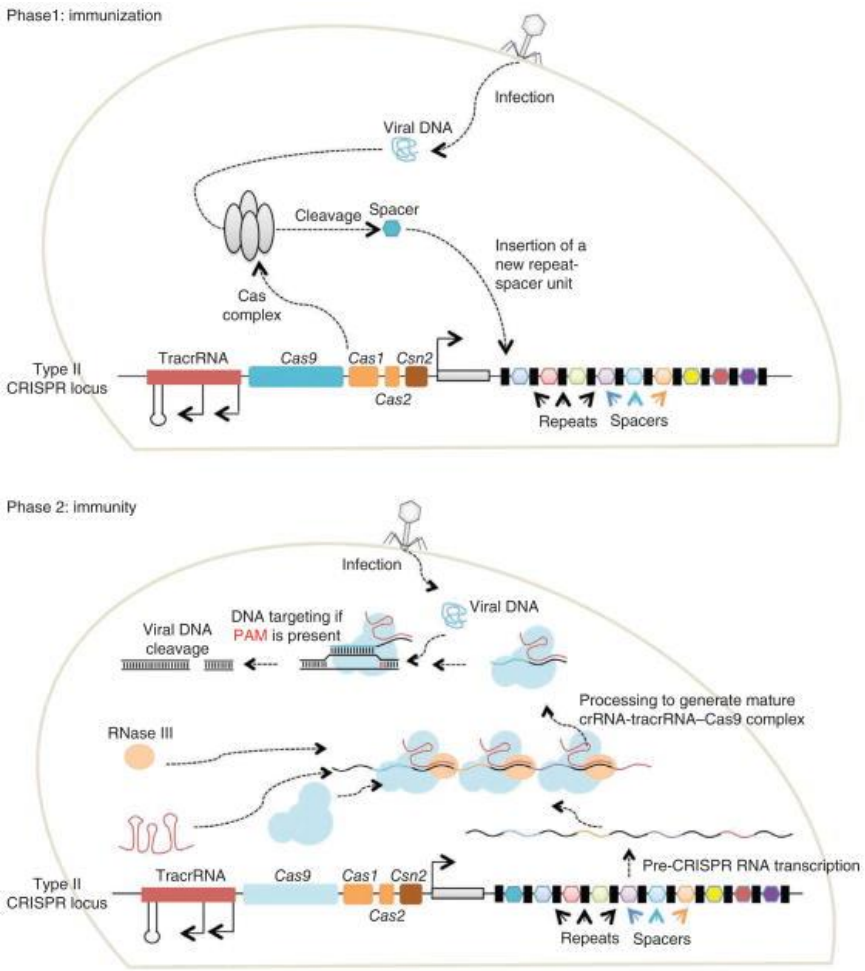


Figure 1-12 CRISPR-Mediated Adaptive Immune Systems

A diverse set of CRISPR-associated (cas) genes encode proteins required for new spacer sequence recognition, CRISPR RNA biogenesis and target interference. Each CRISPR locus consist of a series of direct repeats separated by unique spacer sequences acquired from invading genetic elements known as protospacers. These sequences are flanked by a short motif called protospacer adjacent motif (PAM) that is located on the 5' (type I) or 3' (type III) side of the foreign DNA. Taken from (Mali, Esvelt et al. 2013).

Cas9 proteins constitute a family of enzymes that require base-paired structure formed between the activating tracrRNA and the targeting crRNA to cleave target dsDNA. At sites, complementary to the crRNA-guide sequence, the Cas9 HNH nuclease domain cleaves the complementary strand, whereas the Cas9 RuvC-like domain cleaves the non-complementary strand. The dual tracrRNA:crRNA can be engineered as a single artificial RNA chimera, known as single guide RNA or gRNA, to direct sequence-specific Cas9 dsDNA cleavage (Jinek, Chylinski et al. 2012, Mali, Esvelt et al. 2013, Cong and Zhang 2015).

The type II CRISPR system from *Streptococcus pyogenes* has been adapted for inducing sequence-specific DSBs and targeted genome editing (Jinek, Chylinski et al. 2012,

Doudna and Charpentier 2014, Barrangou and Doudna 2016). In the simplest and most widely used form of this system, two components must be introduced into and/or expressed in cells or an organism to perform genome editing: the Cas9 nuclease and the gRNA, consisting of a fusion of a crRNA and a fixed tracrRNA (Jinek, Chylinski et al. 2012, Fu, Reyon et al. 2014) **(Fig. 1-13).**

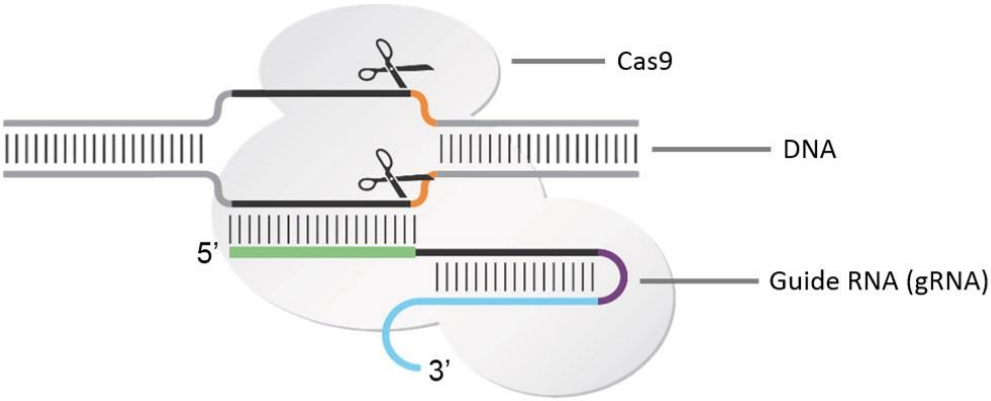


Figure 1-13 CRISPR/Cas9 Gene Editing Components

CRISPR-Cas9 system is a rapid gene editing technology. Cas9 is a CRISPR-associated endonuclease that can be programmed to mediate a double-stranded break in the DNA. The location at which the Cas9 can cut the DNA is specified by short RNA molecule called guide RNA (gRNA).

Twenty nucleotides at the 5′ end of the gRNA (corresponding to the protospacer portion of the crRNA) directs Cas9 to a specific target DNA site using standard RNA-DNA complementarity base-pairing rules. These target sites must lie immediately 5′ of a PAM sequence that matches the canonical form 5′-NGG. N stands for any nucleotide. Thus, with this system, Cas9 nuclease activity can be directed to any DNA sequence of the form N20-NGG simply by altering the first 20 nucleotides of the gRNA to correspond to the target DNA sequence.

Type II CRISPR systems from other species of bacteria that recognize alternative PAM sequences and that utilize different crRNA and tracrRNA sequences have also been used for targeted genome editing (Heler, Samai et al. 2015). Following the initial demonstrations in 2012 that Cas9 could be programmed to cut various DNA sites *in vitro* (Jinek, Chylinski et al. 2012, Mali, Esvelt et al. 2013, Mali, Yang et al. 2013). CRISPR-Cas9 platform has rapidly developed into a remarkable genome editing tool. Of the designer nuclease systems, currently available for precision genome engineering, the CRISPR-Cas system is by far the simplest and user friendly, showcasing high efficiency as well as versatility across multiple platforms – plants (Bortesi and Fischer 2015), yeasts (DiCarlo, Norville et al. 2013), insects (Bassett, Tibbit et al. 2013, Hammond, Galizi et al.

2016), monogenic human diseases (Cox, Platt et al. 2015, Prakash, Moore et al. 2016), and human embryos and zygotes (Liang, Xu et al. 2015, Tang, Zeng et al. 2017).

1.4.4 Limitations and Perspective

Classical gene therapy approaches have centred on the delivery of DNA to augment endogenous gene expression. Predominantly, these approaches rely on the transfer of functional genes using viral vectors. Retroviral vectors provided the first clear demonstrations of therapeutic benefit in primary immunodeficiencies, and they also highlighted the risk of adverse events attributable to insertional mutagenesis due to genomic integration of proviruses. Some success stories of gene therapy include Glybera, the first clinically approved gene therapy in the European Union, which uses an AAV vector drug for lipoprotein lipase deficiency – although this has now been retracted from the market; Strimvelis, a retroviral vector based drug for ADA-SCID; Spinraza,, an antisense oligonucleotide based drug targeting RNA transcript for treatment of all types of Spinal Muscular Atrophy; and in the case of cystic fibrosis, repeated nebulization of liposomes encoding the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. These success stories and ongoing research on gene augmentation has paved the way for gene editing technologies that hold tremendous promise. The past few years have seen notable demonstrations of genome editing being applied across a multitude of disease models. Whilst the application of engineered nucleases holds significant therapeutic promise, optimum progress can only be achieved by examining the advancement of gene editing holistically. Several ubiquitous challenges need to be considered, mostly relating to efficacy of genome editing at the target sequence, safety concerns related to nuclease-associated off-target effects and delivery of gene editing tools.

Editing efficiency is dependent upon the DNA repair process being relied upon. In instances where the desired effect can be achieved by NHEJ, the correction will most likely occur at a relatively high frequency as NHEJ is the major repair pathway in mammalian cells, although the usefulness of this approach may be limited by the stochastic nature of the Indels being formed. As discussed earlier, NHEJ has been used to mediate disruption of coding and regulatory sequences, targeted deletions of exons or large intervening sequences and disruption of splice sites. Methods that can predict and evaluate micro-homology sites can be used to bias the repair towards frameshift mutations in protein coding sequence (Bae, Kweon et al. 2014). This would partially address the potential reduction in efficiency caused by micro-homology-mediated end

joining, a secondary end-joining pathway with a bias for in-frame deletions (Morton, Davis et al. 2006).

Precise HDR-based locus alterations allow targeted insertion or *in situ* correction of mutated DNA sequence that are suitable for a large subset of disease-causing mutations. However, they are reliant upon homologous recombination, which in turn is normally limited to S and G2 phases of the cell cycle and therefore requires a DNA template and inherently occurs at lower frequencies. HDR-based strategies may also require enrichment and expansion of corrected cells, normally restricted to *ex vivo* approaches. While *ex vivo* manipulation may be possible in diseases like those treated by correcting bone marrow HSCs, it would limit applications in diseases with multi-organ involvement or those where transplantation is not an option. Further progress in enabling HDR with higher efficacy would therefore be beneficial. In this respect, a recent report has demonstrated that it may be possible to transiently activate HDR in G1 cells by restoring BRCA1-PALB2 interaction (Orthwein, Noordermeer et al. 2015), possibly facilitating HDR genome-editing in quiescent cells.

The specificity of genome editing is one of the major safety concerns for translational research. Owing to sequence similarities within the genome, endonucleases can cleave and modify off-target regions that are distinct from the site of interest. Off-target effects can lead to unwanted genetic modifications causing cellular stress, functional impairment or enhancement, and oncogenicity, all of which could have detrimental effects clinically (Maggio and Goncalves 2015). Considerable work is being undertaken to increase fidelity through better design of nuclease components, which has led to improved variants such as megaTALs (Boissel, Jarjour et al. 2014), dead Cas9-*Fok*1 fusion nucleases (Tsai, Wyvekens et al. 2014), Cas9 nickases (Ran, Hsu et al. 2013), and Cas9 nucleases with truncated guide RNAs (Fu, Reyon et al. 2014). Furthermore, screens of bacterial strains have led to discovery of several alternative Cas9-nucleases with varying specificities (Sapranauskas, Gasiunas et al. 2011, Esvelt, Mali et al. 2013, Hou, Zhang et al. 2013, Ran, Cong et al. 2015). More recently, Cpf1, a prominent CRISPR variant that requires a shorter RNA and generates a staggered cut which could improve HDR (Zetsche, Gootenberg et al. 2015, Kleinstiver, Tsai et al. 2016), a high-fidelity engineered Cas9 (Kleinstiver, Pattanayak et al. 2016) that showed high on target specificity, and the smallest Cas9 orthologue derived from *Campylobacter jejuni*, called CjCas9 (Kim, Koo et al. 2017), have also been described. These variants highlight the progress in the field but still require extensive examination prior to their application in a translational research setting. Specificity of modification can also be helped by careful target site selection and use of delivery methods that would allow for efficient but transient expression of

nucleases. New methods, such as GUIDE-seq (Tsai, Zheng et al. 2015) and BLESS (Crosetto, Mitra et al. 2013), have also been developed for unbiased evaluation of off-target modifications on a genome-wide scale.

The final challenge pertains to the delivery of gene editing reagents including nucleases and a donor template in case of HDR. A variety of delivery approaches are being explored depending on cell types to be targeted. Cells that can be cultured and engrafted under *ex vivo* conditions are amenable to delivery via nucleic acids, proteins and viral vector systems; mRNA and protein delivery of the nucleases are now well-established procedures. However, for *in vivo* gene editing applications the most promising delivery systems are viral vectors. Both integrating and non-integrating viral vector systems have been explored in this context, although the latter are favoured due to their safety profile. AAV vectors, with a wide range of serotypes and ability to transduce a variety of tissue types are promising candidates.

Despite the outlined challenges, genome editing is advancing at fast pace, with continued focus on pioneering and improving strategies. The safe use of ZFNs targeting CCR5, the co-receptor necessary for HIV to infect T-cells, to control AIDS, remains a milestone of gene-editing in a therapeutic setting (Tebas, Stein et al. 2014). In this case, the infusion of autologous CD4 T cells in which the CCR5 gene was rendered permanently dysfunctional by ZFNs was deemed safe. 12 patients who had been taking antiretroviral drugs were treated with gene edited T cells. After the treatment, six of the 12 participants stopped their antiretroviral drug therapy and their HIV levels rebounded more slowly than normal, and their T-cell levels remained high for weeks. This first small trial implicated that gene editing approach seemed to be safe and is an important advance in the direction for this kind of research.

1.5 *PRKDC* Severe Combined Immunodeficiency

1.5.1 Overview

Severe combined immunodeficiencies (or SCID) are a group of rare genetic disorders that are characterised by lack of T or B immune cells. There are several forms of SCID with variable underlying genetic causes. Broadly these can be classified into two categories: i) T-cell signalling defects giving rise to the commonly occurring (up to 70% of all cases) T^-B^+ form of SCID, and ii) V(D)J recombination defects that causes T^-B^- phenotype. T^-B^+ SCID patients show complete lack of T cells while B lymphocytes are present and functional. T^-B^- patients on the other hand show complete absence of circulating B as well as T lymphocytes.

Patients with T^-B^- SCID can have variable genetic defects affecting genes involved in different steps of V(D)J recombination. In approximately 70% of T^-B^- patients, mutations are found in the RAG1 and RAG2 gene; while the remaining cases exhibit mutations in genes encoding NHEJ proteins – Artemis, LIG4, XLF/Cerunnos and DNA-PKcs. This subset of T^-B^- SCID is known as radiosensitive SCID (or RS-SCID). People with RS-SCID thus suffer from immune deficiency and are sensitive to ionizing radiations. Radiation sensitivity is seen in both bone marrow cells and primary skin fibroblasts. The prognosis for RS-SCID patients is poor, with lethality resulting typically within the first year of life. Furthermore, patients in this group present with a broad spectrum of clinical conditions ranging from radiosensitive leukaemia due to mutations in LIG4, a classical form of SCID, to growth retardation, microcephaly and immunodeficiency caused by profound T and B cell lymphocytopenia arising from mutations in XLF/Cernunnos or PRKDC (van der Burg, Ijspeert et al. 2009, van der Burg, van Dongen et al. 2009, Woodbine, Neal et al. 2013). Mutations in *Artermis* are often null and confer RS-SCID, but no overt growth or microcephaly.

DNA-PK severe combined immunodeficiency is a type of radiosensitive SCID that is caused by a defect in DNA PKcs which causes dual effects – difficultly in DNA damage repair, for example in response to ionising radiations and lack of immune cells owing to faulty V(D)J recombination (Kurimasa, Kumano et al. 1999).

Biochemical and genetic data have revealed DNA–PK to be composed of three components: a catalytic subunit and two regulatory subunits, DNA-binding proteins, Ku80 and Ku70. DNA PKcs, product of the *PRKDC* gene, is a large polypeptide of over 4000 amino acids. The N-terminal ~250 kDa of DNA-PKcs contains a putative DNA-binding domain, a leucine-rich region and a series of HEAT [huntingtin, elongation factor

3, A subunit of protein phosphatase 2A and TOR1 (target of rapamycin 1)] repeats **(Fig. 1-14)**. The C-terminal region contains a FAT [FRAP (FKBP12-rapamycin-associated protein), ATM (ataxia telangiectasia mutated), TRRAP (transactivation/transformation domain-associated protein)] domain that is characterized by weak amino acid similarity to other members of the PIKK family, followed by a kinase domain and a C-terminal FATC domain (Sibanda, Chirgadze et al. 2010, Davis, Chen et al. 2014, Sibanda, Chirgadze et al. 2017) **(Fig. 1-14)**.

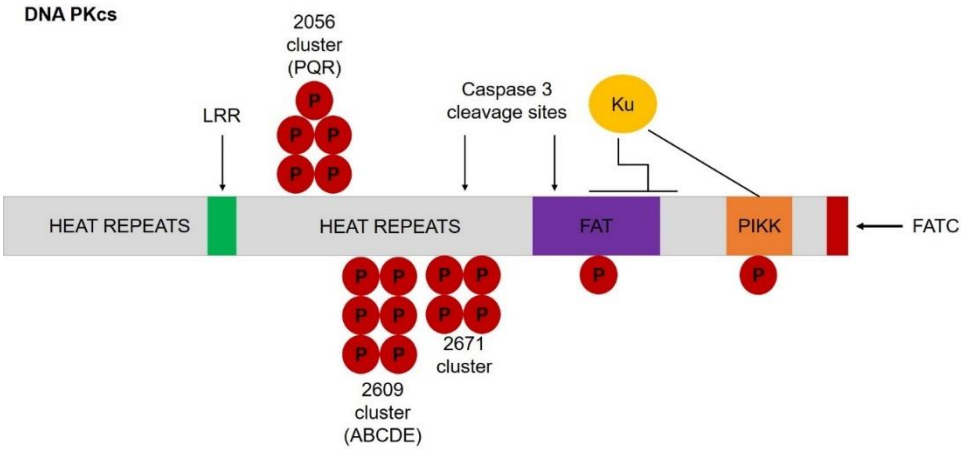


Figure 1-14 Major Features of DNA-PKcs

The N-terminal domain containing HEAT repeats and a leucine-rich region (LRR) extends from amino acids 1–2908, the FAT domain is from amino acids 2908–3539, the PIKK domain from amino acids 3645–4029 and the FATC domain from amino acids 4906–4128. ABCDE cluster represents the in vivo phosphorylation sites between Thr2609 and Thr2647 while the PQR cluster represents the in vivo phosphorylation sites between Ser2023 and Ser2056. The 2671 cluster contains four sites between Thr2671 and Thr2677. Ku interacts at amino acids 3002–3850. HEAT repeats: huntingtin, elongation factor 3, A subunit of protein phosphatase 2A and TOR1 (target of rapamycin 1)] repeats; FAT: FRAP (FKBP12-rapamycin-associated protein), ATM (ataxia telangiectasia mutated), TRRAP (transactivation/transformation domain-associated protein) domain; PIKK: phosphoinositidine 3-kianse-like family of protein kinases; FATC: FRAP, ATM, TRRAP C-terminal domain.

1.5.2 V(D)J Recombination

V(D)J recombination is the process by which T cells and B cells randomly assemble different gene segments – known as variable (V), diversity (D) and joining (J) genes – to generate unique receptors (known as antigen receptors) that can collectively recognize many different types of molecule (Jung and Alt 2004, Schatz and Ji 2011, Schatz and Swanson 2011).

V(D)J recombination mediates immunoglobulin and T cell receptor gene assembly from variable (V), diversity (D), and joining (J) gene segments; a crucial step in differentiation of broad range of antigen-specific B and T cells. It is a tightly regulated mechanism involving several factors and takes place both in bone marrow and thymus during early T and B cell differentiation. Recombination is initiated by the introduction of DSBs at recombination signal sequences (RSS) by lymphoid specific RAG 1 or RAG 2 endonucleases. The DNA ends, including the sequences encoding the antigen receptors (termed as coding ends), have hairpin termini; while the RSSs (or signal ends) are blunt ended. Re-joining of these recombination intermediates occurs via non-homologous end joining pathway of DNA repair (Jung and Alt 2004, Schatz and Ji 2011, Schatz and Swanson 2011).

1.5.3 Murine DNA-PK Deficiency

Mice lacking DNA-PKcs are experimentally viable, with no overt phenotype other than SCID and are classical models for studying mechanisms of DNA damage and repair (Bosma, Custer et al. 1983, Bosma and Carroll 1991). Located on chromosome 16, *Prkdc* is a large gene comprising of 86 exons extending across about 193 kb (Araki, Fujimori et al. 1997). In mice, SCID is caused by a single T to A transversion in *Prkdc* exon 85 which switches the Tyrosine reside at codon-4046 into a stop codon (Blunt, Finnie et al. 1995, Blunt, Gell et al. 1996, Beamish, Jessberger et al. 2000). This results in truncation of 83 amino acids at the C-terminal of the protein, thereby losing the highly conserved catalytic domain (Beamish, Jessberger et al. 2000).

1.5.4 DNA-PK Deficiency in Humans

DNA-PK SCID (OMIM 600899) is an ultra-rare condition in humans with only two patients showing mutations in *PRKDC* identified so far. Recent work strongly suggests that DNA PK complex is essential and that a *PRKDC* null mutation would be incompatible with life in humans (Ruis, Fattah et al. 2008).

The first human *PRKDC* SCID patient, a girl (IMD26; 615966) of Turkish descent born of consanguineous parents, manifested infantile SCID with absent B and T cells. (van der Burg, van Dongen et al. 2009) identified a homozygous c.9185T>G transversion on *PRKDC*, resulting in the substitution of Leuine at pos.3062 to Arginine (L3062R) substitution at a highly conserved residue in the FAT domain. The unaffected parents were heterozygous for the mutation. The patient also carried a homozygous deletion of Gly2113, but this residue is not well conserved and was demonstrated to be non-pathogenic. Studies of patient cells showed normal DNA-PK kinase and autophosphorylation capacity. Patient bone marrow cells showed increased long palindromic (P)-nucleotide stretches in the immunoglobulin coding joints, indicating a defect in hairpin opening and insufficient Artemis activation. *PRKDC*-deficient cells showed abnormal junctional pattern during V(D)J recombination, as well as impaired non-homologous end-joining that could not be restored to normal by mutant L3062R. L3062R mutation, which retains kinase and autophosphorylation activity but fails to activate Artermis, differs substantially from the spontaneous *Prkdc* mutations described in SCID horses, mice, and dogs, all of which result in truncated proteins.

(Woodbine, Neal et al. 2013) identified the second *PRKDC* SCID patient, a boy (IMD26; 615966) with profound neurologic abnormalities. He carried two compound heterozygous mutations in the *PRKDC* gene: a c.10721C>T transversion, resulting in a substitution of Alanine at pos.3574 to Valine (A3574V) inherited from the unaffected mother on 1 allele, and a cDNA that lacked exon 16 on the other allele. Genomic sequencing of the patient's DNA showed a 1-bp insertion (IVD16+1510insA) 700-bp upstream of the intron 16 splice site on the other allele, but it was unclear if this change caused the in-frame skipping of exon 16. Immortalized patient cells showed decreased but detectable protein, but no detectable kinase activity. Patient cells showed a defect in DNA DSB repair following irradiation, which could be rescued by expression of wild type *PRKDC*. The A3574V substitution occurred at a highly-conserved residue within the FAT domain, outside the kinase domain. Cells transfected with the mutation showed impaired DNA-PKcs function in response to irradiation and a less severe defect in V(D)J end-joining, suggesting that the missense mutation retained some functional capacity. Functional studies of cells lacking exon 16 suggested that it represented a null allele. The overall findings were consistent with a loss of function. In addition to SCID, the patient had microcephaly, brain malformations, hearing loss, visual impairment, and little developmental progress. The patient died at an age of 31 months with intractable seizures (Woodbine, Neal et al. 2013).

1.5.5 Treatment Approaches

Currently, human leukocyte antigen (HLA)-identical bone marrow (BM) transplantation is the only choice for effective treatment of SCID. Success of this procedure is highly dependent on the availability of a suitable matched donor. Gene transfer to haematopoietic stem cells (HSCs) first shown to have major therapeutic effects in SCID over 10 years ago however continues to show promise. Initially, gamma retroviral gene therapy demonstrated long-term clinical efficacy for SCID-X1 and ADA-SCID by complementation of a correct copy of the defective gene. However, SCID-X1 trials also highlighted oncogenic risks requiring improved vector design for clinical safety (Hacein-Bey-Abina, Von Kalle et al. 2003). Despite the serious adverse events associated with retroviral integration into chromosomal DNA, gene therapy remains an important approach for the treatment of SCIDs. The long-term data in over 90 patients with several inherited primary immune deficiencies who have received gene therapy using conventional gamma retroviral vectors over the last decade show >90% overall survival, with the clear majority experiencing clinical benefits, despite insertional mutagenesis in around ten patients (Gennery, Slatter et al. 2010). The observed toxicities in these trials shared a common mechanism, namely upregulated expression of oncogenes induced by powerful enhancer sequences of viral vectors used. Moreover, several regulatory agencies have recommended a move away from the continued use of retroviral vectors and the development of safer vector designs. This has led to the development of later generation vectors with improved efficiency, specificity and safety. Over the last years, vector design has been largely modified, and gamma retroviral vectors have been replaced by self-inactivating systems for added safety (Seymour and Thrasher 2012, Cicalese, Ferrua et al. 2016). Self-inactivating vectors, based on the HIV-1 lentiviral vector, in which the HIV LTR is deleted and transgene expression placed under the control of an internal promoter with minimal or no enhancer activity have received considerable attention. Separately, *ex vivo* genome editing and HSC transplantation is being investigated as a therapeutic approach (Naldini 2015). Currently, human gene therapy clinical trials are under way using lentivectors in a wide range of human diseases. The proof-of-principle of the therapeutic efficacy of gene therapy targeting the immune system has been established by the “successful” clinical trials for the treatment of X-SCID (Cavazzana-Calvo et al., 2000). Gene and cell therapy research recently reached a fundamental milestone toward the goal to deliver new medicines for orphan diseases. In 2016, the European Commission granted market approval to GlaxoSmithKline for *ex vivo* hematopoietic stem cell gene therapy for the treatment of ADA-SCID. The medicine, Strimvelis, is an autologous CD34+ enriched cell fraction that

contains CD34+ cells transduced with retroviral vector that encodes for the human adenosine deaminase cDNA sequence from human haematopoietic stem/progenitor (CD34+) cells (Aiuti et al., 2017). It is the first *ex vivo* stem cells gene therapy to receive regulatory approval anywhere in the world and demonstrates the ever-growing profile of gene therapy research.

1.6 Ataxia Telangiectasia

1.6.1 Overview

Ataxia Telangiectasia (A-T) is a complex multi-system neurodegenerative disorder caused by defects in a single gene (OMIM 607585). Inherited as an autosomal recessive trait A-T has an estimated prevalence of 1 in 40,000-400,000 people (Lavin 2008). This frequency varies greatly from country to country. A-T is categorised as a rare disease with no cure at present (http://rarediseases.org/rare-diseases/ataxia-telangiectasia/).

A-T is characterised by neurological and cerebellar manifestations, with ataxia (difficulty controlling movement) being evident from first years of life. This gradually progresses to affect speech, eye and involuntary movements. Most affected children generally require a wheelchair from their early teenage years. Telangiectatic or 'spider-like' vessels in the eye and skin is the second most common manifestation of the disorder. A-T also presents with variable immunodeficiency leading to recurrent infections usually of respiratory system, accompanied with immunoglobulin deficiencies and poor response to vaccines. Affected individuals also present with hypersensitivity to ionising radiations, chromosomal instabilities and highly increased risks to lymphoid malignancies or breast cancer in case of women (Chun and Gatti 2004, Lavin 2008).

The causative gene, known as ataxia telangiectasia mutated (*ATM*) localises to chromosome 11q22-23 and is constitutively expressed in all eukaryotic cells. It encodes for A-T mutated (ATM) protein which is a serine/threonine kinase and is a member of the phosphoinositide 3-kinase (PIK3)-related protein kinase (PIKK) family, which also includes ATM and Rad3-related protein (ATR) and DNA PKcs. ATM protein is an apical kinase with central roles in cellular response to DNA double strand breaks, apoptosis and cell-cycle checkpoint control (Taylor, Harnden et al. 1975, Jeggo, Carr et al. 1998, Lavin 2008). ATM interacts with multiple substrates – an estimated 700 substrates have been identified through proteomic analysis (Matsuoka, Ballif et al. 2007)

ATM has multiple domains - the kinase domain is located close to the C terminus like in other members of PIKK family, a FAT domain, an extreme C-terminal FATC domain, N-terminal substrate binding domain, a leucine zipper and a proline-rich region **(Fig. 1.15)**. Protein post-translational modifications includes activation of ATM protein by various kinases and phosphatases (PP2A, PP5 and WIP1); auto-phosphorylation at residues Ser367, Ser1893, Ser1981; and acetylation by TIP60 on Lys3016, crucial for ATM activation **(Fig. 1-15)** (Lavin 2008).

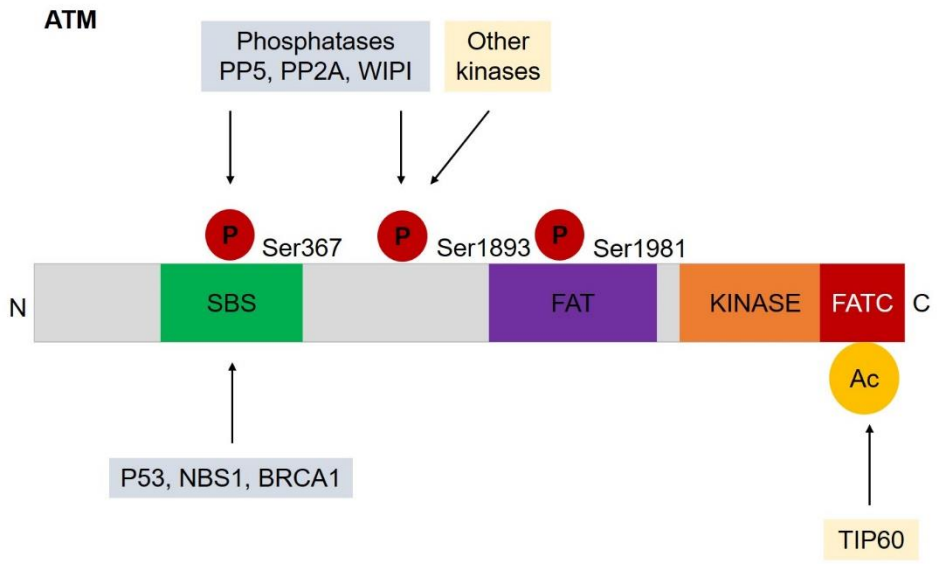


Figure 1-15 Schematic Representation of ATM Protein

ATM is a member of phosphoinositide 3-kinase (PI3K)-related protein kinase (PIKK) family and like other members contains a FAT FRAP [(FKBP12-rapamycin-associated protein)] domain, a protein kinase domain and a FATC [FRAP (FKBP12-rapamycin-associated protein), ATM (ataxia telangiectasia mutated), TRRAP (transactivation/transformation domain-associated protein)] domain. Three autophosphorylation sites Ser367, Ser1983 and Ser1981, have been identified on ATM. Three phosphatases, protein phosphatase-2A (PP2A), protein phosphatase-5 (PP5) and WIPI, have been implicated in the control of ATM activation. In addition to the binding in the kinase domain, several ATM substrates binds to a region near the N terminus of the protein at the substrate binding site (SBS). ATM FATC domain functions as a binding site for the acetyltransferase TIP60. Ac: acetylation; Nbs1: Nibrin.

1.6.2 A-T Related Immunodeficiency

ATM kinase is involved in various cellular responses to DNA damage – especially to DSBs. DSBs can arise due to exogenous factors or as part of cellular processes. Lymphocytes for instance undergo DSBs in the process of V(D)J recombination and Ig class switch recombination for their development and maturation. In the absence of *ATM*, DSBs are difficult to repair. This causes a developmental block in the lymphoid compartment leading to immune defects associated with A-T (Jeggo, Carr et al. 1998, Chun and Gatti 2004).

A-T related immunodeficiency is one of the key risks for A-T patients. It is characterised by reduced numbers of lymphocytes (especially T cells) and immunoglobulin deficiency. Thus, patients succumb to recurrent infections and show poor responses to vaccines. The immunodeficiency is particularly common in children, with pulmonary failure associated with or without infections, being the major cause of failing health and mortality (Chopra, Davies et al. 2014). The immunodeficiency of A-T is rarely progressive, but

remains stable in most cases without deterioration in symptoms with time. This can have a significant impact on the patient's quality of life and requires long-term management (Nowak-Wegrzyn, Crawford et al. 2004, Staples, McDermott et al. 2008, Chopra, Davies et al. 2014).

In its classical form, A-T is caused by the presence of two truncating mutations, which results in either a complete lack of ATM protein or a mutant protein with no kinase activity. In contrast, subsets of patients harbour mutations that allow the reduced production of the normal protein, or a mutant protein with limited kinase activity. Patients with this type of A-T, also known as variant A-T, show a milder phenotype with later disease onset, often in adulthood, and slower rate of neurodegeneration. A UK study of 80 A-T patients showed that patients suffering from variant A-T exhibited less recurrent sinopulmonary infections and had a lesser need for prophylactic treatments (Staples, McDermott et al. 2008). Another study of 296 A-T patients (from the UK and Netherlands) reported that in childhood, total absence of ATM kinase activity was associated, exclusively, with development of lymphoid tumours. Their findings also suggested that expression of some residual ATM kinase activity has a strong protective effect against tumour development in A-T in childhood. The residual ATM kinase activity was shown due to the presence of leaky splice mutation (IVS40-1050A>G) expressing a low level (~5%) of normal ATM (Reiman, Srinivasan et al. 2011). These suggests that therapeutic benefits, including normal development of immunological function, might be achieved in A-T patients even if only modest increases (~5%) in functional ATM protein levels could be obtained (Taylor and Byrd 2005). Treatment of A-T related immune defects could therefore have a significant impact on improving the quality of life, including reduced rates of infection and perhaps lower cancer risks.

1.6.3 Treatment Approaches

The classical form of A-T typically presents in early childhood. Systemic complications and progressive neurological decline lead to a reduced life expectancy and affected individuals often die of respiratory complications or a malignancy (Lavin 2008, Taylor, Lam et al. 2015). At present, there is no therapy to cure A-T or slow disease progression. Typically, patients are managed by a multidisciplinary team which consists of respiratory physicians, immunologists, geneticists and neurologists (http://www.atsociety.org.uk/clinical-guidance). Disease management is thus merely symptomatic and supportive (Lavin, Gueven et al. 2007). Current A-T related immunodeficiency is managed by: (i) early antibiotic treatment and continuous prophylactic therapy; (ii) use of vaccines; and (iii) regular immunoglobulin therapy administered intravenously every 3-4 weeks (Chopra, Davies et al. 2014). Allogeneic bone marrow transplantation with a low conditioning regime has also been reported in the treatment and long term survival of one patient with A-T (Ussowicz, Musial et al. 2013).

In this context, approaches that could allow permanent modification of the ATM gene and thus restore ATM protein kinase production could be beneficial. The wide variety of ATM mutations, which occur throughout the whole gene without hot spots, makes the development of therapies challenging. However, several mutation-targeted therapies are already under investigation. Since most A-T causing mutations lead to protein truncations due to presence of stop codons, small molecular read through (SMRT) compounds have been primarily tested. SMRT compounds can read through premature termination codons allowing translation of full-length protein, and have been used in A-T patient cell lines to restore ATM expression (Du, Jung et al. 2013). Such SMRT compounds have been previously used in the context of Duchenne muscular dystrophy and cystic fibrosis and are being evaluated in pre-clinical stages. However, most SMRT molecules do not cross the blood-brain barrier (BBB), a critical factor for treating a neurological disorder like A-T. Promising work in this field aims at the development of SMRT compounds with optimised efficiency, low toxicity, and ideally the ability to cross the BBB (Lavin 2008, Du, Jung et al. 2013, Lee, Martin et al. 2013).

Another significant set of mutations identified amongst A-T patients are those affecting splicing. These constitute approximately 30% of mutations, leading to a variety of consequences such as exon skipping, intron retention, and activation or creation of new splice sites. In many cases, splicing defects lead to loss of ATM protein (Teraoka, Telatar et al. 1999). In this context, antisense morpholino oligonucleotides that conceal aberrant

splice sites have been used to demonstrate restoration of ATM mRNA and protein in A-T patient cell lines (Du, Pollard et al. 2007). Again, while this method provides proof-of-principle, its effectiveness and safety remains to be validated in animal models harbouring appropriate mutations. However, peptide-conjugated antisense oligonucleotides capable of crossing the BBB have been reported, and novel chemistries are under active investigation, so there is mileage for this type of therapy in A-T.

Delivery of full-length *ATM* cDNA via viral vectors has also been explored as an alternative. Due to the large size of the cDNA (~9.2 kb, plus regulatory sequences) (Zhang, Chen et al. 1997), herpes simplex virus (HSV)-based vectors that have a large packaging capacity have been tested. One study reported successful introduction of ATM cDNA using a HSV amplicon vector in cultures of A-T fibroblasts and *Atm*-/- mice brain with expression in cerebellum (Qi, Shackelford et al. 2004). Separately, a HSV/AAV (adeno-associated virus) hybrid was shown to mediate insertion of full-length *ATM* cDNA into the AAV insertion site (AAVS1) in normal and A-T human cells and in transgenic *Atm*-/- mouse cells *in vivo*, producing functional expression of ATM. No adverse immune responses were noted in treated mice although the frequency of insertion was low (Cortes, Oehmig et al. 2008). Research on an *ATM* mini-gene has been previously funded but no published outcomes are available so the feasibility of the approach is unclear (http://www.atsociety.org.uk/data/files/Research/Molina_WD.pdf; http://actionforat.org/gene-therapy-for-the-treatment-of-ataxia-telangiectasia/).

1.6.4 Cell and Gene Therapy for A-T

The studies summarised in the previous section suggest that gene therapy could be utilised for the permanent modification of *ATM* to restore its function. With regards to immunodeficiency related to A-T, transplantation of haematopoietic stem and progenitor cells with functional ATM can be advantageous; especially since low levels of ATM activity have been associated with milder phenotypes. HSCs have great therapeutic potential because of their ability to self-renew and differentiate. Given their unique properties, a small number of genetically modified HSPCs could accomplish lifelong, corrective restoration of entire haematopoietic system. *Ex vivo* gene therapies - based on isolation of HSCs from a patient's bone marrow, expansion and genetic modification in culture, followed by transplantation – are one of the best developed and most successful forms of cells therapies. In fact, allogenic bone marrow transplantation from a matched donor is the only treatment for several hematologic, immunological and metabolic conditions. However, due to the risks involved with allogenic transplant and

limited availability of matched donors, autologous transplants using patient's own cells have gained momentum.

Despite BMT being a routine treatment for many immunodeficiencies, it is currently not considered safe for treating A-T (http://primaryimmune.org/treatment-information/stem-cell-and-gene-therapy/). This is mainly because A-T patients present with a very high sensitivity to radiotherapy which can be fatal in case a full cytotoxic conditioning regime is used to ablate the bone marrow prior to transplantation. There has been one reported case of death because of full conditioning BMT in an A-T patient who was undiagnosed at the time of treatment (Ghosh, Schuster et al. 2012). On the contrary, long-term survival was demonstrated in another A-T patient who received a lower dose conditioning regime (Ussowicz, Musial et al. 2013). Likewise, *Atm* mice studies where BMT was carried out using non-myeloablative host-conditioning regimes yielded promising results. BMT using *Atm*-positive cells significantly increased the longevity of *Atm*-/- mice and overcame the high tumour incidence. Moreover, these studies reported reduction in T-cell developmental/maturation block, and a reduction in risks of developing lymphoid tumours (Bagley, Cortes et al. 2004, Pietzner, Baer et al. 2013). Furthermore, BMT also led to the regeneration of lung tissue (Bagley, Cortes et al. 2004).

A promising avenue for improving the safety of conditioning is the use of biological agents based drugs, such as antibodies, that are specifically targeted to HSCs and other haematopoietic cells in the bone marrow niche and that spare non-haematopoietic cells. Recently a study described the combined use antibodies targeting HSC surface markers, anti c-Kit and anti-CD47, along with T-cell depleting antibodies to eliminate >99% of host HSCs and enable robust autologous as well allogenic HSCT in adult immunocompetent mice, with no geno-toxic side-effects (Chhabra, Ring et al. 2016). The anti-CD47 biological agents used in this targeted conditioning regimen is currently being tested for safety and efficacy in two separate Phase I human clinical trials (ClinicalTrials.org NCT02216409 and NCT02367196). Although the combinatorial use of these antibodies has not been tested in humans, there is mileage in the approach. Separately, another study pre-clinical described the use of an immunotoxin targeting the HSC restricted CD45 receptor to effectively condition immunocompetent mice (Palchaudhuri, Saez et al. 2016).

Given the impact of HSC gene therapy in other immunodeficiencies and due to promising results from BMT in *Atm* mice studies, targeted gene therapy approach can be beneficial for AT. Similar success has been reported for Fanconi anemia (FA) which like A-T arises due to chromosomal instability and defects in DNA damage signalling and RS-SCID

(Rahman, Kuehle et al. 2015). Targeted gene correction of FA mutations in patient cells led to rescue of FA phenotype as well as increased efficacy of reprogramming and generation of HSCs (Peffault de Latour, Porcher et al. 2013) In case of RS-SCID, correction of underlying mutation led to efficient reprogramming of HSCs followed by the restoration of T-cell development and maturation in *in vitro* assays (Rahman, Kuehle et al. 2015). These reports suggest that there is mileage in the study and genome editing for DNA repair disorders.

1.7 Aims and Objectives

Programmable nucleases allow defined alterations in the genome with ease-of-use, efficiency and specificity. Their availability has led to accurate and widespread genome engineering, with multiple applications in basic research, biotechnology and therapy. With regards to human gene therapy, nuclease-based gene editing has facilitated the development of a broad range of therapeutic strategies based on both non-homologous end-joining and homology-dependent repair.

Diseases of monogenic origin can particularly benefit from such gene correction therapies. There are 5,000-8,000 monogenic diseases, defined as inherited conditions arising from mutations in a single gene (Rodwell C July 2014). These often manifest during childhood and lead to morbidity and sometimes premature death. While each monogenic disease is rare, it has been estimated that together they will affect about 6% of people at some point in their lives (Rodwell C July 2014). Diagnosis and treatment for these diseases remain largely insufficient and the care is primarily palliative, focusing on disease management without addressing the underlying genetic defects. The realisation of the social and economic importance of rare diseases and the acute need for diagnostics and treatments has led to initiatives like the International Rare Disease Research Consortium (IRDiRC; http://www.irdirc.org), the Undiagnosed Diseases Network (UDN; http://www.genome.gov/27562471) and Syndromes Without a Name UK (SWAN UK; http://www.geneticalliance.org.uk/projects/swan.htm). This study focuses on gene editing in two rare disease models namely, *Prkdc* severe combined immunodeficiency (*Prkdc* SCID) and Ataxia Telangiectasia (A-T). Both conditions arise due to defects in genes encoding core proteins involved in DNA damage response repair pathways, thereby presenting with immunodeficiency and sensitivity to ionising radiation in patients.

The first part of this project builds up from previous work in the laboratory using a mouse model of the ultra-rare disease *PRKDC* SCID, selected for the expected selected advantage of corrected haematopoietic progenitors *in vivo*. Engineered zinc finger nucleases (ZFNs) and *Prkdc* repair matrices were generated to demonstrate homology-directed repair in polyclonal SCID mice fibroblasts and haematopoietic stem cells which led to rescue of T-cell compartment upon transplantation in immune-deficient SCID mice (Abdul-Razak 2013). However, demonstration of gene editing events at a clonal level remained unexplored, which is an important hallmark of the gene editing workflow. The main aims of this section are as follows:

1. Molecular analysis of gene editing events in ZFN treated Prkdc SCID clones.

2. Use of *in vitro* assays demonstrating DNA damage repair to correlate the genomic correction of SCID mutation with rescue of the encoded DNA PKcs protein.

With the advancement in the field of genome editing, the second part of this study will aim for proof-of-concept targeting of *Prkdc* mutation using CRISPR-Cas9 technology. The main aims of this section are as follows:

1. *In vitro* gene editing of *Prkdc* in SCID mouse fibroblasts using *Streptococcus pyogenes* Cas9 (spCas9) nuclease.
2. Homology-mediated repair of *Prkdc* mutation using Cas9 and plasmid repair templates in SCID mouse fibroblasts.
3. Characterisation of the effects of gene editing through molecular analyses in clonal populations. This will allow me to overcome some of the limitations observed with the earlier efforts employing ZFNs.

After demonstrating proof-of-principle efficacy in *Prkdc* SCID model, this study will progress towards gene editing of *ATM* gene as a therapeutic strategy for Ataxia Telangiectasia, a clinically relevant DNA repair disease model. The specific aims for this investigation are as follows:

1. Investigation of *ATM* gene structure to generate a map of known disease-causing mutations.
2. Development of therapeutic gene editing strategies targeting both A-T specific mutations and universal cDNA knock-ins.
3. Identification of A-T mutations to envision generation of an A-T mouse model.
4. Production of *ATM* spCas9 gene editing reagents for *ATM* gene editing in human embryonic kidney (HEK-293T) cells.

This early pre-clinical work will pave the way for future therapeutic strategies of application to human haematopoietic stem cells and the central nervous system (CNS), which are the two major tissue targets in A-T. Prime candidates for genome editing in the bone marrow compartment will be CD34+ progenitor cells or Purkinje cells for proof-of-principle genome editing in the context of the CNS.

2 Materials and Methods

2.1 Bioinformatics

2.1.1 Web Tools

- Benchling https://benchling.com
- Cas-Designer http://www.rgenome.net/cas-designer/
- CHOPCHOP https://chopchop.rc.fas.harvard.edu
- CHOPCHOPv2 http://chopchop.cbu.uib.no
- EMBOSS Matcher http://www.ebi.ac.uk/Tools/psa/emboss_matcher/
- Ensembl Genome Browser http://www.ensembl.org/index.html
- Human Gene Mutation Database (HGMD®) http://www.hgmd.cf.ac.uk/ac/index.php
- MIT CRISPR Designer http://crispr.mit.edu
- Primer-BLAST https://www.ncbi.nlm.nih.gov/tools/primer-blast/
- TIDE https://tide.nki.nl

2.1.2 Software

- Chromas lite (version 2.6.2) http://technelysium.com.au/wp/
- GraphPad Prism
- Rotor-Gene 6000 (Qiagen)
- SnapGene Viewer http://www.snapgene.com/products/snapgene_viewer/
- Vector NTI ® Express (Thermo Fisher Scientific)
- ZEN lite (Zeiss)

2.1.3 Selecting Guide RNA Targets

Identification of guide RNAs (gRNAs) is the first step of a CRISPR-Cas9 gene editing experiment. 20 nucleotide gRNAs for *SpCas9* nuclease were designed using four different wed-based tools – MIT CRISPR Designer, CHOPCHOP, Benchling and Cas-Designer. To design guides, DNA sequence, genomic location or name of gene of interest, and species were input in the design tool. Algorithms specific to each tool generated an output with a list of candidate guide sequences with corresponding

predicted off-target sites for each input. Programs would find any candidate sequences based on the $N_{20}NGG$ sequence pattern principle, where NGG represents the PAM region. Guide sequences with minimal likelihood of off-target effects and greater on-target specificity were chosen for experiments described. Since a 5' guanine is required for transcription from U6 promoters, target sites that lack this feature were extended in the 5' direction by a single guanine. 5' extensions do not affect gRNA function (Ran, Hsu et al. 2013).

2.1.4 TIDE

TIDE or Tracking Indels by DEcomposition is a bioinformatic web-tool that is designed to detect genome editing by CRISPR-Cas system. TIDE requires as input a control sequence data file (obtained from cells transfected without Cas9-gRNA), a sample sequence data file (DNA from a pool of cell treated with Cas9-gRNA) and a character string representing the sgRNA sequence (20 nt). The sequencing data files were imported (*.abif* format) into TIDE software. Additional parameters were adjusted, if necessary. Version 3.1.1 of TIDE code was used for analysis described. TIDE code was written in R (Brinkman, Chen et al. 2014).

2.2 Molecular Biology

2.2.1 Media and Buffer Recipes

2.2.1.1 Luria-Bertani (LB) Broth

- 2% (w/v) LB broth powder (Sigma Aldrich #28713) in deionized water
- Supplemented with 100 µg/ml Ampicillin (Sigma Aldrich #A9393)

2.2.1.2 LB-Agar Media

- 2% (w/v) LB powder in deionized water
- 1.5% Agar (Sigma Aldrich #A1296)
- Supplemented with 100 µg/ml Ampicillin

2.2.1.3 NZY$^+$ Broth

- NZ amine broth (Sigma Aldrich #N3518) in deionized water
- 5 g of NaCl (Sigma Aldrich #S7653)
- 1 M $MgCl_2$ (Sigma Aldrich #M8266)
- 1 M $MgSO_4$ (Sigma Aldrich #M2643)
- 2 M glucose (Sigma Aldrich #G8270)

2.2.1.4 50x TAE Buffer

- 40mM Trizma® base (Sigma Aldrich #T1503)
- 20mM Glacial acetic acid (Sigma Aldrich #1005706)
- 1mM Ethylenedinitrilotetraacetic acid (EDTA; Sigma Aldrich #EDS)

2.2.2 Cell lines

- Top-10: Chemically competent genetically modified *E.coli* strain for high-efficacy cloning and plasmid propagation (Thermo Fisher Scientific #C404010)
- XL10-Gold: Ultracompetent cells for transformation of large DNA molecules with high efficiency (Stratagene #200315)

2.2.3 Plasmids

pRY refers to plasmid record in Dr Rafael Yáñez's laboratory, RHUL.

2.2.3.1 Lentiviral Vector Construction Plasmids

- pRY397 pMDLg/pRREintD64V: Lentiviral packaging plasmid encoding group specific antigen (GAG), polymerase (POL) and integrase (IN) gene with D64V mutation. Generated by Dr Rafael Yáñez (Yanez-Munoz, Balaggan et al. 2006).
- pRY398 pMDLg/pRRE: Lentiviral packaging plasmid. Contains Gag and Pol and wild-type IN. Kind gift from Prof Luigi Naldini.
- pRY399 pMD2.VSV-G: Envelope plasmid encoding the glycoprotein from the vesicular stomatitis virus (VSV-G); viruses pseudo-typed with this envelope have extensive host range transduce. Kind gift from Prof Luigi Naldini.
- pRY400 pRSV-REV: Plasmid encoding the Rev protein that is essential in transfer of lentiviral RNAs from the nucleus to cytosol. Kind gift from Prof Luigi Naldini.

2.2.3.2 CRISPR-Cas9 Plasmids

- pRY557 pSpCas9(BB)-2A-GFP; PX458: Plasmid encoding Cas9 from S. pyogenes with 2A-EGFP, and cloning backbone for sgRNA. Kind gift from Feng Zhang lab (Addgene plasmid #48138)
- pRY566: PX458 plasmid with *Prkdc* gRNA-1
- pRY567: PX458 plasmid with *Prkdc* gRNA-2
- pRY582: PX458 plasmid with *Prkdc* gRNA-3
- pRY503: *Prkdc-neo* donor template (previously generated by H. Abdul-Razak in Yáñez lab).
- pRY583: *Prkdc-neo* template PAM mutated for *Prkdc* gRNA-1
- pRY584: *Prkdc-neo* template PAM mutated for *Prkdc* gRNA-3
- pRY568 lentiCRISPRv2: Lentiviral transfer plasmid vector encoding Cas9 from S. pyogenes with 2A-Puro, and a cloning backbone for chimeric gRNA. Kind gift from Feng Zhang lab (Addgene plasmid #52961)
- pRY562: lentiCRISPRv2 with *Prkdc* gRNA-1

2.2.3.3 Other Plasmids

- pMC-Cre: Plasmid encoding Cre recombinase for mammalian expression (Gu, Zou et al. 1993).

- pRY572 pCAG-GFP:Cre: Plasmid encoding Cre recombinase fused to GFP for mammalian expression. Kind gift from Connie Cepko (Addgene plasmid # 13776).
- pRY396 pCCL.CMV.eGFP.WPRE: Plasmid encoding enhanced green fluorescent protein reporter gene (eGFP). Kind gift from Prof Luigi Naldini.

2.2.4 Bacterial Transformation

To propagate plasmids, plasmid DNA was transformed and grown in chemically competent Top-10 or ultracompetent XL-10 Gold (Stratagene #200315) *E. coli* bacterial strains. Competent cells were stored at -80°C. 50 µl of bacterial cells were used for each transformation reaction. Cells were pipetted in a microcentrifuge tube and thawed on ice before adding 2 µl (10 pg to 100 ng) of DNA to be transformed. The reaction was gently mixed by flicking the bottom of the tube before being incubated on ice for 15 minutes. The bacteria were then heat shocked by placing the bottom half or two-thirds of the tube in a 42°C heat block for 30 seconds and placed back on ice for another 10 minutes. 900 µl of warm antibiotic-free LB broth (NZY^+ broth for XL-10 Gold) was added to the microcentrifuge tube. The mixture was then transferred to a sterile 14 ml vent cap tube (Greiner Bio One #187261) which was then placed on a shaker (225 rpm) at 37°C for 30 minutes for the outgrowth of bacteria. Transformation mixture was then mixed and plated on warm LB-agar plates with 100 µg/ml of ampicillin. Plates were incubated at 37°C overnight.

2.2.5 Mini-prep of Plasmid DNA

Low quantities of plasmid DNA were prepared using the QIAprep spin Miniprep kit (Qiagen #27104). Single bacterial colonies from the agar plates were picked and grown up in a starter culture of 5 ml of 2% LB Broth with 100 µg/ml of ampicillin at 37°C in a shaking incubator (225 rpm) overnight. For long term storage, a glycerol stock was made by mixing 500 µl of mini-culture with sterile glycerol (Sigma Aldrich #G5516) at a final concentration of 15% and stored at -80 °C in a cryovial. 4 ml of each starter culture was then centrifuged at 8,000 rpm for 3 minutes at room temperature. The supernatant was removed and the cells resuspended in 250 µl Buffer P1 before being transferred to a microcentrifuge tube. 250 µl of Buffer P2 was added and the tube mixed thoroughly through multiple inversions. 350 µl of Buffer N3 was then added and mixed immediately through inversions followed by centrifugation for 10 minutes at 13,000 rpm. 800 µl of the supernatant was then added to a QIAprep spin column and centrifuged at 13,000 rpm for 30 seconds. The flow-through was discarded and the spin column washed by the addition of 500 µl of Buffer PB and centrifuged at 13,000 rpm for 30 seconds. 750 µl of

Buffer PE was then added to the column and centrifuged at 13,000 rpm for 30 seconds. The flow-through was discarded and the column centrifuged at 13,000 rpm for 1 minute to remove residual wash buffer. The QIAprep column was then placed in a sterile microcentrifuge tube and the DNA eluted in 50 µl of Buffer EB by incubation at room temperature for 1 minute followed by centrifugation at 13,000 rpm for 1 minute. The plasmid DNA was quantified using NanoDrop ND-1000.

2.2.6 Maxi-prep of Plasmid DNA

Once the plasmid was confirmed through restriction digests, the Endo-Free Plasmid Maxi Kit (Qiagen #12362) was used for large scale plasmid DNA preparation. From a starter culture, 250 µl was placed in 250 ml (1:1000 dilution) of fresh, sterile LB broth and left to grow on a shaker (225 rpm) at 37°C overnight. Bacterial cells were harvested by centrifugation at 3,500 rpm at 4°C for 30 minutes. The supernatant was discarded and the pellet thoroughly re-suspended in 10 ml Buffer P1. 10 ml of Buffer P2 was mixed by multiple inversions and incubated at room temperature for 5 minutes. 10 ml of chilled Buffer P3 was added and mixed by inverting. The lysed cell mixture was then applied to the barrel of QIAfilter Cartridge and incubated for 10 minutes. The lysate was filtered by inserting a plunger into the barrel. 2.5 ml of Buffer ER was added to the filtered lysate collected from the QIAfilter Cartridge and mixed before incubating on ice for 30 minutes. The QIAGEN-tip 500 was equilibrated by applying 10 ml of Buffer QBT and allowing the column to empty via gravity flow. The filtered lysate was added to the QIAGEN-tip 500, which was then washed with 30 ml of Buffer QC twice. DNA was eluted into 15 ml of Buffer QN and precipitated using 10.5 ml (0.7 volumes) of isopropanol followed by centrifugation at 6,000 rpm for 60 minutes at 4°C. The supernatant was carefully decanted and the pellet washed with 5 ml of endo-toxin free room temperature 70% ethanol, before another centrifugation at 6,000 rpm for 30 minutes at 4°C. The supernatant was removed and the pellet allowed to air dry in a laminar flow hood before being dissolved in endotoxin-free Buffer TE. An aliquot of plasmid DNA was then quantified using ND-1000. The stock plasmid DNA was kept endotoxin-free and only used under a laminar flow hood.

2.2.7 Mammalian DNA Extraction

Genomic DNA was extracted using DNasy® Extraction Kit (Qiagen #69504). First, cells were harvested from the tissue culture vessel using 0.25% Trypsin-EDTA (Thermo Fisher Scientific #25200056) and pelleted by centrifugation at 300 $\times g$ for 5 minutes. The pellets were then washed with PBS to remove excess growth media and suspended in 200 µl PBS. From here, DNA was extracted as per manufacturer's protocol. Genomic

DNA was eluted in TE buffer or sterile water and stored at -20°C. DNA concentration was quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) with a 0.2 mm path length.

2.2.8 Gel Electrophoresis

Gel electrophoresis was carried out in 1x TAE Buffer. 0.7-3.5% w/v agarose (Bioline #41026) gels were prepared using 1x TAE buffer by heating in a microwave in 30 seconds short-bursts until agarose was fully dissolved. The mixture was cooled before adding 0.5 µg/ml ethidium bromide (Sigma Aldrich #E1510). The gel was poured into electrophoresis tanks and cooled. Gel electrophoresis was carried out in 1x TAE buffer containing 0.5 µg/ml ethidium bromide at 1 Volt per cm for 60 mins. Gel images were visualised and captured under UV trans illumination using Bio-Rad Gel Doc 2000 system.

2.2.9 Restriction Digests

Restriction digests were carried out in 25 to 50 µl reactions. Each reaction consisted of sterile water, restriction enzyme (New England Biolabs), corresponding buffer solution, and genomic or plasmid DNA. The reaction was gently mixed and incubated at enzyme specific temperature. Incubation time was based upon the quantity of DNA to be digested (3 hours for DNA 1 µg DNA or overnight for 10 µg DNA). Digested DNA fragments were analysed using gel electrophoresis.

2.2.10 PCR Purification

For sequencing, PCR products were first run on a gel to test for correct amplification followed by purification using QIAquick PCR Purification Kit (Qiagen #28104) using a microcentrifuge. 5 volumes of Buffer PB 1 were added to 1 volume of the PCR sample and mixed. The pH was adjusted using 3 M sodium acetate, pH 5.0 (Sigma Aldrich #W302406), as required. The DNA was applied to a QIAquick column and placed in a 2 ml collection tube and centrifuged. Each centrifugation step was carried out at 13,000 rpm for 1 minute. The flow through was discarded and the spin column was washed by addition of 750 µl Buffer PE by centrifugation. The flow through was discarded again and the column centrifuged for an additional 1 minute to completely remove residual ethanol. Lastly, the QIAquick column was placed in a 1.5 ml microcentrifuge tube and the bound DNA was eluted in 30 µl Buffer EB (10 mM Tris-Cl, pH 8.5) by centrifugation.

2.2.11 Sequencing

Sequencing was carried using the cycle sequencing technology (dideoxy chain termination) on Applied Biosystems 3703XL automated DNA sequencing machines (MWG Eurofins, UK). Data were analysed using Chromas Lite software.

2.3 Cell Culture

2.3.1 Media and Buffer Recipes

2.3.1.1 Growth Media

- Dulbecco's Modified Eagle Medium (DMEM) high glucose (4.5 g/L) with stable glutamine and sodium pyruvate (Sigma Aldrich #41966052)
- 10% Heat-inactivated foetal bovine serum (Sigma Aldrich #10270106)
- 100 U/ml Penicillin and 100 µg/ml Streptomycin (Sigma Aldrich #15070063)

2.3.1.2 Freezing Media

- 40% DMEM
- 50% Heat-inactivated foetal bovine serum
- 10% Dimethylsulfoxide (DMSO; Sigma Aldrich #D8418)

2.3.1.3 2x HEPES-Buffered Saline (HBS) Buffer, pH 7.12

- 100 mM HEPES (Sigma Aldrich #H3375)
- 281 mM NaCl
- 1.5 mM Sodium phosphate dibasic ($Na2HPO_4$; Sigma Aldrich #S3264)

2.3.1.4 Detergent Lysis Buffer

- 0.45% Nonidet P40 (NP40; Sigma Aldrich #NP40)
- 0.45% Tween-20 (Sigma Aldrich #P1379)

2.3.2 Cell lines

Most of the cell lines described in this Thesis were obtained as frozen stocks with low passage numbers (p0-p3) from liquid nitrogen storage in Dr Yáñez's laboratory at Centre for Biomedical Sciences, School of Biological Sciences, Royal Holloway University of London (RHUL), unless otherwise stated.

- Balb/c 3T3: Primary skin fibroblasts obtained from C57 balb/c mice. Courtesy of Prof John Thacker, MRC Harwell.
- Balb/c mTert 3T3: Balb/c 3T3 fibroblasts immortalised using *mTert* gene

- SCID mTert 3T3: Primary skin fibroblasts obtained from C57 balb/c SCID mice and immortalised using *mTert* gene. SCID 3T3 fibroblasts (Courtesy of Prof John Thacker, MRC Harwell), immortalised by H. Abdul-Razak in Yáñez lab.
- HEK-293T/HEK-293T clone 17: Variant of human embryonic kidney 239 cell line containing a temperature-sensitive SV40 large T-antigen (ATCC #CRL-3216/11268)
- HEK-293T-Cas9: HEK-293T cells stably expressing Cas9 from AAVS1 (GeneCopoeia Inc. #SL502)
- HeLa: Human cell line derived from cervical cancer (ATCC #CCL2)

2.3.3 Storage

For long term storage, cells were washed with 1x phosphate buffered saline (PBS) solution, pH 7.4 (Thermo Fisher Scientific #10010023) and detached from tissue culture vessel using Trypsin-EDTA. Trypsin was inactivated by the addition of 9 ml of pre-warmed growth media. Cells were pelleted by centrifugation at 300 x g or 5 minutes and re-suspended in cold freezing medium. Suspended cells were aliquoted into cryovials (Thermo Fisher Scientific # 10654721) and cooled slowly to -80°C in isopropanol containers overnight before long-term storage in liquid nitrogen (gaseous phase).

2.3.4 Thawing of Frozen Cells

To thaw cells, frozen cell aliquots were retrieved from liquid nitrogen and rapidly warmed in 37 °C water bath. Defrosted cells were mixed with 9 ml pre-warmed growth medium and pelleted by centrifugation at 1000 rpm for 5 mins to remove all traces of DMSO. The cell pellets were then re-suspended in warmed growth media, transferred to tissue culture flasks and maintained at 37°C under 5% CO_2 atmospheric conditions.

2.3.5 Cell Maintenance and Sub-Culture

All cell culture work was performed under Class II laminar flow hoods using sterile tissue-culture grade plastic ware. HEK293T, HEK293T-Cas9, HeLa, C57 Balb/c mTert 3T3, and C57 Balb/c SCID mTert 3T3 cells were maintained in growth medium at 37°C under 5% CO_2 atmospheric conditions. C57 Balb/c mTert 3T3 and C57 Balb/c SCID mTert 3T3 cells were grown with 3 µg/ml puromycin (Sigma Aldrich #P8833).

Cells were sub-cultured every 2-3 days or upon reaching 80-90% confluency. For sub-culture, the growth media was aspirated and cells were washed with 1X PBS (Thermo Fisher Scientific #10010023) to remove traces of media. Cells were detached from surface of flask by incubation with Trypsin-EDTA for 1 min at 37°C. Trypsin was

inactivated by adding growth media and cells were pelleted by centrifugation at 3,000 rpm for 5 minutes. The media was removed thereafter and cells were re-suspended in fresh pre-warmed growth media. 1:5 to 1:10 of the suspended cells were placed in a fresh flask with additional growth and allowed to grow in the incubator.

2.3.6 Generation of Clonal Populations

Single-cell colonies were obtained by plating cells at a low seeding density, approximately 100-1000 cells, in 10 cm tissue culture plates. Cells were continued to grow for 7-10 days until small circular colonies (with a diameter of roughly 1 mm) were visible. Colonies were picked up using 3 mm sterile cloning discs (Sigma Aldrich #Z374431). To do so, first colonies were marked at the bottom of the dish. Next, the culture medium was gently aspirated and the plate was washed with 1x PBS to remove traces of growth media. Using sterile forceps, cloning discs were briefly dipped in Trypsin-EDTA and gently placed on marked colonies for 30-60 seconds. After 1 minute, cells adhered to the discs were transferred to 96-well plates. On reaching confluency, cells were again detached using Trypsin-EDTA and transferred to larger culture vessels for expansion, leaving the cloning discs behind in 96-well plates.

2.3.7 Crystal Violet Staining

Cells to be stained were first washed with 1 ml 1x PBS followed by 5 minute incubation in 1 ml diluted (1:10 in water) crystal violet. A stock solution of crystal violet (Sigma Aldrich #C0775) is 1% (w/v) in ethanol, kept at room temperature. Once stained, crystal violet was discarded and the tissue culture vessel was gently washed with water to remove excess crystal violet. Finally, the tissue culture plates were left upside-down to air-dry and the colonies were scored by eye.

2.3.8 Mammalian Cell Transfection

2.3.8.1 Lipofectamine-2000

Purified plasmid DNA or RNA were mixed with Lipofectamine®2000 (Thermo Fisher Scientific #11668019) in growth media without selection drugs. The mixture was incubated at RT for 30 mins to facilitate formation of lipid-DNA complexes. After incubation, the transfection mix was dropwise added on to plated cells. The solution was evenly distributed by gently rocking the plates before putting the plates in a cell culture incubator. Transfected cells were harvested 24 to 48 hours post-transfection.

2.3.8.2 Calcium-Phosphate Co-Precipitation

For 15 cm plates, a DNA mix was prepared by mixing 25 µg of purified plasmid DNA made up to 112.5 with µl 1x Buffer TE, 1,012.5 µl of tissue-culture quality grade water (Thermo Fisher Scientific #15230188) and 125 µl of 2.5 M $CaCl_2$ (Sigma Aldrich #C5670). The mix was then vortexed and incubated for 5 min at RT. This was followed by the addition of 1,250 µl of 2x HBS drop wise to the DNA mix while vortexing at high speed. The mixture was immediately added on to the plated cells (80-90% confluent) and gently mixing with the growth medium. 14 to16 hours post-transfection, the growth medium was aspirated and cells were twice washed with 1x PBS, before 25 ml fresh growth media being added. Cells were grown for 2-3 days before harvesting. Based on surface area ratios of tissue culture vessels, the protocol was scaled up and down as per need.

2.3.9 Cell Viability Assay

MTT (3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide) assay was used to detect the cell viability after exposure to Melphalan. The assay depends on the conversion of the water soluble MTT into insoluble MTT-formazan crystals by mitochondrial dehydrogenases of living cells. The crystals are then solubilised and the concentration of solution is determined by optical density at 570 nm (Price and McMillan 1990, Hansen and Bross 2010, van Meerloo, Kaspers et al. 2011). A 5 mg/ml MTT (Sigma Aldrich #M2003) stock solution was prepared in 1x PBS then filtered and kept at 4 °C in dark till use. Cells were seeded in 24-well plates. After 24 hours, the culture medium was replaced with growth medium with added10 µM Melphalan (Sigma Aldrich #M2011). After 1 hour incubation at 37°C, medium containing the drug was replaced with fresh growth media. The plates were incubated for 4 days prior to determining cell viability by measurement of MTT conversion. MTT solution at a final concentration of 1 mg/ml was added to each sample, incubated for 4 hours and then the medium removed gently without disturbing the formazan crystals. The plates were left for 1 hour to dry then DMSO was added to dissolve the crystals, and incubated with agitation for 10 min. Later, 100 µl from each sample were transferred into 96-well plate to measure the optical density at 570 nm using GloMax®-Microplate Multimode Reader (Promega) The optical density from control untreated cells was considered as 100 % viability.

2.3.10 Clonogenic Survival Analysis

At day 1 of the analysis, one T25 cm flask containing 2 x 10^5 cells per sample cell line and one 10 cm dish containing 6 x 10^4 feeder cells were set up per sample cell line to be tested. Feeder cells were a wild-type cell line irradiated using 35 Gγ before being plated into growth medium supplemented with new born calf serum. The following day, cells from T25 flasks were trypsinizied and counted. These were then diluted, treated and plated on to 10 cm feeder plates. Cell numbers used were:

- 0 Gγ - 200 cells
- 1 Gγ - 400 cells
- 3 Gγ - 2,000 cells
- 5 Gγ - 10,000 cells
- 7 Gγ - 30,000 cells

Cells were stained and counted between days 12 to 21 by adding 1 ml of Methylene blue (Sigma Aldrich #M9140) directly to the growth medium of each dish before incubating them for 45 minutes at room temperature. After incubation, the growth medium was poured, plates washed with water and left to air dry before scoring.

2.3.11 Immunofluorescence

Slides were prepared by placing coverslips into 3cm dishes and 2 x 10^5 cells were added. For G0/G1 analysis, the cells were grown at 37°C until confluent or starved of serum by using 0.5% FBS for 3 to 4 days. For G2 analysis, cells were cultured until they reached 60-80% confluency. To monitor repair kinetics after IR exposure, each dish was irradiated with 3 Gγ (9 seconds/Gγ) at 0.25, 2, 6, 24, 48, and 72 hour intervals. All dishes were processed together. Cells were fixed for 10 minutes at room temperature using 3% paraformaldehyde and 2% sucrose in 1x PBS. Next, the cells were permeabilized using 0.2% TritonX-100 (Sigma Aldrich #X100) in PBS for 2.5 minutes at room temperature and washed with 1x PBS three times.

Primary and secondary antibodies were diluted in PBS supplemented with 2% bovine serum fraction V albumin (Sigma Aldrich #A2153). Primary incubations were performed for 30 minutes at 37°C and secondary for 20 minutes at 37°C followed by three washes with 1x PBS. Nuclei were counterstained by DAPI (0.000025% in PBS) for 10 minutes at room temperature, followed by three PBS washes. Coverslips were mounted onto microscope slides in Vectashield and sealed with clear nail varnish before counting H2AX foci using fluorescence microscope. A total of 15 views including 10 cells were counted per sample.

2.3.12 Fluorescence Activated Cell Sorting (FACS)

Zinc-finger nuclease targeted fibroblast clones treated with a plasmid encoding *Cre* and *GFP* were subjected to cell sorting. Two days after transient transfection, cells were sorted using FACSAria III (BD Biosciences). GFP-positive clones were collected and maintained in conditioned growth media in 96-well plates.

2.4 Prkdc Zinc Finger Nuclease Gene Editing

2.4.1 Inside-Out PCR

The donor-genome 3' junction was amplified using an inside-out PCR where the forward primer lied within the region of homology (that is sequence present in the repair template) while the reverse primer was designed to bind in the genomic sequence lying outside the region of homology, thus absent from the repair template. This amplicon would only be present in targeted cells and absent from non-targeted cells where the donor template may integrate at random genomic loci. For PCR, genomic DNA was extracted from 1 x 10^6 G418-R cells and the concentration was measured using a Nanodrop ND-1000. Following primer sequences (Sigma Aldrich) were used for the inside-out PCR. The amplicon length was 1,335 bp.

GT1-F: 5'- TCGCCTTCTTGACGAGTTCT-3'

GT1-R: 5'- TTTTCCCCCTCATGTCACTC -3'

For PCR amplification, 50 µl reactions were prepared containing 10 µl 5 × GoTaq® reaction buffer, 1.5 µl of 10 mM dNTPs, 4 µl of 25 mM MgCl2, 1.5 µl of 10 µM of each forward and reverse primer, 200 ng DNA template (genomic DNA extracted from gene-targeted cells), and 0.5 µl of 5 u/µl GoTaq DNA polymerase (Promega #M7650). Thermo-cycling conditions were: initial denaturation at 95°C for 2 min, then 35 cycles of 95°C for 45 sec, 59°C for 60 sec, 72°C for 72 sec, followed by final extension at 72°C for 5 min.

2.4.2 Mutation Analysis

A single PCR amplifying a short (304 bp) region around ZFN cleavage and SCID mutation loci was used to analyse sequence mutations at both sites. The following primers (Sigma Aldrich) were used for the reaction.

mDNA-PK-F: 5' GCAGACAATGCTGAGAAAAGG 3'

mDNA-PK-R: 5' GCACAAAACAGACAAGGGTGT 3'

The PCR reaction consisted of: 1 x Go Taq® reaction buffer, 2 mM MgCl2, 0.4 mM dNTPs, 0.5 µM of forward and reverse primers, and 0.05 U/µl Taq DNA polymerase (Promega #M7650). Thermo-cycling conditions were: initial denaturation at 95°C for 5 min, then 35 cycles of 95°C for 30 sec, 60°C for 30 sec, 68°C for 40 sec, followed by final extension at 68°C for 2 min, holding the reaction at 4°C.

10 µl of the PCR product was tested on 1% agarose gel. On determining the correct size, the remaining PCR mixture was purified using QIAquick PCR purification kit protocol, mixed with 2 µl of 10 µM forward primer and sequenced (MWG Eurofins, UK).

2.4.3 Cre-LoxP Recombination

2.4.3.1 *Neo* Removal PCR tests

Two PCRs were designed to screen cells for loss of neo cassette. The first reaction was designed to amplify a 1,000 bp region within the neo gene using the following primers (Sigma Aldrich).

Neo-Removal-F: 5' CCGGCATTCTGCACGCTTCA 3'

Neo-Removal-R: 5' AGTGGCACCTTCCAGGGTCA 3'

The PCR mix contained: 1 x Go Taq® reaction buffer, 2 mM MgCl2, 0.4 mM dNTPs, 0.3 µM of forward and reverse primers, and 0.05 U/µl of Taq DNA polymerase (Promega #M7650). Thermo-cycling conditions were: initial denaturation at 94°C for 2 min, then 40 cycles of 94°C for 30 sec, 60.4°C for 30 sec, 72°C for 60 sec, followed by final extension at 72°C for 5 min, holding the reaction at 4°C.

The second PCR was designed to amplify the region around left LoxP site using the following primers (Sigma Aldrich).

Left-loxp-F: 5' AATCCTTGGGCCATTGTGAGT 3'

Left-loxp-R: 5' AGAGGCCACTTGTGTAGCG 3'

The PCR mix contained: 1 × Go Taq® reaction buffer, 1.5 mM MgCl2, 0.3 mM dNTPs, 0.3 µM of forward and reverse primers, and 0.05 U/µl of Taq DNA polymerase (Promega #M7650). Thermo-cycling conditions were: initial denaturation at 95°C for 2 min, then 35 cycles of 95°C for 45 sec, 65°C for 60 sec, 72°C for 15 sec, followed by final extension at 72°C for 5 min, holding the reaction at 4°C.

2.4.3.2 *Neo* removal using pMC-Cre

2 x 10^6 cells from ZFN targeted clones (c.5, c.11 and c.12) were seeded in 10 cm plates containing 10 ml growth medium with G418 (800 µg/ml; Thermo Fisher Scientific #10131027) and puromycin (3 µg/ml). Two hours before transfection the media was discarded to remove the selection drugs and fresh growth medium was added. 10 µg of purified, endo-toxin free pMC-Cre-15 plasmid was used for transfection using a calcium phosphate co-precipitation method. 16 hours after transfection the growth medium was

discarded again and cells were washed with 1x PBS two times to remove excessive calcium precipitate. Fresh growth medium (without G418) was added and cells incubated for further 48 hours at 37°C. Subsequently, the cells were harvested and seeded for generation of monoclonal populations.

To screen for the removal of neo, 2 x 10^5 cells from each of 20 clones per sample were isolated in 200 µl PBS. Cells were lysed by adding 25 µL of detergent lysis buffer followed by addition of 1 µl of Pronase (Sigma Aldrich #PRON-RO). The reaction was then incubated at 50°C for 1 hour to facilitate cell lysis followed by 10 minutes at 94°C to inactivate the enzyme and then stored on ice. 25 µl of the cell lysate was directly mixed with 25 µl PCR mix to screen for removal of neo (Section 2.5.3.1).

2.4.3.3 *Neo* removal using pCAG-Cre:GFP

5 x 10^5 cells from ZFN targeted clones (c.3, c.5, c.10 and c.11) were seeded in 6-well plates containing 2 ml growth medium with G418 (800 µg/ml) and puromycin (3 µg/ml). Two hours before transfection the media was discarded to remove the selection drugs and fresh growth media was added. 2 µg of purified endo-toxin free pCAG-GFP:Cre plasmid was used for transfection using 12 µl Lipofectamine-2000® (6:1 Lipofectamine: DNA ratio) and plates were incubated at 37°C for 48 hours. After 48 hours, cells were harvested for FACS. Sorted cell clones (10 per sample) were expanded in culture in standard growth medium without G418. Genomic DNA was harvested from 1 x 10^6 cells for downstream PCRs to screen for loss of *neo* (Section 2.5.3.1).

2.4.4 Indel Detection Using Amplicon Analysis (IDAA)

IDAA was carried out at University of Copenhagen as reported previously (Yang, Steentoft et al. 2015). The following primer sequences were used for IDAA in ZFN targeted fibroblast clones. Primers were obtained from TAG Copenhagen, Denmark (http://tagc.com/).

mDNA-PK-F: 5’ GCAGACAATGCTGAGAAAAGG 3’

mDNA-PK-R: 5’ GCACAAAACAGACAAGGGTGT 3’

Amplicons were fluorophore labelled by tri-primer amplification using a universal 6-FAM 5′-labelled primer FamF and primers flanking the gene editing target site of which the sense primer carried an FamF target sequence extension. PCR was performed in 25 µl using AmpliTaq Gold (ABI/Life Technologies), 0.5:0.05:0.5 µM (FamF:F:R) primers and a touchdown thermocycling profile using an initial 72°C annealing temperature ramping down by 1 degree/cycle to 58°C followed by an additional 25 cycles using 58°C annealing

temperature. Denaturation and elongation was performed at 95°C for 45 sec and 72°C for 30 sec respectively. 1 µl of the PCR reaction or dilutions hereof was mixed with 0.5 µl LIZ600 or LIZ500 size standard (ABI/Life Technologies) and applied to fragment analysis on ABI3010 sequenator (ABI/Life Technologies) using conditions recommended by the manufacturer. Raw data obtained were analysed using Peak Scanner Software V1.0 (ABI/Life Technologies).

2.5 *Prkdc* CRISPR-Cas9 Gene Editing

2.5.1 Insert Oligo Design

To clone the guide sequence into the sgRNA scaffolds of PX458 and lentiCRISPRv2 plasmids, two oligos containing appropriate overhangs and 5' guanine residue was designed for each guide, as reported previously (Sanjana, Shalem et al. 2014). Oligo sequences for each guide RNA are described in **Table 2-1**.

Sequences were obtained as standard de-salted DNA oligos (Sigma Aldrich) and suspended as 100 µM stock solutions. Oligos were phosphorylated and hybridised in a single reaction. 1 µl each of the forward and reverse oligos were mixed with 1 µl of 10 x T4 ligation buffer (New England Biolabs #B0202) and 0.5 µl of T4 polynucleotide kinase (New England Biolabs #M0201) made up to a final volume of 10 µl with sterile water. The mix was heated to 37°C for 30 minutes then 95°C for 5 minutes with a temperature ramp down to 25°C at 5°C per minute (Techne-512).

Table 2-1 Prkdc gRNA oligos sequences.

Overhangs are highlighted in grey. 5' guanine and complementary cytosine residues are indicated in bold.

	Annealed oligo sequence with overhangs
***Prkdc* gRNA-1 oligos**	
Forward	5′ CACC**G**TGCTAAGAGAAAGTTAGCAG 3′
Complement	3′ **C**ACGATTCTCTTTCAATCGTCCAAA 5′
Reverse complement	5′ AAACCTGCTAACTTTCTCTTAGCA**C** 3′
***Prkdc* gRNA-2 oligos**	
Forward	5′ CACC**G**GCATAGCGTATTTTATGTTG 3′
Complement	3′ **C**CGTATCGCATAAAATACAACCAAA 5′
Reverse complement	5′ AAACCAACATAAAATACGCTATGC**C** 3′

2.5.2 Production of Cas9-gRNA Plasmids

Cloning of annealed *Prkdc* gRNAs 1 and 2 into PX458 plasmid backbone was carried out using a single-step digestion-ligation protocol as previously reported (Cong, Ran et al. 2013). The annealed oligo mix was diluted 1:250 in sterile water. 100 ng of the uncut PX458 plasmid was mixed with 2 µl of the diluted annealed oligos, with 1 µl of FastDigest *Bpil*I (*BbS*I; Thermo Fisher Scientific #FD1014), 2 µl of FastDigest buffer, 0.5 µl of T4 Ligase (New England Biolabs #M0202S), 1 µl of DTT (100 mM stock; Sigma #DTT-RO), 1 µl of ATP (10 mM stock; New England Biolabs #P0756S) and sterile water made up to

a final volume of 20 µl. The mix was placed in a heat cycle of 37°C for 5 mins then 23°C for 5 mins for 6 cycles and stored at 4°C. The ligated plasmid was then transformed into Top-10 competent cells. Transformed cells were streaked on LB agar plates with ampicillin and incubated at 37°C for 24 hours. Single colonies were picked and grown in a 5 ml starter culture at 37°C overnight before a mini-prep was performed.

To check for successful ligation the recombinant plasmid was sequenced (MWG Eurofins, UK) using the following primer (Sigma Aldrich) located in the upstream U6 promoter.

hU6-F: 5' TACGATACAAGGCTGTTAGAGAG 3'

2.5.3 Indel Detection by TIDE

To analyse indels introduced by Cas9 the genomic region surrounding PAM regions at the targeted sites were amplified using PCR and subjected to sequencing. Sequence traces from control and edited cells were analysed using TIDE software to quantify indels.

The following primers were used for the amplification of regions (304 bp) around gRNA-1 and 2 PAM.

mDNA-PK-F: 5' GCAGACAATGCTGAGAAAAGG 3'

mDNA-PK-R: 5' GCACAAAACAGACAAGGGTGT 3'

The PCR reaction consisted of: 1 x Q5 reaction buffer, 0.2 mM dNTPs, 0.5 µM of forward and reverse primers, and 0.02 U/µl Q5 Hot Start polymerase (New England Biolabs #M0493). Thermo-cycling conditions were: initial denaturation at 98°C for 30 sec, then 35 cycles of 9°C for 10 sec, 60°C for 30 sec, 72°C for 10 sec, followed by final extension at 72°C for 2 min, holding the reaction at 4°C.

The following primers were used for the amplification of region (805 bp) around gRNA-3 PAM.

mDNA-PK-F: 5' GCAGACAATGCTGAGAAAAGG 3'

SCID-G3-R: 5' CTTCTGGGGTCATGCCTAGC 3'

The PCR reaction consisted of: 1 x Go Taq® green reaction buffer, 1.5 mM MgCl2, 0.2 mM dNTPs, 0.5 µM of forward and reverse primers, and 0.05 U/µl Taq DNA polymerase (Promega #M7650). Thermo-cycling conditions were: initial denaturation at 95°C for 5

min, then 35 cycles of 95°C for 30 sec, 59°C for 30 sec, 72°C for 48 sec, followed by final extension at 72°C for 5 min then holding the reaction at 4°C.

10 µl of the PCR product was tested on 1% agarose gel. On determining the correct size, the remaining PCR mixture was purified using QIAquick PCR purification kit protocol, mixed with 2 µl of 10 µM forward primer and sequenced (MWG Eurofins, UK).

2.5.4 Modification of Repair Template

PAM regions for *Prkdc* gRNA-1 and 3 present in *Prkdc-neo* donor DNA template were mutated using QuickChange II XL Site-Directed Mutagenesis Kit (Agilent #200521). QuikChange Primer Design Program http://www.genomics.agilent.com/primerDesignProgram.jsp was used to design the following primers, which were obtained as PAGE-purified oligos (Sigma Aldrich).

G117T-F: 5' AAGAGAAAGTTAGCAGG**T**GCCAACCCAGCTGTTAT 3'

G117T-R: 5' ATAACAGCTGGGTTGGC**A**CCTGCTAACTTTCTCTT 3'

G213A-F: 5' ACACTTCTGAGCAGTATG**A**CACTTCACTGTGTAGATG 3'

G213A-R: 5' CATCTACACAGTGAAGTG**T**CATACTGCTCAGAAGTGT 3'

Each mutant strand synthesis PCR reaction comprised of 1x Reaction buffer, 40 ng of dsDNA template, 125 ng of primer # forward, 125 ng of primer # reverse, 1 µl of proprietary dNTP mix, 3 µl of QuickSolution reagent and 1 µl of *PfuUltra* HF DNA polymerase (2.5 U/µl) in a total volume of 50 µl. Thermo-cycling conditions used were: denaturation at 95°C for 1 min, followed by 18 cycles of 95°C 50 sec, annealing at 95°C for 50 sec, 68°C 14 mins (1 min/kb of plasmid length), and final extension at 68°C for 7 mins. Parental DNA was digested by incubating the PCR mix with 1 µl of *Dpn* I restriction enzyme (10 U/µl) at 37°C for 90 mins. 2 µl of *Dpn* I-treated was transformed into XL10-Gold ultracompetent cells.

The plasmid DNA was verified for mutagenesis by sequencing (MWG Eurofins, UK) using the following primer (Sigma Aldrich).

mDNA-PK-F: 5' GCAGACAATGCTGAGAAAAGG 3'

2.5.5 Gene Targeting

1 x 10^6 SCID mTert 3T3 cells were seeded in 10 cm plates with 10 ml growth media containing 3 ug/ml Puromycin. 4 hours later, plasmid constructs encoding genome

editing reagents (10 µg each) were transfected using calcium phosphate co-precipitation method, as described in **Table 2-2.**

Table 2-2 CRISPR-Cas9 gene targeting plasmid transfections

Plate #	CRISPR-Cas9 Plasmid	Donor DNA Plasmid
Plate 1	Cas9-GFP (*pRY557*)	*Prkdc-neo* (*pRY503*)
Plate 2	Cas9-GFP-gRNA-1 (*pRY566*)	*Prkdc-neo*
Plate 3	Cas9-GFP-gRNA-2 (*pRY567*)	*Prkdc-neo*
Plate 4	Cas9-GFP-gRNA-3 (*pRY582*)	*Prkdc-neo*
Plate 5	Cas9-GFP-gRNA-1	*Prkdc-neo-mut-PAM-g1* (*pRY583*)
Plate 6	Cas9-GFP-gRNA-3	*Prkdc-neo-mut-PAM-g3* (*pRY584*)
Plate 7	Cas9-GFP	-
Plate 8	-	*Prkdc-neo*
Plate 9	-	*Prkdc-neo-mut-PAM-g1*
Plate 10	-	*Prkdc-neo-mut-PAM-g3*

Since Cas9 plasmid constructs used contained a GFP reporter gene plasmid expression was monitored by fluorescence microscopy at 24 and 48 hour time-points after transfection. After 48 hours, the cells were washed with 1 x PBS two times and selected for gene targeting by addition of 800 µg/ml G418. The next day the cells were trypsinized and seeded for the generation of monoclonal populations and crystal violet staining. Separately, the polyclonal cell pool was also expanded.

2.5.6 Production of LentiCRISPRv2 Plasmid

Cloning of annealed *Prkdc* gRNA-1 into lentiCRISPRv2 plasmid backbone was carried out using a single-step digestion-ligation protocol as previously reported (Cong et al., 2013). The annealed oligo mix was diluted 1:250 in sterile water. 100 ng of the uncut lentiCRISPRv2 plasmid was mixed with 2 µl of the diluted annealed oligos, with 1 µl of *BsmB*I (Thermo Fisher #FD0454), 2 µl of FastDigest buffer, 0.5 µl of T4 Ligase (New England Biolabs #M0202S), 1 µl of DTT (100 mM stock; Sigma #DTT-RO), 1 µl of ATP (10 mM stock; New England Biolabs #P0756S) and sterile water made up to a final volume of 20 µl. The mix was placed in a heat cycle of 37°C for 5 minutes then 23°C for 5 minutes for 6 cycles and stored at 4°C. The ligated plasmid was then transformed into Top-10 competent cells. Transformed cells were streaked on LB agar plates with ampicillin and incubated at 37°C for 24 hours. Single colonies were picked and grown in a 5 ml starter culture at 37°C overnight before a mini-prep was performed.

To check for successful ligation, the recombinant plasmid was sequenced (MWG Eurofins, UK) using the following primer located in the upstream U6 promoter.

hU6-F: 5’ TACGATACAAGGCTGTTAGAGAG 3’

2.5.7 LentiCRISPRv2 Vector Production

HIV-1-based vectors were produced by transient transfection of HEK-293T cells using the calcium phosphate co-precipitation method. 72 hours prior to transfection, 3 X 10^6 cells were seeded in 15 cm plates containing 25 ml growth media. At about 80% confluency, the medium was replaced with 20 ml fresh growth medium, 2 hours before transfection. HEK-293T cells were co-transfected with four plasmids in molar ratio of 1:1:1:2 (packaging: REV: envelope: transfer). 12.5 µg of integration-proficient or integration-deficient packaging plasmid (pMDLg/pRRE or pMDLg/pRRE-intD64V respectively), 7 µg of VSV-G pseudotyping plasmid (pMD2.VSV-G), 6.25 µg of REV plasmid (pRSV-rev) and 32 µg of transfer plasmid of choice were added to a 50 ml tube. The DNA was mixed with 125 µl of 2.5 M $CaCl_2$, vortexed and incubated for 5 min at RT. To each tube 1,250 µl of 2x HBS Buffer was added drop-wise whilst vortexing at high speed. The DNA mixture was then immediately mixed gently with the media on the plated cells. At 14-16 hours after transfection, the medium was replaced with 18 ml of growth medium. Subsequently, at 48 hours post-transfection, media containing viral particles was harvested and plates replaced with 18 ml growth medium for second harvest done at 72 hours post-transfection. The harvested viral supernatant was cleared by centrifugation at 1000 x g for 10 min at RT and then passed through 0.22 µm-pore-size filters to remove suspended cells or debris. The filtered supernatant was then transferred to polyallomer centrifuge tubes (Beckman Coulter # 326823) and concentrated by ultracentrifugation (50,000 x g for 2 hours at 4°C). After centrifugation, the supernatant was discarded and pelleted viral particles resuspended in 50 µl of serum free DMEM. Viral suspension was then transferred to 1.5 ml microfuge tubes and spun at 1000 x g for 10 min at 4°C. The supernatant was then transferred to a new microfuge tube. The viral stock was adjusted to 10 mM $MgCl_2$ (Sigma Aldrich #M8266), DNase I (Promega #M6101) was added at 5 U/ml and the mixture was incubated at 37°C for 30 min. This step is required to remove any residual DNA. Finally, virus suspensions were aliquoted and frozen at -80 °C.

2.5.8 LentiCRISPRv2 Vector Titration

To assess the titres of lentiviral vectors, HeLa cells were transduced with dilutions of the concentrated vectors and qPCR was performed on total DNA extracted from these cells. 1 x 10^5 cells were seeded in 6-well plates one day before transduction. Cells were transduced in parallel with 1 in 2,000 and 1 in 20,000 dilutions of the prepared virus (each lentiviral vector had two preparations, harvest day 1 and day 2) in 2 ml of growth medium containing 8 µg/ml polybrene (Sigma Aldrich #TR-1003). A mock sample incubated with

only growth medium was also included. 24 hours post transduction, cells were harvested using trypsin and DNA was extracted from these using the QIAGEN DNeasy® tissue kit in a final volume of 400 µl through two sequential column elutions. In order to quantify the vector genomic DNA in transduced HeLa cells, the late reverse transcript (LRT) region, common to all lentiviral vectors cDNAs (Yoder and Fishel 2008) was amplified using LRT specific (Butler, Hansen et al. 2001) and quantified by qPCR (Rotor-Gene 6000). The primer sequences were as follows:

LRT-F: 5' TGTGTGCCCGTCTGTTGTGT 3'

LRT-R: 5' GAGTCCTGCGTCGAGAGAGC 3'

The reaction mixture comprised of: 100 nM of each primer, 1X SensiMixPlus SYBR (Bioline # QT650-05), and 6.25 µl of the sample DNA in a final volume of 20 µl. 10^2-10^7 copies of plasmid pHR'SIN-cPPT-SEW DNA per reaction were used as standards. Samples were analysed in duplicates and standards were in triplicates (including the negative control). The thermocycling conditions were: 50°C for 2 min, 95°C for 10 min and 50 cycles of 95°C for 15 sec/ 60°C for 60 sec. A melting curve calculation was performed at the end with a continuous reading between 50 and 100°C.

The LRT copy number titre was normalised to the number of cells transduced, which was determined though a parallel qPCR reaction. The number of cells was quantified using primers located in the beta actin gene following the same thermocycling conditions as the LRT reaction. The primer sequences were as follows:

Actin-F: 5' TCACCCACACTGTGCCCATCTACGA 3'

Actin-R: 5' CAGCGGAACCGCTCATTGCCAATGG 3'

The reaction was performed under the following conditions: 100 nM of each primer, 1 X SensiMixPlus SYBR, and 5 µl of the sample DNA in a final volume of 20 µl. DNA extracted from 10^6 HeLa cells using QIAGEN DNeasy® tissue kit was used to prepare the standard curve (10^1-10^5 cells) for this reaction. The final titre was expressed as qPCR transducing units/ml vector stock.

2.5.9 LentiCRISPRv2 Vector Transduction

1 x 10^4 balb/c wild-type and SCID mTert fibroblasts were seeded in 48-well plates one day before viral transduction in 1 ml growth media containing puromycin. Cells were transduced with lentiCRISPRv2 vector at a MOI 40 in 500 µl of growth medium containing 8 µg/ml polybrene. The plates were incubated for 48 hours at 37°C before harvesting for genomic DNA extraction.

2.6 *ATM* CRISPR-Cas9 Gene Editing

2.6.1 gRNA Transfection

ATM gRNAs were obtained as purified synthetic RNA oligos at a final concentration of 3 µM (Synthego®, USA). 100,000 HEK293T-Cas9 cells were plated in 24-well plates and transfected with 30 nM gRNAs using 2.5 µl Lipofectamine-2000®. Plates were incubated at 37°C, 5% CO_2 for 72 hours before harvesting for DNA extraction to detect gene editing.

2.6.2 Indel Detection by TIDE

To analyse indels introduced by Cas9 the genomic region surrounding PAM regions at the targeted sites were amplified using PCR, subjected to sequencing and analysed using TIDE software to quantify indels. Details of primer sequences, concentrations and annealing temperatures (Tm) are mentioned in **Table 2.3**. All primers were obtained from Sigma Aldrich.

The PCR reaction consisted of: 1 x Go Taq® clear reaction buffer, 1.5 mM $MgCl_2$, 0.4 mM dNTPs, forward and reverse primers, and 0.05 U/µl Taq DNA polymerase (Promega #M7650). Thermo-cycling conditions were: initial denaturation at 95°C for 5 min, then 35 cycles of 95°C for 30 sec, Tm°C **(Table 3)** for 30 sec, 72°C for 40-60 sec (calculated as 1 min/kb), followed by final extension at 72°C for 5 min, then holding the reaction at 4°C.

10 µl of the PCR product was tested on 1% agarose gel. On determining the correct size, the remaining PCR mixture was purified using QIAquick PCR purification kit protocol, mixed with 2 µl of 10 µM forward primer, except for ATM intron 32 where reverse primer was used for sequencing, and sequenced (MWG Eurofins, UK).

Table 2-3 List of oligos for *ATM* CRISPR-Cas9 targets

Name	Sequence (5'-3')	Conc.	Amplicon	Tm °C
2T>C-F	CTACTACTGCAAGCAAGGCAAA	0.5 µM	289 bp	51
2T>C-R	TGGTTCACAATTTCAGAACACAC			
103C>T-F	ACAGTGATGTGTGTTCTGAA	0.3 µM	225 bp	51
103C>T-R	GTTTCAGGATCTCGAATCAG			
5762ins-F	ATTGGGGTAGAATGGTGTTGAC	0.4 µM	211 bp	62.8
5762ins-R	TTTGTTTTTGCCCAACATACTG			
Int3-F	ATCCTTGAGTGCTCATTTCCTT	0.3 µM	768 bp	58.2
Int3-R	TGTTCTAGTTGACGGCAGCA			
Int32-F	TTACAGAGCACTTGGTACTTTTGA	0.5 µM	223 bp	51
Int32-R	CCCAGCTTTTAGTTTCTGTGTTTT			

2.6.3 Whole-Genome Sequencing

After 72 hours of *ATM* gRNA transfections in HEK293T-Cas9 cells, genomic DNA was isolated using a DNeasy Blood & Tissue kit and on-target loci were amplified for targeted deep sequencing. Deep sequencing libraries were generated by PCR. TruSeq HT Dual Index primers were used to label each sample. Pooled libraries were subjected to paired-end sequencing (LAS, Inc.).

2.7 Statistics

Mean and standard deviation were calculated using Microsoft Excel. GraphPad Prism software was used for statistical analyses. Standard data was presented as mean ± standard error of the mean (SEM). One- or two-way ANOVA with appropriate post hoc tests were used in most cases. In most analyses, differences were considered statistically significant when the P value was ≤ 0.05.

2.8 Collaborations

Several experiments described in this study were outsourced and/or done in conjunction with several national and international collaborators **(Table 2-4)**.

Table 2-4 List of Collaborators

Collaborator	Contribution
Ayad Eddaoudi Flow Cytometry Core Facility University College London, UK	Fluorescence activated cell sorting of ZFN *Prkdc* fibroblast clones
Pramila Gurung Third-year undergraduate student RHUL, UK	Primer design and execution of PCRs to screen for loss of neomycin selection cassette in genomic DNA from ZFN *Prkdc* fibroblast clones
Lisa Woodbine Genome Stability and Damage Centre University of Sussex, UK	H2AX phosphorylation immunofluorescence analysis and clonongenic cell survival assays following gamma-radiation ZFN *Prkdc* fibroblast clones
Eric Paul Bennett and Camilla Andersen Copenhagen Centre for Glycomics University of Copenhagen, Denmark	Indel detection amplicon analysis (IDAA) in ZFN *Prkdc* fibroblast clones
Daniyal Asghar Third-year undergraduate student RHUL, UK	Design of gRNAs and primers for *ATM* gene editing. Execution of PCRs and TIDE for indel detection in genomic DNA from CRISPR-Cas9 edited HEK293T cells.
Taeyoung Koo Centre for Genome Engineering Institute of Basic Science, South Korea	Next-generation sequencing of ATM targets in genomic DNA from CRISPR-Cas9 edited HEK293T cells.

3 Gene Editing in *Prkdc*-SCID Using Zinc Finger Nucleases

3.1 Introduction

Mice homozygous for the *Prkdc* SCID mutation are severely deficient in functional T and B lymphocytes. The mutation leads to a block in V(D)J recombination, preventing rearrangement of immunoglobulin and antigen receptor genes and thus causing an arrest in the early development of B and T lineage-committed cells, while other haematopoietic cells appear to develop and function properly. SCID mice readily support lymphocyte development and can be reconstituted with normal lymphocytes from other mice (Bosma, Custer et al. 1983, Bosma and Carroll 1991, Vladutiu 1993). Thus, these mice are of interest for studies of both normal and abnormal lymphocyte development and function. In addition, SCID mice also present as a good model for an *ex vivo* genome editing approach, where transplantation of gene corrected haematopoietic progenitors with selective growth advantage could reconstitute and restore the immune system.

A previous study undertaken in this lab demonstrated proof-of-principle editing of *Prkdc* gene in fibroblasts and haematopoietic stem/progenitor cells (HSC/HSPCs) obtained from *Prkdc*-SCID mouse. Using zinc finger nucleases and a corrective repair matrix, the study showed correction of the SCID mutation as well as partial rescue of the T-cell phenotype in mice transplanted with gene-edited HSCs (Abdul-Razak 2013).

In fibroblasts, gene editing led to the generation of a polyclonal population, which was originally analysed as a single sample. This chapter describes work undertaken to investigate gene editing events in single-cell clones derived from the ZFN-targeted cell pool. The main aims of this section were to verify repair of *Prkdc* DNA sequence and whether the repair correlated with functional rescue of the DNA PK enzymatic activity *in vitro*.

3.1.1 Homology-directed repair strategy

A homology-directed repair strategy was employed to target the *Prkdc* gene and modify the mutated nucleotide (a T-to-A transversion) in cultured, *mTert*-immortalised SCID mouse fibroblasts (Abdul-Razak, 2013). HDR relies on the engineered introduction of a DSB break in the DNA close to the region to be modified; and on the intrinsic ability of cells to repair breaks in their DNA using homologous recombination. Two obligatory heterodimeric zinc-finger nucleases (Sangamo BioSciences, California USA) were designed to target *Prkdc* and introduce a double-stranded break 86-bp downstream of the SCID mutation site located in exon 85. The homologous sequence containing the wild-type nucleotide (T) at the mutation site was ~7.5 kb-long, and a selection marker was supplied on a plasmid as a donor template (referred to as the '*Prkdc-neo*'). The homology region was derived from a BAC clone containing *Prkdc* sequence, and conceptually was divided into homology arms left (HAL) and right (HAR), joined at the SCID mutation site. The 5.1 kb HAL encompassed sequences from intron 79 until SCID mutation site in exon 85; HAR included 2.4 kb of homologous DNA sequence downstream of SCID site covering sequences from exon and intron 85, as well as an additional neomycin-resistance gene (*neo*) selection cassette (Neo-R). The 1.6 kb Neo-R consisted of *neo* floxed between two unidirectional *LoxP* sites and was inserted in intron 85 **(Fig. 3-1)**.

Recombination of the donor template with the cellular genome could lead to correction of the mutated nucleotide and incorporation of floxed *neo* cassette in the intron. Alternatively, the donor template can also recombine through NHEJ with the genomic DNA, generating random integrants **(Fig. 3-2)**. In both circumstances, *neo* would be introduced in the modified cells, conferring resistance to antibiotic G418. Antibiotic selection after gene targeting allows for positive selection and confers growth advantage to cells that have undergone genetic modification.

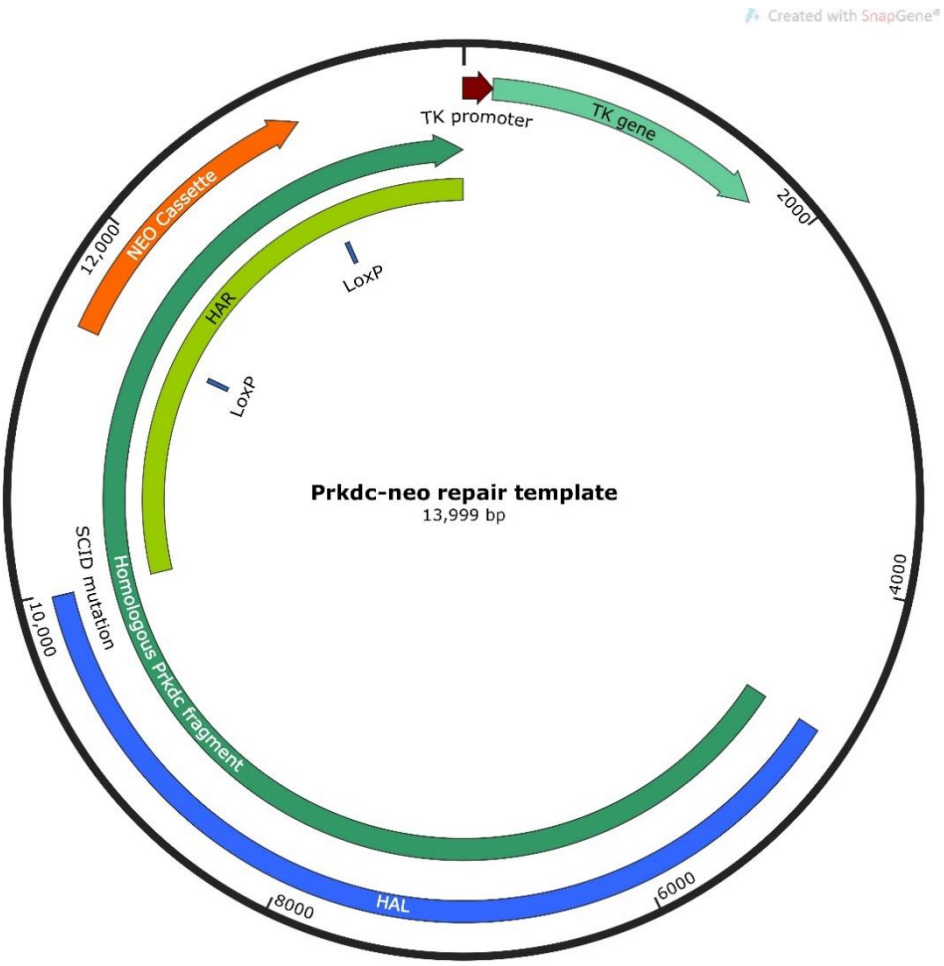


Figure 3-1 Schematic Representation of *Prkdc-neo* Donor Template

Prkdc-neo repair template was cloned as a double stranded circular plasmid DNA. Wild-type *Prkdc* sequences spanning homology arms left (HAL; 5.1 Kb) and right (HAR; 2.4 Kb) were obtained from a BAC clone. A selection marker cassette (1.6 Kb) containing a neomycin gene (neo) floxed between two loxP sites was incorporated in the HAR, as close as possible to the SCID mutation. Plasmid map created with SnapGene®.

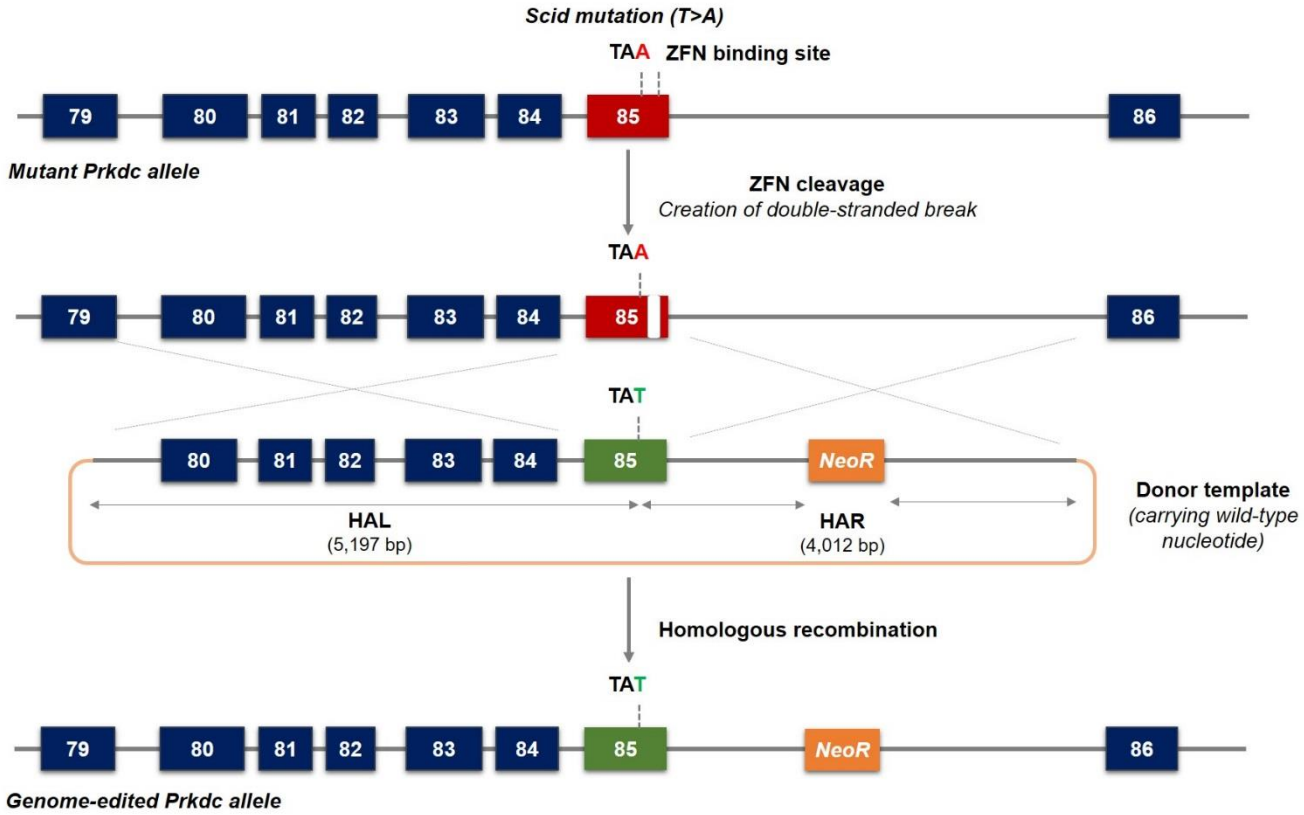


Figure 3-2 Strategy for HDR of *Prkdc* SCID Mutation Using ZFNs

Prkdc SCID is caused by a truncating mutation (T>A) in *Prkdc* exon 85. A homology-directed repair strategy using zinc finger nucleases was designed to repair the mutated nucleotide. ZFNs were designed to mediate a double stranded break downstream the mutation within exon 85. Wild-type *Prkdc* sequence containing correct nucleotide sequence was supplied on a plasmid donor template with left and right homologous arms (HAL, HAR). As a selection marker, neomycin gene floxed between two unidirectional loxP sites (*NeoR*) was included in the right arm of homology. Homologous recombination of the donor template with the cellular genome would lead to correction of mutated nucleotide and incorporation of floxed *neo* cassette conferring selective advantage to gene edited cells.

To validate the gene editing strategy, fibroblasts isolated from a *Prkdc*-SCID mouse and immortalised using mouse Tert (*mTert*) gene were utilised. First, ZFN genes were delivered on integration-proficient lentivectors followed by transfection of a repair template (10 µg) using calcium phosphate co-precipitation. Subsequently, G418 selection was applied, leading to selective growth of a G418-resistant (G418-R) cell population, arising from either HDR- or NHEJ-mediated recombination. In order to determine the efficiency of gene targeting, single-cell clones were derived from the diversified G418-R polyclonal population **(Fig. 3-3)** (Abdul-Razak 2013)**.** The clones were genotyped for successful recombination events using PCR and Sanger sequencing. Phenotype analyses was carried out by assessing DNA-PK activity.

Subsequently, Cre-LoxP recombination was used to remove *neo* and generate selection marker free G418-sensitive (G418-S) cell populations. The Cre-lox system is a technology that can be used to induce site-specific recombination events in the genome. The system consists of two components derived from the P1 bacteriophage namely, the Cre recombinase and a LoxP recognition site (Sauer and Henderson 1988, Sauer 1998). Cre (or *c*yclization) recombinase is a 38 kDa protein responsible for intra- and inter-molecular recombination at the loxP recognition sites. Cre acts independently of any other accessory proteins or co-factors. LoxP (*l*ocus *o*f *X* (cross)-over in *P*1) sites are 34 bp-long recognition sequences consisting of two 13 bp long palindromic repeats separated by an 8 bp-long asymmetric core spacer sequence. The LoxP sequence does not occur naturally in any known genome other than P1 phage, and is long enough that there is virtually no chance of it occurring randomly. Therefore, inserting LoxP sites at deliberate locations in a DNA sequence allows for very specific manipulations. The Cre recombinase catalyses the site-specific recombination event between two LoxP sites, which can be located either on the same or on separate pieces of DNA (Sauer and Henderson 1988, Sauer 1998).

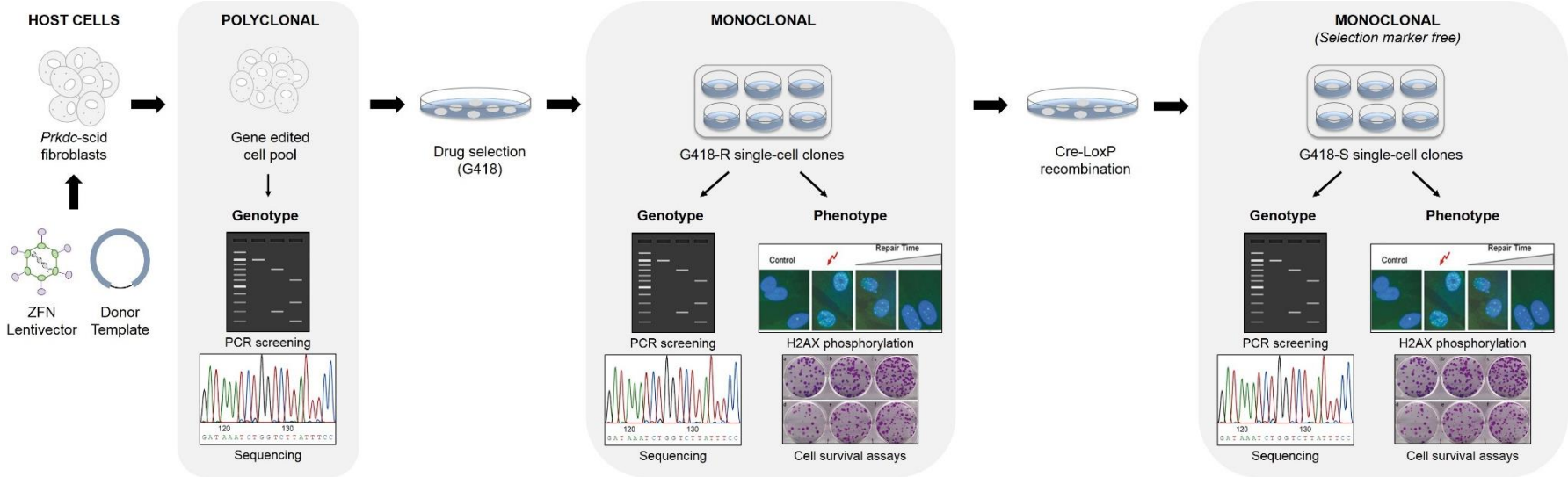


Figure 3-3 Workflow for Clonal Analysis of Genetically Modified SCID Fibroblasts

Gene editing reagents zinc finger nucleases and donor repair template were introduced into cultured SCID fibroblasts using lentiviral transduction and calcium phosphate transfection, respectively. Drug (G418) selection was used to isolate recombinant fibroblasts and generate a G418-resistant (G418-R) cell pool. PCR and Sanger sequencing were used to characterise polyclonal population. Subsequently, monoclonal populations were derived using dilution cloning. Gene editing events were analysed in individual cell clones using PCR and Sanger sequencing. For phenotype analysis, the activity of DNA-dependent protein kinase was assessed using H2AX phosphorylation and cell survival assays in response to DNA damage. Lastly, Cre-LoxP recombination was utilised to remove neo from the genome of edited cells and G418-sensitive (G418-S) clones were derived. These clones were then again subjected to genotype and phenotype analyses.

3.1.2 Validation of genome edits in monoclonal populations

Introducing a nuclease in cells (with or without a donor template) results in a mixed population of cells. Following the introduction of a double strand break in mammalian cells, cellular machinery repairs the DSB by non-homologous end joining or homology directed repair. Repair via the NHEJ pathway predominates in mammalian cells resulting in the creation of indel errors, short heterogeneous insertions and deletions of nucleic acid sequences, at the site of the DSB. The HDR pathway requires the presence of a repair template, which is used to fix the DSB in a more specific manner. HDR faithfully copies the sequence of the repair template to the cut target sequence. Some cells will not be edited, some will have one allele edited, and some will have both alleles edited.

In order to validate genome editing in ZFN targeted SCID fibroblasts, monoclonal populations derived from the targeted polyclonal cell pool were analysed individually. This involved amplification and sequencing of a region around the SCID mutation and ZFN cleavage sites.

3.1.3 Rescue of DNA-PK enzymatic activity

Loss of DNA PKcs enzymatic activity is associated with the impaired ability of cells to repair their DNA in response to DNA damage. The truncating mutation in *Prkdc* SCID fibroblasts abolishes most of the catalytic activity of DNA PKcs. Hence, correction of the mutated nucleotide should lead to restoration of DNA PKcs activity. In order to demonstrate that correction of SCID mutation leads to phenotypic rescue, DNA PKcs activity was assayed in ZFN-targeted clones. This was assessed using cell survival assays or analysis of H2AX phosphorylation in response to DNA damage in gene-edited clones. DNA damage assays were conducted using the cytotoxic drug Melphalan or by subjecting the cells to γ-radiation.

3.2 Results

3.2.1 Genotyping ZFN targeted clones

An ‘inside-out’ PCR strategy amplifying the 3’ donor-genome junction was utilised to screen for HR in ZFN targeted clones. A successful HR event can be verified by PCR amplification of a region from inside the repair construct extending to the genome outside of the region of homology; cells resistant to G418 as a result of a non-homologous random integration event will not undergo PCR amplification with such primers (Abdul-Razak 2013). 14 out of 36 (38%) G418-R clones screened demonstrated gene targeting **(Fig. 3-4)**.

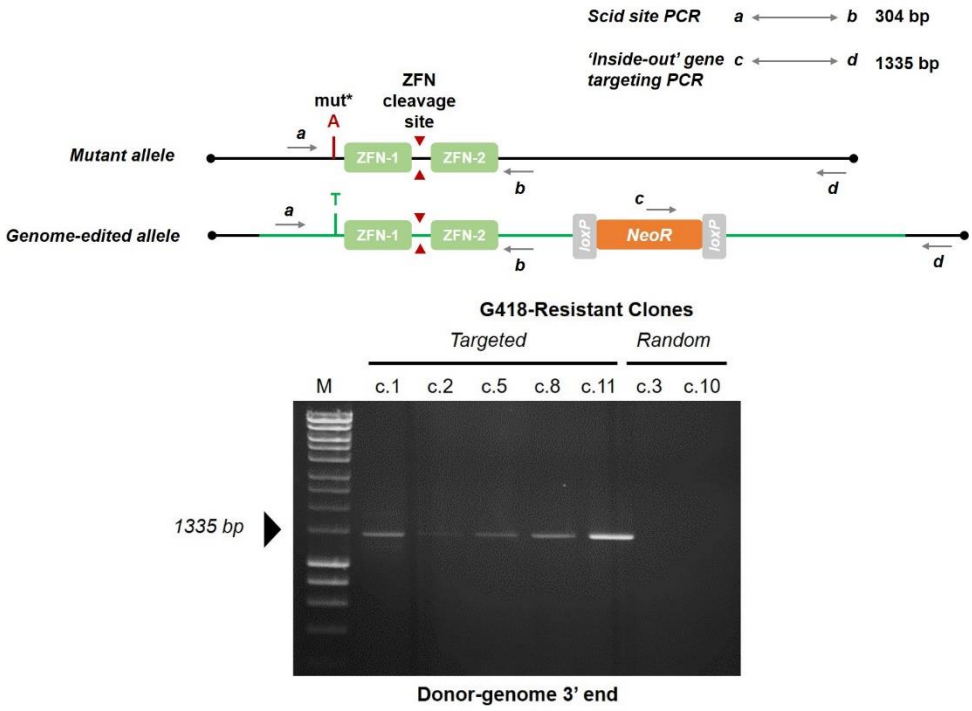


Figure 3-4 Verification of Gene Targeting in ZFN *Prkdc* Fibroblast Clones

Schematic at top represents the PCRs used to validate targeting in G418-resistant (G418-R) cells generated using gene editing with zinc finger nucleases. The PCR product obtained using primers ‘a’ and ‘b’, generating a 304 bp amplicon (represented by arrows), was used to determine genotype of the genomic region spanning the SCID mutation site and ZFN cut site. Primers ‘c’ and ‘d’ were used to amplify the donor-genome 3’ junction. Forward primer ‘c’ binds neo inside the repair template while the reverse primer ‘d’ lies outside the region of homology. Band ‘c-d’ will only be present in targeted cells and absent from cells where the repair template randomly integrated. Gel image at the bottom shows 1,335 bp PCR band in gene targeted clones. Gel, 1% agarose, 1x TAE buffer; M, Hyperladder 1 kb; ZFN, zinc finger nuclease; Neo-R, neomycin selection cassette.

The recombination event at the targeted locus should restore the correct nucleotide sequence at the mutation site. To precisely determine the genotype and whether one or both alleles had been edited, standard sequencing was performed on PCR products

amplified from genomic DNA spanning the ZFN cleavage site **(Fig. 3-4)** from the 14 targeted clones. A successful targeting event would lead to correction of at least one allele showing a double-peak, indicating mutant (A) and wild-type (T) nucleotides, at the mutation site. Majority of clones (9 out of 14) obtained from the ZFN targeted population showed presence of a double peak indicating mono-allelic correction of *Prkdc* gene at the SCID site. However, this was also followed by the presence of indels at the ZFN cleavage site. This region located in exon 85 of *Prkdc* downstream of mutation locus contained highly variable heterozygous indels. Only one clone (c.11) showed mono-allelic correction at the SCID site along with a 3-bp (TAT) insertion at the ZFN cut site, as confirmed by Sanger sequencing **(Fig. 3-5)**. The data produced by capillary electrophoresis allowed identification of sequence changes as well as detection of mixed sequences at the SCID mutation and ZFN cleavage sites. Genome modification at the mutation site is reported as SCID, heterozygous (ht) when both wild-type (wt) and SCID nucleotides are present, or wt. At ZFN cleavage site heterozygous genotype is indicated when there are indels on at least one allele **(Fig. 3-6)**.

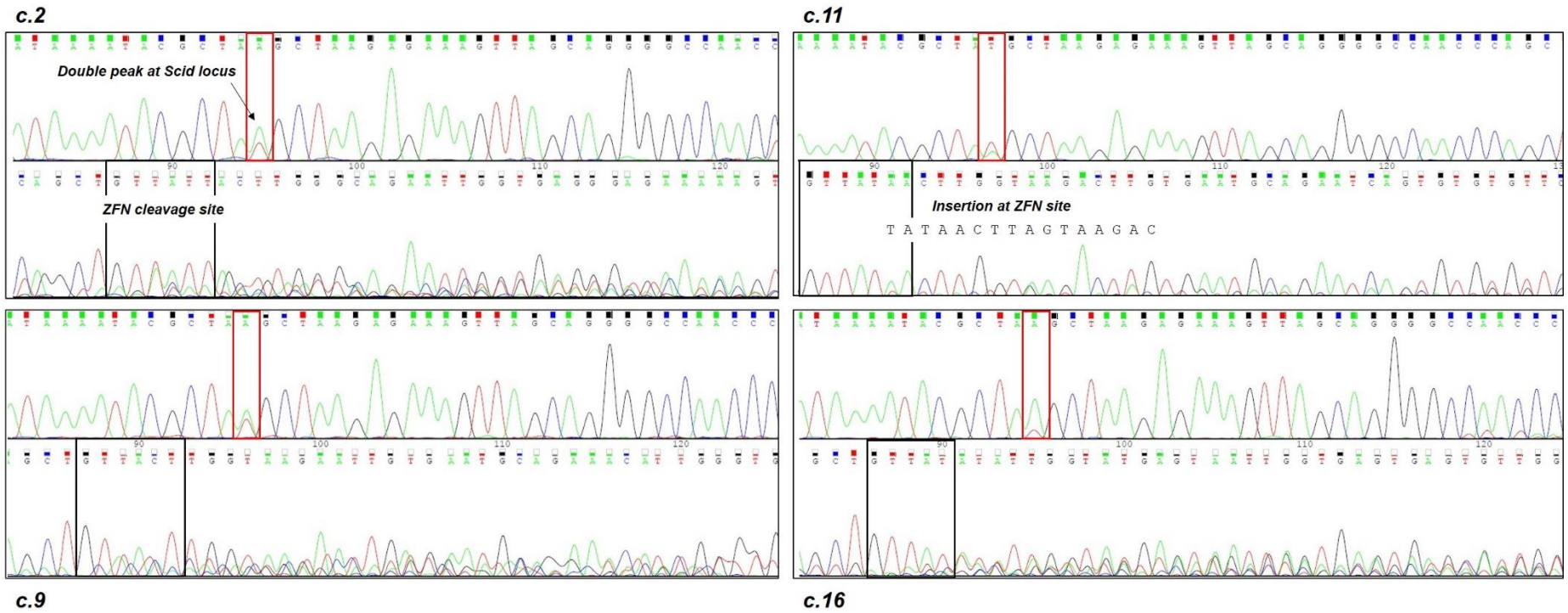
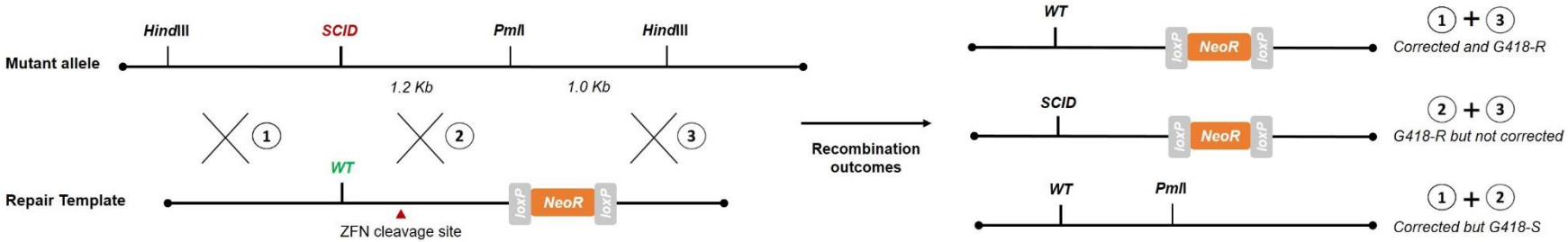


Figure 3-5 Genotyping ZFN *Prkdc* Fibroblast Clones

Chromatographs obtained from sequencing PCR products amplifying the region spanning the SCID mutation site from gene targeted clones 2, 9, 11 and 16. All four clones display mono-allelic correction at SCID locus, indicated by presence of double peaks corresponding to SCID (A) and wild-type (T) nucleotides. ZFN cut site shows heterozygous mixed sequences in all clones. Clone 11 showed a 3-bp insertion (TAT) at the ZFN site. ZFN, zinc finger nuclease.



Clonal Line #	HDR at Mutation Site		Genotype at ZFN Target Site
	Y/N	*Genotype*	
1	N	Scid	Wt
2	Y	Ht	Ht
5	Y	Ht	Ht
8	Y	Ht	Ht
11	Y	Ht	Ht - 3 Bp TAT Insertion
12	Y	Ht	Ht
23	Y	Ht	Ht
24	Y	Ht	Ht
25	N	Scid	Ht
26	Y	Ht	Ht
27	N	Scid	Ht
29	N	Scid	Ht
30	N	Scid	Ht
31	Y	Ht	Ht

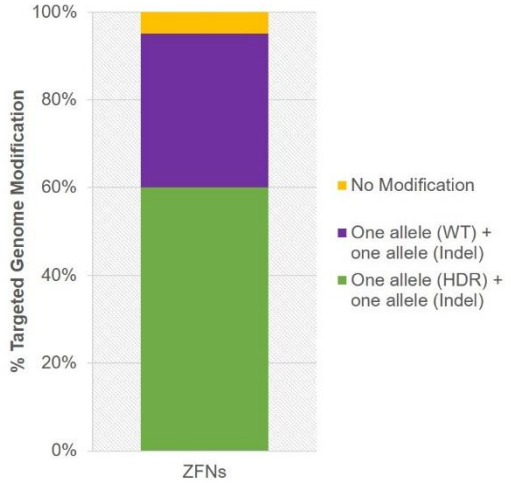


Figure 3-6 Characterization of Genome Modifications in ZFN *Prkdc* Fibroblast Clones

Top shows a schematic depicting possible outcomes of HR between *Prkdc* allele targeted with ZFNs and donor template. Event 1+3 would correct the SCID mutation and incorporate the selection cassette (NeoR) giving rise to G418-resistant populations. Event 2+3 would also incorporate Neo-R but without repair of mutation. Event 1+2 would repair the mutation but these cells would be sensitive to G418 sensitive (G418-S). Bottom panel shows gene editing events in ZFN targeted clones (n=14). Majority of clones (64%; green) showed mono-allelic correction of *Prkdc* mutation indicated by the presence of a heterozygous (ht) sequence at the SCID site. However, these also have ht indels at the ZFN site. These clones were generated because of recombination event 1+3. Only one clone (#1), shown in yellow, showed no correction of the SCID mutation followed by wild-type (wt) sequence at ZFN site. Approximately 20% (purple) clones showed no correction of SCID site while ZFN site contained indels (event 2+3). HR, homologous recombination; ZFN, zinc finger nucleases.

3.2.2 DNA-PK activity in gene edited G418-R fibroblasts

3.2.2.1 Clonogenic cell survival in response to Melphalan

Clonogenic assays examine the potential of single cells to grow and reproduce into large colonies in the presence of cytotoxic factors (Munshi, Hobbs et al. 2005, Franken, Rodermond et al. 2006) . SCID fibroblasts are sensitive to the radiomimetic drug Melphalan. To verify that successful gene editing could abrogate radiosensitivity, colony survival assays with Melphalan were conducted. Cells tested included 4 gene edited G418-R clones (c.3, c.5, c.10 and c.11) and SCID fibroblasts. The fibroblast cell line Balb/c NIH-3T3 served as positive control. All cells were plated in varying seeding densities, treated with Melphalan and allowed to grow until colonies appeared. Under standard conditions, which is in the absence of Melphalan, all clones yielded low rates of colony formation with an average of 10% of total cells inoculated generating colonies, indicating the inability of these cells to form colonies. Cells treated with Melphalan yielded further reduced plating efficiency values, with <1% cells forming colonies (not shown).

3.2.2.2 Cell survival in response to Melphalan using MTT assay

Since the fibroblasts failed to grow as colonies, results from clonogenic assays were inconclusive. As an alternative, the colorimetric MTT assay was used for studying proliferation and determination of survival of cells after cytotoxic damage. MTT assay is based on the formation of a dark-coloured formazan dye by reduction of the tetrazolium salt MTT by metabolically active cells. After some incubation, the water-insoluble formazan dye forms crystals, which are dissolved in an organic solvent and their amount is determined semi-automatically using a microplate reader (Price and McMillan 1990, Hansen and Bross 2010). Absorbance readings are related to the number of cells therefore providing the possibility to use the MTT assay as a proliferation assay to assess cell growth (Price and McMillan 1990). Relative cell proliferation or viability was determined in cells treated with or without Melphalan.

Cell viability of ZFN targeted clones was determined 4 days after treatment with 10µM Melphalan. As expected, the control wild-type fibroblasts recovered from DNA damage showing 95% viability post-treatment while SCID fibroblasts showed up to 2-fold decrease ($p \leq 0.05$, two-tailed T-test, $n=9$). However, the gene edited targeted clones (clone 5 and 11) remained sensitive to DNA damage despite the repair of mutated nucleotide **(Fig. 3-7)**.

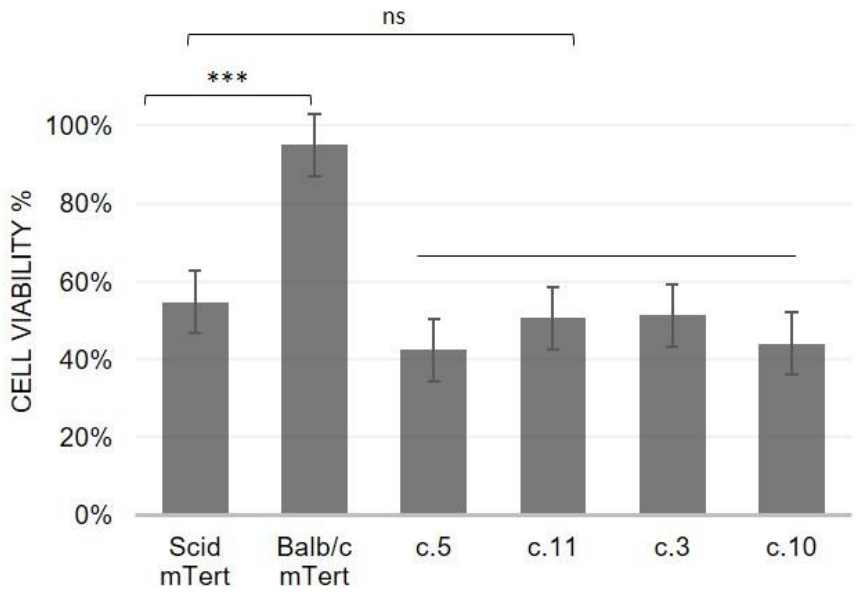


Figure 3-7 Cell Viability of G418-Resistant Clones

Cell viability of G418-resistant gene targeted (clones 5 and 11) or random integrants (clones 3 and 10) in response to 10 µM Melphalan using MTT assay. Data are represented as mean ± SD. There is a 2-fold difference between wild-type and SCID fibroblasts (*** = $p \leq 0.05$, n=9), but no significant difference between SCID fibroblasts and gene edited clones ($p \leq 0.05$, n=9). SCID mTert, immortalised SCID fibroblasts; Balb/c mTert, immortalised wild-type fibroblast from Balb/c mice.

3.2.3 Cre-LoxP recombination for excision of neo

Intronic presence of the neomycin selection marker can have unwanted effects on the expression of the gene-edited locus. To remove the *neo* selection marker from *Prkdc* locus, Cre-LoxP recombination was employed. The *Prkdc-neo* repair template was designed to contain *neo* floxed between two unidirectional *LoxP* sites. Expression of Cre recombinase would therefore lead to recombination between the two LoxP sites, excising *neo*, leaving a single 34 bp LoxP 'scar' sequence in *Prkdc* intron 85 **(Fig 3-8)**.

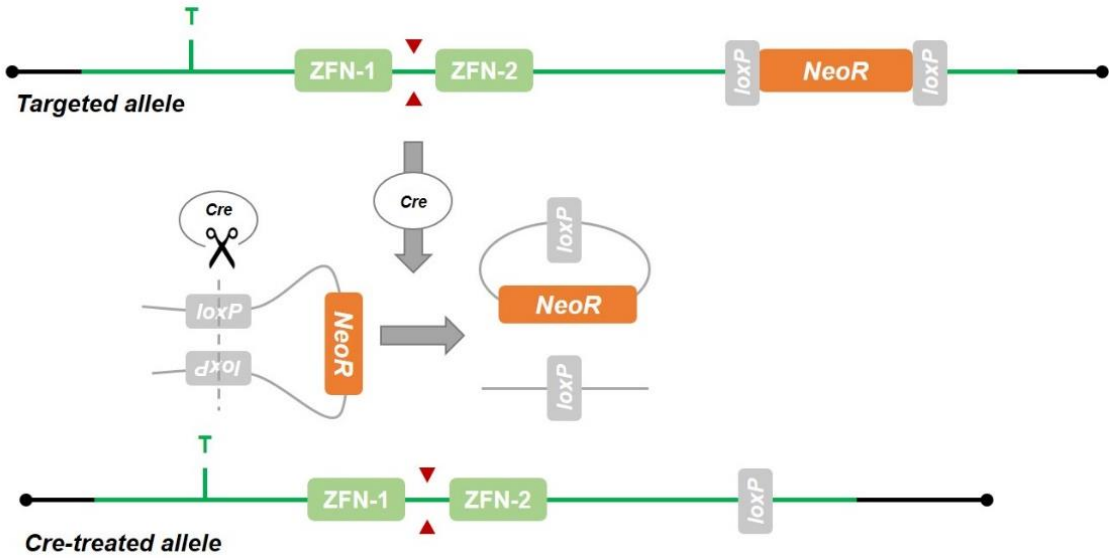


Figure 3-8 Cre-LoxP Recombination at *Prkdc* Locus

Expression of *cre* recombinase in gene-edited fibroblasts would mediate the removal of *neo* from the target locus by site-specific recombination, leaving a single *loxP* site left in intron 85 of *Prkdc*.

3.2.3.1 Inefficient removal of neo

Cre was transiently introduced in ZFN targeted clones (c.5, c.11 and c.12) by transfection of pMC-Cre plasmid using calcium phosphate co-precipitation method. A plasmid encoding *eGFP* used in parallel indicated poor transfection efficiency in SCID fibroblasts. Analysis of genomic DNA from Cre-treated polyclonal as well as monoclonal populations showed inefficient removal of neo **(Fig. 3-9).**

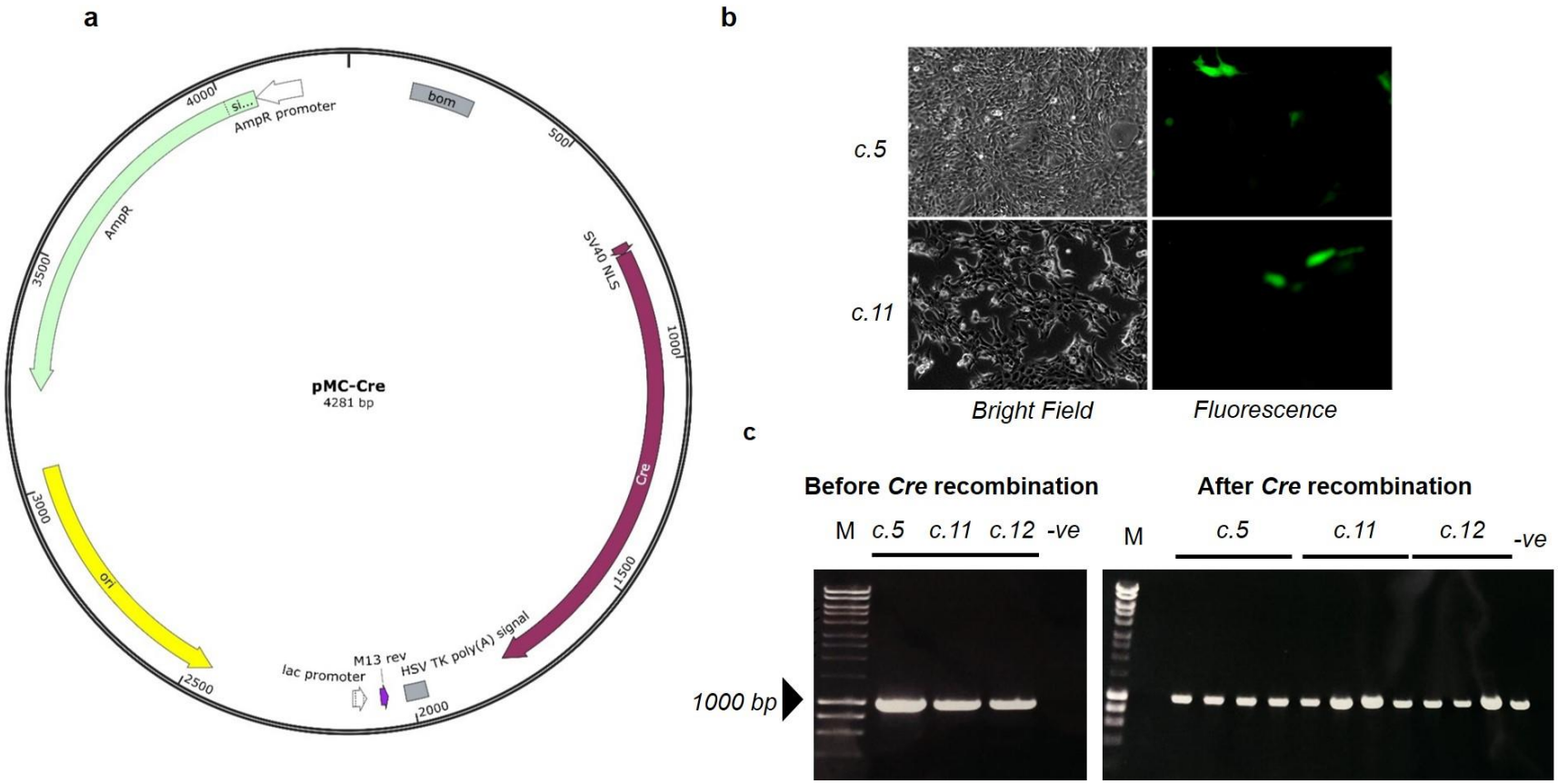


Figure 3-9 pMC-Cre Transfections to Remove Neo Selection Marker

(a) Schematic plasmid map of pMC-Cre plasmid created with SnapGene® (b) Bright field and fluorescence microscopy images of ZFN *Prkdc* fibroblast clones (c. 5 and 11) 48 hours post- transfection with plasmid encoding eGFP. (c) Shows neo positive bands in gene targeted SCID fibroblast clones (c. 5, 11 and 12) before and after transfection with pMC-Cre plasmid. Products were run on a 1% agarose gel. M, Hyperladder 1 kb, negative control (-ve), SCID fibroblasts.

3.2.3.2 Enrichment of Cre-treated cells

To address the low efficiency of *neo* removal by transfection of *cre* recombinase gene, flow cytometry sorting was used. Cre was transiently introduced in cells from ZFN *Prkdc* fibroblast gene targeted clones (c.5 and c.10) and random integrants (c.3 and c.11) by transfecting a plasmid encoding *eGFP-Cre* fusion gene (pCAG-GFP:Cre) **(Fig. 3-10)**. Post-transfection cells expressing GFP were enriched using flow cytometry, expanded in culture and screened for the loss of *neo*.

Three different PCRs were designed to screen for the removal of the *neo* cassette from the three possible *Prkdc* alleles present in the ZFN *Prkdc* fibroblast clones. These are illustrated in **(Fig. 3-11*a*)**. Treatment with Cre led to removal of *neo* in all four clones, leaving an intronic 34 bp LoxP 'scar' site **(Fig. 3-11*b*)**. To isolate cells where *neo* was removed and the SCID mutation was corrected, the Cre-treated clones 5 and 11 were subjected to flow cytometry sorting for *eGFP* expression, followed by limiting dilution, and 20 further clones were generated. PCR screening in these cells showed efficient removal of neo and presence of a single intronic LoxP site at the Prkdc locus **(Fig. 3-11*c*).** As expected, following the removal of neomycin these clones became sensitive to G418 (G418-S) showing cell death when grown in presence of G418 **(Fig. 3-12)**.

Downstream sequence analysis of genomic DNA using Sanger sequencing and Indel Detection by Amplicon Analysis (IDAA) from G418-S clones displayed the same pattern as the parent clones – mono-allelic correction of the mutated nucleotide along with Indels (3-bp insertion) at the ZFN cut site **(Fig. 3-13)**. An exception to this was clone 11.2, where the ZFN target sequence remained unaffected while SCID mutation was corrected on one allele **(Fig. 3-14)**

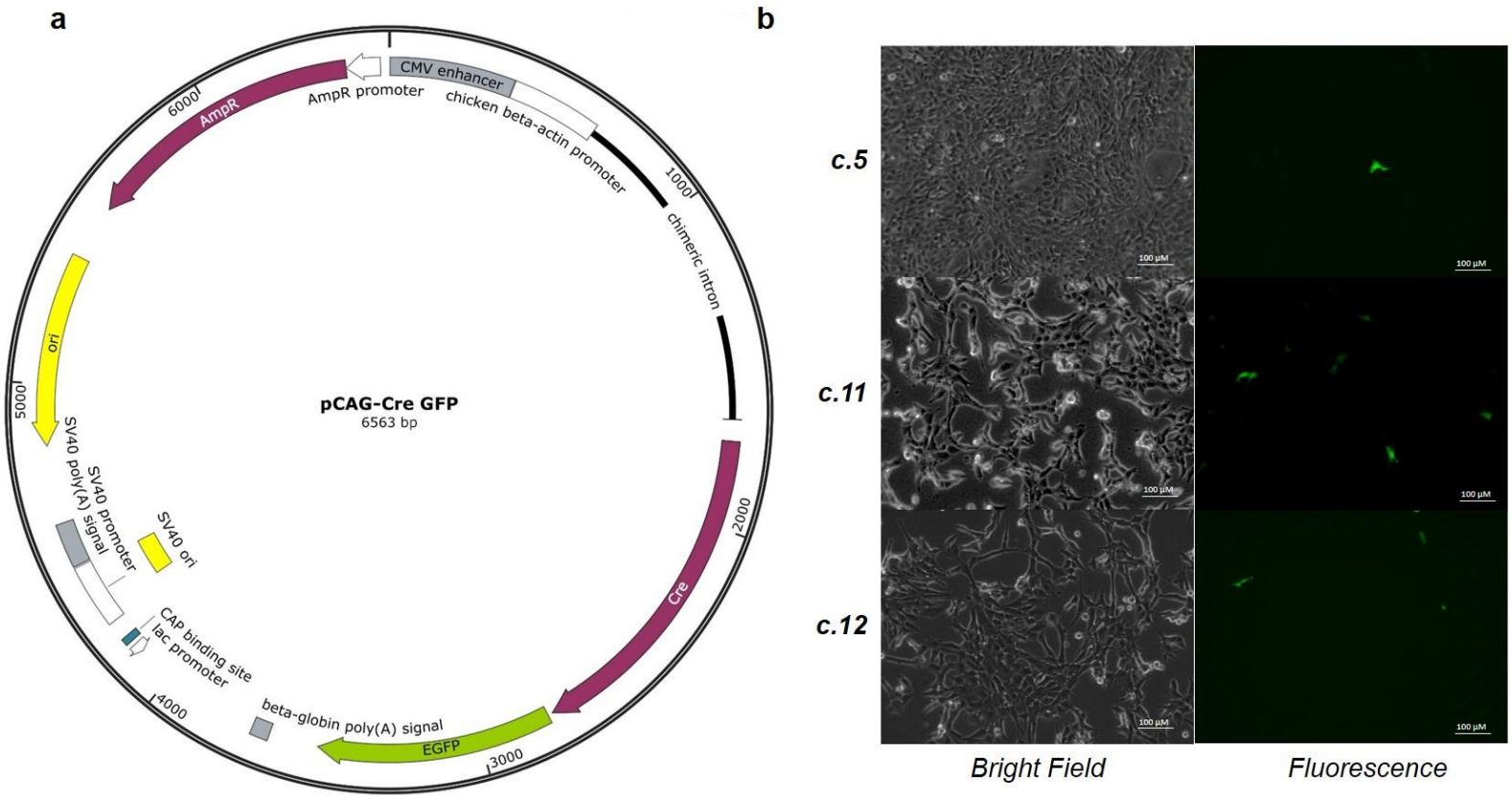


Figure 3-10 pCAG-eGFP:Cre Transfections to Remove Neo Selection Marker

(a) Schematic plasmid map of pCAG-eGFP:Cre created with SnapGene® (b) Bright field and fluorescence microscopy images of ZFN *Prkdc* fibroblast clones (c. 5, c.11 and 12) 48 hours post-transfection with pCAG-eGFP:Cre plasmid.

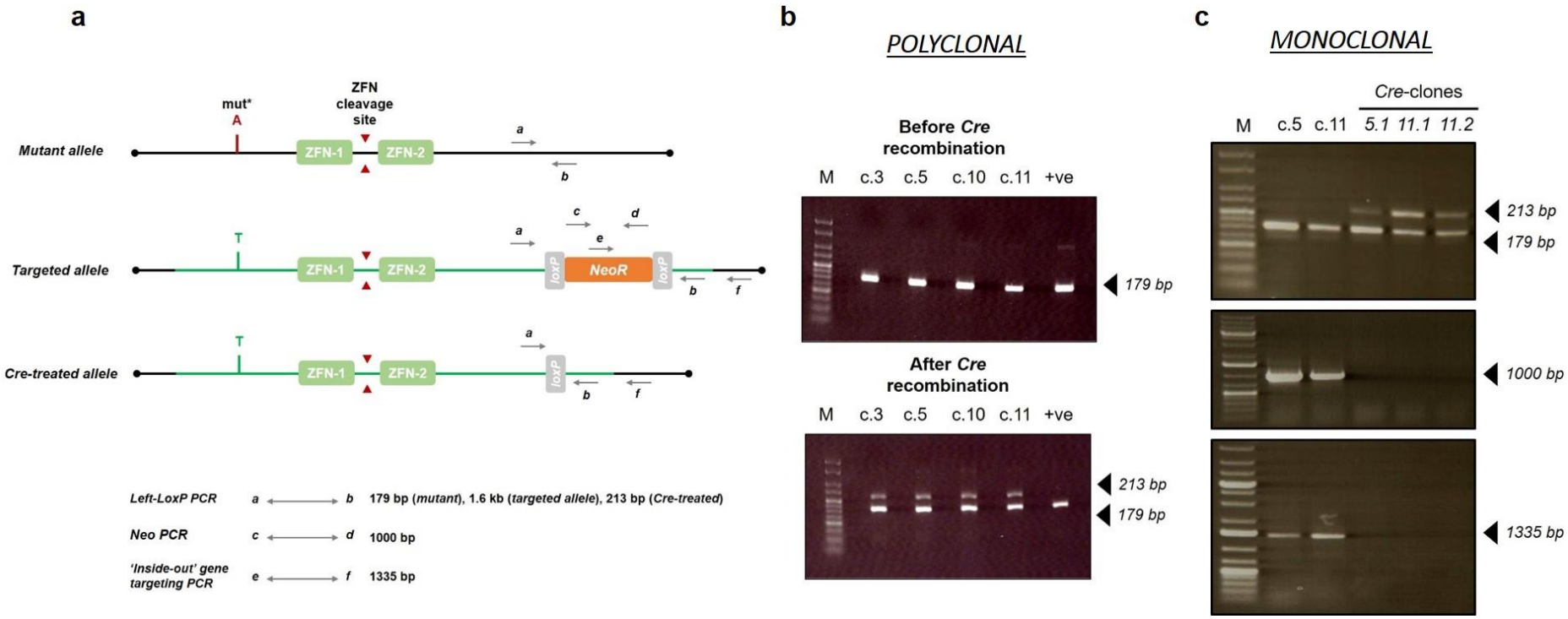


Figure 3-11 Cre-LoxP Mediated Removal of Neomycin Selection Marker

(a) Schematic of PCRs used to screen for the presence of neomycin selection marker cassette indicating primer sites and amplicon sizes. Arrows represent PCR primers and their direction. (b) PCR products ran on 3% agarose gel from ZFN *Prkdc* fibroblasts clones (gene targeted clones 5 and 11; random integrant clones 3 and 10) before and after Cre-LoxP recombination indicating loss of neomycin cassette. (c) PCR products ran on 1% agarose gel from clones derived from Cre-treated ZFN *Prkdc* clone 5 (clones 5.1, 5.11 and 5.12). M, Hyperladder 1 kb, Neo-R, neomycin-resistance selection marker cassette.

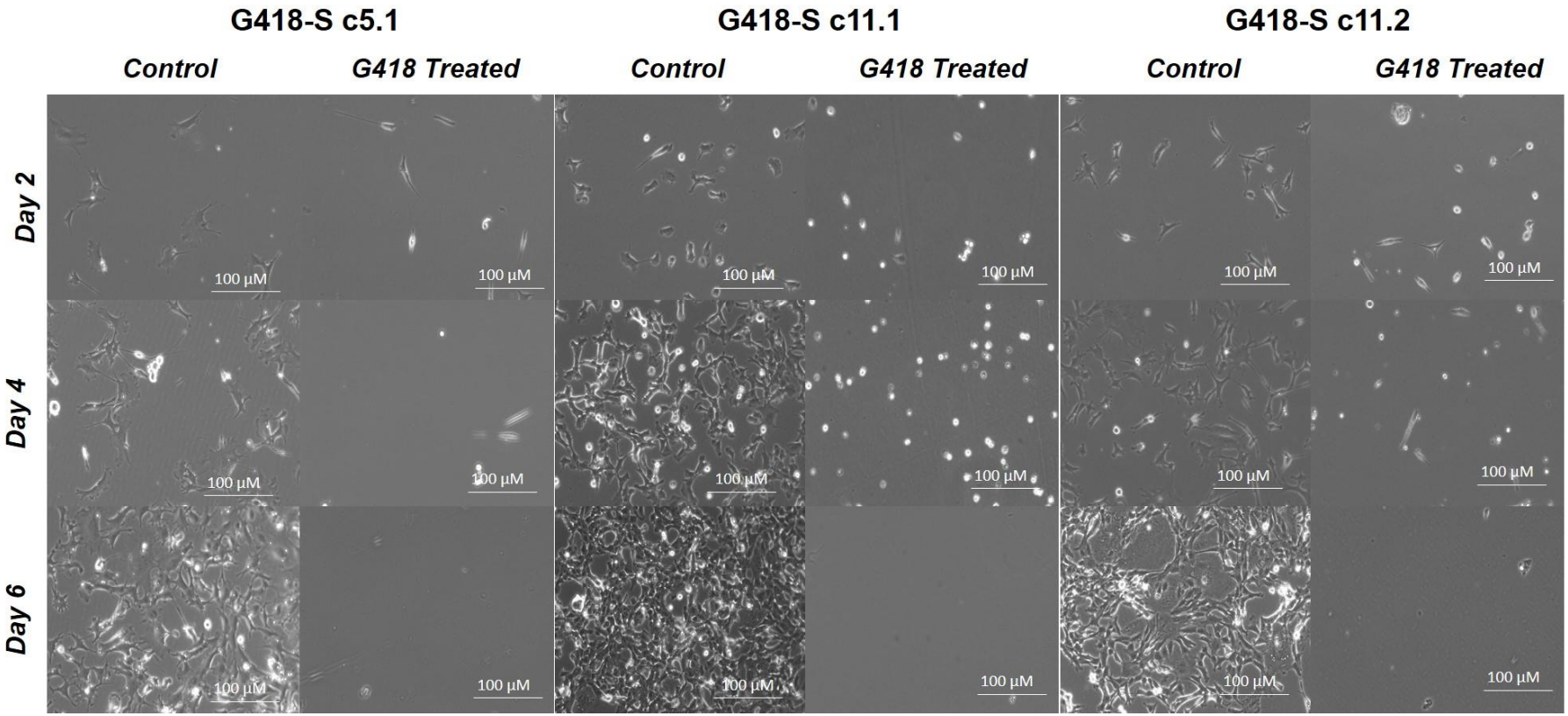


Figure 3-12 G418 Treatment in Cre-Treated *Prkdc* Fibroblasts

Microscopy images from control and Cre-treated ZFN *Prkdc* fibroblast clones (c.5.1, 11.1 and 11.2) cultured in presence of antibiotic G418 (800 µg/ml). Cre-treated clones showed increased sensitivity to G418 showing total cell death by day 6.

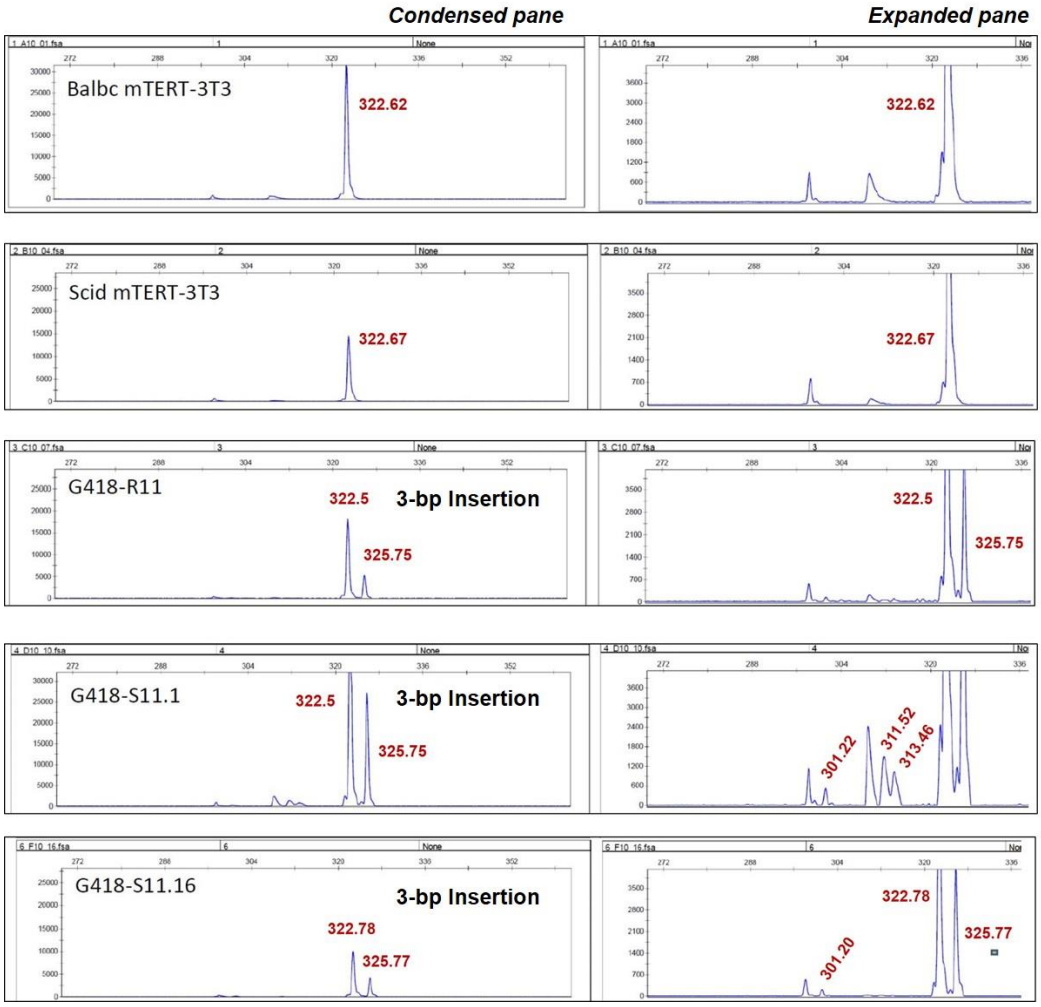


Figure 3-13 Indel Detection Amplicon Analysis (IDAA) in ZFN *Prkdc* Fibroblasts

Fragment analysis showing fluorescently labelled amplicons containing the indels. Peaks represent amplicon size in base pairs (denoted in red) based on fluorescent intensity. SCID and wild-type Balb/c fibroblast show one peak representing a single amplicon while gene targeted fibroblast clones show multiple peaks indicating presence of a mixture of alleles. Cells derived from gene targeted clone 11 show a second allele with a 3-bp insertion before (G418-R11, middle panel) and after Cre recombination (G418-S clones). Indels, insertions or deletions.

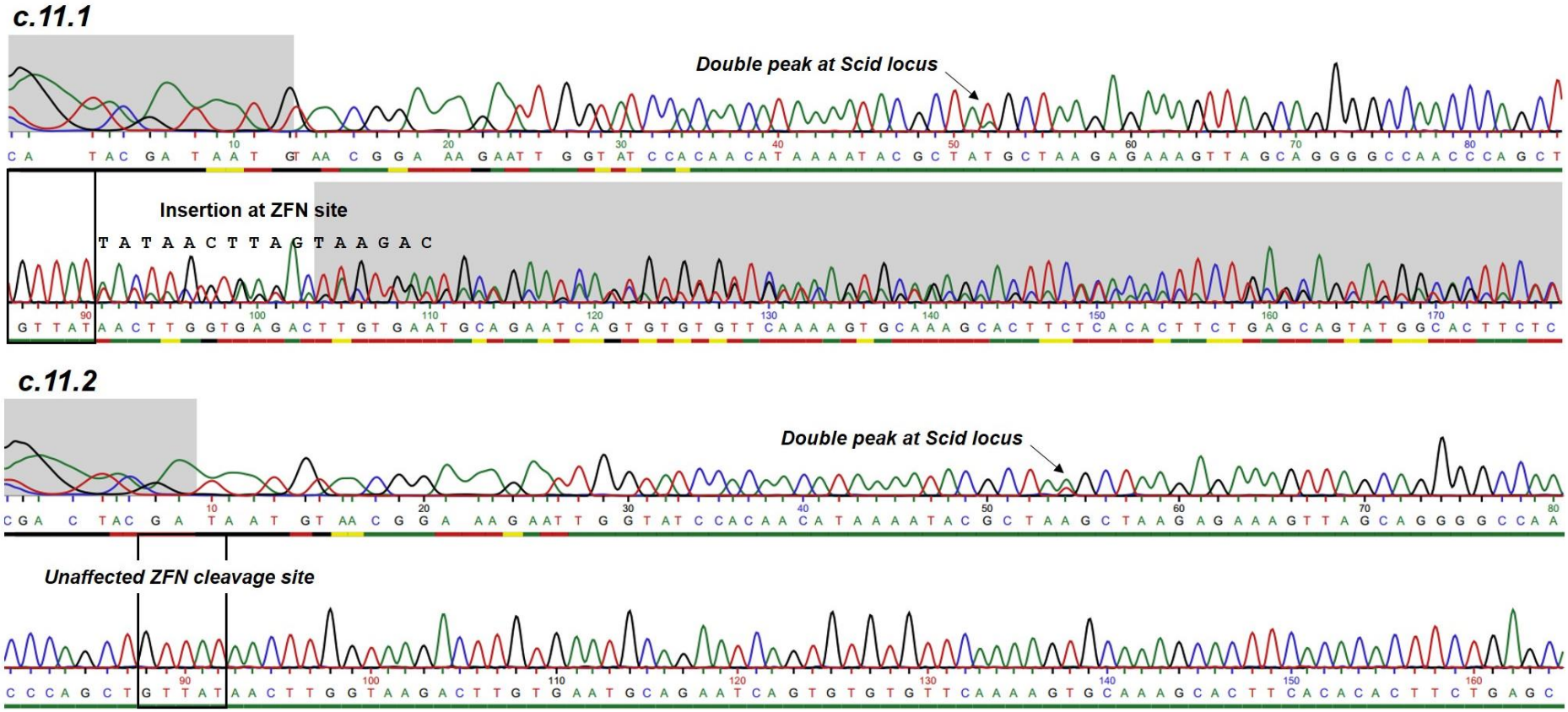


Figure 3-14 Genotype Analyses of G418-S Clones by Sanger Sequencing

Chromatographs obtained from Sanger sequencing the region around *Prkdc* ZFN target site in Cre-treated ZFN *Prkdc* fibroblast clones. Both clones show mono-allelic correction of SCID mutation indicated by presence of double peaks at SCID site. Additionally, clone 11.1 show a 3-bp insertion at the ZFN target site while in clone 11.2, ZFN site is not affected.

3.2.4 DNA-PK activity in gene edited G418-S fibroblasts

3.2.4.1 Cell survival in response to Melphalan

G418-S clones (c.5.1, c.11.1 and c.11.2), SCID mTert fibroblasts, and wild-type Balb/c mTert fibroblasts were subjected to 10 µM Melphalan and cell viability was determined using MTT assay. As expected, treatment with Melphalan drastically reduced cell viability in DNA-PKcs deficient SCID fibroblasts by up to 30% ($p<=0.05$ two-tailed T-test; n=9) **(Fig. 3-15)**. In comparison, the gene edited clones showed no significant difference in cell viability upon cytotoxic treatment. This suggests that the cells remained sensitive to DNA damage despite the repair of mutated nucleotide as well as removal of *neo*.

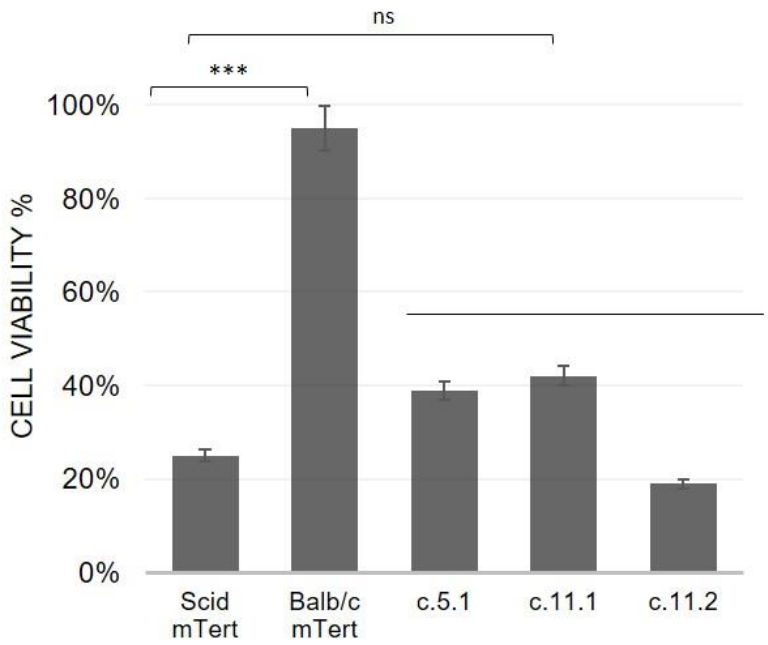


Figure 3-15 Cell Viability of G418-S Clones

Cell viability of G418-sensitive (Cre-treated) ZFN *Prkdc* fibroblast clones 5.1, 11.1 and 11.2 in response to 10 µM Melphalan using MTT assay. Data are represented as mean ± SD. 4-fold difference between wild-type and SCID fibroblasts. No significant difference between SCID fibroblasts and gene edited clones ($p\leq0.05$, n=9). SCID mTert, immortalised SCID fibroblasts; Balb/c mTert, immortalised wild-type fibroblast from Balb/c mice.

3.2.4.2 Clonogenic cell survival in response to γ-radiation

SCID fibroblasts are sensitive to γ-radiation showing a dose-dependent decrease in cell survival as compared to wild-type fibroblasts with 3 Gγ gamma irradiation being the optimal dose to kill SCID cells. Gene editing should rescue this phenotype however this was not observed, with both G418-R c.11 and G418-S c.11.2 displaying decreased cell survival in response γ-radiation (n=2) **(Fig. 3-16)**.

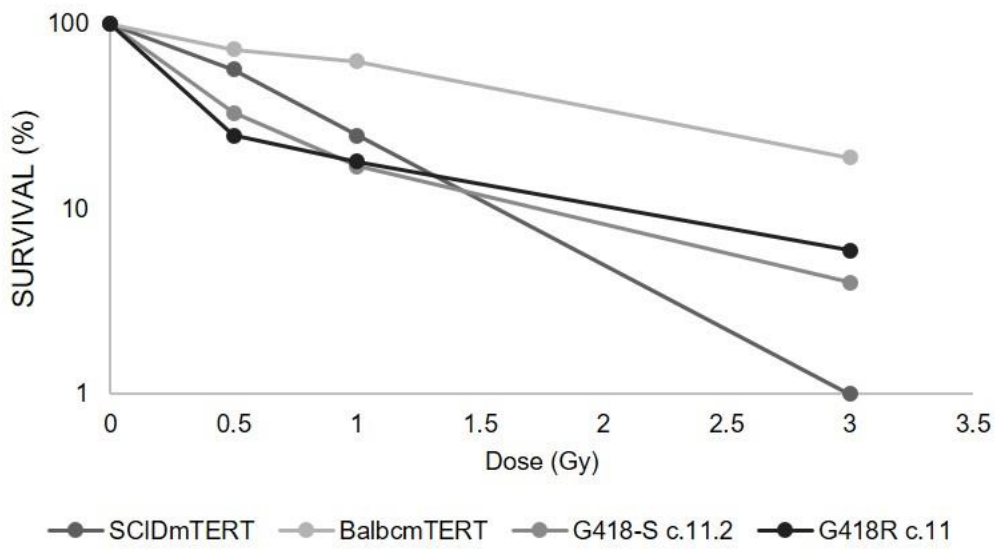


Figure 3-16 Clonogenic Cell Survival in Response to γ-Radiation Following *neo* Removal

Fibroblasts were subjected to increasing dosage of γ-radiation (3 Gγ being the maximal dose to kill *Prkdc* fibroblasts). Cell survival was quantified by counting colonies relative to untreated samples. Data are represented as mean (n=2). SCID mTert, immortalised SCID fibroblasts; Balb/cmTert, wild-type immortalised fibroblasts; G418-S c.11.2, targeted SCID fibroblast clone with *neo* removed; G418R c.11, targeted SCID fibroblast clone with *neo.*

3.2.4.3 H2AX phosphorylation

When DNA DSBs are induced by ionizing radiation in cells, histone H2AX is quickly phosphorylated into gamma-H2AX around the DSB site. DNA-PKcs regulates the phosphorylation of H2AX and cell cycle progression in response to DNA damage. To confirm that repair of mutated SCID nucleotide restored DNA-PKcs activity, cells from G418-S clones (c.5.1, c.11.1, c.11.2, c.11.16, and c.11.18), SCID mTert fibroblasts, and wild-type Balb/c mTert fibroblasts were treated with γ-radiation. Under these experimental conditions, H2AX was phosphorylated. The level of gamma H2AX foci in all cells increased rapidly with a peak level at 0.25 - 1.0 h after 3 Gγ gamma irradiation (n=3) **(Fig. 3-17)**. With ongoing DNA repair, the number of H2AX foci rapidly decreased in wild-type fibroblasts within 24 hours. SCID fibroblasts, deficient in DNA PKcs, showed slower decline in the number of H2AX foci. Compared to SCID fibroblasts, gene edited clones showed decrease in H2AX foci, however this was not significant.

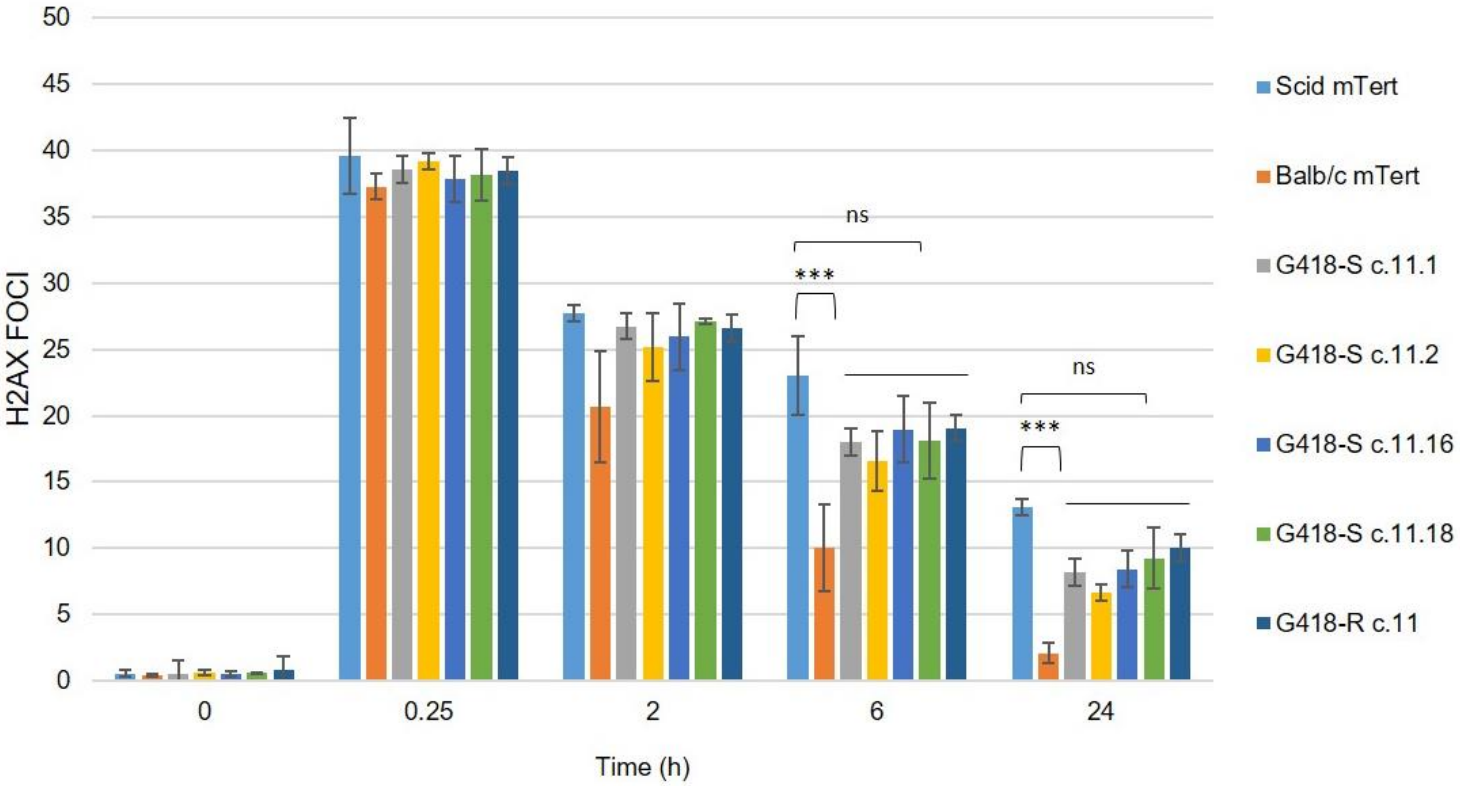


Figure 3-17 H2AX Phosphorylation in Response to γ-Radiation

Number of H2AX foci formed at 0, 0.25, 2, 6, and 24 hour intervals post γ-radiation (3 Gγ) in ZFN *Prkdc* fibroblasts. Data are represented as mean ± SD (n=3). SCID mTert, immortalised SCID fibroblasts; Balb/c mTert, wild-type immortalised fibroblasts; G418-R c.11, targeted *Prkdc* fibroblast clone with *neo*; G418-S c.11.1, 11.2,11.16 and 11.18, targeted SCID fibroblast clones with *neo* removed.

3.3 DISCUSSION

This chapter described experiments that were undertaken to obtain formal proof of gene editing and repair of the *Prkdc* SCID mutation in murine SCID fibroblasts. Use of zinc finger nucleases and a homologous repair template previously reported successful gene targeting and phenotypic rescue of DNA PKcs activity in a polyclonal population of SCID fibroblasts (Abdul-Razak 2013). However, the clonal lines generated from this polyclonal repertoire were not studied, which is an important step of the genome editing workflow.

The fibroblast clones were previously generated using positive drug (G418) selection. Any cell with a copy of the repair template would therefore show resistance to the drug. To rule out clones that contained randomly integrated copies of the repair template, I used a PCR specific for determining gene targeting at the Prkdc locus. This showed that up to 40% of the clones were targeted at the *Prkdc* locus using homologous recombination. These levels were line in with a similar study in SCID mouse fibroblasts where ZFNs were used to target the *Prkdc*-SCID mutation (Rahman, Kuehle et al. 2015) It is worthwhile to note here that SCID fibroblasts may have a bias towards HDR. As seen in Fanconi anaemia, which is also a DNA repair disorder, in the absence of NHEJ, cells preferentially employ the HDR arm of DNA repair, especially when a homologous donor template is present (Rio, Banos et al. 2014).

The gene targeting however did not correlate with functional rescue of DNA-PKcs as determined by a preliminary *in vitro* enzymatic assay that was previously carried out in the parent polyclonal populations (not shown) and the poor cell survival of fibroblasts observed upon cytotoxic DNA damage. It is noteworthy to mention primary fibroblasts do not form colonies in the absence of a feeder layer, especially upon cytotoxic treatment (Dr Lisa Woodbine, *personal communication*).

These setbacks led to further genotyping of the targeted clones using Sanger sequencing and IDAA. Sequence analysis showed that while gene targeting led to the repair of mutated SCID nucleotide in the majority of clonal lines tested, additional heterozygous sequence alterations were introduced at the ZFN target site. There was only one cell clone (c.11) that showed an in-frame 3-bp insertion in the coding sequence, which would lead to addition of an isoleucine residue in the protein sequence. These sequence alterations could explain the loss in DNA PKcs activity observed. Since no targeted clone without an unaffected sequence at the ZFN cut site was identified, further experiments were carried out using select few cell lines. Clone 11 that showed an in-frame change in its genomic sequence was analysed as the primary candidate.

Before carrying out further phenotypic experiments, I sought to remove the selection marker present in the repaired locus. This was done because the intronic presence of a neo selection marker cassette has been reported to perturb gene function previously (Pham, MacIvor et al. 1996). The HDR strategy was designed with this in consideration and hence *neo* was introduced floxed between two unidirectional LoxP sites. To remove *neo*, Cre recombinase was transiently expressed in a select few targeted clonal cell lines using plasmid transfection. Unfortunately, primary SCID fibroblasts were found extremely challenging to transfect using a variety of lipid-based transfection reagents (data not shown) To bypass this, a new strategy based on selective enrichment of transfected cells was developed. Cells were transfected with a new plasmid which encoded *Cre* fused to *GFP* under the regulation of a stronger, CAG, promoter. Fluorescence was used as a measure of transfection. Since GFP/Cre was not a true fusion protein, being separated by a T2A polypeptide, low transfection rates observed were unlikely due to GFP quenching The small population of transfected cells were then enriched using flow cytometry sorting based on expression of *eGFP*. Use of the new plasmid and cell sorting was effective at isolating transfected cells and led to the generation of *neo* free, G418-sensitive cells containing a single *loxP* scar sequence in the downstream intron. Removal of *neo* was confirmed by genotyping the cell sorted clones as well as by subjecting them to G418 drug treatment for 8 to 10 days, demonstrating the removal of the marker at the genomic and phenotypic levels.

Subsequently, these clonal populations were genotyped to characterize the sequence around the ZFN target site. G418-sensitive cells derived from clones 5, 11 and 12 retained the same signature deletion of the coding region. All cells derived from clone 11 showed the 3-bp insertion except clone 11.2, which interestingly showed wild-type sequence at the ZFN site along with repair of the SCID mutation. This suggested that perhaps clone 11 was not a completely homogenous cell population to begin with. However, identification of a cell clone with the desired edit – correction of SCID nucleotide and unaffected ZFN target site – was encouraging.

Following genotype analysis, G418-sensitive cells were assayed for the rescue of DNA PKcs activity using two different techniques – cell survival assays and H2AX phosphorylation. The first method looked at cell survival in response to damage by cytotoxic drug Melphalan and by γ-radiation. Cell survival assays not only demonstrate DNA damage repair but also indicate the potential of cells to recover from the damage and grow as per normal (Munshi, Hobbs et al. 2005, Franken, Rodermond et al. 2006). Initial colony forming assays demonstrated the poor colony forming abilities of scid fibroblasts. To facilitate the formation of colonies, the cells were plated on a feeder layer

of primary embryonic fibroblasts, upon DNA damage. This led to formation of intact colonies which were easier to score than in the previous attempt.

The second technique used was based on direct detection of DSB damage. The formation of DSBs triggers activation of many factors, including phosphorylation of the histone variant H2AX, producing gamma-H2AX. Phosphorylation of H2AX plays a key role in DNA damage response and is required for the assembly of DNA repair proteins at the sites containing damaged chromatin as well as for activation of checkpoints proteins which arrest the cell cycle progression (An, Huang et al. 2010, Podhorecka, Skladanowski et al. 2010). Unfortunately, neither of the assays used showed significant rescue of DNA PKcs enzymatic activity in G418-sensitive clonal cell lines in comparison to wild-type control cells. This was consistent even in clone 11.2, which had shown the genomic profile of a corrected clone.

It is not clear why clone 11.2 failed to show rescue of DNA-PK activity when it has the expected genomic structure from a repair event: mono-allelic correction of SCID mutation, and wild-type sequence at the ZFN target site. As we have not sequenced the full *Prkdc* gene or analysed epigenetic marks in this clone, it is possible that *de novo* mutations or epigenetic changes introduced during genetic manipulation or the extended period in culture are responsible for the inability to display evidence of normal DNA-PK activity.

In any case, given that only one clonal population carried the desired edit highlights the potential problems in genome editing procedures caused by intrinsic DNA repair defects in *Prkdc* SCID fibroblasts. NHEJ is the primary DNA damage repair pathway in vertebrates. Although it is known to be error-prone introducing insertions and deletions at the repair site, NHEJ can perfectly join the two broken ends of DNA in some cases. Studies have shown that NHEJ is not as error-prone as perceived; in fact, in the absence of NHEJ, cells utilise an alternative and highly mutagenic repair pathway, to process the damaged DNA ends (Betermier, Bertrand et al. 2014). This pathway is known as alternative (alt-NHEJ) or microhomology-mediated end-joining (MMEJ) (Wang, Perrault et al. 2003, Wang and Xu 2017). Since SCID fibroblasts are defective in NHEJ, the DSB introduced by ZFNs could have been inaccurately repaired by alt-NHEJ causing the sequence alterations observed. This however is only speculative and would require definitive and quantitative assessment. A possible method to overcome this limitation would be to carry out a larger screen of edited cells. Out of the 30 clones initially tested in this study, only one sub-clone with the desired edit was identified. Furthermore, since the ZFN target site was present in the repair template, the template could have been re-

cut by the nuclease even after a correct homologous recombination effect. To avoid this, a repair template lacking the nuclease cut site should be utilised. A simple method of doing that is introducing a silent mutation at the cut site in the homologous template (Hollywood, Lee et al. 2016).

Given the nature of the DNA repair defect in SCID cell lines, the results obtained here were disappointing but not completely unexpected. Nonetheless this study highlighted that genome editing can be carried out in cells defective in DNA damage repair. In fact, a high proportion of cells targeted showed mono-allelic repair of the genetic mutation.

Lastly this study demonstrated the necessity of laborious and exhaustive genome editing technology workflow. It is crucial to thoroughly analyse genomic modification at a clonal level as a single unidentified genetic change can lead to deleterious effects on protein expression and function, and potential artefacts in downstream experiments.

Building up on the conclusions from this Chapter, we decided to progress this work using the CRISPR-Cas9 technology. CRISPR-Cas9 system is a robust genome editing platform that is increasingly being used across various disease models. However, it has still not been used in the context of *Prkdc*-SCID. We decided to switch to this system in order to generate new clonal populations and interrogate the questions that remain unanswered here. CRISR-Cas9 based gene editing of *Prkdc*-SCID is described in the next Chapter.

4 Gene Editing in *Prkdc*-SCID Using CRISPR-Cas9

4.1 Introduction

The CRISPR-Cas9 genome editing platform is a versatile and efficient technology that enables a variety of gene targeting strategies (Cox, Platt et al. 2015). Building upon the work using engineered ZFNs described in the previous Chapter, this section will investigate the use of CRISPR-Cas9 technology to target and correct the murine *Prkdc* SCID mutation in cultured *Prkdc* SCID fibroblasts. CRISPR-Cas9 technology has shown progress in efficient genome editing of many genetic diseases with ever growing applications. However, the system has not been employed in the context of *Prkdc*-SCID so far. This study would thus provide novel proof-of-concept information on the targeting and recombination at the *Prkdc* locus. Moreover, since Cas9 nuclease and ZFNs create different double-strand breaks cuts in the DNA – Cas9 generates a blunt end cut while ZFNs induce a staggered cut – the recombination output could vary. The previous Chapter aimed to rescue DNA PKcs enzymatic activity due to the repair of underlying mutation; however, this was met with challenges pertaining to mutations being introduced at the nuclease target site in all clones tested. Therefore, the main aims of this section were to develop CRISPR-Cas9 gene editing reagents to target *Prkdc,* correct the SCID point mutation using a homologous recombination strategy, and generate single-cell clones for downstream analysis.

The CRISPR-Cas9 system comprises of two components - a nuclease (Cas9) and a short RNA sequence (guide RNA or gRNA), which directs Cas9, through its complementary bases, to a specific DNA sequence which then mediates a double-stranded break (Jinek, Chylinski et al. 2012). Three independent gRNAs were designed to target the *Prkdc* SCID mutation – two in exon 85 close to the previously investigated ZFN site described in *Chapter 3* and one in the downstream intron 85. These guides were then separately used in conjunction with the Cas9 nuclease and the previously described *Prkdc-neo* donor template for homology-directed repair of the SCID point mutation **(Fig. 4-1)**.

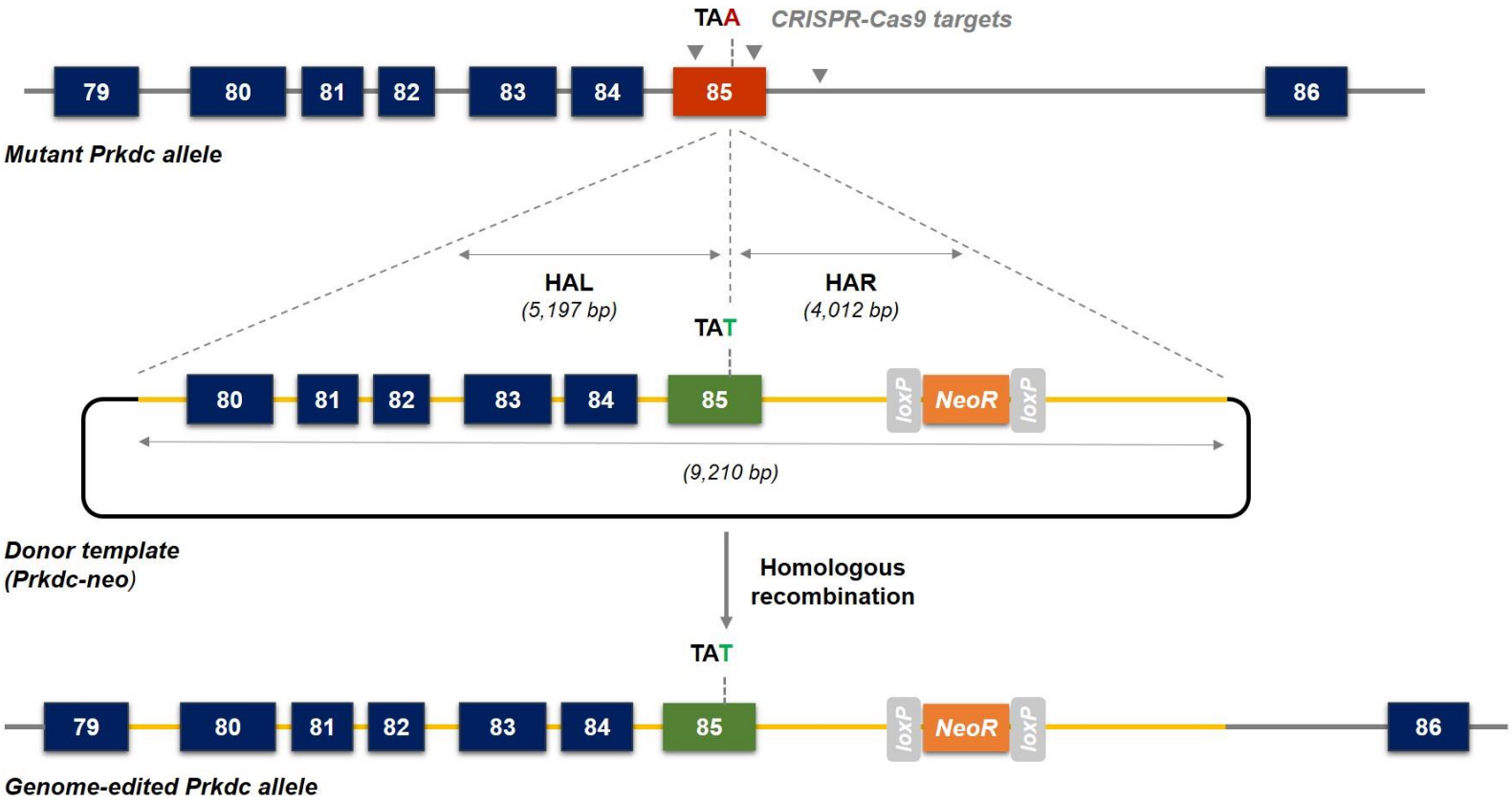


Figure 4-1 Homology-Directed Repair of *Prkdc* SCID Mutation Using CRISPR-Cas9

Prkdc SCID is caused by a truncating mutation (T to A transversion) in exon 85 of *Prkdc* gene. A homology-directed repair strategy using CRISPR-Cas9 was designed to repair the mutated nucleotide. Three different gRNAs were designed to introduce a double stranded break close to the SCID mutation. Wild-type *Prkdc* sequence containing correct nucleotide was supplied on a plasmid donor template flanked between homology arms left (HAL) and right (HAR). As a selection marker, neomycin gene floxed between two unidirectional loxP sites (NeoR) was included in the right arm of homology. Homologous recombination of the donor template would lead to correction of mutated nucleotide and incorporation of floxed Neo-R conferring selective advantage to gene targeted cells.

Separately, a modified *Prkdc-neo* donor DNA template was also investigated, where Cas9-recognition sites (known as protospacer adjacent motifs or PAM regions) were altered using site-directed mutagenesis. Having intact CRISPR PAM targets in the homologous donor template can cause degradation of donor plasmids by Cas9 activity in cells or modification of genomic sequence upon recombinant of the template (Hollywood, Lee et al. 2016). This can be detrimental especially if the change affects the coding region of a gene. A possible way to avoid this is to mutate the PAM regions into silent mutations within the donor template (Hollywood, Lee et al. 2016).

To carry out gene targeting, *Prkdc* gRNAs were cloned in spCas9 expression plasmids. These were then transfected into immortalized SCID fibroblasts along with donor DNA plasmids containing *neo* selection marker. G418 antibiotic selection was used to pool targeted cells, leading to the generation of a polyclonal G418-resistant (G418-R) cell population. To determine the efficiency of gene editing, single-cell clones were derived from the diversified G418-R population**.** The clones were genotyped for successful recombination events using standard PCRs and sequencing. The overall gene targeting approach is illustrated in **(Fig. 4-2).**

The previous chapter showed a high rate of sequence alterations at the nuclease cut site in ZFN targeted cells suggesting a lack of DNA damage repair in SCID fibroblasts. To test this, editing events at *Prkdc* locus will be compared in SCID and Balb/c wild-type fibroblasts. For this, a *Prkdc* guide RNA will be developed into a third-generation lentiviral vector and transduced into cells. Downstream editing events will be validated via standard PCRs and bioinformatics tools.

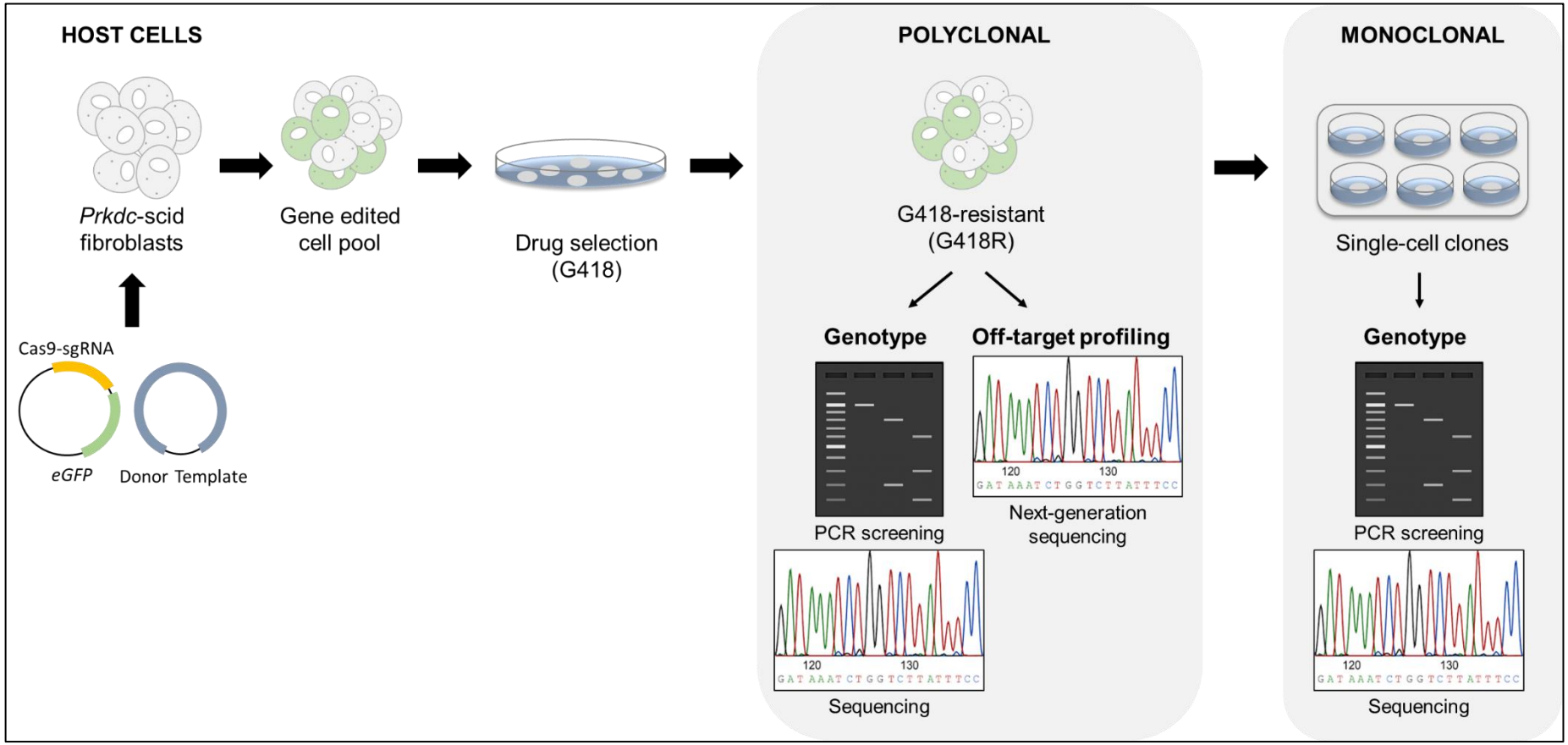


Figure 4-2 Experimental Design of CRISPR-Cas9 *Prkdc* Gene Editing

Gene editing reagents Cas9 and gRNA, and a donor repair template were introduced into cultured SCID fibroblasts on plasmid backbones using calcium phosphate transfection. Drug G418 selection was used to select targeted fibroblasts and generate a G418-resistant (G418-R) cell pool. PCR and Sanger sequencing were used to characterise polyclonal population. Off-target profiling to be carried out using next-generation sequencing of predicted off-target genomic sites. Consecutively, monoclonal populations were derived using dilution cloning and genotyped

4.2 Results

4.2.1 *In-silico* design of *Prkdc* guide RNAs

Single spCas9 guide RNAs (gRNA-1 and -2; 20 nucleotides each) targeting *Prkdc* exon 85 were designed using the web tool CHOPCHOP v1. The guides were designed to match a 20-nucleotide target in the input DNA sequence from *Prkdc* exon 85 adjacent to a 'NGG' PAM specific for spCas9 nuclease. For target selection, guides were designed with cut sites within 100 bp of the junction of the homology within exon 85.

gRNA sequence targeting *Prkdc* intron 85 (gRNA-3; 20-nt) was designed using Cas-Designer. It is a web-based bioinformatics tool which scores Cas9 targets in each input sequence with potential off-target numbers with 2 nucleotide mismatches and out-of-frame scores via microhomology-predictor. gRNA sequences, PAM and distance from SCID mutation site are mentioned in **Table 4-1** and **(Fig. 4-3)**.

Table 4-1 *Prkdc*-SCID gRNAs

gRNA	gRNA Sequence (5' - 3')	SpCas9 PAM	Location of Cut Site
gRNA-1	AGCTAAGAGAAAGTTAGCAG	GGG	19 bp downstream
gRNA-2	GCTTAGCGTATTTTATGTTG	TGG (- strand)	17 bp upstream
gRNA-3	AGCTAAGAGAAAGTTAGCAG	GGG	117 bp downstream

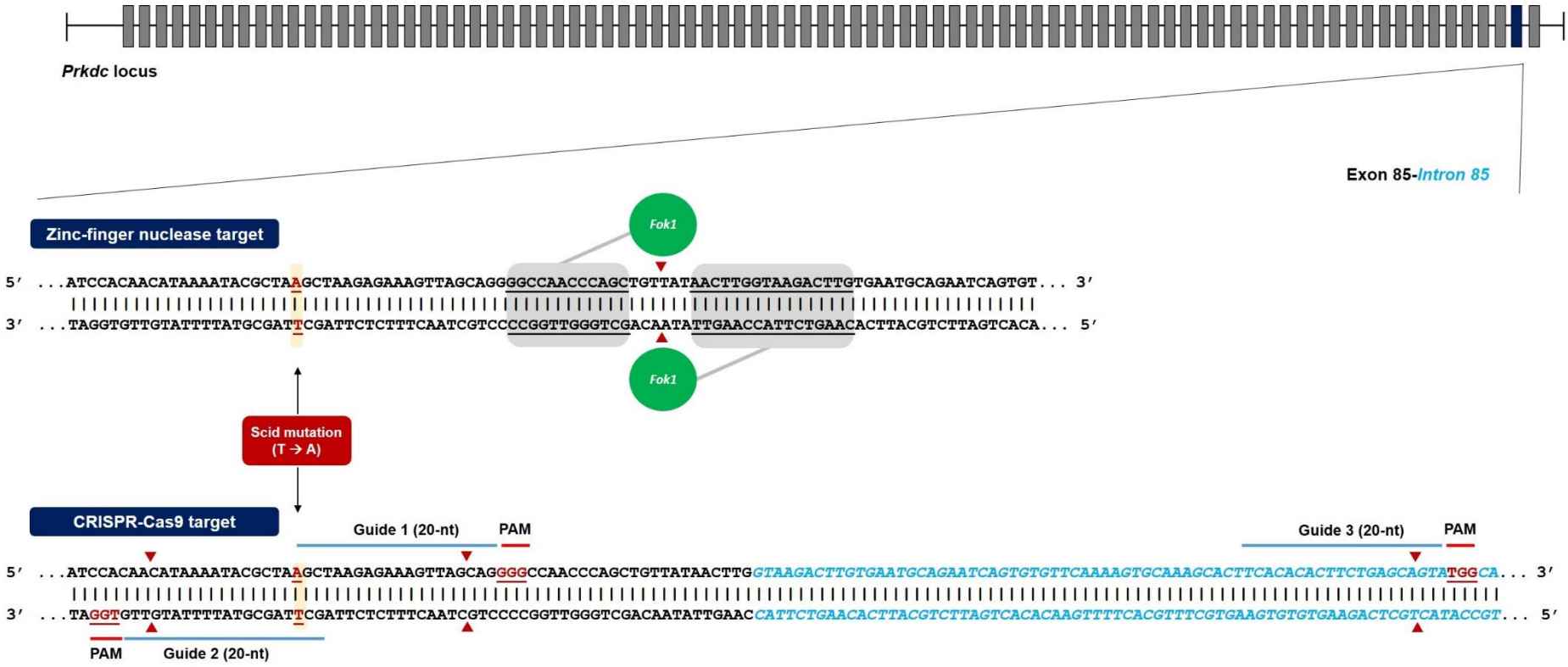


Figure 4-3 Schematic of *Prkdc* CRISPR-Cas9 gRNAs

Three different spCas9 ('NGG' PAM) gRNAs were designed adjacent to SCID mutation on *Prkdc*. gRNA 1 and 2 target regions close to SCID mutation site (T>A) on exon 85. gRNA-1 target site overlaps with the previously used zinc finger nucleases. gRNA-3 targets a region in the downstream intron.

4.2.2 Cloning of guide RNAs in Cas9 expression plasmid

For expression in culture, gRNAs 1, 2 and 3 and their complementary sequences with overhangs for the *Bbs*I (or *BpiI)* enzyme were obtained as single-stranded oligonucleotides, annealed and cloned into a plasmid backbone encoding *sp*Cas9-GFP. The plasmid PX458 contains a short filler region flanked by two *Bbs*I restriction sites downstream of human U6 promoter, which drives the expression of gRNA (plasmid map in (**Fig. 4-4)**. The filler region was removed and replaced by double-stranded gRNA nucleotides containing appropriate overhangs. Plasmid cloning was carried out using a single-step digestion-ligation protocol where the plasmid backbone was digested with *Bbs*I restriction enzyme and ligated with annealed oligo-duplex at the same time. Correct cloning of gRNA nucleotides was verified by sequencing the recombinant plasmid with a primer binding within the human U6 promoter. The cloning strategy and resulting sequence of gRNAs cloned in the Cas9 backbone is illustrated in **(Fig. 4-4)**.

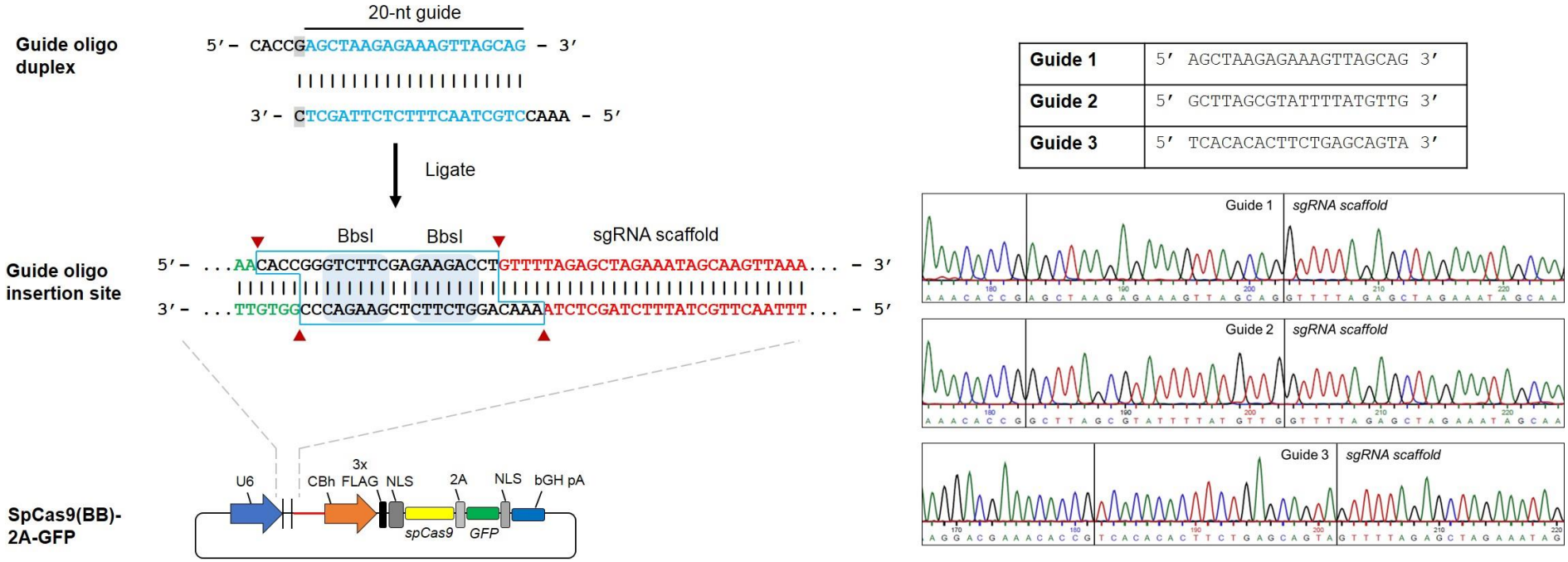


Figure 4-4 Cloning of *Prkdc* gRNAs in Cas9 Expression Plasmid

Prkdc-SCID gRNAs were cloned on a Cas9 plasmid co-expressing *GFP* using a single-step cloning protocol. (Left) First, single-stranded gRNAs were annealed to form a duplex. Next, the donor Cas9 plasmid was digested with *Bbs*I restriction enzyme and ligated with the oligo duplex at the same time. (Right) Cloning was verified by sequencing of the recombinant plasmid. Chromatographs obtained after Sanger sequencing show gRNA sequences inserted in the Cas9 backbone.

4.2.3 *Prkdc* gRNA cleavage efficiency

Guide RNA cleavage efficiency is determined by the detection of indels at the Cas9 cut site. Plasmids encoding *sp*Cas9-GFP and sgRNAs 1, 2 or 3 were introduced in cultured SCID fibroblasts using lipid-based transfection and green fluorescence was used to monitor plasmid expression. In all cases, SCID fibroblasts showed poor transfection efficiencies. **(Fig. 4-5)** shows the low transfection efficiency observed in SCID fibroblasts 24- and 48-hours post- transfection. Indels were detected at 48 hours post-transfection time point using the Tracking Indels by Decomposition (TIDE) method. To carry out TIDE, regions around the Cas9 cleavage sites for all three guides were amplified by PCR from control (unedited) and edited cells. These were then subjected to standard capillary Sanger sequencing. Subsequently, the obtained sequences were analysed using the TIDE software to quantitate the indel spectrum. TIDE analysis showed cleavage efficiency of 8.4%, 3% and 3.8% for gRNAs 1, 2 and 3 respectively with R^2=0.99. The indel profiles for all three gRNAs are illustrated in **(Fig 4-6)**.

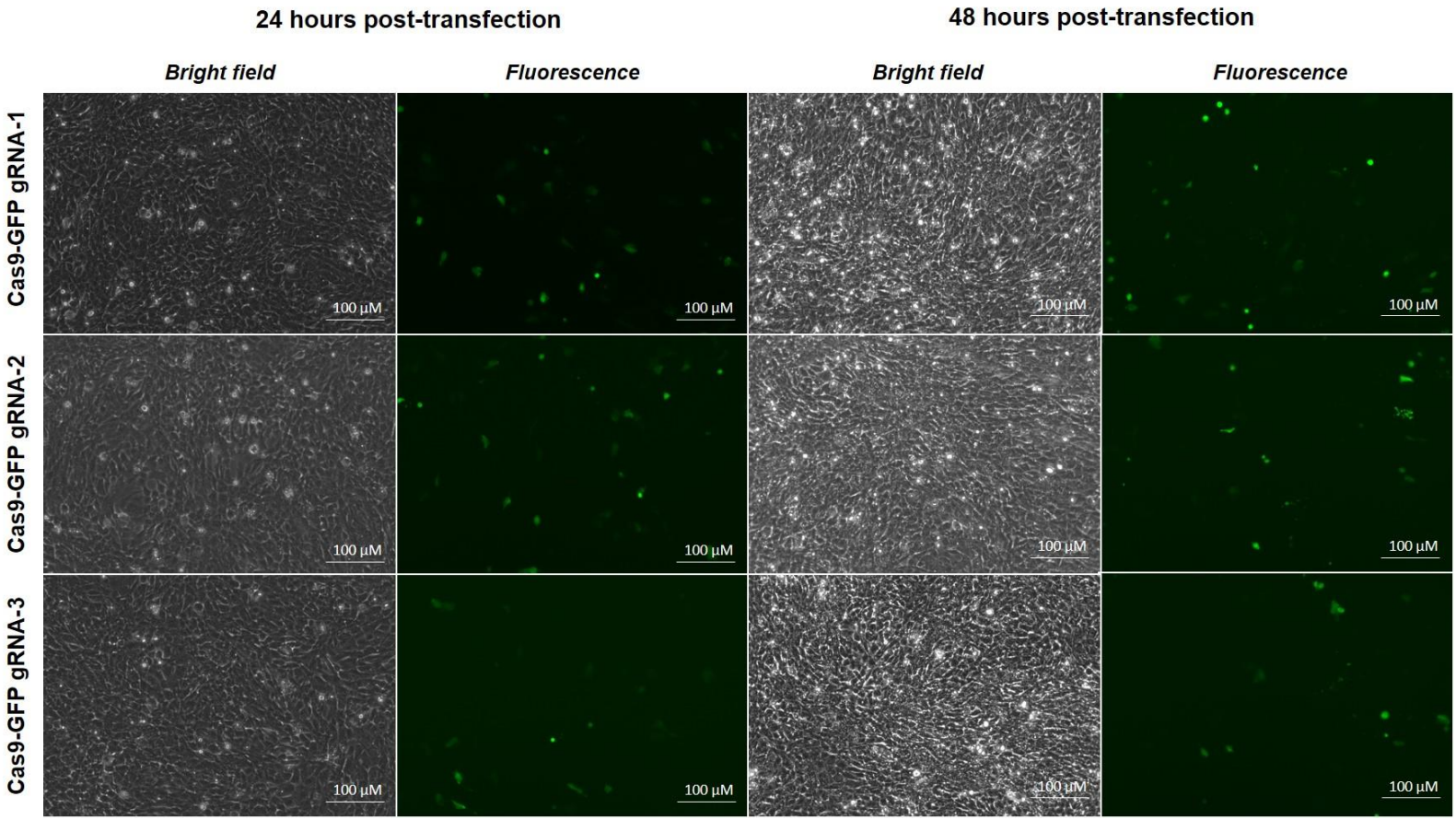


Figure 4-5 Transfection of *Prkdc* gRNAs in SCID Fibroblasts

Microscopy images showing transfection of Cas-GFP encoding *Prkdc* gRNA 1, 2 and 3, 48 hours post-transfection. Bright field and fluorescence images were captured in the same field.

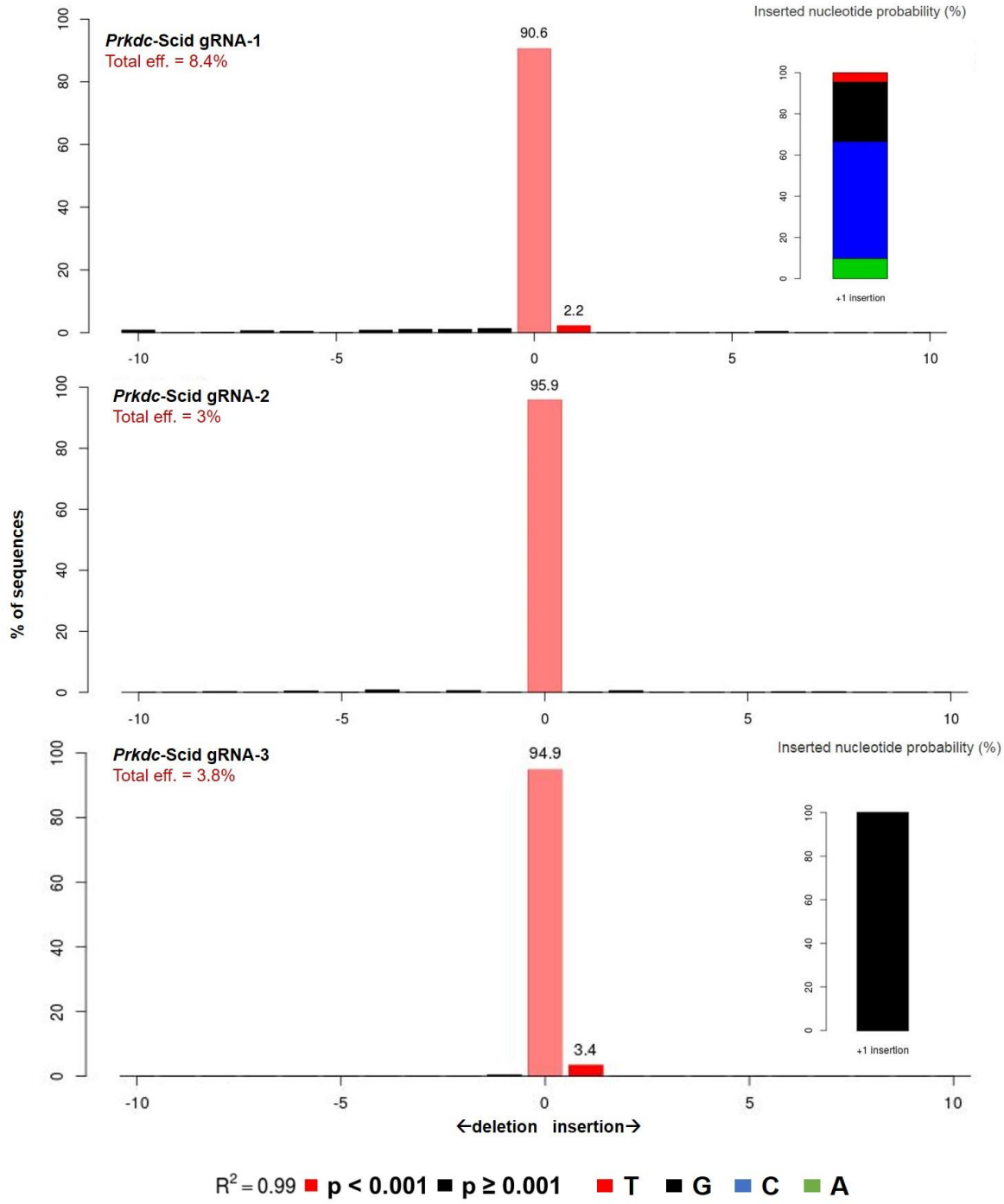


Figure 4-6 Indel Detection using TIDE

Charts showing spectrum of indels generated by *Prkdc* gRNAs 1, 2 and 3 in immortalised SCID fibroblasts. Results are shown as total efficiency of gRNA indicated by % of sequences containing insertions or deletions (indels) around the Cas9 cut site (represented by 0), and probability of nucleotide inserted at +1 bp location. R^2 represents the goodness of fit. The P-value is associated with the estimated abundance of each indel calculated by a two-tailed *t*-test.

4.2.4 Gene targeting of *Prkdc* exon 85

A homology-mediated repair strategy was used to target and repair the SCID mutation on *Prkdc* exon 85. For HDR, Cas9-GFP expression plasmids encoding gRNA 1 or 2 were transfected into immortalised SCID fibroblasts in the presence or absence of the *Prkdc-neo* donor template. Plasmid expression was monitored using the fluorescence of GFP from the marker gene. In all cases, SCID fibroblasts showed poor transfection efficiencies **(Fig. 4-7)**. Following the introduction of gene editing reagents, cells were subjected to G418 drug selection for 5 days to generate a G418-resistant targeted cell population. To validate gene targeting using HR, an inside-out PCR strategy amplifying the 3' HR junction was used as an initial screen **(Fig. 4-8)**. All samples transfected with Cas9 in the presence of gRNAs and the donor template showed gene targeting despite low transfection levels and led to the generation of G418-resistant colonies. No gene targeting was observed in cells transfected only with a donor template or with a Cas9 in absence of a gRNA **(Fig. 4-8)**.

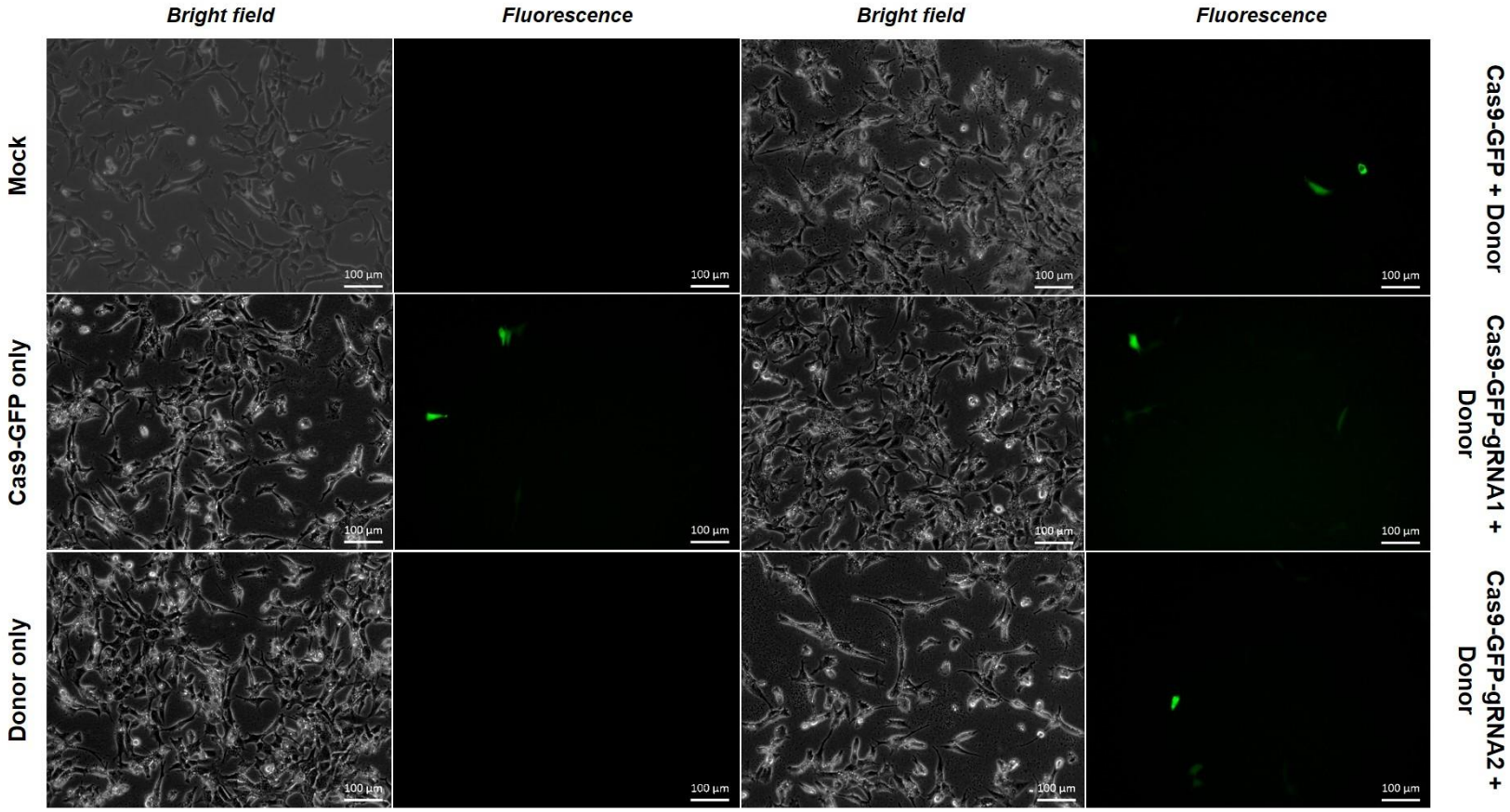


Figure 4-7 Transfection of Genome Targeting Reagents

Gene targeting was carried out by introduction of Cas9 and *Prkdc* gRNAs (1 and 2) along with a donor template in immortalised SCID fibroblasts. Bright field and fluorescence microscopy images were captured in the same field 48-hours post-transfection.

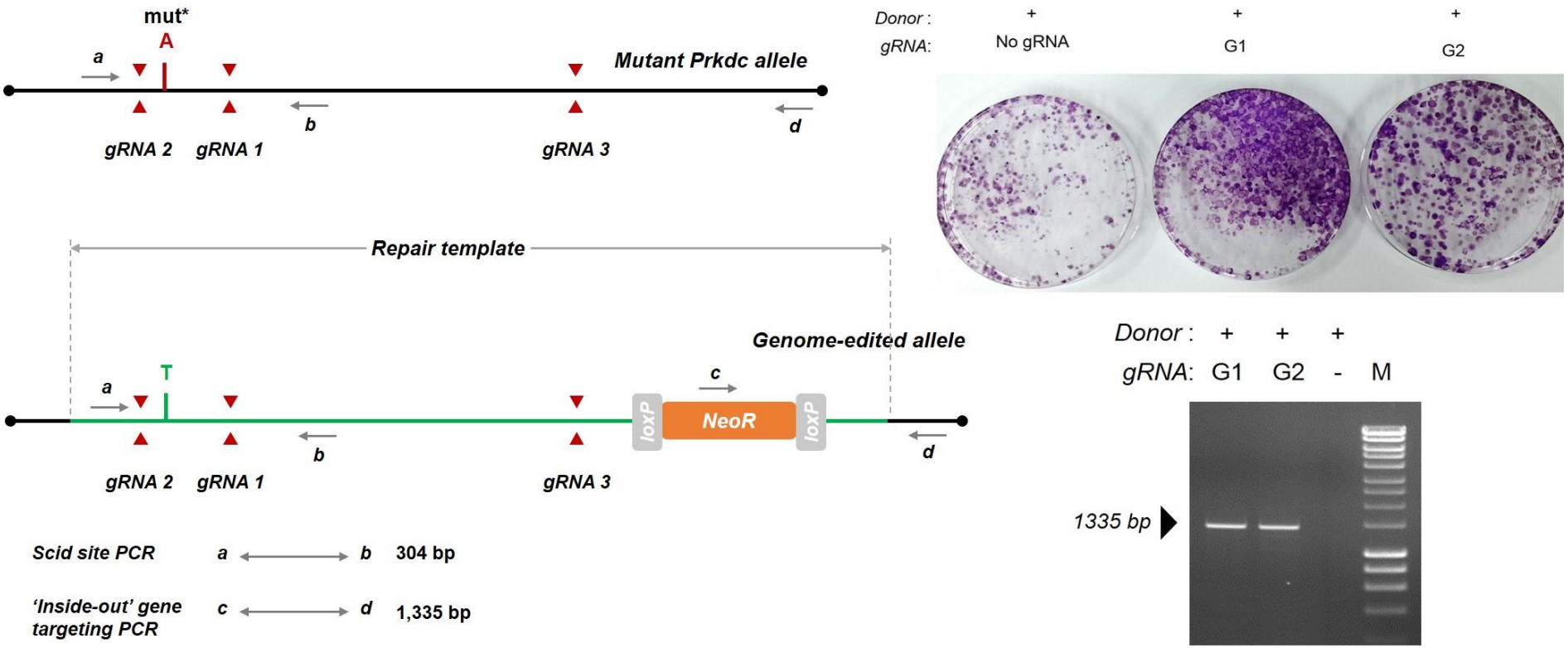


Figure 4-8 Gene Targeting of *Prkdc* Exon 85

On the left is shown a schematic representing PCRs that were used for genotyping gene edited cells. Arrows indicate PCR primers and their orientation. In order to verify gene targeting, an inside out PCR amplifying the genome–donor 3´-junction was used. Top right panel shows G418-resistant colonies stained with crystal violet, that were generated because of gene targeting using Cas9 and *Prkdc* gRNAs 1 and 2. Gel image shows the diagnostic gene targeting band in polyclonal SCID fibroblasts obtained by an inside-out PCR amplification of the genome–donor 3´-junction. No band was seen in cells targeted with donor in the absence of CRISPR-Cas9 reagents. M, Hyperladder 1 Kb; G1, gRNA-1; G2, gRNA-2, Gel, 1% agarose, 1x TAE buffer.

To determine the efficiency of the gene targeting approach and to analyse the types of gene edits, 30 single-celled clones from each gRNA targeted polyclonal cell pool were isolated. These monoclonal populations were expanded in culture and genomic DNA was extracted for genotype analysis. Firstly, an inside-out PCR amplifying the 3' HR junction was used to determine gene targeting in independent clones. 36% (i.e. 11 out of 30) of cells clones obtained from gRNA-1 cell pool and 33% (10 out of 30) clones obtained from gRNA-2 population showed the diagnostic band indicative of gene targeting **(Fig. 4-9)**. In comparison, no gene targeting was observed in cell clones derived from G418-R population obtained using donor template and Cas9 without a gRNA.

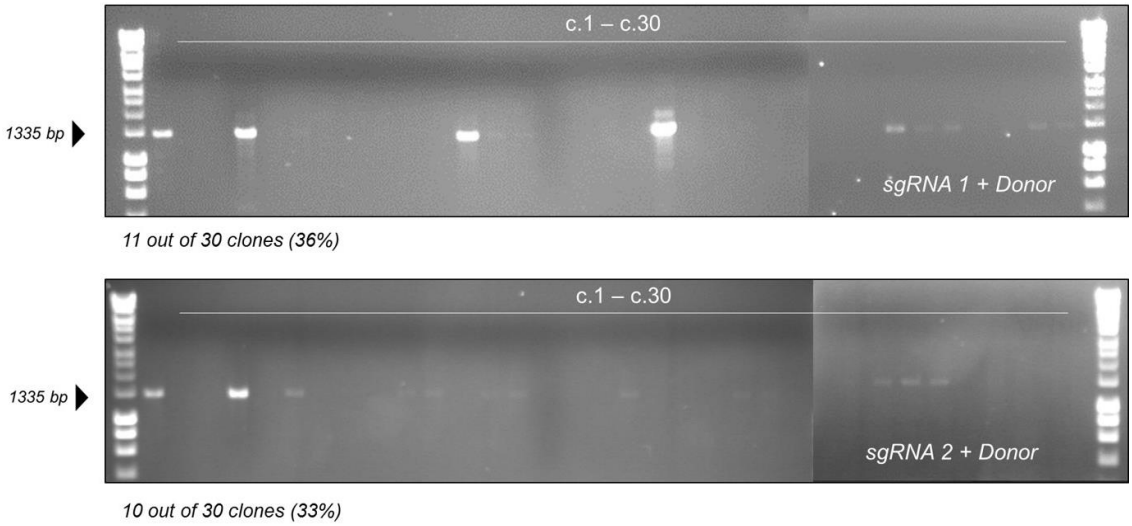


Figure 4-9 Gene Targeting in *Prkdc* CRISPR-Cas9 Clones

Gel images showing diagnostic gene targeting band (3' donor-genome junction) in clones isolated from gene edited polyclonal cell pools using *Prkdc* gRNA 1 and 2. Gene targeting was verified using an inside-out amplification of genome-donor 3' junction. (Top) Gel image showing PCR amplicons from 30 targeted clones derived from gRNA-1 and donor template treated cell pool. (Bottom) PCR amplicons from gRNA-2 clones. Overall approximately 30% of clones analysed showed targeted recombination of donor template in the presence of CRISPR-Cas9. M, Hyperladder 1 kb; Gel, 1% agarose, 1x TAE buffer.

To validate genome editing, a short region spanning the mutation and sgRNA target site (304 bp) was amplified and sequenced from clones positive for HR. Possible outcomes of HR between the *Prkdc* allele and the donor template are illustrated in **(Fig. 4-10)**. The data produced by capillary electrophoresis led to identification of sequence changes as well as detection of mixed sequences at the SCID mutation and Cas9 cleavage sites. Genome modification at the mutation site is reported as heterozygous (ht) for mono-allelic correction indicating both wild-type and SCID sequences, corrected for bi-allelic correction, or as SCID if there was no repair. At the Cas9 cleavage site, the genotype is

indicated as heterozygous when there were indels on at least one allele, or as wild-type in case no indels were observed **(Table 4-2)**. Targeting using gRNA-1 indicated mono- or bi-allelic correction of the SCID nucleotide in 60% of clones (n=10). However, this was also followed by the presence of heterogeneous indels at the Cas9 cleavage site, as observed previously in experiments using ZFNs. 30% of the clones showed no HDR at the SCID mutation, however they had indels at the Cas9 cut site. Only one clone (clone #7) did not show evidence of indels at the Cas9 cut site. As for gene targeting using gRNA-2, only 25% clones (n=8) indicated HDR (mono- or bi-allelic) at the SCID site along with indels at the Cas9 cut site. HDR at the mutation locus could not be determined in the remaining 75% clones while indels were identified at the Cas9 cut site **(Fig. 4-10)**.

Table 4-2 Mutational Profile of CRISPR-Cas9 *Prkdc* Fibroblast Clones

Genotype at the mutation site and Cas9 cut site were assessed using Sanger sequencing. At the mutation site, the genotype is reported as *SCID* for presence of SCID nucleotide, *corrected* for wild-type nucleotide, or *heterozygous (ht)* for presence of both SCID and wild-type nucleotides. At Cas9 cut site, *ht* indicates presence of insertions or deletions while *wt* represents unaffected sequence. y: yes, n: no, ht: heterozygous, wt: wild-type, nd: not determined.

Clonal line #	HDR at mutation site		Genotype at Cas9 cut site
	Yes/No	*Genotype*	
Prkdc-gRNA1			
c-2	y	ht	ht
c-7	n	SCID	wt
c-9	n	SCID	ht – deletion
c-10	y	ht	ht
c-14	y	ht	ht
c-18	y	ht	ht
c-19	y	corrected	ht – 3 bp *GCA* deletion
c-23	y	corrected	ht
c-24	n	SCID	ht – deletion
c-25	n	SCID	ht – deletion
Prkdc-gRNA2			
c-2	y	ht	ht
c-8	n	nd	ht
c-9	n	nd	ht
c-10	y	corrected	ht
c-15	n	nd	ht
c-25	n	nd	ht
c-29	n	nd	ht
c-30	n	nd	ht

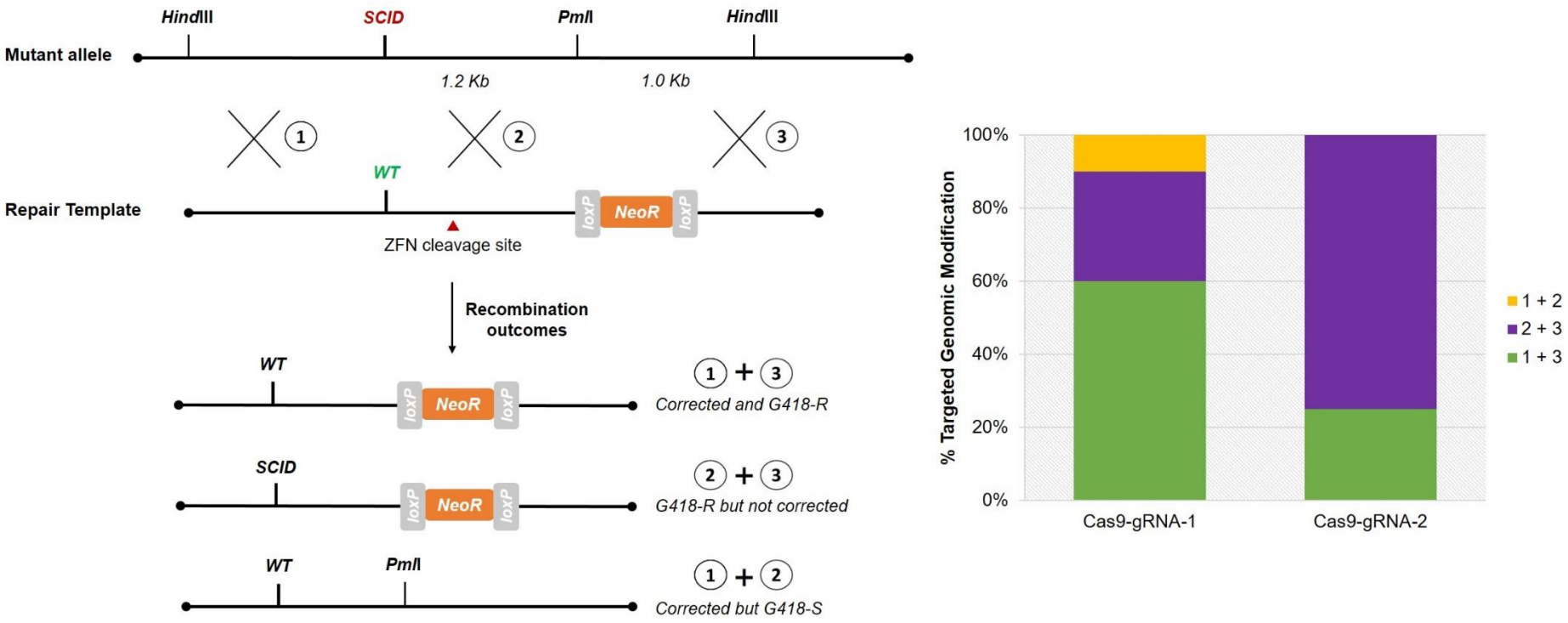


Figure 4-10 Genome Modification in CRISPR-Cas9 *Prkdc* Clones

(Left) Schematic depicting possible outcomes of homologous recombination between *Prkdc* targeted with Cas9 and *Prkdc* gRNAs and a donor template. Event 1+3 would lead to correction of SCID mutation and incorporate the selection cassette (NeoR) giving rise to G418-resistant population. Event 2+3 would also create a G418-R population but without the repair of mutated nucleotide. Recombination 1+2 would also repair the mutation however these cells would be sensitive to G418 (G418-S). (Right) Gene editing in *Prkdc* gRNA 1 or 2 targeted clones (n = 10 or 8, respectively) was by Sanger sequencing. Approximately 60% and 25% (green) clones from gRNA-1 and 2 showed mono-allelic correction by HDR of *Prkdc* mutation indicated by the presence of a heterozygous sequence at the SCID site. However, this was followed by the presence of indels at the Cas9 cleavage site. These clones were generated because of recombination event 1+3. Only one gRNA-1 clone, depicted in yellow, showed mono-allelic correction of the SCID mutation followed by wild-type sequence at Cas9 cut site. No such clone was obtained for gRNA-2. Approximately 20% (purple) clones obtained from RGEN-1 showed no correction of SCID site while Cas9 site contained indels (recombination event 2+3). For RGEN-2, majority of clones showed this profile.

4.2.5 Gene targeting of *Prkdc* intron 85

Previous gene targeting experiments employed nucleases with target sites close to the SCID mutation, lying within *Prkdc* exon 85. Downstream analysis showed that while effective at repair of SCID mutation, targeting exonic regions introduced indels in SCID fibroblasts. To overcome this problem, a third gRNA (gRNA-3) was designed that while being near (117 base pairs downstream) to the SCID mutation targeted the downstream intron. gRNA-3 was used, along with the *Prkdc-neo* repair template, to target the SCID mutation using the previously described gene targeting approach **(Fig. 4-2)**.

Genome editing reagents were transfected into SCID fibroblasts using calcium phosphate coprecipitation **(Fig. 4-11)**. The gene targeted Gene targeted cells were selected with G418 and analysed for gene targeting using the inside-out PCR approach. Cells transfected with Cas9 in presence of gRNA-3 and donor template showed gene targeting (despite a low transfection efficiency) and led to the generation of G418-resistant cells. No gene targeting was evident in cells transfected with donor template and Cas9 in absence of the gRNA **(Fig. 4-12)**.

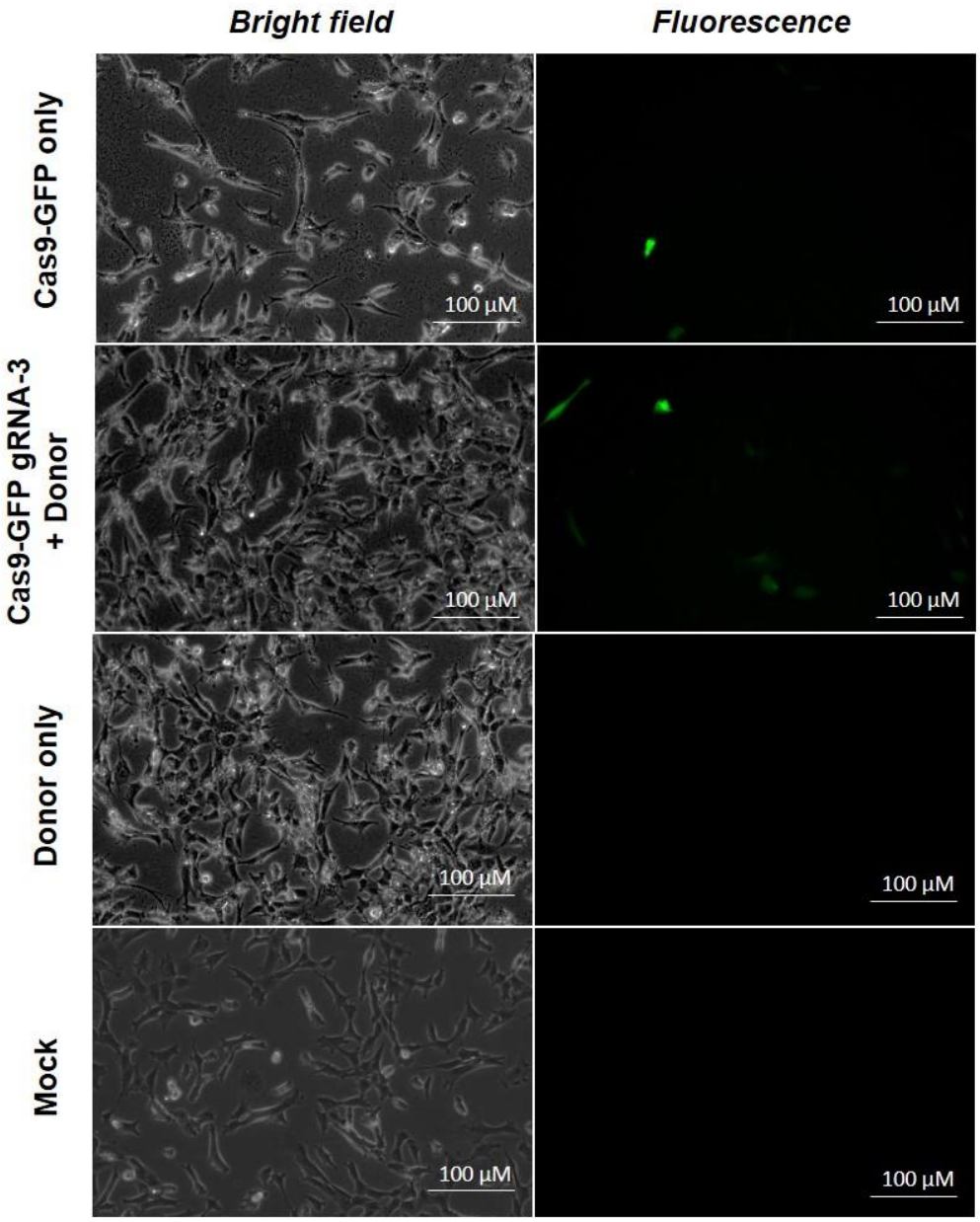


Figure 4-11 Transfection of *Prkdc* gRNA-3 Cas9 Plasmid

Microscopy images showing expression of *Prkdc* gRNA-3 Cas9-GFP plasmid 48-hours post-transfection. Bright field and fluorescence images were captured in the same field of cells.

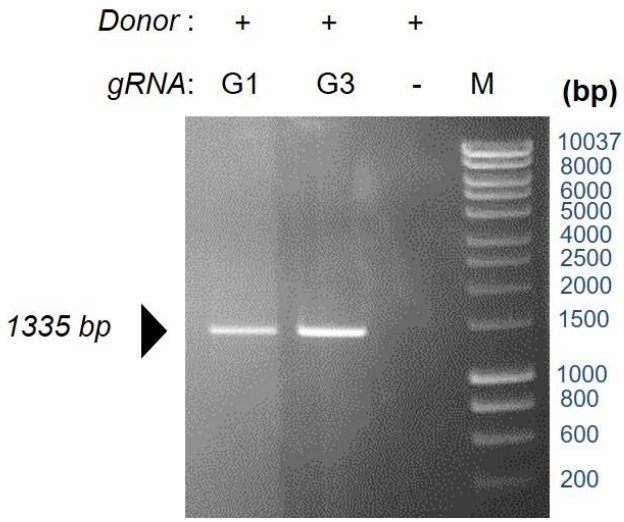


Figure 4-12 Demonstration of Gene Targeting Using *Prkdc* gRNA-3

Gel image showing diagnostic gene targeting band in polyclonal cells targeted with *Prkdc* gRNA-3 and *Prkdc-neo* donor template. No band was seen in cells targeted with donor in absence of CRISPR-Cas9 reagents. gRNA-1 cell population was used as a positive control. M, Hyperladder 1 Kb, G1: gRNA-1, G3: gRNA-3

4.2.6 Modification of donor DNA template

The initial gene editing experiments targeted *Prkdc* exon 85 with the HR donor template carrying intact PAM sequences for both guides tested. Genotyping of clonal populations showed that while HR led to mono- or even bi-allelic correction of the mutation, the HR template was modified at PAM sites and incorporated in the *Prkdc* locus with variable indels, thereby modifying the coding sequence of *Prkdc.* This would be unfavourable for gene expression or on the activity of the protein produced. To avoid re-cutting of the donor template, PAM regions for gRNA-1 and gRNA-3 were altered by creating single nucleotide substitutions. The mutations were designed using the web-tool Benchling. Since the PAM for gRNA-1 was present in the coding region of *Prkdc* (exon 85), a silent mutation was designed. The PAM for gRNA-3, which was in intron 85, was designed to be converted into a stop codon. PAM regions in the existing HR template *Prkdc-neo* were modified using site-directed mutagenesis. Site-directed mutagenesis is an *in vitro* procedure that uses custom designed oligonucleotide primers to confer a desired mutation in a double-stranded plasmid. The primer design results in a product that will re-circularize to form a doubly-nicked plasmid, which can then be transformed into *E.coli*. For site-directed mutagenesis of *Prkdc-neo*, it was amplified using primers containing the intended mutation. The modified plasmid was then verified by sequencing the plasmid DNA **(Fig. 4-13)**.

gRNA-1 PAM modification

GG**G** (Glycine) to GG**T** (Glycine) silent mutation.

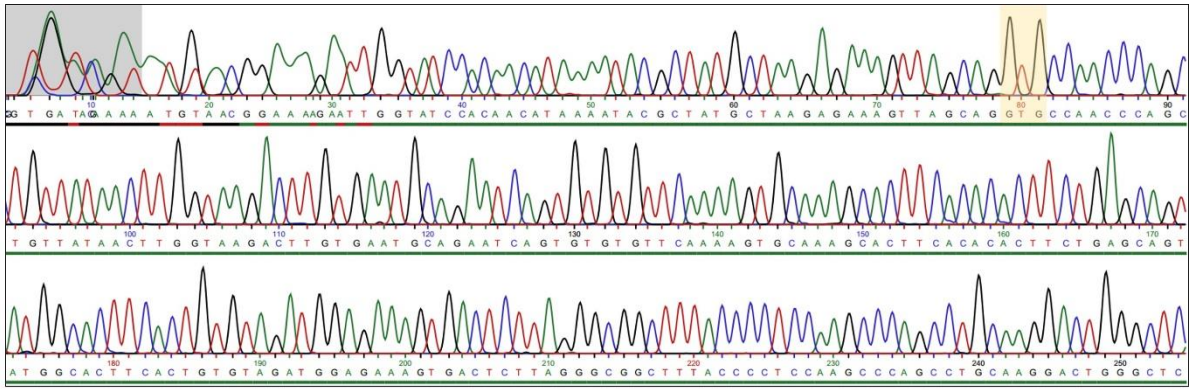
A/T Scid Locus

GGG PAM

GTG Mutated PAM

Glycine Silent mutation

gRNA-1	...TAA GCT AAG AGA AAG TTA GCA GGG GCC...
Original Target Site in *Prkdc-neo*	...TAT GCT AAG AGA AAG TTA GCA GGG GCC...
After Site Removal in *Prkdc-neo*	...TAT GCT AAG AGA AAG TTA GCA GGT GCC...



gRNA-3 PAM modification

TG**G** (Tryptophan) to TG**A** (Stop Codon).

TGG PAM

TGA Mutated PAM

Stop Codon (Intronic)

gRNA-3	...TTC ACA CAC TTC TGA GCA GTA TGG...
Original Target Site in *Prkdc-neo*	...TTC ACA CAC TTC TGA GCA GTA TGG...
Mutated Site in *Prkdc-neo*	...TTC ACA CAC TTC TGA GCA GTA TGA...

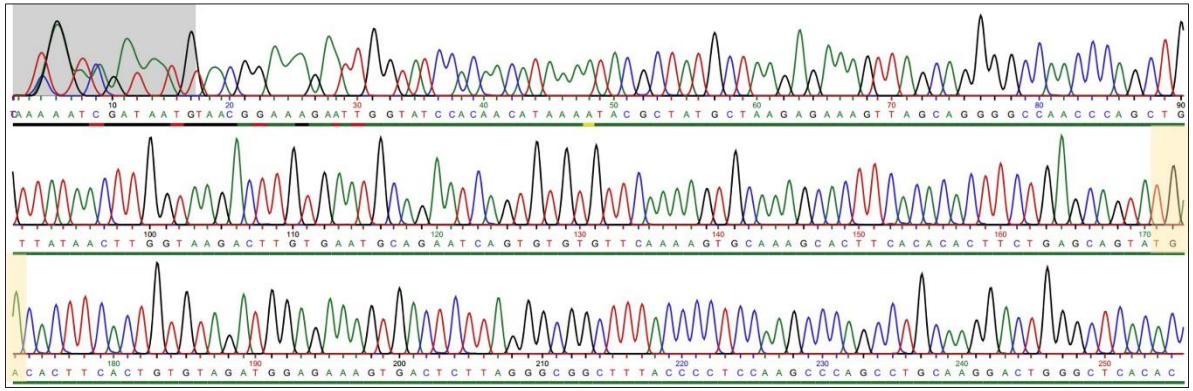


Figure 4-13 Site-Directed Mutagenesis of *Prkdc-Neo* Donor Template

Left panels show designs of *Prkdc* gRNA 1 and 3 PAM modification. gRNA-1 PAM GGG was modified to GTG (silent glycine mutation). gRNA-3 PAM TGG was modified to TGA (stop codon). Right panels show chromatographs obtained from Sanger sequencing of the mutated plasmid. Modified codons are highlighted in pale orange.

4.2.7 Gene targeting using modified donor DNA

The modified donor DNA templates with PAM regions mutated for gRNA-1 and 3 were used in conjunction with *Prkdc* guides gRNA-1 or gRNA-3 respectively for HDR of SCID point mutation. The gene targeting approach used was the same as described previously. Analysis of polyclonal G418 selected population showed successful targeting in SCID fibroblasts in both cases **(Fig. 4-14).**

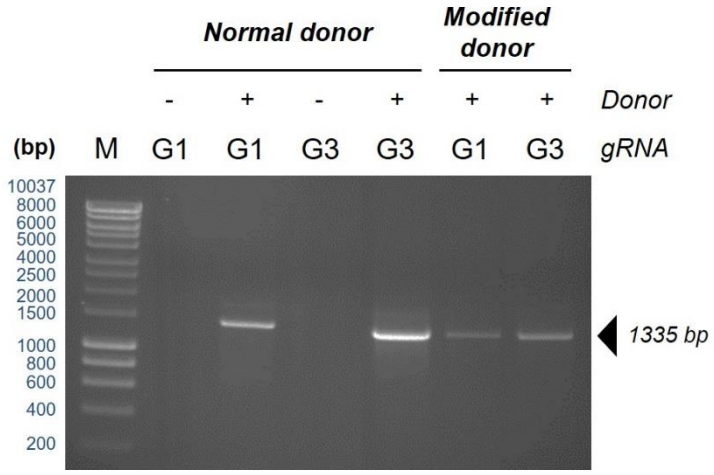


Figure 4-14 Gene Targeting Using Modified *Prkdc-neo* Donor

Gel image showing diagnostic gene targeting band (amplifying 3' donor-genome junction) in polyclonal SCID fibroblasts obtained after gene editing using *Prkdc* gRNA 1 or 3 along with normal or modified donor DNA. M, hyperladder 1 Kb; G1, gRNA-1; G3, gRNA-3

4.2.8 Generation of CRISPR lentiviral vectors

For the generation of lentiviral vectors, the designed gRNAs 1 was obtained as single-stranded oligos and cloned into lentiCRISPRv2 plasmid backbone. The LentiCRISPRv2 plasmid contains two expression cassettes, hSpCas9 and the chimeric guide RNA. *Prkdc*-SCID gRNA-1 was cloned into the chimeric gRNA scaffold with appropriate overhangs using a single-step restriction digestion-ligation method. This involved digestion of the vector with the *BsmB*I restriction enzyme and ligation with the annealed oligo-duplex simultaneously. Correct cloning of gRNA nucleotides was verified by sequencing the recombinant plasmid with a primer binding within the human U6 promoter region. The cloning strategy and resulting sequence of gRNA cloned in the lentiCRISPRv2 plasmid backbone is illustrated in **(Fig.15)**.

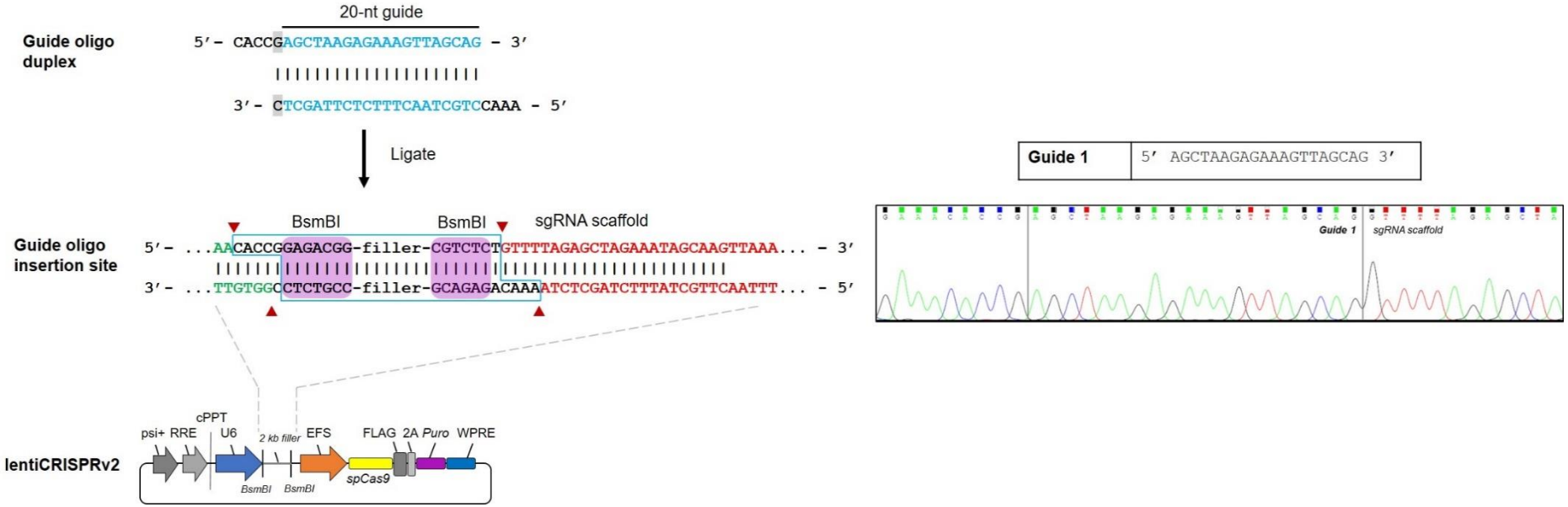


Figure 4-15 Cloning of *Prkdc* gRNA-1 on LentiCRISPR Backbone

Prkdc-SCID gRNA-1 targeting exon 85 was cloned in lentiCRISPRv2, a dual-expression lentiviral transfer plasmid encoding human spCas9, gRNA scaffold and lentiviral vector components. Left panel shows the cloning strategy. Single-stranded gRNA oligos with overhangs were annealed to form a duplex and subsequently ligated in the LentiCRISPRv2 plasmid backbone digested with BsmBI restriction enzyme. Right panel shows chromatograph from Sanger sequencing of the recombinant plasmid showing correct insertion of the gRNA. cPPT, central poly purine tract; WPRE, woodchuck hepatitis post-transcriptional regulatory element; RRE, Rev-responsive element; Puro, puromycin gene; psi, packaging signal.

Third-generation integration-proficient lentiCRISPR viral vector encoding *Prkdc* -SCID gRNA-1 and gRNA-2 was generated in HEK293T cells. Viral particle titrations were carried out in HeLa cells and quantified as transducing units per ml (TU/ml) using real-time PCR. Day 1 and day 2 titres of generated lentiCRISPR viral vector are shown in **Table 4-4.**

Table 4-3 LentiCRISPR Viral Vector Titres

Lentiviral Vector	Transfer Plasmid	Viral Titre (TU/ml)	
		Day 1	*Day 2*
lentiCRISPRv2-Prkdc-gRNA-1	pRY562	3.4E+07	1.5E+07
lentiCRISPRv2-Prkdc-gRNA-2	pRY563	3.4E+07	7.1E+06

4.2.9 Comparison of gene editing in wild-type and SCID fibroblasts

The alternative-NHEJ (Alt-NHEJ) DNA damage repair pathway is suggested to take place in absence of NHEJ (Betermier, Bertrand et al. 2014). To test this, DNA damage response was tested in SCID and wild-type fibroblasts using Cas9 nuclease targeting *Prkdc* gene. An integrating lentiCRISPR viral vector encoding Cas9 and *Prkdc* gRNA-1 was transduced into wild-type and SCID fibroblasts. To compare the effect of gene editing, the region around the Cas9 cleavage site was amplified and subjected to sequencing. **(Fig. 4-16*a*)** shows cutting by Cas9 in wild-type fibroblasts led to fewer indels with low editing efficiency (3.8%), while a wide spectrum of indels were noted in SCID fibroblasts (editing efficiency 30.8%) suggesting the inability of SCID cells to carry out DNA repair using NHEJ.

To quantitate the efficiency of guides in both cell types and identify the resulting indels, sequences from PCR products were analysed using TIDE software. Gene editing in SCID cells led to creation of a wide spectrum of indels with editing in 30% of cells. In comparison, only a very small proportion of wild-type cells were edited thus generating fewer indels **(Fig. 4-16*b*)**. This suggests that after cutting, the DNA was repaired in wild-type cells, while the same repair mechanism is absent in SCID cells. However, it has to be noted that gRNA-1 target site in wild type cells differed by 1-bp at the 5' terminal.

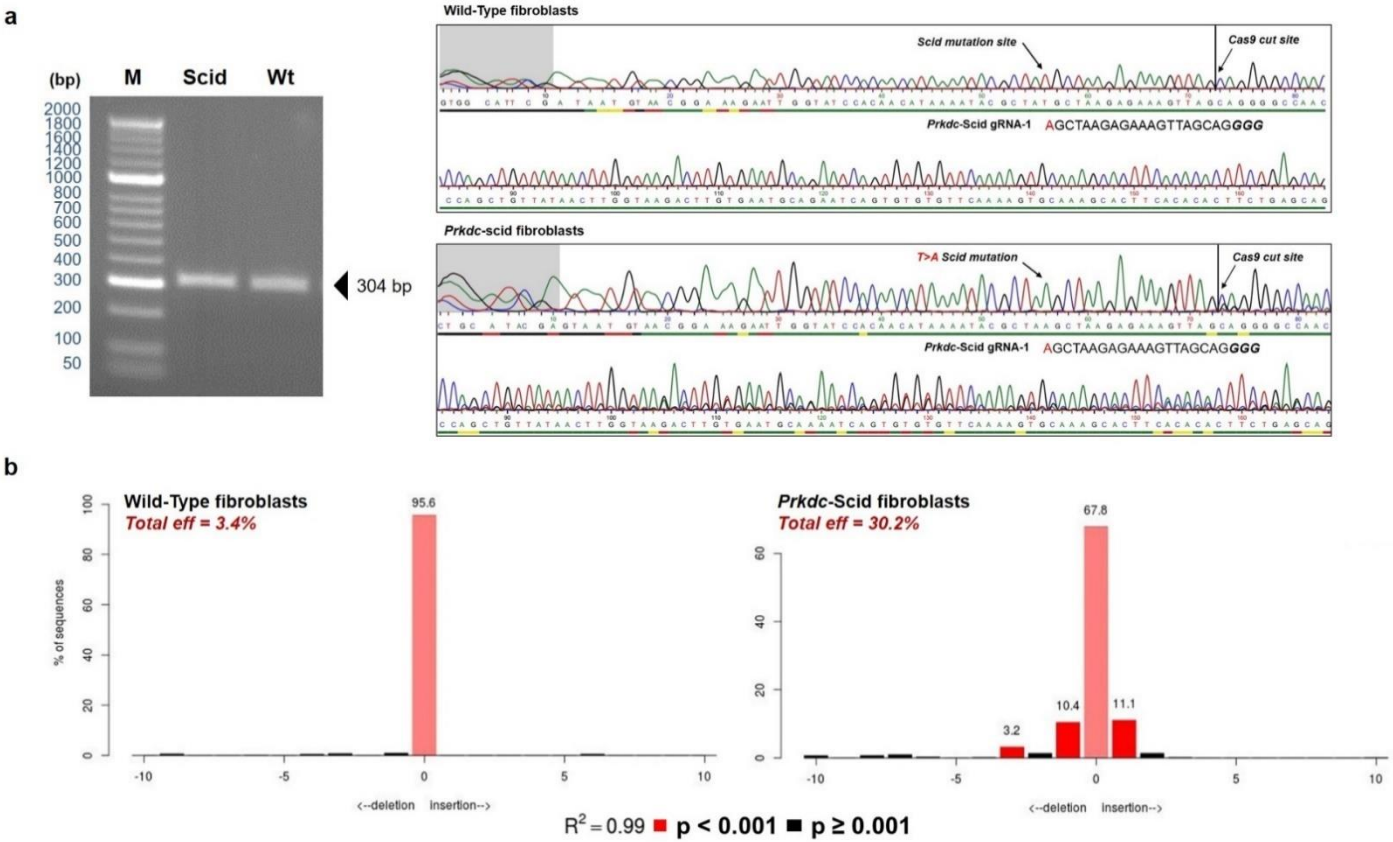


Figure 4-16 Comparison of Gene Editing in Wild-Type and *Prkdc* Fibroblasts

Prkdc-SCID guide RNA (gRNA 1) and Cas9 encoding lentiviral vectors (MOI 40) were transduced in wild-type (wt) and SCID mouse fibroblasts. (a) On the left is shown a gel image indicating band obtained from amplifying region around Cas9 cleavage site in wt and SCID fibroblasts. Chromatographs on the right were obtained from sequencing the PCR band and show the location of SCID mutation (T>A), sequence of *Prkdc* gRNA-1 followed by NGG PAM, and location of Cas9 cut site. (b) Chart showing spectrum of indels generated upon gene editing by *Prkdc* gRNA-1 in wt and SCID fibroblasts. Spectrum is shown as % of modified sequences around Cas9 cut site (represented by 0). M, Hyperladder 50 bp; Gel, 1.5% agarose, 1x TAE buffer.

4.3 Discussion

This chapter described targeting of the *Prkdc* gene and repair of the SCID point mutation using CRISPR-Cas9 technology. Gene targeting was carried out in murine immortalised fibroblasts. Three independent guides targeting the SCID mutation on the *Prkdc* gene were designed and validated. Two guides were identified within exon 85 near to the SCID mutation site. gRNA-1 indeed overlapped with the previously described ZFN target site. A third guide was identified in the downstream intron. All three guides were obtained as DNA oligonucleotides and their efficiency to edit *Prkdc* locus was tested in SCID fibroblasts. Unfortunately, mouse fibroblasts were found to be very difficult to transfect using lipid-based transfection reagents. A standard method to detect indels, formed by repair following editing, is to use enzyme-mismatch assays, namely, T7E1 or Surveyor assay (Ran, Hsu et al. 2013). The sensitivity of these assays however is around 5% (Abdul-Razak 2013, Ran, Hsu et al. 2013) and did not yield conclusive results in preliminary experiments (data not shown). This was not surprising given the poor efficiency of transfection of gene editing reagents in murine SCID fibroblasts. Thus, a different technique, TIDE, was used to assess gene editing. TIDE is a bioinformatic tool that aligns genomic sequences from edited and unedited cells along with a gRNA sequence to identify indels. TIDE analysis led to robust and rapid detection of cutting. Moreover, it also provided information about the cut site, giving a spectrum of indels generated. Use of TIDE analysis showed editing in SCID fibroblasts using all three gRNAs albeit with low efficiencies. This could largely be due to the poor transfection to begin with. Only three guides were tested in this section, however, usually several different gRNAs must be screened to identify the best candidate. In such scenarios, gRNA screening can be done using *in vitro* cell reporter assays, where the target region is synthetically introduced in an easy to transfect cell line (Kim, Koo et al. 2017). Moreover, such kits also overcome the limitation of poor transfection in primary cells as seen in this case. Alternatively, the genome editing reagents could be delivered into cells as mRNA or protein. Delivery of gRNA as purified RNA complexed with Cas9 protein is being increasingly used with high efficiencies for manipulation in mammalian cells (Kim, Kim et al. 2014). Having established gene editing at *Prkdc* with an all-plasmid system, future work related should employ the use of mRNA/protein delivery systems.

Despite the low editing efficiency, the guides were separately used for homology-mediated repair of SCID mutation. Gene targeting was carried out using an all-plasmid approach, where gRNA and Cas9 and the repair template were expressed on plasmid constructs. Calcium phosphate co-precipitation was used to transiently introduce the gene editing reagents in to the cells. This method was used as it showed comparable

efficiency of transfection in SCID fibroblasts as obtained using lipid-based reagents in preliminary experiments. It also has the added advantage of low costs, with reagents produced in-house, and can be easily up-scaled for larger number of cells. Both gene targeting experiments - targeting exon 85 using gRNA 1 and 2 or targeting intron 85 using gRNA 3 - led to generation of G418-resistant cell populations. These cell populations can also be generated due to random integration of the donor template conferring drug resistance. A method of validating gene editing is to obtain and screen several clones from the primary transformed pool of cells. This necessitates two rounds of screening. First, a primary screen to determine the relative fraction of cells containing the edit. This is important because the nuclease cleavage and repair process is not completely efficient or accurate (Ran, Hsu et al. 2013, Hollywood, Lee et al. 2016). Hence, resistant populations were genotyped for targeted integration using a PCR that amplified the donor-genome junction. Successful gene targeting was obtained in all three cases only when gRNA, Cas9 and donor template were present in combination. Knowing the efficiency of the edit determined the number of single-cell clones that were isolated for expansion. The secondary screen was performed on the 30 single-cell derived monoclonal populations from gRNA 1 and 2 targeted cell pools. Around 30% of clones from both cell pools showed gene targeting, comparable to gene targeting with ZFNs. Furthermore, mutational analysis of the clones showed a similar profile as observed in ZFN-targeted clones – repair at mutation site followed by heterogenous indels at the nuclease target site. Three different genotype profiles were observed in gene targeted clones. These were identified to be because of HR between the *Prkdc* allele and Prkdc-neo donor DNA. Possible outcomes of recombination along with a comparison of genotypes in ZFN and CRISPR-Cas9 clones is shown in **(Fig. 4-17)**.

A major difference between the ZFN and CRISPR-Cas9 targeted cells were that while all clones from ZFN population only showed mono-allelic correction, cells targeted with Cas9 displayed both mono- and bi-allelic correction of the mutated nucleotide. The presence of indels at the nuclease cut site was consistent in both ZFN and Cas9 experiments. Only one clone (Cas9-gRNA-1 targeted clone #7) showed targeting of *Prkdc* followed by an intact nuclease site. This suggests an intrinsic defect in DNA repair in DNA-PK deficient SCID fibroblasts and that a wider screen with greater number of clones would be required to identify cells with the desired edit – that is corrected SCID nucleotide followed by an intact nuclease target site. Nevertheless, these experiments demonstrate that CRISPR-Cas9 system can be used to carry out genome editing and HDR in cell models that show defects in DNA repair processes.

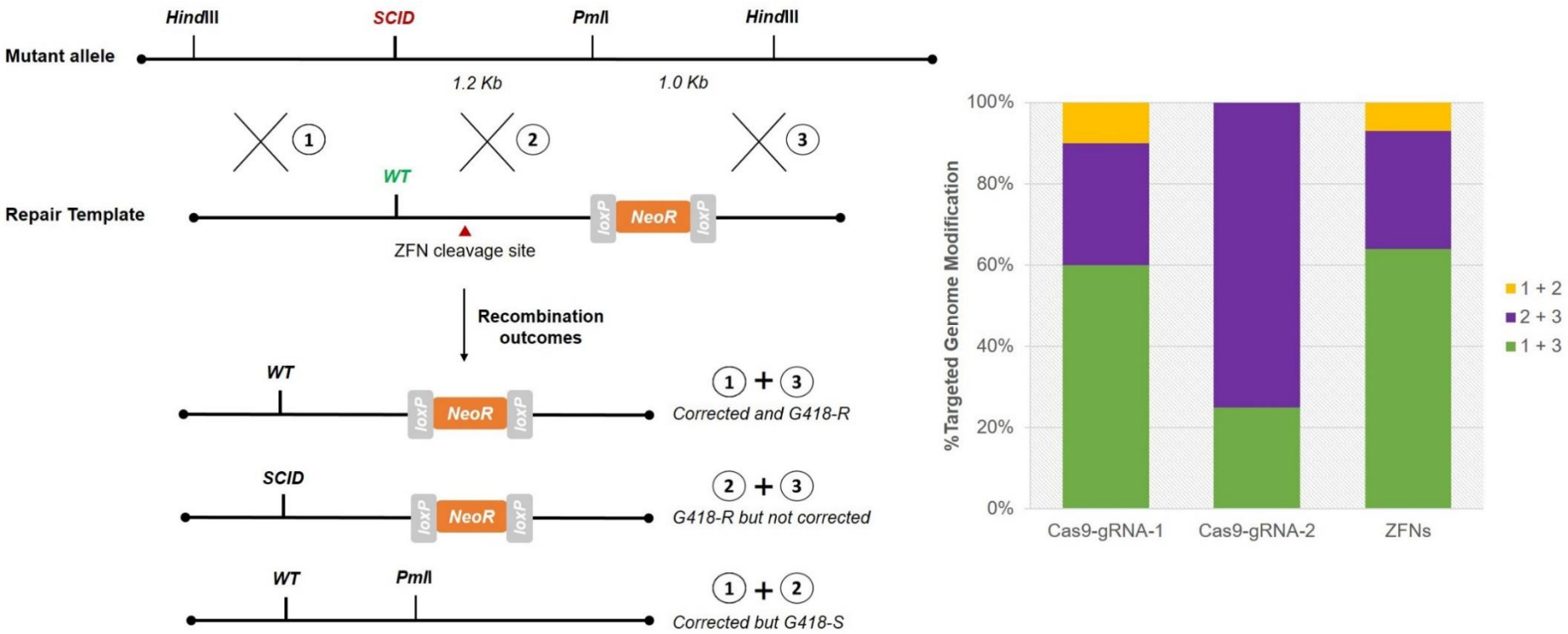


Figure 4-17 Comparison of Genome Editing in *Prkdc* ZFN And Cas9 Clones

(*Left*) Schematic representation of possible recombination outcomes between *Prkdc* allele and *Prkdc-neo* donor template. Event 1+3 would lead to correction of SCID mutation and incorporate the selection cassette (NeoR) giving rise to G418-resistant population. Event 2+3 would also create a G418-R population albeit repair of mutated nucleotide. Recombination 1+2 would also repair the mutation however these cells would be sensitive to G418 (G418-S). (*Right*) Comparison of gene editing in ZFN (n=14) and CRISPR-Cas9 (gRNA 1, n=10; gRNA 2, n=8) gene edited clones determined by Sanger sequencing.

Furthermore, these experiments highlighted two major limitations of the gene targeting approach used. Firstly, the location of Cas9 nuclease target site. Cutting in the coding region can introduce unwanted mutations. In case of the *Prkdc*-SCID fibroblasts, where the NHEJ pathway is defective, the inaccurate alt-NHEJ prominently takes place for DNA damage repair (Betermier, Bertrand et al. 2014). Because of this, mutations were detected in the coding region of *Prkdc* at the Cas9 cut site in all cells analysed. This can have a detrimental effect on gene expression and protein activity even though the original mutation was repaired. To overcome this limitation, the gene targeting approach was modified by using a gRNA that targeted the SCID mutation by introducing the DSB in the downstream intron. For effective HDR to occur, the DSB should lie within 100-150 bp of the mutation site (Ran, Hsu et al. 2013). gRNA-3 was identified within this range and shown to mediate genome editing in SCID fibroblasts. When used in conjunction with the existing donor DNA template, it showed gene targeting in polyclonal cells. However, further validation of gene editing using this guide RNA is required by demonstrating gene editing events in monoclonal populations.

The second limitation of the gene targeting approach was the presence of nuclease target sites in the donor DNA. Presence of intact target sites in the donor DNA can cause it to be degraded by the target-specific nuclease present in the cells, in this case Cas9; thus, modifying the donor DNA as well as the genomic locus upon recombination. The initial gene editing experiment targeted *Prkdc* exon 85 with a HR template carrying intact PAM sequences for both guides. Genotyping of clonal populations showed that while HR led to correction of the mutation, the Cas9 site included modifications. One method to overcome this limitation is to modify PAM regions in the donor DNA by creating silent mutations (Hollywood, Lee et al. 2016). Since, guides 1 and 3 showed higher editing efficiencies, PAM regions for these were modified in the donor DNA using site-directed mutagenesis, and the modified donor DNA plasmids were used for a repeated gene targeting experiment. Introduction of these modified donor DNA plasmids in SCID fibroblasts led to significant cell death. However, the cells were recovered with continued culture and showed successful gene targeting by PCR. This however requires further validation by genotyping of monoclonal populations followed by next-generation sequencing.

Nuclease specific gene editing in SCID fibroblasts continually showed inaccurate repair of DNA. Since SCID cells carry a defect in NHEJ, it was hypothesised that in the absence of NHEJ, alt-NHEJ was the dominant DNA damage repair process (Betermier, Bertrand et al. 2014, Shibata and Jeggo 2014). To test this, CRISPR-Cas9 system was used to introduce DSBs, a form of DNA damage, in SCID and wild-type Balb/c fibroblasts. *Prkdc* gRNA-1 sequence lies in both wild-type and SCID cells, with 1-bp difference at the 5' end, and it has been shown to mediate cutting in SCID cells. Hence this guide was chosen to introduce DSBs. Since these cells were found to be difficult to transfect, gRNA-1 was developed into a third-generation lentiCRISR viral vector. It is noteworthy to mention that the CRISPR-Cas9 reagents were developed in both integrating and non-integrating lentiviral vectors. In fact, the non-integrating lentiviral vector was better suited to allow for transient expression of Cas9 nuclease, long-term expression of which could be disadvantageous in cultured systems. However, the viral vector titres obtained for the non-integrating lentiCRISPR vectors were very low for experimental use (data not shown). I therefore performed the preliminary experiments with the integrating lentiviral vector.

Transduction of a lentiCRISPR vector, encoding Cas9 and gRNA-1, led to cutting at the SCID locus in both wild-type and SCID cells, albeit at a very low frequency in the wild-type cells. This could mean that in wild-type cells the locus was cut but efficiently repaired or that there was no editing to begin with. In SCID cells, cutting by Cas9 however led to a spectrum of indels. We hypothesise that this could be due to the presence of an inefficient DNA repair mechanism, alternative NHEJ (alt-NHEJ), that takes place in the absence of NHEJ. Unlike it's classical counterpart, the alt-NHEJ is more prone to causing mutations at the DSB site. Although this experiment gave an insight in DNA damage repair in both cell types, further work investigating and assessing DNA repair damage pathways is required.

5 CRISPR-Cas9 Gene Editing as a Therapeutic Approach for Ataxia Telangiectasia

5.1 Introduction

Previous chapters described genome editing in *Prkdc* SCID, a DNA repair disorder characterised by severe defects of the immune system and radio-sensitivity. Although the SCID mouse is an archetype for proof-of-principle gene repair, clinical translation of genome editing as a therapy for DNA-PK radiosensitive SCID is limited. This is primarily because of the rarity of the condition in humans. *PRKDC*-SCID is an ultra-rare disease, with only two patients identified so far. Ataxia telangiectasia (shortened as A-T) is a related, rare, genetic DNA repair disorder. Inherited as an autosomal recessive trait, A-T has an estimated prevalence of 1 in 40,000-400,000 people (Lavin 2008). People affected with A-T carry mutations on the gene known as *ataxia telangiectasia mutated* (*ATM;* OMIM 607585*)*, which encodes the ATM protein, an apical kinase with critical roles in cellular response to DNA damage, apoptosis and cell cycle checkpoint control. Characterised largely as a neurological disorder, A-T is a complex condition that affects multiple systems within the body, including nervous, immune and respiratory systems. The hallmark feature of A-T is progressive neurodegeneration in cerebellum accompanied with immunodeficiency and predisposition to cancers. People with A-T are also hypersensitive to the effects of radiation and incur chromosomal instabilities (Lavin 2008). Currently, the disease has no cure. Gene and cell therapies however hold promise for treating aspects of the disease. This chapter describes the work undertaken to utilise the CRISPR-Cas9 system to target human *ATM* locus with the fundamental aim to identify and develop gene editing approaches that could potentially be therapeutic for A-T models. Firstly, bioinformatics tools were used to investigate human and mouse *ATM* gene structures and

all A-T causing mutations known-to-date were systematically catalogued. Based on this analysis, various genome editing strategies were designed. These included NHEJ- and HR-mediated gene repair strategies applicable to select few *ATM* mutations with either a higher incidence rate in A-T patients or for which primary cell models already exists; and, universal approaches that could theoretically be applicable to A-T models irrespective of underlying *ATM* mutations. Furthermore, a strategy to generate a humanised A-T mouse model was examined *in silico*. Lastly, to establish proof-of-principle, CRISPR-Cas9 gRNAs targeting various regions across *ATM* were designed and genome editing was validated in human cells.

5.2 Results

5.2.1 Analyses of human and mouse ATM genes

Human and mouse *ATM* genes were analysed using Ensembl genome browser to construct gene maps indicating intron-exon boundaries and protein domains. **Fig. 5-1 and 5-2** illustrate human *ATM* and mouse *Atm* maps, respectively.

Similarity between human and mouse *ATM* cDNA was analysed using local alignment tool EMBOSS Matcher. The alignment revealed 72.2% sequence similarity between the two loci. Furthermore, the exon-intron structures were also highly conserved. An exception was the in-frame deletion of a very short exon (exon 45) in mouse *Atm*. Exons before and after exon number 45 however were alike to the human *ATM* counterparts. Major differences were noted in untranslated regions of both genes. The 5'UTR of human *ATM* is very complicated with two leading exons, the same is not true for mouse *Atm*. Mouse *Atm* further had a shorter 3'UTR as opposed to the human counterpart. Details of human and mouse *ATM* genomic loci are summarised in **Table 5-1.**

Table 5-1 Description of Human and Mouse *Atm* Locus

Features	*ATM*	*Atm*
Genomic Location (GRCh38)	Chromosome 11: 108,222,484-108,369,102 *Forward strand*	Chromosome 9: 53,439,149-53,536,70 *Reverse strand*
Exons	63 exons (62 coding)	64 exons (63 coding)
ORF	9,171 bp	9, 201 bp
Transcripts	25	5
Primary Transcript	ATM-201 (13,147 bp) *NM_000051*	Atm-001 (10,006 bp) *NM_007499*
Protein	3,056 aa	3,066 aa

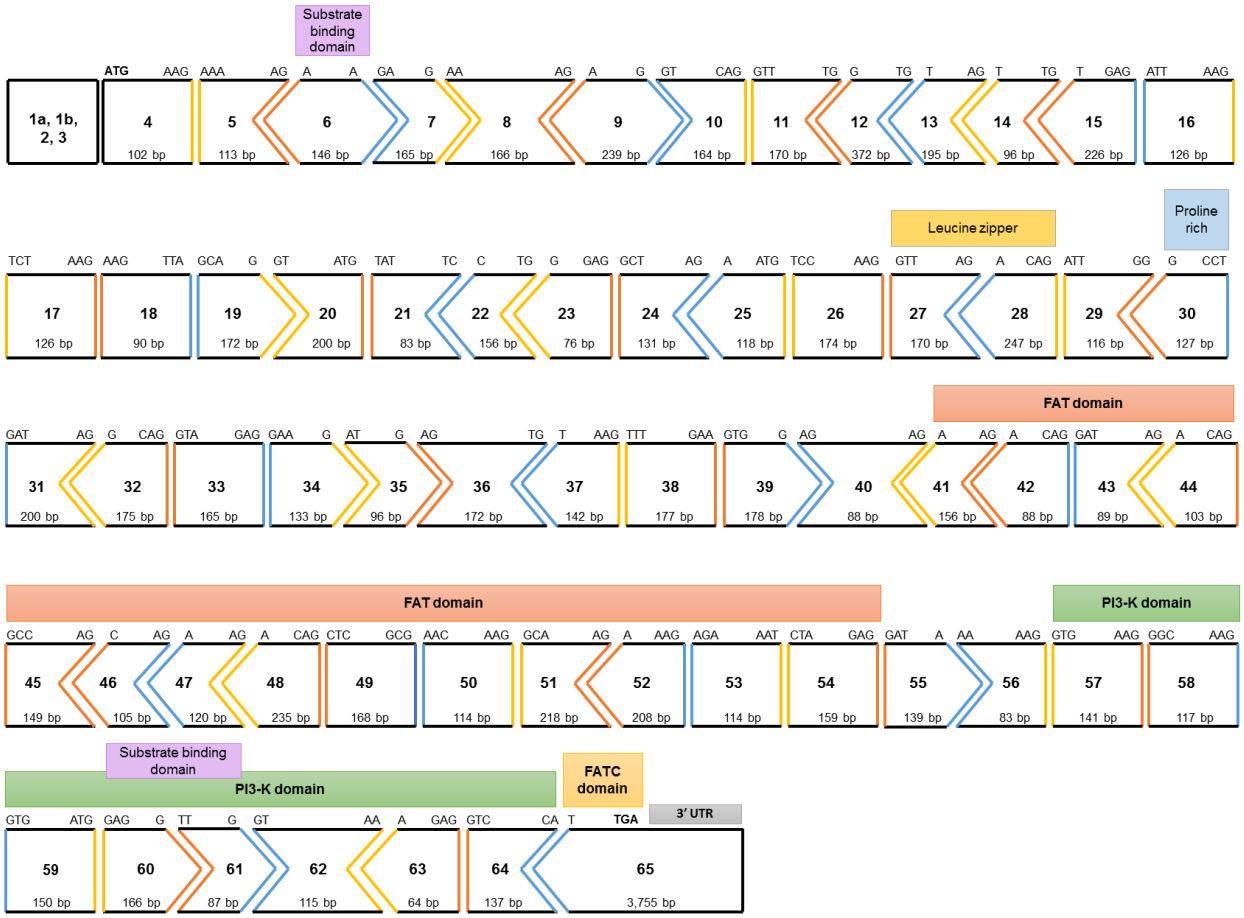

Substrate binding domain
1a, 1b, 2, 3
ATG AAG AAA AG A A GA G AA AG A G GT CAG GTT TG G TG T AG T TG T GAG ATT AAG
4 102 bp
5 113 bp
6 146 bp
7 165 bp
8 166 bp
9 239 bp
10 164 bp
11 170 bp
12 372 bp
13 195 bp
14 96 bp
15 226 bp
16 126 bp
Leucine zipper
Proline rich
TCT AAG AAG TTA GCA G GT ATG TAT TC C TG G GAG GCT AG A ATG TCC AAG GTT AG A CAG ATT GG G CCT
17 126 bp
18 90 bp
19 172 bp
20 200 bp
21 83 bp
22 156 bp
23 76 bp
24 131 bp
25 118 bp
26 174 bp
27 170 bp
28 247 bp
29 116 bp
30 127 bp
FAT domain
GAT AG G CAG GTA GAG GAA G AT G AG TG T AAG TTT GAA GTG G AG AG A AG A CAG GAT AG A CAG
31 200 bp
32 175 bp
33 165 bp
34 133 bp
35 96 bp
36 172 bp
37 142 bp
38 177 bp
39 178 bp
40 88 bp
41 156 bp
42 88 bp
43 89 bp
44 103 bp
FAT domain
PI3-K domain
GCC AG C AG A AG A CAG CTC GCG AAC AAG GCA AG A AAG AGA AAT CTA GAG GAT A AA AAG GTG AAG GGC AAG
45 149 bp
46 105 bp
47 120 bp
48 235 bp
49 168 bp
50 114 bp
51 218 bp
52 208 bp
53 114 bp
54 159 bp
55 139 bp
56 83 bp
57 141 bp
58 117 bp
Substrate binding domain
PI3-K domain
FATC domain
3' UTR
GTG ATG GAG G TT G GT AA A GAG GTC CA T TGA
59 150 bp
60 166 bp
61 87 bp
62 115 bp
63 64 bp
64 137 bp
65 3,755 bp

Figure 5-1

Figure 5-1 Schematic of Human *ATM* Exonic Structure and ATM Protein Domains

ATM consists of 66 exons (exon 1 is an alternative exon 1a or 1b), spanning an open reading frame of 9,171 bp, with the translation initiation codon (ATG) lying in exon 4. Indicated for each exon are number and length in base pairs. Also shown are the sequences of codons adjacent to or overlapping exon boundaries, with graphical depiction of the codon location: vertical lines between exons indicate that the codons are fully contained within the adjacent exons, while left or right pointing exon boundaries reflect the split of the codon at the junction (right: one nucleotide at the 5' exon; left, two nucleotides at the 5' exon). ATM protein domains include the N-terminal substrate binding domain (encoded within exon 6), leucine zipper (exons 27-28), a proline rich region (exon 30), FAT domain (exons 41-54), PI3-K domain (exons 57-64), C-terminal substrate binding domain (exons 60-61), and FATC domain (exon 65).

Source: Ensembl http://www.ensembl.org

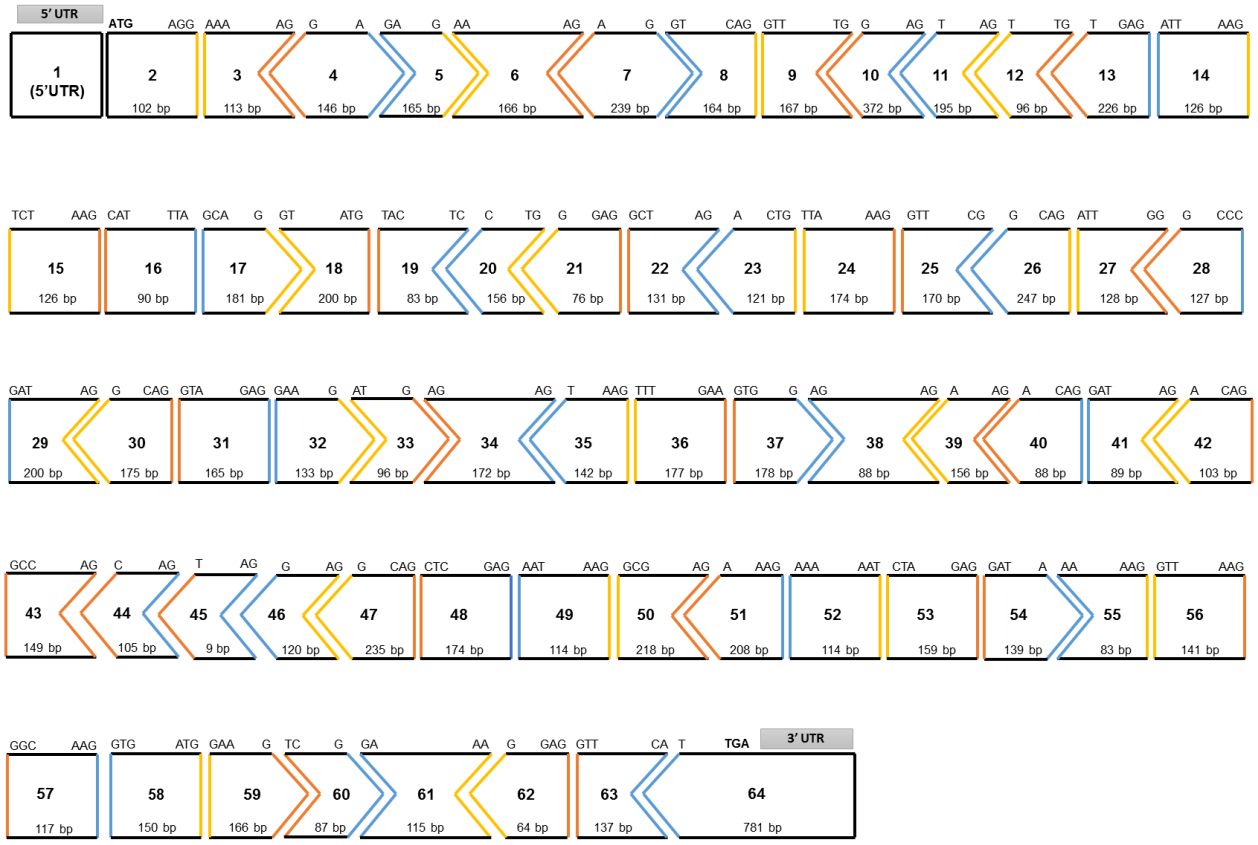

5' UTR
ATG AGG AAA AG G A GA G AA AG A G GT CAG GTT TG G AG T AG T TG T GAG ATT AAG
1 (5'UTR)
2 102 bp
3 113 bp
4 146 bp
5 165 bp
6 166 bp
7 239 bp
8 164 bp
9 167 bp
10 372 bp
11 195 bp
12 96 bp
13 226 bp
14 126 bp
TCT AAG CAT TTA GCA G GT ATG TAC TC C TG G GAG GCT AG A CTG TTA AAG GTT CG G CAG ATT GG G CCC
15 126 bp
16 90 bp
17 181 bp
18 200 bp
19 83 bp
20 156 bp
21 76 bp
22 131 bp
23 121 bp
24 174 bp
25 170 bp
26 247 bp
27 128 bp
28 127 bp
GAT AG G CAG GTA GAG GAA G AT G AG AG T AAG TTT GAA GTG G AG AG A AG A CAG GAT AG A CAG
29 200 bp
30 175 bp
31 165 bp
32 133 bp
33 96 bp
34 172 bp
35 142 bp
36 177 bp
37 178 bp
38 88 bp
39 156 bp
40 88 bp
41 89 bp
42 103 bp
GCC AG C AG T AG G AG G CAG CTC GAG AAT AAG GCG AG A AAG AAA AAT CTA GAG GAT A AA AAG GTT AAG
43 149 bp
44 105 bp
45 9 bp
46 120 bp
47 235 bp
48 174 bp
49 114 bp
50 218 bp
51 208 bp
52 114 bp
53 159 bp
54 139 bp
55 83 bp
56 141 bp
GGC AAG GTG ATG GAA G TC G GA AA G GAG GTT CA T TGA 3' UTR
57 117 bp
58 150 bp
59 166 bp
60 87 bp
61 115 bp
62 64 bp
63 137 bp
64 781 bp

Figure 5-2

Figure 5-2 Schematic of *Atm* (Mouse) Exonic Structure

Atm gene consists of 64 exons (exon 1 comprised of the 5'UTR) with the transcription initiation codon (ATG) lying in exon 2. *Atm* cDNA depicted here contains 62 coding exons spanning an open reading frame of 9, 201 bp. Indicated for each exon are number and length in base pars. Also shown are the sequences of codons adjacent to or overlapping exon boundaries, with graphical depiction of the codon location: vertical lines between exons indicate that the codons are fully contained within the adjacent exons, while left or right pointing exons boundaries reflect the split of the codon at the junction (right: one nucleotide at the 5; exon; left, two nucleotides at the 5; exon).

Source: Ensembl http://www.ensembl.org

5.2.2 Characterisation of A-T causing mutations

Mutations known to cause Ataxia Telangiectasia were obtained from the Human Gene Mutation Database (HGMD®) to construct a mutation map **(Fig. 5-3)**. As of March 2016, over 600 different A-T causing mutations had been identified and published. As indicated in the map, these mutations span across entire genomic region of *ATM* without specific mutational 'hotspots'. Details of mutations depicted on the map including their type and location on the *ATM* are listed in ***Appendix I.***

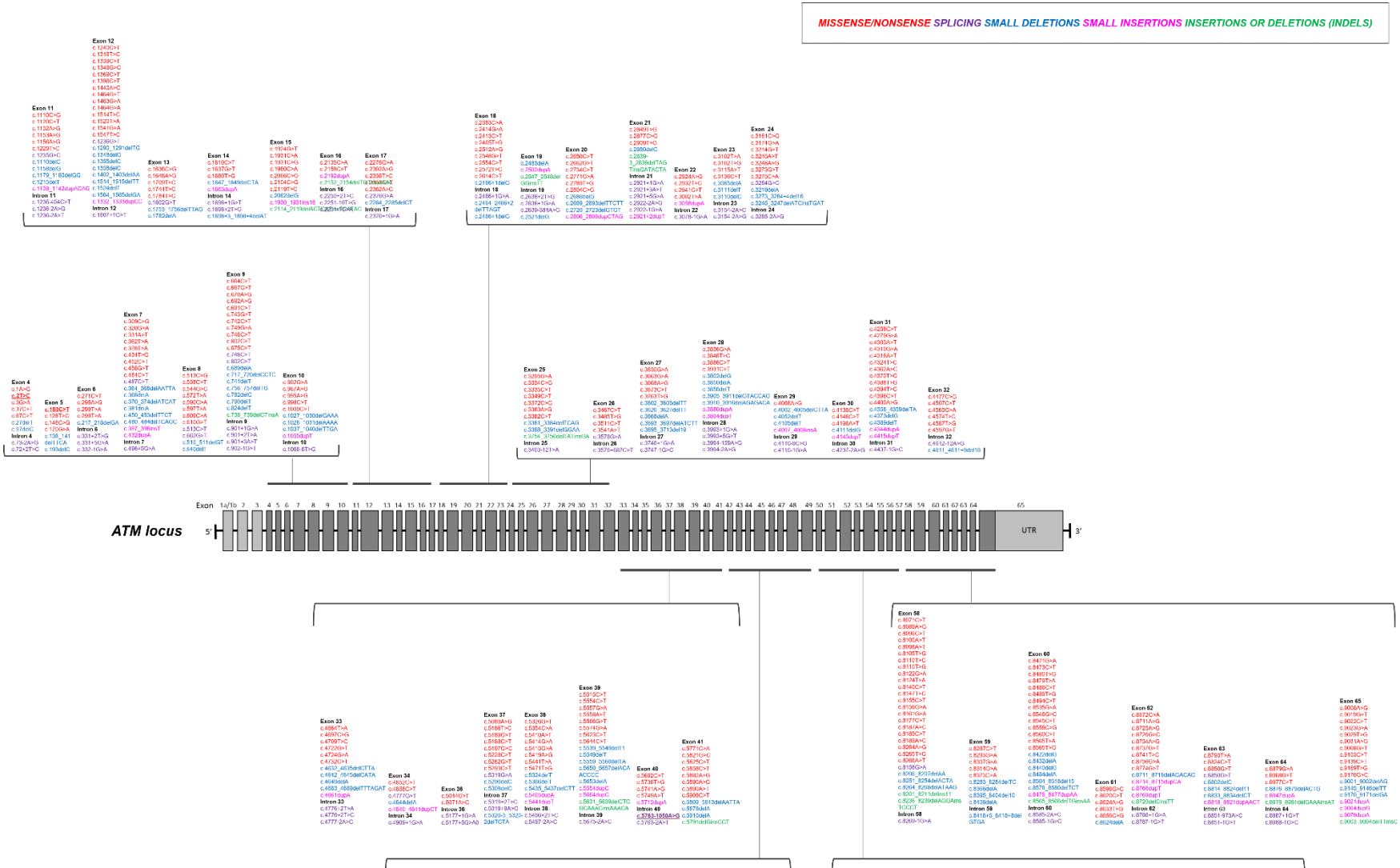

MISSENSE/NONSENSE SPLICING SMALL DELETIONS SMALL INSERTIONS INSERTIONS OR DELETIONS (INDELS)
ATM locus
5'
3'
UTR
Exon 1a/1b 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65

Figure 5-3

Figure 5-3 *ATM* Mutation Map

Over 600 mutations across the *ATM* gene locus have been identified as disease-causing leading to ataxia telangiectasia (A-T) or A-T variant phenotype, which is clinically milder. Depicted is the ATM locus indicating the exonic structure. Different A-T mutation types are colour coded, including nonsense/missense and splicing mutations, small deletions, small insertions, and insertions/deletions. Regulatory mutations, gross insertions or deletions and complex rearrangements have not been included. Disease-causing mutations are spread across ATM with no evident hot-spots. The majority are missense/nonsense mutations that either lead to truncation of ATM protein, or production of an unstable protein that is rapidly degraded or a mutant protein with abolished or residual protein kinase activity. Splicing mutations lead to loss of exons or intron retention. Mutations described in chapter to be targeted using CRISPR/Cas have been highlighted bold and italic, and include: (i) missense ATM2T>C located in exon 4, (ii) missense ATM103C>T located in exon 5, (iii) splicing ATM5762ins137 c.5763-1050A>G located in intron 40, and (iv) small deletion ATM7638del9 c.7638_7646del located in exon 54.

Source: The Human Gene Mutation Database (HGMD) http://www.hgmd.cf.ac.uk/ac/index.php (updated 26 March 2016)

5.2.3 *In silico* design of genome editing strategies

5.2.3.1 A NHEJ-mediated strategy to target the *ATM* 5762ins137 mutation

ATM 5762ins137 is a disease-causing mutation that lies in the intron 40 of *ATM*. The A to G substitution mutation at the intronic position -1050 activates an intronic cryptic splice donor site resulting in aberrant splicing of a 137-bp intronic sequence into the mature *ATM* transcript. The aberrant transcript is thought to produce an unstable protein, which then rapidly gets degraded. The 5762ins137 mutation however is “leaky”, which causes low level production of the correct transcript. This causes residual kinase activity in patients with this mutation. 5762ins137 mutation thus confers the variant A-T phenotype and is prevalent in 10-15% of all UK patients (Stankovic, Kidd et al. 1998). To target the *ATM* 5762ins137 A>G point mutation, CRISPR-Cas technology can be used to induce NHEJ indels at or near the ectopic intronic splice site. Presence of NHEJ-mediated indels would disrupt the cryptic splice site allowing for the formation of normal transcript without intervening sequence **(Fig. 5-4).**

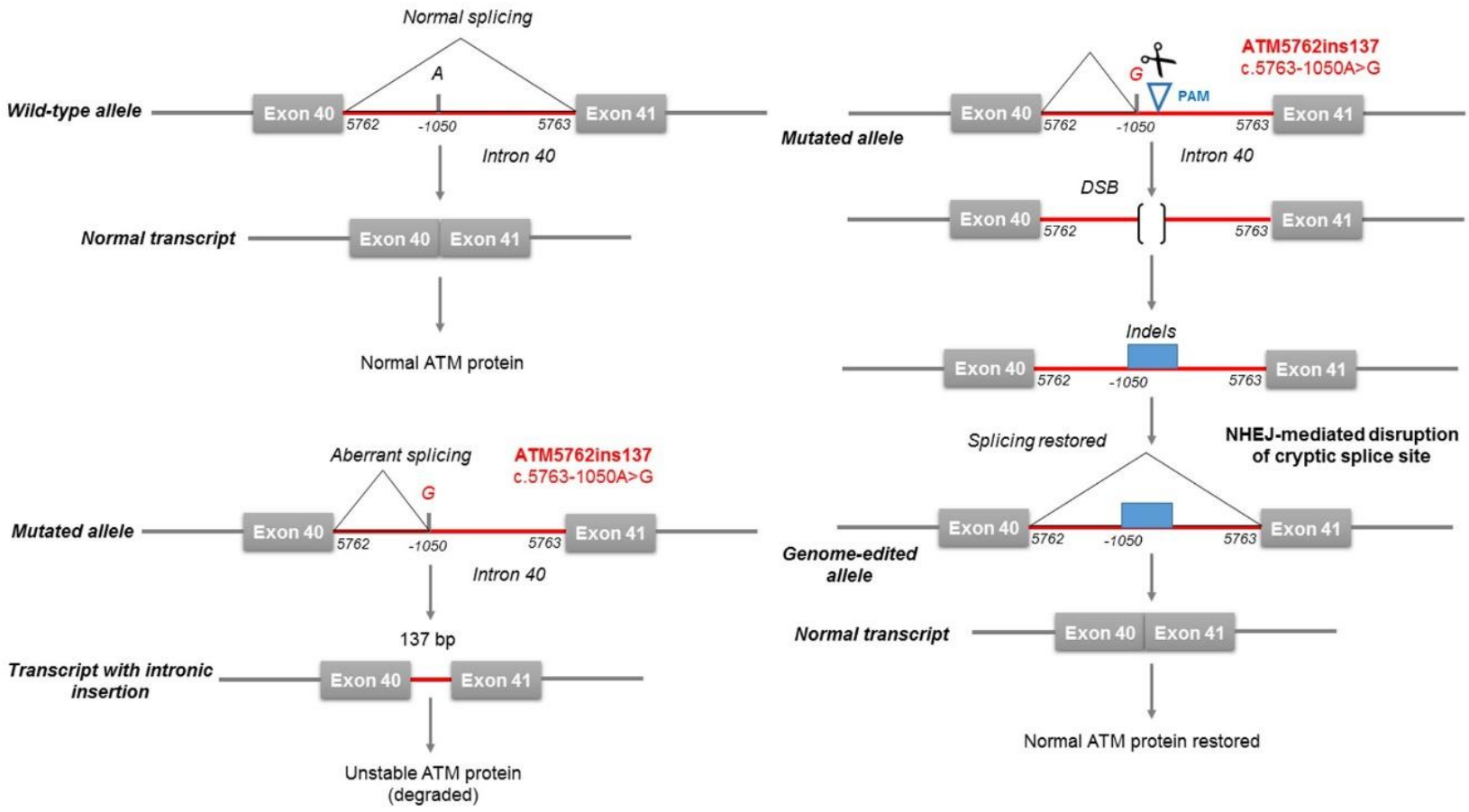

Normal splicing
A
Wild-type allele
Exon 40
5762
-1050
5763
Exon 41
Intron 40
Normal transcript
Exon 40
Exon 41
Normal ATM protein
Aberrant splicing
ATM5762ins137
c.5763-1050A>G
G
Mutated allele
Exon 40
5762
-1050
5763
Exon 41
Intron 40
137 bp
Transcript with intronic insertion
Exon 40
Exon 41
Unstable ATM protein
(degraded)
G
PAM
ATM5762ins137
c.5763-1050A>G
Mutated allele
Exon 40
5762
-1050
5763
Exon 41
Intron 40
DSB
Exon 40
5762
5763
Exon 41
Indels
Exon 40
5762
-1050
5763
Exon 41
Splicing restored
NHEJ-mediated disruption of cryptic splice site
Genome-edited allele
Exon 40
5762
-1050
5763
Exon 41
Normal transcript
Exon 40
Exon 41
Normal ATM protein restored

Figure 5-4

Figure 5-4 NHEJ-Based Disruption of ATM5762ins137A>G Mutation

(Left) ATM5762ins137A>G is a commonly occurring mutation that affects 10-15% of all A-T patients in the UK. Located in intron 40 of *ATM,* A>G transversion at c.5763-1050 point mutation leads to the activation of a cryptic splice site, which causes aberrant splicing and insertion of a 137-bp intronic sequence in the transcript. This transcript produces an unstable ATM protein which is rapidly degraded.

(Right) Genome editing can be used to disrupt this mutation and restore wild-type protein production. gRNAs designed in close proximity to this mutation are denoted by the protospacer adjacent motif (PAM), which is where CRISPR/Cas nuclease is directed to cleave the DNA and create a double-stranded break (DSB). Creation of a DSB activates the error-prone NHEJ DNA repair pathway which re-joins the DNA ends by creating small insertions/deletions (indels) at the break site. Presence of indels causes frameshift mutation which can lead to disruption of the cryptic site allowing restoration of normal splicing between exons 40 and 41, and production of normal ATM transcript and protein.

5.2.3.2 HDR strategies for repair small *ATM* mutations: applicable to 2T>C, 103C>T and 7638del9

Small mutations can be corrected using homology-directed repair with single-stranded oligonucleotide donor DNA templates. Such a strategy could therefore be applicable to correction of a variety of nonsense, missense, and small insertions as well as deletions. Commonly reported nonsense mutations *ATM* 2T>C, 103C>T and the deletion mutation 7638del9 were identified for HDR. ATM 2T>C mutation, which lies on exon 4, is observed several times in the UK population. The ATM 103C>T mutation in exon 5 is a founder mutation commonly reported in Jewish families. Both mutations lead to absence of protein (Gilad, Bar-Shira et al. 1996). The *ATM* 7638del9 mutation, is an in-frame deletion of 9 nucleotides in exon 54 (codons 2546-2548), which is known to destabilize the protein and abolish its kinase activity (Laake, Telatar et al. 1998). HDR of these mutations would require induction of a double-stranded break near the mutation site and the correct or deleted sequence to be supplied on a short single-stranded DNA template. **(Fig. 5-5)** illustrates HDR approaches for all three mutations.

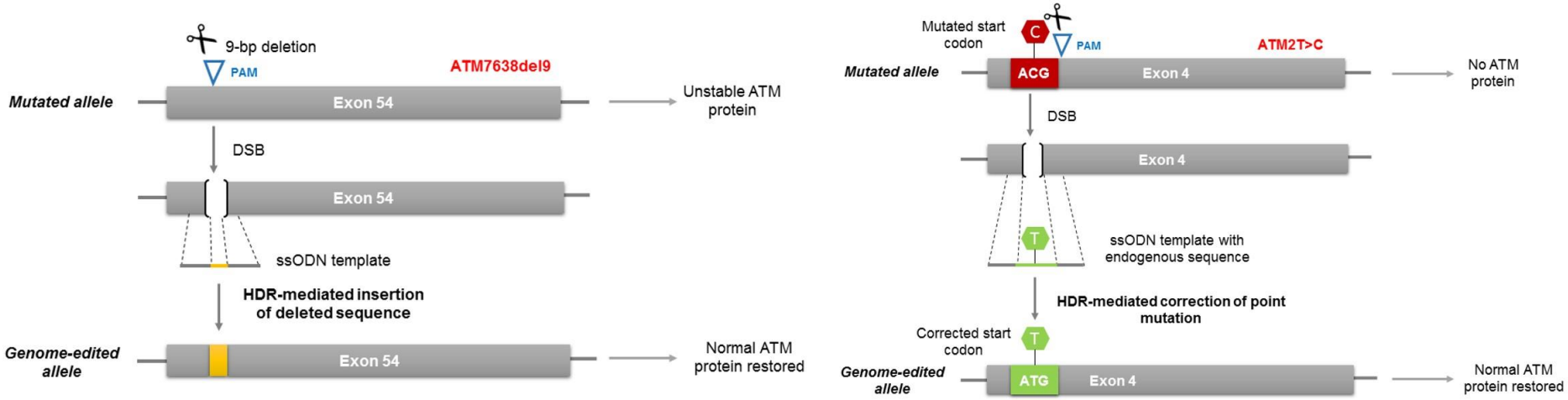


Figure 5-5 HDR Mediated Correction of Small *ATM* Mutations

ATM7368del9 and ATM2T>C are two commonly reported A-T causing mutations. ATM7368del9 is a 9-bp (TAG AAT TTC) deletion in exon 54 which leads to the formation of a non-functional ATM protein. ATM2T>C is a missense mutation that disrupts the translation initiation codon (ATG-to-ACG) in exon 4, as a result of which no ATM protein is produced. Homology-dependent genome editing can be used to permanently correct these mutations. In order to do so, gRNAs are designed in close proximity to each mutation (denoted by PAM), which direct CRISPR/Cas nuclease to introduce a DSB. The correct endogenous sequence can be supplied on a synthetic single-stranded oligonucleotide (ssODN) repair template. Creation of a DSB and presence of the repair template would allow homologous recombination to take place, thus correcting the mutation leading to normal protein production. Such a strategy could also be applied across a variety of similar patient-specific mutations like ATM103T>C.

5.2.3.3 Universal *ATM* genome editing strategies

Two different knock-in approaches were designed for universal gene editing of *ATM*. (1) knock-in of 5' or 3' halves of an *ATM* cDNA into the endogenous ATM locus, and (2) knock-in of a full-length *ATM* expression cassette into the AAVS1 safe harbour locus. Both strategies would require the use of large *ATM* cDNA fragments, providing either half or the whole *ATM* open reading frame.

The 5' knock-in would be applicable to repair mutations in the 5' end of the endogenous ATM locus. This can be achieved by inducing a double cut at introns 3 (i.e. just before the start codon) and 32, which lies mid-way of the gene, using two separate gRNAs and Cas9 nuclease. Use of dual gRNAs would enable the excision of the 5' end of the gene. A repair cassette including arms of homology, *ATM* exons 4-32, a synthetic splice donor, and a downstream intronic splicing enhancer could then be used to knock-in the 5' cDNA block **(Fig. 5-6)**. Likewise, mutations in the 3' half of the endogenous ATM locus could be repaired by 3' knock-in. This can be achieved by inducing a single cut in intron 32 using single gRNA. The synthetic repair template would include arms of homology, a synthetic splice acceptor, *ATM* exons 33-65, and a transcription terminator (a standard sequence incorporating three stop codons and the bovine growth hormone polyA element) **(Fig. 5-6)**. With either approach, the repaired *ATM* locus will include a superexon providing the 5' or 3' half of the synthetic cDNA generating a chimeric transcript.

Another universal approach would be utilising the targeting of human AAVS1 (or adeno-associated virus integration site 1) "safe harbour" locus for the knock-in of full synthetic *ATM* cDNA cassette by HDR. AAVS1 is a human locus which encodes the constitutively expressed *PPP1R12C* gene, in which transgenes can be knocked-in resulting in stable transgenic expression and no obvious detrimental effect. This can be achieved by inducing a CRISPR-Cas cut in AAVS1 in addition to introduction of a synthetic template including arms of homology, a promoter, *ATM* exons 4-65, a mutated WPRE and a transcription terminator. HDR with the synthetic donor would result in transgenic expression of a full-length *ATM* from the AAVS1 locus **(Fig. 5-7)**.

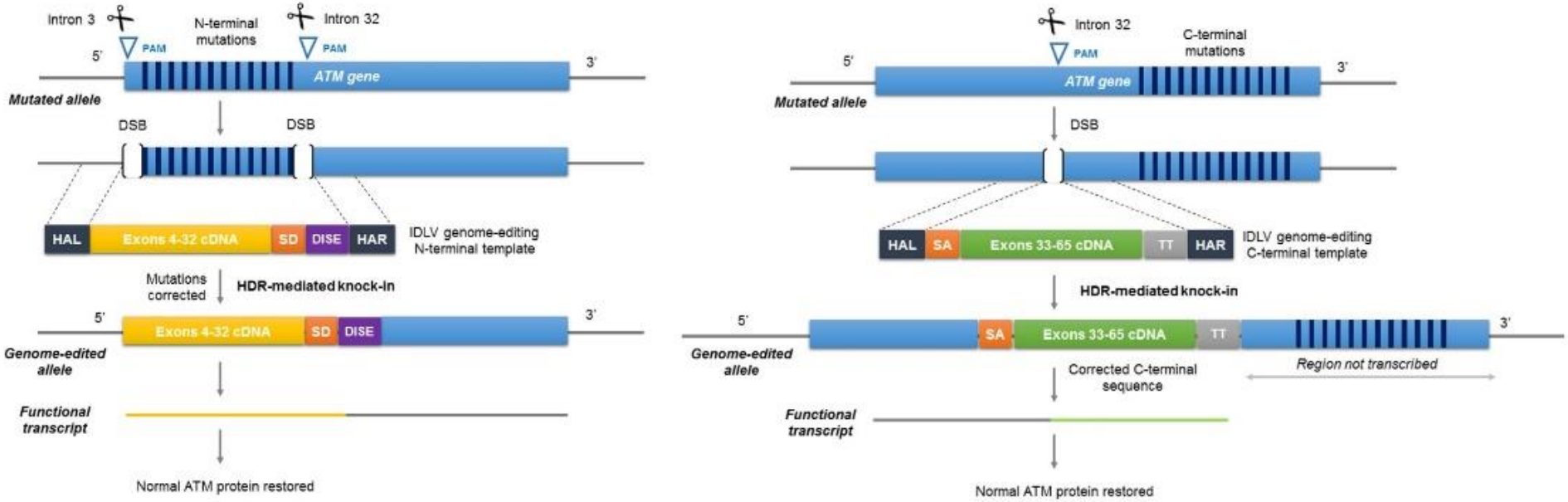

Intron 3
N-terminal mutations
Intron 32
PAM
PAM
5'
3'
ATM gene
Mutated allele
DSB
DSB
HAL
Exons 4-32 cDNA
SD
DISE
HAR
IDLV genome-editing N-terminal template
Mutations corrected
HDR-mediated knock-in
5'
3'
Exons 4-32 cDNA
SD
DISE
Genome-edited allele
Functional transcript
Normal ATM protein restored
Intron 32
C-terminal mutations
PAM
5'
3'
ATM gene
Mutated allele
DSB
HAL
SA
Exons 33-65 cDNA
TT
HAR
IDLV genome-editing C-terminal template
HDR-mediated knock-in
5'
3'
SA
Exons 33-65 cDNA
TT
Genome-edited allele
Corrected C-terminal sequence
Region not transcribed
Functional transcript
Normal ATM protein restored

Figure 5-6

Figure 5-6 Partial Knock-In of *ATM* cDNA Cassettes

ATM cDNA will be divided into two blocks – exons 4 to 32 and exons 33-65 - containing sequences encoding for the N- and C- terminal domains of ATM protein, respectively. Targeted insertion of these cDNA cassettes would allow for correction of any A-T causing mutation lying within these regions; a strategy that could be applied across multitude of patients harbouring different A-T mutations. Exogenous N- or C-terminal terminal synthetic cDNA cassettes supplied on an integration-deficient lentiviral vector (IDLV) will be inserted into the 5' or 3' end of the ATM locus by a homology-dependent genome editing strategy. (Left) For 5' knock-in to take place, firstly, the intervening genomic sequence between introns 3 and 32 will be excised. This will be done by designing gRNAs in these introns (denoted by PAM) which will direct the CRISPR/Cas nuclease to create two DSBs. Simultaneously, an exogenous N-terminal gene editing template consisting of cDNA encoding exons 4-32 (~4.6 kb) flanked by homology arms left (HAL) and right (HAR), a splice donor (SD) and DISE element will be supplied. This would allow homologous recombination between the genome-editing template and endogenous *ATM* locus, and targeted insertion of exons 4-32 at the 5' end of the locus. Presence of downstream splicing elements would ensure formation of a functional transcript and normal ATM protein. (Right) For targeted insertion of C-terminal cassette, a single DSB will be introduced in intron 32. A C-terminal template consisting of HAL, a splice acceptor (SA), exons 33-65 cDNA (~4.6 kb), a transcription terminator (TT), and HAR will be supplied exogenously on an IDLV. Successful homologous recombination would lead to insertion of this genome-editing template in intron 32. The resulting transcript will encode sequences from the endogenous 5' end and supplied 3' cDNA block resulting in normal protein production.

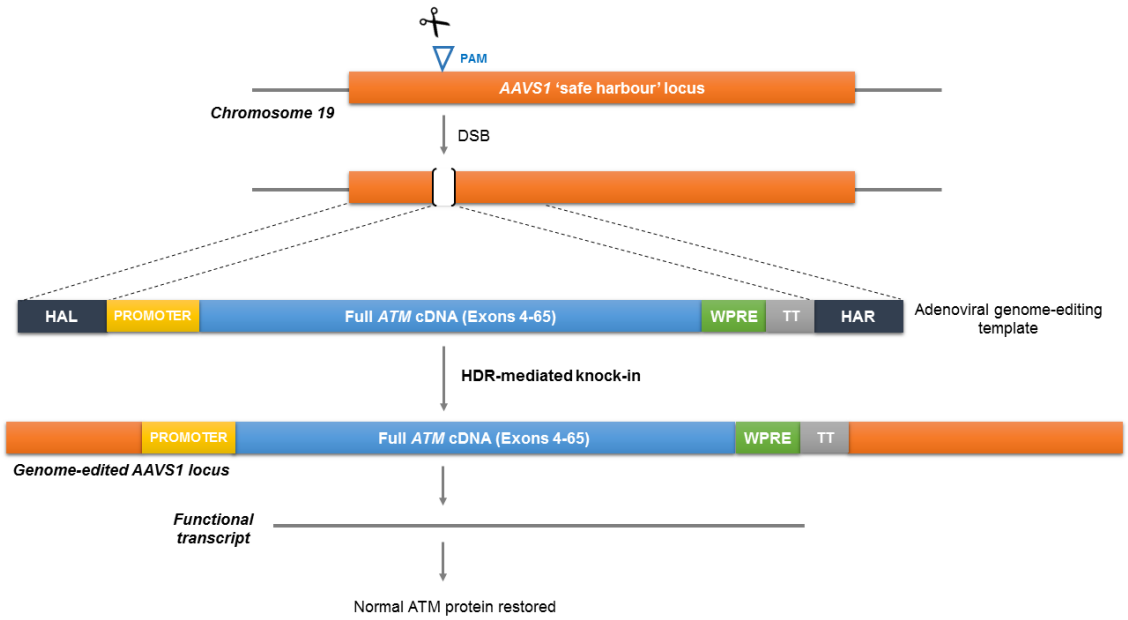


Figure 5-7 Targeted knock-in of full-length *ATM* cDNA into the AAVS1 'safe harbour' locus

AAVS1, located in chromosome 19 in the human genome, encodes for the constitutively expressed PPP1R12C gene. This region is known as a 'safe harbour' because it is amenable to the integration of a transgene and allows for its stable expression. Targeted integration of a full-length *ATM* cDNA cassette into the AAVS1 can be used as a universal therapeutic strategy of benefit to A-T patients with recessive mutations. HDR-mediated knock-in will be achieved by introducing a DSB in the human AAVS1 locus using the CRISPR/Cas system (denoted by PAM) and by simultaneously supplying *ATM* cDNA and regulatory sequences on an exogenous synthetic template. The genome-editing template will include arms of homology corresponding to the AAVS1 locus, a short promoter (*ATM*-specific or hPGK), *ATM* cDNA encoding exons 4-65 (9.2 kb), Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) and a transcription terminator (TT). Owing to its large size, the genome-editing template can be supplied on an adenoviral vector. Successful insertion of the template into the AAVS1 locus would lead to the generation of a functional *ATM* transcript restoring ATM protein levels.

5.2.4 *In silico* design of humanised A-T mouse model

Animal models are essential for the future validation of the genome editing strategies described thus far. However, the current A-T animal models are not suitable to answer how many cells need to be corrected to achieve clinical benefit. There are at least two further problems with the existing A-T mouse models: they are not humanised hence they do not carry a relevant section of the human gene as the target for therapeutic correction, and their phenotype does not adequately reflect the neurodegeneration observed in patients. To begin to address these issues, we decided to pursue generation of a humanised mouse model of A-T, as a stepping stone towards the future generation of a set of humanised A-T models. We have explored in silico the creation of such model for the human ATM5762ins137 mutation. This mutation has been chosen as it can be corrected both by NHEJ as well as HDR.

A transgenic A-T mouse model can be created by partial knock-in of a portion of *ATM* human sequence into the mouse *Atm* locus using HR. In order to develop a humanised A-T mouse model, the intronic splicing mutation *ATM* 5762ins137 A>G was chosen. The mutation lies in intron 40 of human ATM gene. Sequence analysis of human and mouse ATM gene sequences around the mutation site revealed high levels of similarity in the structure of these introns and exons **(Table 5-2)**. Local alignment of *ATM* exons 40 and 41 with *Atm* exons 37 and 38 showed 90.9% and 82.3% sequence similarity, respectively **(Fig. 5-8)**.

a

```
  1 ACCTTCTTCAGGAACAATTTTTAATGATGCTTTCTGGCTGGATTTAAATT     50
    ||||||.||.||||||..||||.||||.||||||||||||||||.||||
  1 ACCTTCCTCGGGAACAGCTTTTGATGACGCTTTCTGGCTGGATTTGAATT     50

 51 ATCTAGAAGTTGCCAAGGTAGCTCAGTCTTGTGCTGCTCACTTTACAGCT    100
    ||||.||.|||||.|||||.||||||||.||..||||||||||.||.||.
 51 ATCTTGAGGTTGCGAAGGTGGCTCAGTCCTGCTCTGCTCACTTCACGGCC    100

101 TTACTCTATGCAGAAATCTATGCAGATAAGAAAAGTATGGATGATCAAGA    150
    ||.|||||.|||||.||||||.||||||||||||||.|..||.||.|||||
101 TTGCTCTACGCAGAGATCTATTCAGATAAGAAAAGCACAGACGAGCAAGA    150

151 GAAAAG    156
    ||||||
151 GAAAAG    156
```

b

```
  1 AGTCAGAGCACTTTTTCCGATGCTGTTTGGATAAAAAATCACAAAGAACA     50
    ||||||||.|||||.||||||||||||||||||||||.||||||||||.
  1 AGTCAGAGAACTTTCTCCGATGCTGTTTGGATAAAAAGTCACAAAGAACC     50

 51 ATGCTTGCTGTTGTGGACTACATGAGAAGACAAAAGAG     88
    ||||||||||||||.|||||..|||||||.||||||||
 51 ATGCTTGCTGTTGTCGACTATCTGAGAAGGCAAAAGAG     88
```

Figure 5-8 Pairwise Sequence Alignment of Human and Mouse *ATM* Locus

(a) ATM exon 41 (top) and Atm exon 39 (bottom) showed 82.3% similarity. (b) *ATM* exon 40 (top) and *Atm* exon 38 (bottom) showed 90.9% similarity. Matches are indicated by straight line. Dots denote mismatch between the sequences aligned. Source: EMBOSS Matcher

Table 5-2 Comparison of Human and Mouse *ATM* Around 5762insA>G

Human *ATM*		Mouse *Atm*	
Intron 39	*3,044 bp*	*Intron 37*	*5,877 bp*
Exon 40	88 bp	Exon 37	88 bp
Intron 40	*2,175 bp*	*Intron 38*	*2,108 bp*
Exon 41	156 bp	Exon 38	156 bp
Intron 41	*2,095 bp*	*Intron 38*	*3,299 bp*

In order to carry out HDR, human ATM sequence can be supplied on an exogenous template containing a portion of human *ATM* harbouring the 5762ins137 point mutation (intron 39 to intron 41), flanked by homologous arms derived from *Atm*. The knock-in strategy including design of the exogenous knock-in DNA template is depicted in **(Fig. 5-9)**.

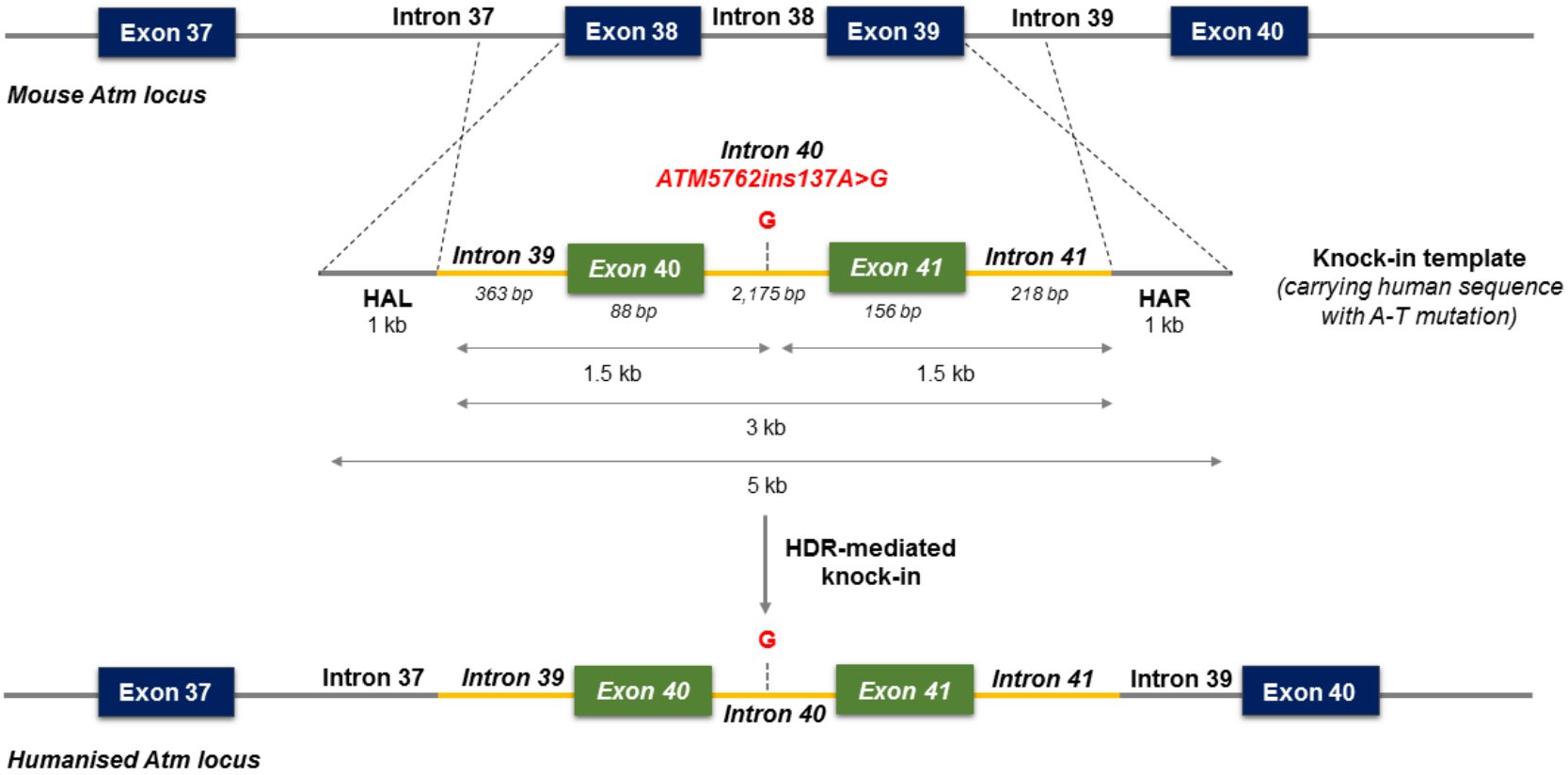

Exon 37
Intron 37
Exon 38
Intron 38
Exon 39
Intron 39
Exon 40
Mouse Atm locus
Intron 40
ATM5762ins137A>G
G
Intron 39
Exon 40
Exon 41
Intron 41
HAL
1 kb
363 bp
88 bp
2,175 bp
156 bp
218 bp
HAR
1 kb
Knock-in template
(carrying human sequence
with A-T mutation)
1.5 kb
1.5 kb
3 kb
5 kb
HDR-mediated
knock-in
G
Exon 37
Intron 37
Intron 39
Exon 40
Intron 40
Exon 41
Intron 41
Intron 39
Exon 40
Humanised Atm locus

Figure 5-9

Figure 5-9 Proposed Knock-In Strategy for Humanised A-T Mouse Model

A humanised A-T mouse model can be created by using homology-directed repair (HDR) strategy to knock-in a portion of human *ATM* gene into the mouse Atm locus. The exogenous ATM donor template would contain a 3 kb human *ATM* sequence carrying the A-T causing ATM5762ins137 mutation flanked between 1 kb homology arms left (HAL) and right (HAR) each derived from mouse Atm locus. The ATM5762ins137 mutation is a commonly occurring mutation that affects 10-15% of all A-T patients in the UK.

5.2.5 *In silico* design of gRNAs

Single spCas9 gRNAS targeting regions across *ATM* locus were designed using web-based bioinformatics tools CHOPCHOP v2, Cas-Designer or Benchling. The guides were designed to match a 20-nt target in the input DNA sequence from the target genomic region adjacent to the 'NGG' PAM specific for spCas9 nuclease. All guides were designed with cut sites within 100 bp of the region to be edited. A single gRNA targeting the human AAVS1 locus was designed using Benchling.

Details of *ATM* and h*AAVS1* guides are mentioned in **Table 5-3 and Table 5-4** respectively.

Table 5-3 Human *ATM* CRISPR-Cas gRNAs

ATM Target	gRNA Name	gRNA Sequence (5' - 3')	PAM (*sp*Cas9)	Location of Cut Site	Design Tool
2T>C (Exon 4)	G.2T>C	ATGCATATATAGAGAGAAAG	AGG (-strand)	66 bp upstream (intron 3)	Cas-Designer
103C>T (Exon 5)	G.103C>T	TACTAATCACACTTATTTCA	AGG (-strand)	68 bp upstream (intron 4)	Cas-Designer
5762ins137 (Intron 40)	G.5762insA>G	TGAATGGGATATAGAAAAAC	GGG	9 bp upstream (intron 40)	CHOPCHOP v2
7368del9 (Exon 53)	G.del9	TACTGTCTAGATACTGCAGT	GGG (-strand)	112 upstream bp (intron 52)	Cas-Designer
Intron 3	G.int3	GCCTTGCTTGCAGTAGTAGC	AGG (-strand)	*NA*	Cas-Designer
Intron 32	G.int32	GTAGAGAGGTAGTCAAAACT	AGG	*NA*	CHOPCHOP v2

Table 5-4 Human AAVS1 gRNA

Target Region	gRNA Name	gRNA Sequence (5' - 3')	PAM (*sp*Cas9)	Location of Cut Site	Design Tool
Human *AAVS1*	G.AAVS1	GTCACCAATCCTGTCCCTAG	TGG	*NA*	Benchling

5.2.6 *ATM* gene editing in human cells

ATM gene editing was carried out in HEK293T cells stably expressing Cas9 from the AAVS1 locus. gRNAs targeting *ATM* regions – 2T>C, 103C>T, 5762ins137, intron 3 and intron 30 - were transfected as synthetic RNA molecules and genomic DNA was harvested post-transfection. In order to detect indels generated as a result of gene editing, genomic regions spanning the Cas9 cut site were amplified and sequenced before and after transfection. (**Fig. 5-10)** shows PCR amplicons from all five *ATM* target regions.

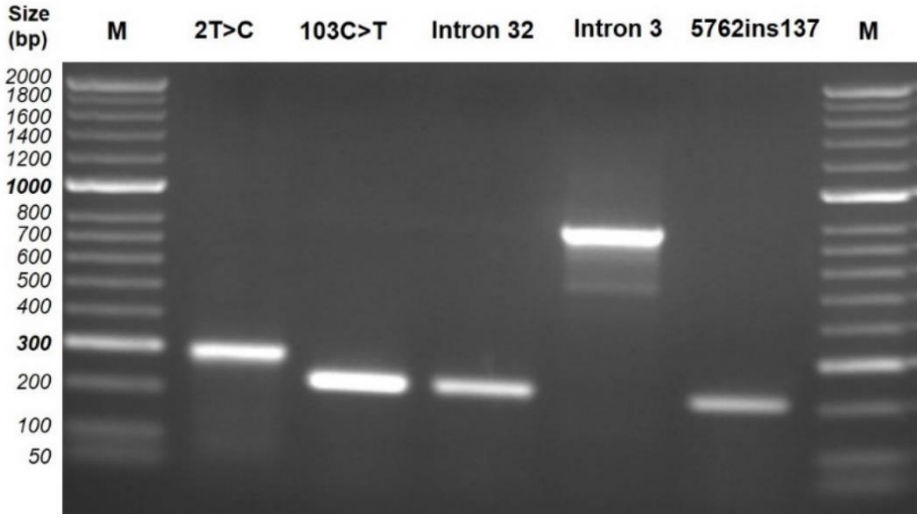


Figure 5-10 Cas9 cleavage site amplifications from *ATM* mutations

Genomic regions spanning Cas9 cut sites at different ATM target sites were amplified using PCR from HEK293T DNA. Gel image shows the resulting amplicons from ATM targets: 2T>C (289 bp), 103C>T (225 bp), intron 32 (223 bp), intron 3 (768 bp) and 5762ins137 (211 bp). M, Hyperladder 50 bp; Gel, 1% agarose, 1X TAE buffer.

Sequence read from control cells showed a clean sequence; except for 103C>T mutation due to a naturally occurring polymorphism in the cell line used, and hence was excluded from the analysis. Gene editing was carried out as normal at the other targets. Upon gene editing, the traces consisted of a mixture of sequences indicating a heterozygous cell population consisting of edited and unedited cells. In case of ATM5762ins mutation locus, cutting by Cas9 led to creation of NHEJ-mediated indels disrupting the mutation site **(Fig. 5-11)**.

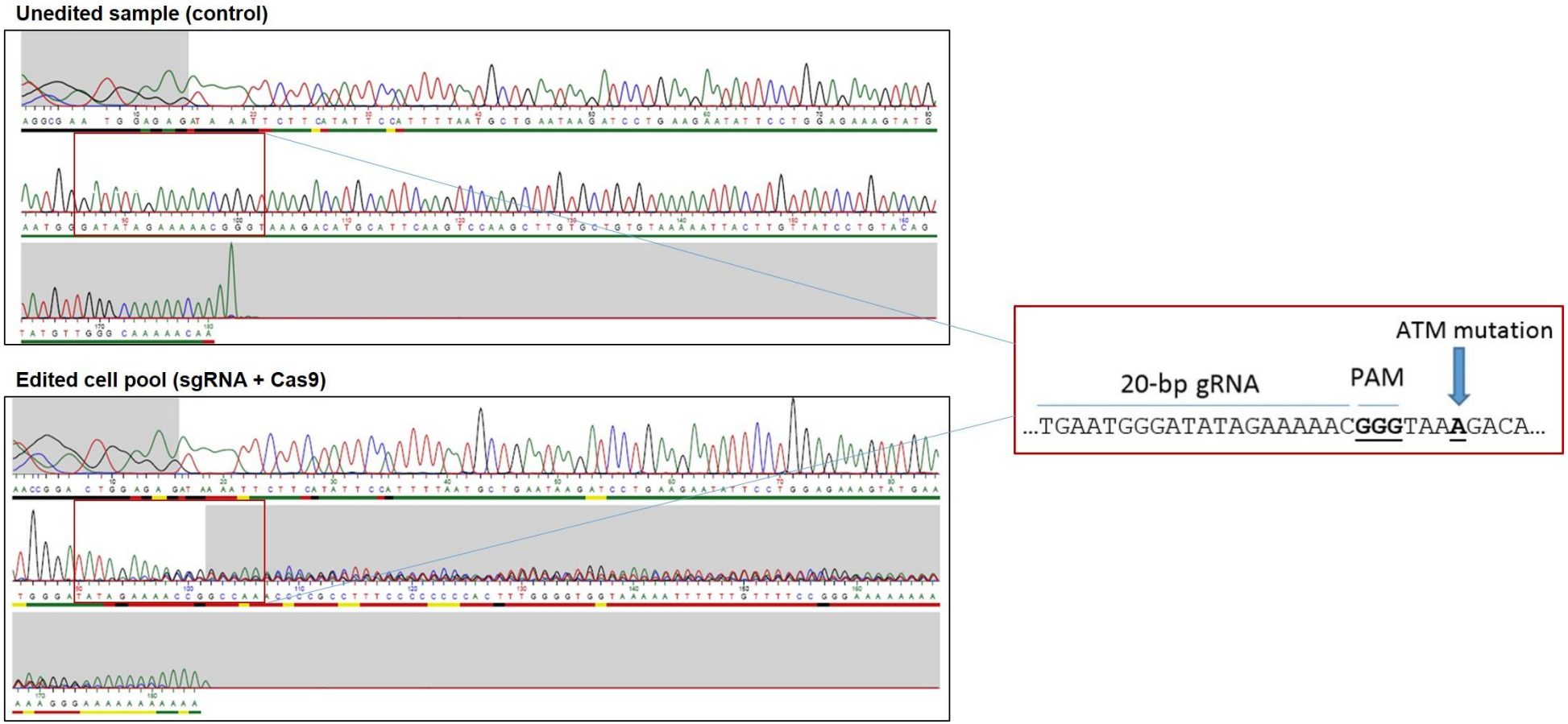

Unedited sample (control)
Edited cell pool (sgRNA + Cas9)
ATM mutation
20-bp gRNA
PAM
...TGAATGGGATATAGAAAAACGGGTAAAGACA...

Figure 5-11

Figure 5-11 Gene editing of ATM5762ins137 mutation

ATM5762ins137 splicing mutation was targeted by introducing a specific gRNA in Cas9 expressing HEK-293T cells. Genomic region around the Cas9 cleavage site was amplified from edited and unedited cells and subjected to Sanger sequencing. Sequence obtained from unedited or control cells showed a uniform sequence (top left) while sequence obtained from edited cell pool consisted of a mixture of signals (bottom left). Red rectangle reflects the Cas9 target and cut sites and location of *ATM* mutation. Cutting by Cas9 led to introduction of insertions and deletions causing frameshift mutations at the ATM5762ins137 mutation locus.

In order to quantify the indels generated, sequence traces from edited and control samples were analysed using the bioinformatics tool TIDE. **Table 5-5** summarises the overall gene editing frequency of various *ATM* target regions obtained using TIDE. Apart from overall efficiency of a gRNA, TIDE further gives a profile of types of indels generated **(Fig. 5-12)**.

Table 5-5 Editing Efficiency of *ATM* gRNAs Using TIDE

***ATM* Target**	**Indels (%)**
2T>C	79.8
5762ins137A>G	51.6
Intron 3	22.9
Intron 32	21.9

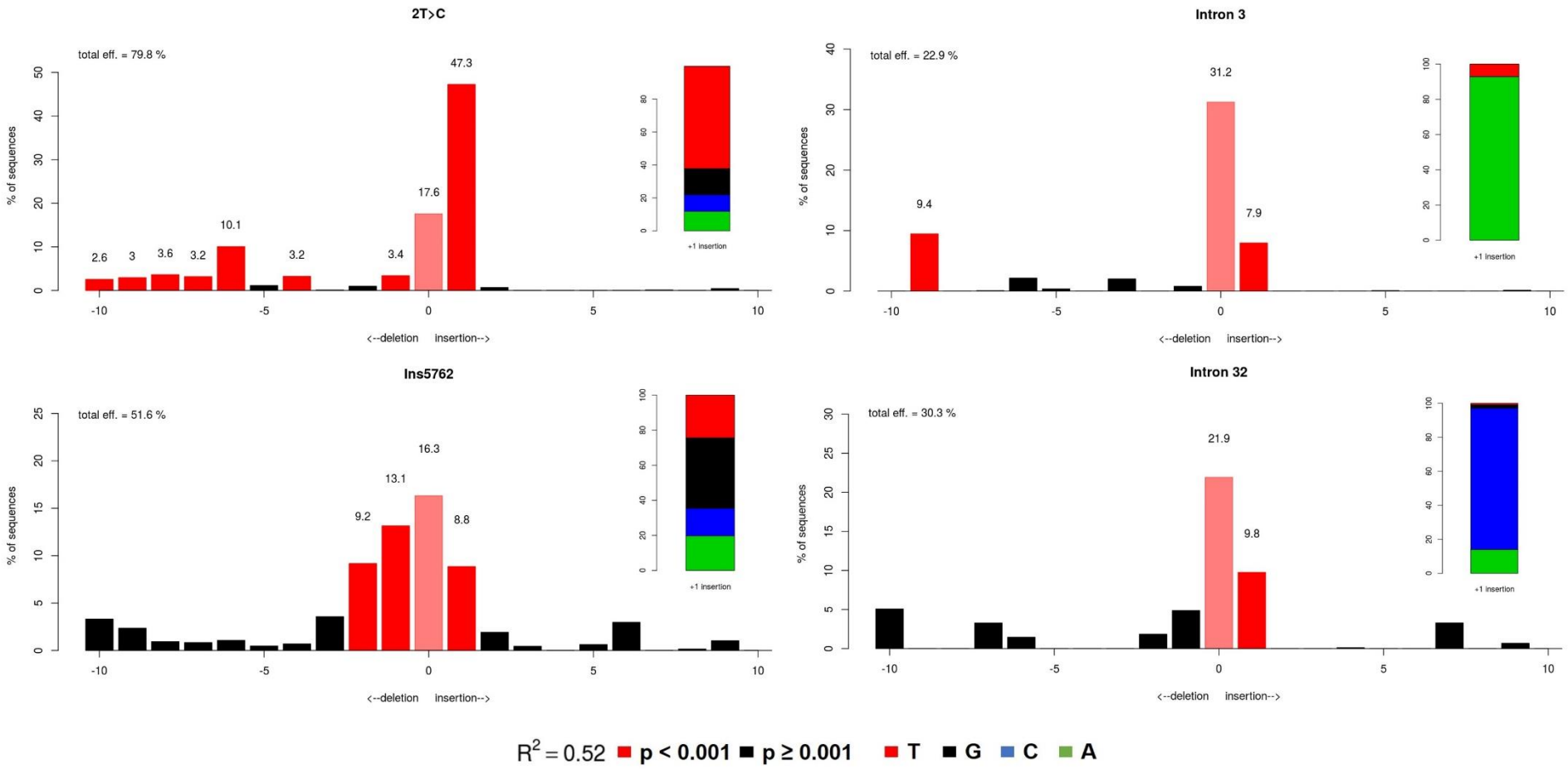

2T>C
total eff. = 79.8 %
% of sequences
2.6
3
3.6
3.2
10.1
3.2
3.4
17.6
47.3
+1 insertion
<--deletion insertion-->
Intron 3
total eff. = 22.9 %
9.4
31.2
7.9
Ins5762
total eff. = 51.6 %
9.2
13.1
16.3
8.8
Intron 32
total eff. = 30.3 %
21.9
9.8
R2 = 0.52 p < 0.001 p ≥ 0.001 T G C A

Figure 5-12

Figure 5-12 Indel Profiles for *ATM* Targets Using TIDE

Various *ATM* regions (2T>C, 5762ins137A>G; intron 40, intron 3 and intron 32) were targeted using CRISPR-Cas9 in HEK293T-Cas9 cells. Efficiency of gene editing was assessed using bioinformatics tool TIDE. Results are obtained as total efficiency of guide RNA, profile of insertions or deletions (indels) generated around the cut site (0 represents the Cas9 cleavage site), and probability of nucleotide inserted as +1bp location. 0 represents the Cas9 cleavage site.

5.3 Discussion

This chapter established the foundation for genome editing of the *ATM* gene for the first time. *ATM* is one of the largest human genes with a plethora of disease-causing mutations. Computational analysis of its genomic sequence and mapping of mutations showed hundreds of different kinds of mutations spanning the entire genomic structure of *ATM*. No mutation rich areas (hotspots) or commonly occurring large deletions, insertions or gene re-arrangements could be identified. While many founder mutations within families have been reported (Gilad, Bar-Shira et al. 1996, McConville, Stankovic et al. 1996, Laake, Telatar et al. 1998, Stankovic, Kidd et al. 1998, Chun and Gatti 2004, Coutinho, Mitui et al. 2004, Fares, Axelord Ran et al. 2004, Stray-Pedersen, Jonsson et al. 2004, Birrell, Kneebone et al. 2005, Mitui, Bernatowska et al. 2005, Chessa, Piane et al. 2009, Carranza, Vega et al. 2017), there are no frequently occurring mutations that affect a large proportion of the population. A bulk of (~50%) of A-T mutations are nonsense/missense mutations that lead to production of either an unstable or non-functional ATM protein. About 30% of mutations are due to splicing defects. Current genetic treatments primarily target these two categories of mutations using compounds that skip pre-termination codons that occur because of nonsense/missense mutations or antisense-oligonucleotides to restore splicing errors. Both forms of treatment show promise as therapies but are limited by the need of repeated dosage, and are mutation-specific (Lavin, Gueven et al. 2007). Targeted genome editing can be used to permanently modify the ATM sequence. Given the widespread profile of *ATM* mutations, this section explored a variety of gene editing strategies applicable to different kinds of mutations, which could ultimately be further developed as therapies for Ataxia Telangiectasia.

For proof-of-principle, both NHEJ and HDR targeting specific ATM mutations were designed. Nonsense mutations – 2T>C and 103C>T – were chosen based on the availability of primary cell models from patients. The 9-bp deletion mutation (del7638) is a commonly known mutation reported in the UK population. The ATM5762ins137A>G splicing defect was identified as one of the major targets, it is a relatively common mutation that affects 10-15% of all A-T patients in the UK (Stankovic, Kidd et al. 1998). This mutation is known to create a cryptic splice site which then leads to insertion of an aberrant intronic splicing in the ATM transcript. However, due to the leaky nature of the mutation, cells from patients heterozygous for 5762ins137 mutation are estimated to produce 4% of the normal amount of protein. Patients homozygous with this mutation have also been described with ~10% ATM kinase activity, which is sufficient to moderate the phenotype although not prevent it (Taylor, Lam et al. 2015). Thus, targeted disruption

of this splice site using either NHEJ or HDR could be advantageous. A NHEJ-disruption strategy would be particularly useful as the process is effective in both dividing and quiescent cells. This is particularly critical in the context for A-T as the non-dividing Purkinje neurons located in the cerebellum are the primary therapeutic targets (Taylor and Byrd 2005).

Separately two different potentially global strategies were also developed which could be used interchangeably to target and correct any *ATM* mutation. Currently, knock-in of the whole *ATM* gene or its cDNA is a challenging prospect given the large genomic size. The open reading frame of ~9.2 kb, exceeds the packaging capacity of many commonly used delivery vectors even in the absence of regulatory sequences and homology arms. Attempts at production and delivery of shorter gene, known as mini-ATM, are in progress however there is no published data available. This chapter explored partial and full ATM cDNA knock-ins. Partial knock-in of N- or C-terminal of the gene would require delivery of either half of the ATM cDNA as exon blocks or “superexons” at the endogenous ATM loci. An N-terminal knock-in would be applicable for any mutation in the first half of the gene and *vice-versa*. The advantages here are a reduced genomic load in delivery vectors and that a single approach could be used across different patients without designing mutation specific treatments. As for full-length knock-in, the human AAVS1 locus was investigated. Consistent delivery, robust transgene expression, the AAVS1 safe harbour site is a preferred target for gene knock-ins. Insertion at this site has been shown to be safe with no phenotypic effects reported, and the surrounding DNA appears to be kept in an open confirmation, enabling stable expression of a variety of transgenes (Sadelain, Papapetrou et al. 2011). Successful targeting of the human AAVS1 locus with site-specific nucleases and homologous recombination of donor DNA bearing selection markers has been previously demonstrated in cells, including human embryonic stem and induced pluripotent stem cells, derived from different disease models (Sadelain, Papapetrou et al. 2011). Similar approaches have been previously reported for Fanconi anemia (Rio, Banos et al. 2014) and RS-SCID (Rahman, Kuehle et al. 2015), both of which are DNA repair disorders showing a developmental block in the generation of immune cells.

Lastly, this chapter explored generation of a humanised mouse model for A-T. Several murine models of A-T harbouring mutations on *Atm* gene have been successfully generated showing some of the clinical manifestation of the disease, however they do not fully recapitulate the hallmark phenotype of neurodegeneration (Barlow, Hirotsune et al. 1996, Lavin 2013). A porcine model of A-T to better phenocopy the disease and bridge the gap between human and current animal models has been recently developed

(Beraldi, Chan et al. 2015). Initial characterisation of A-T pigs revealed early cerebellar lesions including loss of Purkinje cells suggesting a developmental aetiology for A-T. In addition, like patients, A-T these pigs show growth retardation and develop motor skill deficiencies, characteristic to A-T (Beraldi, Chan et al. 2015). Humanised mouse models are powerful tools for studying and treating a variety of human diseases and can provide an alternative to the current models for the study of the mechanism underlying A-T disorder and for development of new therapies (Lavin 2013). The ATM splicing mutation 5762ins137A>G was selected to be introduced in *Atm*. This mutation, as discussed previously, is a good candidate for genome editing because i) it can be corrected through both NHEJ-disruption and HDR; ii) it has a leaky phenotype which means there is some residual ATM protein kinase activity; and iii) the mutation has been frequently reported in 10-15% of the UK A-T patients. To introduce ATM5762ins137 in mouse *Atm*, first, the genomic region around the mutation site were analysed in both human and mouse. In humans, this mutation lies in intron 40, sequence comparisons showed that this region corresponded to mouse *Atm* intron 38. The structure (indicated by exon-intron boundaries) and sequences of human exon 40 and 41 were identified to be identical to mouse Atm exons 37 and 38, respectively. Therefore, a knock-in construct containing relevant human sequences along with the mutation and mouse *Atm* sequence homology arms was envisioned.

CRISPR-Cas gene editing platform was chosen to target *ATM* due to its robustness and efficiency. Several different bioinformatics tools were used to identify *ATM* gRNAs. Top ranked gRNAs were tested for gene editing in HEK293T cells stably expressing Cas9. These cells were chosen as they are easy to transfect with high efficiencies. Second, since Cas9 is endogenously expressed, only a small RNA molecule needs to be introduced in the cells. This makes the screening process quicker and more effective. gRNAs were introduced as commercially produced synthetic RNA oligonucleotides for a faster turnaround.

Gene editing was tested using the software TIDE, which is a simple, rapid and cost-effective R-based tool that accurately quantifies the efficacy and simultaneously identifies the predominant types of indels in a gene edited pool of cells using a decomposition algorithm. Commonly available indel detection methods include enzyme mismatch cleavage assays based on Surveyor and T7E1 endonucleases. These can detect small sequence changes however; they are semi-quantitative and suffer from high background signals in presence of sequence polymorphisms. More importantly, these enzymatic assays do not provide insight into the nature, diversity and frequency of the mutations introduced in a pool of gene edited cells (Brinkman, Chen et al. 2014). TIDE

like high-throughput sequencing can be used to analyse sequence around the induced break site at highly reduced costs and in less time.

Introduction of *ATM* gRNAs in HEK-Cas9 cells showed high levels of gene editing at various *ATM* targets providing promising preliminary data for developing the outlines strategies. A particularly interesting result was the mono-allelic disruption of ATM5762ins137 splicing mutation as a result of NHEJ due to cutting by Cas9 near the mutation locus. Further validation of these edited genomic loci using next-generation sequencing was under progress at the time this Thesis was being written.

6 Conclusions and Further Work

6.1 Summary

This work described genome editing in two rare disease models – *Prkdc* severe combined immunodeficiency or SCID and Ataxia Telangiectasia – that arise due to genetic defects affecting core proteins involved in DNA damage response. The point mutation in *Prkdc*, causative of a radiosensitive form of SCID in mice, was edited using zinc-finger nucleases as well as CRISPR-Cas9 platform in a proof-of-concept *in vitro* model. Various genomic targets on *ATM*, defects on which cause Ataxia Telangiectasia, were targeted using CRISPR-Cas9 technology in human cell lines. This was the first ever demonstration of CRISPR-Cas9 based gene editing for both disorders.

Primary skin fibroblasts from SCID mice were used as the *in vitro* model for *Prkdc* SCID. Molecular analyses of clonal cell lines were carried out by techniques such as PCR, Sanger sequencing and indel detection using amplicon analysis. Clonal analysis of ZFN gene-edited cells demonstrated repair of the underlying SCID mutation. To demonstrate correlation between genomic correction of the mutation and restoration of the encoded DNA-PKcs protein and its activity this study established two different assays. The first assay involved the use of the drug melphalan, a nitrogen mustard commonly used in the treatment of a wide variety of cancers (Ross et al., 1978). Mutant cells deficient in DNA-PKcs activity have previously been reported to exhibit significant hypersensitivity to Melphalan (Caldecott and Jeggo 1991, Muller, Christodoulopoulos et al. 1998). Cells edited using ZFNs were subjected to DNA damage using Melphalan and their potential to survive and grow was investigated in cell survival and clonogenic assays. Separately, phosphorylation of histone H2AX after exposure to ionising radiation (IR) was utilised as an indicator of DNA PKcs activity. H2AX phosphorylation is an early step in response to DNA damage. Under normal growth conditions, IR-induced H2AX phosphorylation is carried out by DNA PK and ATM in a redundant, overlapping manner (Stiff, O'Driscoll et al. 2004). However, perhaps due to the intrinsic deficiency in DNA repair in SCID *Prkdc* fibroblasts, genome editing led to unwanted genomic alterations at nuclease cut sites which prevented the rescue of DNA PKcs activity in gene edited cells.

Separate to ZFNs, the *Prkdc* SCID mutation was also edited using the CRISPR-Cas9 system, leading to mono- and bi-allelic correction of the mutation as demonstrated by clonal analysis. Upon comparison between gene editing using the two nucleases, Cas9 showed higher efficiency in bi-allelic repair of SCID mutation. However, like in the ZFN study, the initially designed nuclease target sites mediated a double stranded break (DSB) within a *Prkdc* exon. This again led to unwanted sequence changes in the coding region of *Prkdc* in all cell clones tested. Given the defects in non-homologous end-joining, the occurrence of the mutagenic alternative-NHEJ pathways was speculated since out of all the clones tested. Perhaps a wider screen of edited cells would allow for identification of a cell containing the desired edit.

Building up from these results, the genome editing reagents were redesigned. First, a new Cas9 nuclease target site located downstream of *Prkdc* intron 85 was identified using bioinformatic tools and successful genome editing at this site was demonstrated in SCID fibroblasts. Introduction of unexpected genome-editing indels at an intronic site is much less likely to have detrimental effects on the mRNA product of the targeted gene. Second, the donor DNA used for homologous recombination was modified by site-directed mutagenesis to no longer include the nuclease target site. A common problem in gene editing can arise if the HR template contains the intact nuclease site. In this case, the nuclease may bind and cut the HR template. Additionally, even after successfully introducing the mutation with the HR template, the nuclease may continue to cut the genomic site if the target site was left intact. To avoid any unintended consequences, the template target site can be modified in advance. In case of CRISPR-Cas9, on-target specificity of Cas9 nuclease is determined by a short 3-bp PAM region adjacent to the guide RNA recognition sequence. Based on scores from codon usage index, PAM regions for the Cas9 PAM sites were modified on the HR donor templates using site-directed mutagenesis. Use of these improved gene editing reagents led to successful gene targeting in *Prkdc* SCID fibroblasts. However, clonal analysis to confirm the genotype of cells could not be completed due to time restrictions. Further work involving isolation, expansion and characterisation of clones from the edited cell pool would be required to substantiate the positive impact of the modifications made to the gene editing reagents.

Given that *Prkdc* SCID is a model of an ultra-rare DNA PKcs deficiency in humans, this study was transitioned towards development of gene editing as a potentially therapeutic strategy for another related DNA repair disorder, Ataxia Telangiectasia. A-T is due to defects on *ATM*. There are several parallels between DNA PKcs and ATM. Both proteins are members of the PIKK family and involved in DNA damage repair. ATM is further

involved in cell cycle checkpoint control, survival and apoptosis. Furthermore, deficiency in either of the proteins leads to sensitivity to IR and immunodeficiency.

Genome editing has not been previously described for A-T to the best of my knowledge. To look for suitable target sites on *ATM*, all the known A-T mutations were mapped to genomic regions on *ATM*. The mutation map produced highlighted over 600 different kinds of mutation span across *ATM* without any mutational hotspots. To demonstrate ATM gene editing, a sub-set of mutations that are prevalent in the UK and could be easily targeted through genome editing strategies in patient cell lines were identified. Given the plethora of mutations, three different gene knock-in strategies were designed in this study. Separately, generation of a mouse model was also envisioned. Proof-of-concept genome editing by CRISPR-Cas9 was shown at all chosen ATM targets in HEK-293T polyclonal cells through the introduction of NHEJ-mediated indels after DNA cutting, confirmed both by TIDE and NGS. These data combined with the *in vitro* assays developed to assess DNA damage set a foundation for further work on genome editing in A-T models.

6.2 Study Limitations

Like with any scientific experiment there were certain limitations in this study. First was the issue of delivery. Throughout the study, we were limited to using plasmids as the delivery platform in immortalised SCID mouse fibroblasts. These cells were however found challenging to transfect using conventional reagents which resulted in low transfection efficiencies. This affected Cre-LoxP recombination which then had to be coupled with FACS to enrich the small proportion of transfected cells. Likewise, poor transfection of Cas9-gRNA plasmids led to low levels of cutting in scid fibroblasts. These limitations could be overcome by using electroporation of existing plasmids or via viral vector transductions. To enhance gene editing, CRISPR-Cas9 reagents could be introduced in cells as protein-mRNA complexes delivered using nucleofection. Alternatively, an easy-to-transfect mouse cell line, such as Neuro2A, could be employed for initial testing of gene editing reagents.

It is worthwhile to note that despite the low levels of indels seen initially, relatively high levels of gene editing by HDR were achieved. This could be due to the inherent defect in NHEJ DNA repair in *Prkdc* fibroblasts. In the absence of high-level NHEJ, HR is favoured. A possible implication of this is that the ratio of NHEJ:HDR is likely to be different in cells that are not defective in DNA repair, and hence a bias towards more effective HDR may have been of benefit in the current experiments.

The other major limitation of the system was the absence of NHEJ in SCID fibroblasts. NHEJ is the prominent DNA repair mechanism and is most of the time more active than HDR. This difference in activity makes treating diseases that require gene correction or insertion of a gene more challenging than those requiring gene inactivation. However, in case of SCID fibroblasts, the ration of NHEJ:HDR is likely to be different than in cells that are not defective in DNA repair. Thus, it is worthwhile not to discount the bias towards effective HDR as seen in current experiments.

The absence of NHEJ also points to another issue. It has been proposed that in case of lack/block in proteins mediating NHEJ, cells employ the alternative or alt-NHEJ pathway. Under normal conditions, NHEJ pathway is prominently employed to repair broken ends. Although it is not error-free and does include insertions and deletions, DNA in some cells can be re-joined without mutations. However, alt-NHEJ is highly mutagenic and is prone to causing mutations at the DSB site. Gene editing relies on the creation of a DSB and exploiting the repair pathways that respond to it. In our experiments, gene editing in SCID cells, led to several mutations at the nuclease target locus (at the genomic site as well as in the repair template) suggesting that caution must be taken when attempting editing in models of DNA repair defects. One way to overcome this problem is to target a non-coding region of the gene, although some degree of impact of genomic alterations in intronic regions should not be completely discounted. Use of repair template devoid of nuclease target site is another strategy. However, this may not be possible for all nuclease target sites selected. Alternatively, HDR could be carried out using a single stranded oligonucleotide to avoid cutting and hence ineffective recombination.

Like in any other gene editing study, the last issue pertains to the understanding and improving specificity of the nucleases. The specificity of genome editing tools is one of the main safety concerns for clinical translation. Genetic modifications are permanent, and deleterious off-target mutations have the potential to create cells with oncogenic potential, reduced fitness, or complete functional impairment. Even low levels of off-target mutagenesis may lead to expansion of edited cells carrying unwanted mutations. Several methods have been developed for genome-wide analysis of off-target analyses of nucleases. Off-target profile of the ZFNs described in this study have previously been studied, indicating no major targets (Abdul-Razak 2013). However, off-target profiling of the Cas9 targets described for *Prkdc* and *ATM* have yet to be undertaken.

6.3 Further Work

Following the work and described in this thesis, work in the immediate future would be molecular analyses of *Prkdc* SCID fibroblasts clones that were generated using the optimised gene editing reagents, followed by assays of phenotypic rescue of DNA PKcs activity. This would be the most accurate model possible to correlate genotype-phenotype correction of *Prkdc* SCID mutation following HDR. This thesis demonstrated the on-target specificity of Cas9 nuclease at *Prkdc*, however, further work assessing off-target implications of the selected genomic targets is required.

Conclusions from zinc finger nuclease study showed that none of the clones generated carried the signature of a true correction and that a larger screen might be required. In this respect, future work can entail use of the nucleases described along with a single-stranded oligonucleotide donor DNA followed by a larger screen of clones.

Beyond validating the repair templates and nucleases described for *Prkdc* in this thesis, an interesting avenue for exploration would be interrogation of DNA repair mechanisms in response to nuclease specific DSB generation in SCID cells. So far, the results from this thesis speculate whether it is possible to carry out genome editing in models of DNA repair defects without a definitive answer and thus requires further evaluation.

For *ATM* gene editing, we used TIDE and NGS to demonstrate targeted modification in a human cell line. Phenotypic analyses of the gene-edited cells are required to validate these events, and off-target analyses to study specificity. Based on our encouraging preliminary results, the genome editing strategies conceptualised in this study will need to be translated into appropriate *in vitro* and *in vivo* models for rescue of haematopoietic as well as CNS A-T defects.

BIBLIOGRAPHY

Abdul-Razak, H. (2013). Correction of the classical scid mouse mutation by gene repair. PhD Thesis, Royal Holloway University of London, Egham, Surrey, United Kingdom.

Aiuti, A., Roncarolo, Roncarolo, M. G., and L. Naldini (2017). "Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products". EMBO Mol Med **9**(6): 737-740.

Aiuti, A., L. Biasco, S. Scaramuzza, F. Ferrua, M. P. Cicalese, C. Baricordi, F. Dionisio, A. Calabria, S. Giannelli, M. C. Castiello, M. Bosticardo, C. Evangelio, A. Assanelli, M. Casiraghi, S. Di Nunzio, L. Callegaro, C. Benati, P. Rizzardi, D. Pellin, C. Di Serio, M. Schmidt, C. Von Kalle, J. Gardner, N. Mehta, V. Neduva, D. J. Dow, A. Galy, R. Miniero, A. Finocchi, A. Metin, P. P. Banerjee, J. S. Orange, S. Galimberti, M. G. Valsecchi, A. Biffi, E. Montini, A. Villa, F. Ciceri, M. G. Roncarolo and L. Naldini (2013). "Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome." Science **341**(6148): 1233151.

Aiuti, A., S. Slavin, M. Aker, F. Ficara, S. Deola, A. Mortellaro, S. Morecki, G. Andolfi, A. Tabucchi, F. Carlucci, E. Marinello, F. Cattaneo, S. Vai, P. Servida, R. Miniero, M. G. Roncarolo and C. Bordignon (2002). "Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning." Science **296**(5577): 2410-2413.

Alton, E. W., A. C. Boyd, D. J. Porteous, G. Davies, J. C. Davies, U. Griesenbach, T. E. Higgins, D. R. Gill, S. C. Hyde and J. A. Innes (2015). "A Phase I/IIa Safety and Efficacy Study of Nebulized Liposome-mediated Gene Therapy for Cystic Fibrosis Supports a Multidose Trial." Am J Respir Crit Care Med **192**(11): 1389-1392.

An, J., Y. C. Huang, Q. Z. Xu, L. J. Zhou, Z. F. Shang, B. Huang, Y. Wang, X. D. Liu, D. C. Wu and P. K. Zhou (2010). "DNA-PKcs plays a dominant role in the regulation of H2AX phosphorylation in response to DNA damage and cell cycle progression." BMC Mol Biol **11**: 18.

Araki, R., A. Fujimori, K. Hamatani, K. Mita, T. Saito, M. Mori, R. Fukumura, M. Morimyo, M. Muto, M. Itoh, K. Tatsumi and M. Abe (1997). "Nonsense mutation at Tyr-4046 in the DNA-dependent protein kinase catalytic subunit of severe combined immune deficiency mice." Proc Natl Acad Sci U S A **94**(6): 2438-2443.

Bae, S., J. Kweon, H. S. Kim and J. S. Kim (2014). "Microhomology-based choice of Cas9 nuclease target sites." Nat Methods **11**(7): 705-706.

Bagley, J., M. L. Cortes, X. O. Breakefield and J. Iacomini (2004). "Bone marrow transplantation restores immune system function and prevents lymphoma in Atm-deficient mice." Blood **104**(2): 572-578.

Barlow, C., S. Hirotsune, R. Paylor, M. Liyanage, M. Eckhaus, F. Collins, Y. Shiloh, J. N. Crawley, T. Ried, D. Tagle and A. Wynshaw-Boris (1996). "Atm-deficient mice: a paradigm of ataxia telangiectasia." Cell **86**(1): 159-171.

Barrangou, R. and J. A. Doudna (2016). "Applications of CRISPR technologies in research and beyond." Nat Biotechnol **34**(9): 933-941.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709-1712.

Bassett, A. R., C. Tibbit, C. P. Ponting and J. L. Liu (2013). "Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system." Cell Rep **4**(1): 220-228.

Beamish, H. J., R. Jessberger, E. Riballo, A. Priestley, T. Blunt, B. Kysela and P. A. Jeggo (2000). "The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity." Nucleic Acids Res **28**(7): 1506-1513.

Beraldi, R., C. H. Chan, C. S. Rogers, A. D. Kovacs, D. K. Meyerholz, C. Trantzas, A. M. Lambertz, B. W. Darbro, K. L. Weber, K. A. White, R. V. Rheeden, M. C. Kruer, B. A. Dacken, X. J. Wang, B. T. Davis, J. A. Rohret, J. T. Struzynski, F. A. Rohret, J. M. Weimer and D. A. Pearce (2015). "A novel porcine model of ataxia telangiectasia reproduces

neurological features and motor deficits of human disease." Hum Mol Genet **24**(22): 6473-6484.
Berns, K. I. and C. Giraud (1995). "Adenovirus and adeno-associated virus as vectors for gene therapy." Ann N Y Acad Sci **772**: 95-104.
Betermier, M., P. Bertrand and B. S. Lopez (2014). "Is non-homologous end-joining really an inherently error-prone process?" PLoS Genet **10**(1): e1004086.
Birrell, G. W., K. Kneebone, M. Nefedov, E. Nefedova, M. N. Jartsev, M. Mitsui, R. A. Gatti and M. F. Lavin (2005). "ATM mutations, haplotype analysis, and immunological status of Russian patients with ataxia telangiectasia." Hum Mutat **25**(6): 593.
Black, S. J., E. Kashkina, T. Kent and R. T. Pomerantz (2016). "DNA Polymerase theta: A Unique Multifunctional End-Joining Machine." Genes (Basel) **7**(9).
Blaese, R. M., K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, J. J. Greenblatt, S. A. Rosenberg, H. Klein, M. Berger, C. A. Mullen, W. J. Ramsey, L. Muul, R. A. Morgan and W. F. Anderson (1995). "T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years." Science **270**(5235): 475-480.
Block, W. D., Y. Yu, D. Merkle, J. L. Gifford, Q. Ding, K. Meek and S. P. Lees-Miller (2004). "Autophosphorylation-dependent remodeling of the DNA-dependent protein kinase catalytic subunit regulates ligation of DNA ends." Nucleic Acids Res **32**(14): 4351-4357.
Blunt, T., N. J. Finnie, G. E. Taccioli, G. C. Smith, J. Demengeot, T. M. Gottlieb, R. Mizuta, A. J. Varghese, F. W. Alt, P. A. Jeggo and S. P. Jackson (1995). "Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation." Cell **80**(5): 813-823.
Blunt, T., D. Gell, M. Fox, G. E. Taccioli, A. R. Lehmann, S. P. Jackson and P. A. Jeggo (1996). "Identification of a nonsense mutation in the carboxyl-terminal region of DNA-dependent protein kinase catalytic subunit in the scid mouse." Proc Natl Acad Sci U S A **93**(19): 10285-10290.
Boissel, S., J. Jarjour, A. Astrakhan, A. Adey, A. Gouble, P. Duchateau, J. Shendure, B. L. Stoddard, M. T. Certo, D. Baker and A. M. Scharenberg (2014). "megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering." Nucleic Acids Res **42**(4): 2591-2601.
Bortesi, L. and R. Fischer (2015). "The CRISPR/Cas9 system for plant genome editing and beyond." Biotechnol Adv **33**(1): 41-52.
Bosma, G. C., R. P. Custer and M. J. Bosma (1983). "A severe combined immunodeficiency mutation in the mouse." Nature **301**(5900): 527-530.
Bosma, M. J. and A. M. Carroll (1991). "The SCID mouse mutant: definition, characterization, and potential uses." Annu Rev Immunol **9**: 323-350.
Boztug, K., M. Schmidt, A. Schwarzer, P. P. Banerjee, I. A. Diez, R. A. Dewey, M. Bohm, A. Nowrouzi, C. R. Ball, H. Glimm, S. Naundorf, K. Kuhlcke, R. Blasczyk, I. Kondratenko, L. Marodi, J. S. Orange, C. von Kalle and C. Klein (2010). "Stem-cell gene therapy for the Wiskott-Aldrich syndrome." N Engl J Med **363**(20): 1918-1927.
Braun, C. J., K. Boztug, A. Paruzynski, M. Witzel, A. Schwarzer, M. Rothe, U. Modlich, R. Beier, G. Gohring, D. Steinemann, R. Fronza, C. R. Ball, R. Haemmerle, S. Naundorf, K. Kuhlcke, M. Rose, C. Fraser, L. Mathias, R. Ferrari, M. R. Abboud, W. Al-Herz, I. Kondratenko, L. Marodi, H. Glimm, B. Schlegelberger, A. Schambach, M. H. Albert, M. Schmidt, C. von Kalle and C. Klein (2014). "Gene therapy for Wiskott-Aldrich syndrome--long-term efficacy and genotoxicity." Sci Transl Med **6**(227): 227ra233.
Brinkman, E. K., T. Chen, M. Amendola and B. van Steensel (2014). "Easy quantitative assessment of genome editing by sequence trace decomposition." Nucleic Acids Res **42**(22): e168.
Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin and J. van der Oost (2008). "Small CRISPR RNAs guide antiviral defense in prokaryotes." Science **321**(5891): 960-964.

Burger, C., O. S. Gorbatyuk, M. J. Velardo, C. S. Peden, P. Williams, S. Zolotukhin, P. J. Reier, R. J. Mandel and N. Muzyczka (2004). "Recombinant AAV viral vectors pseudotyped with viral capsids from serotypes 1, 2, and 5 display differential efficiency and cell tropism after delivery to different regions of the central nervous system." Mol Ther **10**(2): 302-317.
Butler, S. L., M. S. Hansen and F. D. Bushman (2001). "A quantitative assay for HIV DNA integration in vivo." Nat Med **7**(5): 631-634.
Caldecott, K. and P. Jeggo (1991). "Cross-sensitivity of gamma-ray-sensitive hamster mutants to cross-linking agents." Mutat Res **255**(2): 111-121.
Capecchi, M. R. (1989). "Altering the genome by homologous recombination." Science **244**(4910): 1288-1292.
Carranza, D., A. K. Vega, S. Torres-Rusillo, E. Montero, L. J. Martinez, M. Santamaria, J. L. Santos and I. J. Molina (2017). "Molecular and Functional Characterization of a Cohort of Spanish Patients with Ataxia-Telangiectasia." Neuromolecular Med **19**(1): 161-174.
Carroll, D. (2011). "Zinc-finger nucleases: a panoramic view." Curr Gene Ther **11**(1): 2-10.
Carter, B. J. (2005). "Adeno-associated virus vectors in clinical trials." Hum Gene Ther **16**(5): 541-550.
Cartier, N., S. Hacein-Bey-Abina, C. Von Kalle, P. Bougneres, A. Fischer, M. Cavazzana-Calvo and P. Aubourg (2010). "[Gene therapy of x-linked adrenoleukodystrophy using hematopoietic stem cells and a lentiviral vector]." Bull Acad Natl Med **194**(2): 255-264; discussion 264-258.
Cavazzana-Calvo, M., S. Hacein-Bey, G. de Saint Basile, F. Gross, E. Yvon, P. Nusbaum, F. Selz, C. Hue, S. Certain, J. L. Casanova, P. Bousso, F. L. Deist and A. Fischer (2000). "Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease." Science **288**(5466): 669-672.
Cavazzana-Calvo, M., E. Payen, O. Negre, G. Wang, K. Hehir, F. Fusil, J. Down, M. Denaro, T. Brady, K. Westerman, R. Cavallesco, B. Gillet-Legrand, L. Caccavelli, R. Sgarra, L. Maouche-Chretien, F. Bernaudin, R. Girot, R. Dorazio, G. J. Mulder, A. Polack, A. Bank, J. Soulier, J. Larghero, N. Kabbara, B. Dalle, B. Gourmel, G. Socie, S. Chretien, N. Cartier, P. Aubourg, A. Fischer, K. Cornetta, F. Galacteros, Y. Beuzard, E. Gluckman, F. Bushman, S. Hacein-Bey-Abina and P. Leboulch (2010). "Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia." Nature **467**(7313): 318-322.
Chessa, L., M. Piane, M. Magliozzi, I. Torrente, C. Savio, P. Lulli, A. De Luca and B. Dallapiccola (2009). "Founder effects for ATM gene mutations in Italian Ataxia Telangiectasia families." Ann Hum Genet **73**(Pt 5): 532-539.
Chhabra, A., A. M. Ring, K. Weiskopf, P. J. Schnorr, S. Gordon, A. C. Le, H. S. Kwon, N. G. Ring, J. Volkmer, P. Y. Ho, S. Tseng, I. L. Weissman and J. A. Shizuru (2016). "Hematopoietic stem cell transplantation in immunocompetent hosts without radiation or chemotherapy." Sci Transl Med **8**(351): 351ra105.
Chopra, C., G. Davies, M. Taylor, M. Anderson, S. Bainbridge, P. Tighe and E. M. McDermott (2014). "Immune deficiency in Ataxia-Telangiectasia: a longitudinal study of 44 patients." Clin Exp Immunol **176**(2): 275-282.
Chun, H. H. and R. A. Gatti (2004). "Ataxia-telangiectasia, an evolving phenotype." DNA Repair (Amst) **3**(8-9): 1187-1196.
Cicalese, M. P. and A. Aiuti (2015). "Clinical applications of gene therapy for primary immunodeficiencies." Hum Gene Ther **26**(4): 210-219.
Cicalese, M. P., F. Ferrua, L. Castagnaro, R. Pajno, F. Barzaghi, S. Giannelli, F. Dionisio, I. Brigida, M. Bonopane, M. Casiraghi, A. Tabucchi, F. Carlucci, E. Grunebaum, M. Adeli, R. G. Bredius, J. M. Puck, P. Stepensky, I. Tezcan, K. Rolfe, E. De Boever, R. R. Reinhardt, J. Appleby, F. Ciceri, M. G. Roncarolo and A. Aiuti (2016). "Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency." Blood **128**(1): 45-54.

Cockrell, A. S. and T. Kafri (2007). "Gene delivery by lentivirus vectors." Mol Biotechnol **36**(3): 184-204.
Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.
Cong, L. and F. Zhang (2015). "Genome engineering using CRISPR-Cas9 system." Methods Mol Biol **1239**: 197-217.
Cornu, T. I. and T. Cathomen (2010). "Quantification of zinc finger nuclease-associated toxicity." Methods Mol Biol **649**: 237-245.
Cortes, M. L., A. Oehmig, O. Saydam, J. D. Sanford, K. F. Perry, C. Fraefel and X. O. Breakefield (2008). "Targeted integration of functional human ATM cDNA into genome mediated by HSV/AAV hybrid amplicon vector." Mol Ther **16**(1): 81-88.
Coutinho, G., M. Mitui, C. Campbell, B. T. Costa Carvalho, S. Nahas, X. Sun, Y. Huo, C. H. Lai, Y. Thorstenson, R. Tanouye, S. Raskin, C. A. Kim, J. Llerena, Jr. and R. A. Gatti (2004). "Five haplotypes account for fifty-five percent of ATM mutations in Brazilian patients with ataxia telangiectasia: seven new mutations." Am J Med Genet A **126a**(1): 33-40.
Cox, D. B., R. J. Platt and F. Zhang (2015). "Therapeutic genome editing: prospects and challenges." Nat Med **21**(2): 121-131.
Crosetto, N., A. Mitra, M. J. Silva, M. Bienko, N. Dojer, Q. Wang, E. Karaca, R. Chiarle, M. Skrzypczak, K. Ginalski, P. Pasero, M. Rowicka and I. Dikic (2013). "Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing." Nat Methods **10**(4): 361-365.
Davis, A. J., B. P. Chen and D. J. Chen (2014). "DNA-PK: a dynamic enzyme in a versatile DSB repair pathway." DNA Repair (Amst) **17**: 21-29.
Davis, A. J. and D. J. Chen (2013). "DNA double strand break repair via non-homologous end-joining." Transl Cancer Res **2**(3): 130-143.
Daya, S. and K. I. Berns (2008). "Gene therapy using adeno-associated virus vectors." Clin Microbiol Rev **21**(4): 583-593.
DiCarlo, J. E., J. E. Norville, P. Mali, X. Rios, J. Aach and G. M. Church (2013). "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems." Nucleic Acids Res **41**(7): 4336-4343.
Ding, Q., Y. V. Reddy, W. Wang, T. Woods, P. Douglas, D. A. Ramsden, S. P. Lees-Miller and K. Meek (2003). "Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair." Mol Cell Biol **23**(16): 5836-5848.
Dismuke, D. J., L. Tenenbaum and R. J. Samulski (2013). "Biosafety of recombinant adeno-associated virus vectors." Curr Gene Ther **13**(6): 434-452.
Doudna, J. A. and E. Charpentier (2014). "Genome editing. The new frontier of genome engineering with CRISPR-Cas9." Science **346**(6213): 1258096.
Douglas, P., G. P. Sapkota, N. Morrice, Y. Yu, A. A. Goodarzi, D. Merkle, K. Meek, D. R. Alessi and S. P. Lees-Miller (2002). "Identification of in vitro and in vivo phosphorylation sites in the catalytic subunit of the DNA-dependent protein kinase." Biochem J **368**(Pt 1): 243-251.
Du, L., M. E. Jung, R. Damoiseaux, G. Completo, F. Fike, J. M. Ku, S. Nahas, C. Piao, H. Hu and R. A. Gatti (2013). "A new series of small molecular weight compounds induce read through of all three types of nonsense mutations in the ATM gene." Mol Ther **21**(9): 1653-1660.
Du, L., J. M. Pollard and R. A. Gatti (2007). "Correction of prototypic ATM splicing mutations and aberrant ATM function with antisense morpholino oligonucleotides." Proc Natl Acad Sci U S A **104**(14): 6007-6012.
Dull, T., R. Zufferey, M. Kelly, R. J. Mandel, M. Nguyen, D. Trono and L. Naldini (1998). "A third-generation lentivirus vector with a conditional packaging system." J Virol **72**(11): 8463-8471.

Durai, S., M. Mani, K. Kandavelou, J. Wu, M. H. Porteus and S. Chandrasegaran (2005). "Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells." Nucleic Acids Res **33**(18): 5978-5990.
Esvelt, K. M., P. Mali, J. L. Braff, M. Moosburner, S. J. Yaung and G. M. Church (2013). "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing." Nat Methods **10**(11): 1116-1121.
Fares, F., S. Axelord Ran, M. David, N. Zelnik, Y. Hecht, H. Khairaldeen and A. Lerner (2004). "Identification of two mutations for ataxia telangiectasia among the Druze community." Prenat Diagn **24**(5): 358-362.
Franken, N. A., H. M. Rodermond, J. Stap, J. Haveman and C. van Bree (2006). "Clonogenic assay of cells in vitro." Nat Protoc **1**(5): 2315-2319.
Fu, Y., D. Reyon and J. K. Joung (2014). "Targeted genome editing in human cells using CRISPR/Cas nucleases and truncated guide RNAs." Methods Enzymol **546**: 21-45.
Gao, G. P., M. R. Alvira, L. Wang, R. Calcedo, J. Johnston and J. M. Wilson (2002). "Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy." Proc Natl Acad Sci U S A **99**(18): 11854-11859.
Gaspar, H. B., S. Cooray, K. C. Gilmour, K. L. Parsley, S. Adams, S. J. Howe, A. Al Ghonaium, J. Bayford, L. Brown, E. G. Davies, C. Kinnon and A. J. Thrasher (2011). "Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency." Sci Transl Med **3**(97): 97ra79.
Gaspar, H. B., S. Cooray, K. C. Gilmour, K. L. Parsley, F. Zhang, S. Adams, E. Bjorkegren, J. Bayford, L. Brown, E. G. Davies, P. Veys, L. Fairbanks, V. Bordon, T. Petropoulou, C. Kinnon and A. J. Thrasher (2011). "Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction." Sci Transl Med **3**(97): 97ra80.
Gennery, A. R., M. A. Slatter, L. Grandin, P. Taupin, A. J. Cant, P. Veys, P. J. Amrolia, H. B. Gaspar, E. G. Davies, W. Friedrich, M. Hoenig, L. D. Notarangelo, E. Mazzolari, F. Porta, R. G. Bredius, A. C. Lankester, N. M. Wulffraat, R. Seger, T. Gungor, A. Fasth, P. Sedlacek, B. Neven, S. Blanche, A. Fischer, M. Cavazzana-Calvo and P. Landais (2010). "Transplantation of hematopoietic stem cells and long-term survival for primary immunodeficiencies in Europe: entering a new century, do we do better?" J Allergy Clin Immunol **126**(3): 602-610.e601-611.
Georgiadis, A., Y. Duran, J. Ribeiro, L. Abelleira-Hervas, S. J. Robbie, B. Sunkel-Laing, S. Fourali, A. Gonzalez-Cordero, E. Cristante, M. Michaelides, J. W. Bainbridge, A. J. Smith and R. R. Ali (2016). "Development of an optimized AAV2/5 gene therapy vector for Leber congenital amaurosis owing to defects in RPE65." Gene Ther **23**(12): 857-862.
Ghosh, S., F. R. Schuster, V. Binder, T. Niehues, S. E. Baldus, P. Seiffert, H. J. Laws, A. Borkhardt and R. Meisel (2012). "Fatal outcome despite full lympho-hematopoietic reconstitution after allogeneic stem cell transplantation in atypical ataxia telangiectasia." J Clin Immunol **32**(3): 438-440.
Gilad, S., A. Bar-Shira, R. Harnik, D. Shkedy, Y. Ziv, R. Khosravi, K. Brown, L. Vanagaite, G. Xu, M. Frydman, M. F. Lavin, D. Hill, D. A. Tagle and Y. Shiloh (1996). "Ataxia-telangiectasia: founder effect among north African Jews." Hum Mol Genet **5**(12): 2033-2037.
Ginn, S. L., I. E. Alexander, M. L. Edelstein, M. R. Abedi and J. Wixon (2013). "Gene therapy clinical trials worldwide to 2012 - an update." J Gene Med **15**(2): 65-77.
Griesenbach, U., K. M. Pytel and E. W. Alton (2015). "Cystic Fibrosis Gene Therapy in the UK and Elsewhere." Hum Gene Ther **26**(5): 266-275.
Grimm, D. and M. A. Kay (2003). "From virus evolution to vector revolution: use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy." Curr Gene Ther **3**(4): 281-304.
Gu, H., Y. R. Zou and K. Rajewsky (1993). "Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting." Cell **73**(6): 1155-1164.

Hacein-Bey-Abina, S., J. Hauer, A. Lim, C. Picard, G. P. Wang, C. C. Berry, C. Martinache, F. Rieux-Laucat, S. Latour, B. H. Belohradsky, L. Leiva, R. Sorensen, M. Debre, J. L. Casanova, S. Blanche, A. Durandy, F. D. Bushman, A. Fischer and M. Cavazzana-Calvo (2010). "Efficacy of gene therapy for X-linked severe combined immunodeficiency." N Engl J Med **363**(4): 355-364.
Hacein-Bey-Abina, S., S. Y. Pai, H. B. Gaspar, M. Armant, C. C. Berry, S. Blanche, J. Bleesing, J. Blondeau, H. de Boer, K. F. Buckland, L. Caccavelli, G. Cros, S. De Oliveira, K. S. Fernandez, D. Guo, C. E. Harris, G. Hopkins, L. E. Lehmann, A. Lim, W. B. London, J. C. van der Loo, N. Malani, F. Male, P. Malik, M. A. Marinovic, A. M. McNicol, D. Moshous, B. Neven, M. Oleastro, C. Picard, J. Ritz, C. Rivat, A. Schambach, K. L. Shaw, E. A. Sherman, L. E. Silberstein, E. Six, F. Touzot, A. Tsytsykova, J. Xu-Bayford, C. Baum, F. D. Bushman, A. Fischer, D. B. Kohn, A. H. Filipovich, L. D. Notarangelo, M. Cavazzana, D. A. Williams and A. J. Thrasher (2014). "A modified gamma-retrovirus vector for X-linked severe combined immunodeficiency." N Engl J Med **371**(15): 1407-1417.
Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer and M. Cavazzana-Calvo (2003). "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1." Science **302**(5644): 415-419.
Hacein-Bey Abina, S., H. B. Gaspar, J. Blondeau, L. Caccavelli, S. Charrier, K. Buckland, C. Picard, E. Six, N. Himoudi, K. Gilmour, A. M. McNicol, H. Hara, J. Xu-Bayford, C. Rivat, F. Touzot, F. Mavilio, A. Lim, J. M. Treluyer, S. Heritier, F. Lefrere, J. Magalon, I. Pengue-Koyi, G. Honnet, S. Blanche, E. A. Sherman, F. Male, C. Berry, N. Malani, F. D. Bushman, A. Fischer, A. J. Thrasher, A. Galy and M. Cavazzana (2015). "Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome." Jama **313**(15): 1550-1563.
Hammel, M., Y. Yu, B. L. Mahaney, B. Cai, R. Ye, B. M. Phipps, R. P. Rambo, G. L. Hura, M. Pelikan, S. So, R. M. Abolfath, D. J. Chen, S. P. Lees-Miller and J. A. Tainer (2010). "Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex." J Biol Chem **285**(2): 1414-1423.
Hammond, A., R. Galizi, K. Kyrou, A. Simoni, C. Siniscalchi, D. Katsanos, M. Gribble, D. Baker, E. Marois, S. Russell, A. Burt, N. Windbichler, A. Crisanti and T. Nolan (2016). "A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae." Nat Biotechnol **34**(1): 78-83.
Hansen, J. and P. Bross (2010). "A cellular viability assay to monitor drug toxicity." Methods Mol Biol **648**: 303-311.
Heler, R., P. Samai, J. W. Modell, C. Weiner, G. W. Goldberg, D. Bikard and L. A. Marraffini (2015). "Cas9 specifies functional viral targets during CRISPR-Cas adaptation." Nature **519**(7542): 199-202.
Helleday, T., J. Lo, D. C. van Gent and B. P. Engelward (2007). "DNA double-strand break repair: from mechanistic understanding to cancer treatment." DNA Repair (Amst) **6**(7): 923-935.
Hoban, M. D., S. H. Orkin and D. E. Bauer (2016). "Genetic treatment of a molecular disorder: gene therapy approaches to sickle cell disease." Blood **127**(7): 839-848.
Hollywood, J. A., C. M. Lee, M. F. Scallan and P. T. Harrison (2016). "Analysis of gene repair tracts from Cas9/gRNA double-stranded breaks in the human CFTR gene." Sci Rep **6**: 32230.
Hou, Z., Y. Zhang, N. E. Propson, S. E. Howden, L. F. Chu, E. J. Sontheimer and J. A. Thomson (2013). "Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis." Proc Natl Acad Sci U S A **110**(39): 15644-15649.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura and A. Nakata (1987). "Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product." J Bacteriol **169**(12): 5429-5433.
Jackson, S. P. and P. A. Jeggo (1995). "DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK." Trends Biochem Sci **20**(10): 412-415.
Jansen, R., J. D. Embden, W. Gaastra and L. M. Schouls (2002). "Identification of genes that are associated with DNA repeats in prokaryotes." Mol Microbiol **43**(6): 1565-1575.
Jeggo, P. A., A. M. Carr and A. R. Lehmann (1998). "Splitting the ATM: distinct repair and checkpoint defects in ataxia-telangiectasia." Trends Genet **14**(8): 312-316.
Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.
Jones, I. M. and Y. Morikawa (1998). "The molecular basis of HIV capsid assembly." Rev Med Virol **8**(2): 87-95.
Jung, D. and F. W. Alt (2004). "Unraveling V(D)J recombination; insights into gene regulation." Cell **116**(2): 299-311.
Kim, E., T. Koo, S. W. Park, D. Kim, K. Kim, H. Y. Cho, D. W. Song, K. J. Lee, M. H. Jung, S. Kim, J. H. Kim, J. H. Kim and J. S. Kim (2017). "In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni." Nat Commun **8**: 14500.
Kim, H. J., H. J. Lee, H. Kim, S. W. Cho and J. S. Kim (2009). "Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly." Genome Res **19**(7): 1279-1288.
Kim, S., D. Kim, S. W. Cho, J. Kim and J. S. Kim (2014). "Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins." Genome Res **24**(6): 1012-1019.
Kleinstiver, B. P., V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng and J. K. Joung (2016). "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature **529**(7587): 490-495.
Kleinstiver, B. P., S. Q. Tsai, M. S. Prew, N. T. Nguyen, M. M. Welch, J. M. Lopez, Z. R. McCaw, M. J. Aryee and J. K. Joung (2016). "Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells." Nat Biotechnol **34**(8): 869-874.
Kohn, D. B., M. S. Hershfield, D. Carbonaro, A. Shigeoka, J. Brooks, E. M. Smogorzewska, L. W. Barsky, R. Chan, F. Burotto, G. Annett, J. A. Nolta, G. Crooks, N. Kapoor, M. Elder, D. Wara, T. Bowen, E. Madsen, F. F. Snyder, J. Bastian, L. Muul, R. M. Blaese, K. Weinberg and R. Parkman (1998). "T lymphocytes with a normal ADA gene accumulate after transplantation of transduced autologous umbilical cord blood CD34+ cells in ADA-deficient SCID neonates." Nat Med **4**(7): 775-780.
Kotterman, M. A. and D. V. Schaffer (2014). "Engineering adeno-associated viruses for clinical gene therapy." Nat Rev Genet **15**(7): 445-451.
Kunin, V., R. Sorek and P. Hugenholtz (2007). "Evolutionary conservation of sequence and secondary structures in CRISPR repeats." Genome Biol **8**(4): R61.
Kurian, K. M., C. J. Watson and A. H. Wyllie (2000). "Retroviral vectors." Mol Pathol **53**(4): 173-176.
Kurimasa, A., S. Kumano, N. V. Boubnov, M. D. Story, C. S. Tung, S. R. Peterson and D. J. Chen (1999). "Requirement for the kinase activity of human DNA-dependent protein kinase catalytic subunit in DNA strand break rejoining." Mol Cell Biol **19**(5): 3877-3884.
Kwon, I. and D. V. Schaffer (2008). "Designer gene delivery vectors: molecular engineering and evolution of adeno-associated viral vectors for enhanced gene transfer." Pharm Res **25**(3): 489-499.
Laake, K., M. Telatar, G. A. Geitvik, R. O. Hansen, A. Heiberg, A. M. Andresen, R. Gatti and A. L. Borresen-Dale (1998). "Identical mutation in 55% of the ATM alleles in 11 Norwegian AT families: evidence for a founder effect." Eur J Hum Genet **6**(3): 235-244.
Lavin, M. F. (2008). "Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer." Nat Rev Mol Cell Biol **9**(10): 759-769.

Lavin, M. F. (2013). "The appropriateness of the mouse model for ataxia-telangiectasia: neurological defects but no neurodegeneration." DNA Repair (Amst) **12**(8): 612-619.
Lavin, M. F., N. Gueven, S. Bottle and R. A. Gatti (2007). "Current and potential therapeutic strategies for the treatment of ataxia-telangiectasia." Br Med Bull **81-82**: 129-147.
Lee, P., N. T. Martin, K. Nakamura, S. Azghadi, M. Amiri, U. Ben-David, S. Perlman, R. A. Gatti, H. Hu and W. E. Lowry (2013). "SMRT compounds abrogate cellular phenotypes of ataxia telangiectasia in neural derivatives of patient-specific hiPSCs." Nat Commun **4**: 1824.
Lees-Miller, S. P. and K. Meek (2003). "Repair of DNA double strand breaks by non-homologous end joining." Biochimie **85**(11): 1161-1173.
Lheriteau, E., A. M. Davidoff and A. C. Nathwani (2015). "Haemophilia gene therapy: Progress and challenges." Blood Rev **29**(5): 321-328.
Li, X. and W. D. Heyer (2008). "Homologous recombination in DNA repair and DNA damage tolerance." Cell Res **18**(1): 99-113.
Liang, P., Y. Xu, X. Zhang, C. Ding, R. Huang, Z. Zhang, J. Lv, X. Xie, Y. Chen, Y. Li, Y. Sun, Y. Bai, Z. Songyang, W. Ma, C. Zhou and J. Huang (2015). "CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes." Protein Cell **6**(5): 363-372.
Lieber, M. R. (2010). "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway." Annu Rev Biochem **79**: 181-211.
Liu, Y. and S. C. West (2004). "Happy Hollidays: 40th anniversary of the Holliday junction." Nat Rev Mol Cell Biol **5**(11): 937-944.
Maeder, M. L., S. Thibodeau-Beganny, A. Osiak, D. A. Wright, R. M. Anthony, M. Eichtinger, T. Jiang, J. E. Foley, R. J. Winfrey, J. A. Townsend, E. Unger-Wallace, J. D. Sander, F. Muller-Lerch, F. Fu, J. Pearlberg, C. Gobel, J. P. Dassie, S. M. Pruett-Miller, M. H. Porteus, D. C. Sgroi, A. J. Iafrate, D. Dobbs, P. B. McCray, Jr., T. Cathomen, D. F. Voytas and J. K. Joung (2008). "Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification." Mol Cell **31**(2): 294-301.
Maeder, M. L., S. Thibodeau-Beganny, J. D. Sander, D. F. Voytas and J. K. Joung (2009). "Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays." Nat Protoc **4**(10): 1471-1501.
Maggio, I. and M. A. Goncalves (2015). "Genome editing at the crossroads of delivery, specificity, and fidelity." Trends Biotechnol **33**(5): 280-291.
Mahaney, B. L., K. Meek and S. P. Lees-Miller (2009). "Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining." Biochem J **417**(3): 639-650.
Mali, P., K. M. Esvelt and G. M. Church (2013). "Cas9 as a versatile tool for engineering biology." Nat Methods **10**(10): 957-963.
Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." Science **339**(6121): 823-826.
Matrai, J., M. K. Chuah and T. VandenDriessche (2010). "Recent advances in lentiviral vector development and applications." Mol Ther **18**(3): 477-490.
Matsuoka, S., B. A. Ballif, A. Smogorzewska, E. R. McDonald, 3rd, K. E. Hurov, J. Luo, C. E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S. P. Gygi and S. J. Elledge (2007). "ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage." Science **316**(5828): 1160-1166.
McConville, C. M., T. Stankovic, P. J. Byrd, G. M. McGuire, Q. Y. Yao, G. G. Lennox and M. R. Taylor (1996). "Mutations associated with variant phenotypes in ataxia-telangiectasia." Am J Hum Genet **59**(2): 320-330.
Meek, K., S. Gupta, D. A. Ramsden and S. P. Lees-Miller (2004). "The DNA-dependent protein kinase: the director at the end." Immunol Rev **200**: 132-141.
Miller, J. C., M. C. Holmes, J. Wang, D. Y. Guschin, Y. L. Lee, I. Rupniewski, C. M. Beausejour, A. J. Waite, N. S. Wang, K. A. Kim, P. D. Gregory, C. O. Pabo and E. J.

Rebar (2007). "An improved zinc-finger nuclease architecture for highly specific genome editing." Nat Biotechnol **25**(7): 778-785.
Mitui, M., E. Bernatowska, B. Pietrucha, J. Piotrowska-Jastrzebska, L. Eng, S. Nahas, S. Teraoka, G. Sholty, A. Purayidom, P. Concannon and R. A. Gatti (2005). "ATM gene founder haplotypes and associated mutations in Polish families with ataxia-telangiectasia." Ann Hum Genet **69**(Pt 6): 657-664.
Moratto, D., S. Giliani, C. Bonfim, E. Mazzolari, A. Fischer, H. D. Ochs, A. J. Cant, A. J. Thrasher, M. J. Cowan, M. H. Albert, T. Small, S. Y. Pai, E. Haddad, A. Lisa, S. Hambleton, M. Slatter, M. Cavazzana-Calvo, N. Mahlaoui, C. Picard, T. R. Torgerson, L. Burroughs, A. Koliski, J. Z. Neto, F. Porta, W. Qasim, P. Veys, K. Kavanau, M. Honig, A. Schulz, W. Friedrich and L. D. Notarangelo (2011). "Long-term outcome and lineage-specific chimerism in 194 patients with Wiskott-Aldrich syndrome treated by hematopoietic cell transplantation in the period 1980-2009: an international collaborative study." Blood **118**(6): 1675-1684.
Morton, J., M. W. Davis, E. M. Jorgensen and D. Carroll (2006). "Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells." Proc Natl Acad Sci U S A **103**(44): 16370-16375.
Muller, C., G. Christodoulopoulos, B. Salles and L. Panasci (1998). "DNA-Dependent protein kinase activity correlates with clinical and in vitro sensitivity of chronic lymphocytic leukemia lymphocytes to nitrogen mustards." Blood **92**(7): 2213-2219.
Munshi, A., M. Hobbs and R. E. Meyn (2005). "Clonogenic cell survival assay." Methods Mol Med **110**: 21-28.
Naldini, L. (2015). "Gene therapy returns to centre stage." Nature **526**(7573): 351-360.
Nowak-Wegrzyn, A., T. O. Crawford, J. A. Winkelstein, K. A. Carson and H. M. Lederman (2004). "Immunodeficiency and infections in ataxia-telangiectasia." J Pediatr **144**(4): 505-511.
Orthwein, A., S. M. Noordermeer, M. D. Wilson, S. Landry, R. I. Enchev, A. Sherker, M. Munro, J. Pinder, J. Salsman, G. Dellaire, B. Xia, M. Peter and D. Durocher (2015). "A mechanism for the suppression of homologous recombination in G1 cells." Nature **528**(7582): 422-426.
Ott, M. G., M. Schmidt, K. Schwarzwaelder, S. Stein, U. Siler, U. Koehl, H. Glimm, K. Kuhlcke, A. Schilz, H. Kunkel, S. Naundorf, A. Brinkmann, A. Deichmann, M. Fischer, C. Ball, I. Pilz, C. Dunbar, Y. Du, N. A. Jenkins, N. G. Copeland, U. Luthi, M. Hassan, A. J. Thrasher, D. Hoelzer, C. von Kalle, R. Seger and M. Grez (2006). "Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1." Nat Med **12**(4): 401-409.
Paigen, K. and P. Petkov (2010). "Mammalian recombination hot spots: properties, control and evolution." Nat Rev Genet **11**(3): 221-233.
Pala, F., H. Morbach, M. C. Castiello, J. N. Schickel, S. Scaramuzza, N. Chamberlain, B. Cassani, S. Glauzy, N. Romberg, F. Candotti, A. Aiuti, M. Bosticardo, A. Villa and E. Meffre (2015). "Lentiviral-mediated gene therapy restores B cell tolerance in Wiskott-Aldrich syndrome patients." J Clin Invest **125**(10): 3941-3951.
Palchaudhuri, R., B. Saez, J. Hoggatt, A. Schajnovitz, D. B. Sykes, T. A. Tate, A. Czechowicz, Y. Kfoury, F. Ruchika, D. J. Rossi, G. L. Verdine, M. K. Mansour and D. T. Scadden (2016). "Non-genotoxic conditioning for hematopoietic stem cell transplantation using a hematopoietic-cell-specific internalizing immunotoxin." Nat Biotechnol **34**(7): 738-745.
Peffault de Latour, R., R. Porcher, J. H. Dalle, M. Aljurf, E. T. Korthof, J. Svahn, R. Willemze, C. Barrenetxea, V. Mialou, J. Soulier, M. Ayas, R. Oneto, A. Bacigalupo, J. C. Marsh, C. Peters, G. Socie and C. Dufour (2013). "Allogeneic hematopoietic stem cell transplantation in Fanconi anemia: the European Group for Blood and Marrow Transplantation experience." Blood **122**(26): 4279-4286.
Perabo, L., J. Endell, S. King, K. Lux, D. Goldnau, M. Hallek and H. Buning (2006). "Combinatorial engineering of a gene therapy vector: directed evolution of adeno-associated virus." J Gene Med **8**(2): 155-162.

Pham, C. T., D. M. MacIvor, B. A. Hug, J. W. Heusel and T. J. Ley (1996). "Long-range disruption of gene expression by a selectable marker cassette." Proc Natl Acad Sci U S A **93**(23): 13090-13095.
Philpott, N. J. and A. J. Thrasher (2007). "Use of nonintegrating lentiviral vectors for gene therapy." Hum Gene Ther **18**(6): 483-489.
Pietzner, J., P. C. Baer, R. P. Duecker, M. B. Merscher, C. Satzger-Prodinger, I. Bechmann, A. Wietelmann, D. Del Turco, C. Doering, S. Kuci, P. Bader, S. Schirmer, S. Zielen and R. Schubert (2013). "Bone marrow transplantation improves the outcome of Atm-deficient mice through the migration of ATM-competent cells." Hum Mol Genet **22**(3): 493-507.
Podhorecka, M., A. Skladanowski and P. Bozko (2010). "H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy." J Nucleic Acids **2010**.
Porteus, M. H. and D. Carroll (2005). "Gene targeting using zinc finger nucleases." Nat Biotechnol **23**(8): 967-973.
Prakash, V., M. Moore and R. J. Yanez-Munoz (2016). "Current Progress in Therapeutic Gene Editing for Monogenic Diseases." Mol Ther **24**(3): 465-474.
Price, P. and T. J. McMillan (1990). "Use of the tetrazolium assay in measuring the response of human tumor cells to ionizing radiation." Cancer Res **50**(5): 1392-1396.
Qi, J., R. Shackelford, R. Manuszak, D. Cheng, M. Smith, C. J. Link and S. Wang (2004). "Functional expression of ATM gene carried by HSV amplicon vector in vitro and in vivo." Gene Ther **11**(1): 25-33.
Rahman, S. H., J. Kuehle, C. Reimann, T. Mlambo, J. Alzubi, M. L. Maeder, H. Riedel, P. Fisch, T. Cantz, C. Rudolph, C. Mussolino, J. K. Joung, A. Schambach and T. Cathomen (2015). "Rescue of DNA-PK Signaling and T-Cell Differentiation by Targeted Genome Editing in a prkdc Deficient iPSC Disease Model." PLoS Genet **11**(5): e1005239.
Ramirez, C. L., M. T. Certo, C. Mussolino, M. J. Goodwin, T. J. Cradick, A. P. McCaffrey, T. Cathomen, A. M. Scharenberg and J. K. Joung (2012). "Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects." Nucleic Acids Res **40**(12): 5560-5568.
Ran, F. A., L. Cong, W. X. Yan, D. A. Scott, J. S. Gootenberg, A. J. Kriz, B. Zetsche, O. Shalem, X. Wu, K. S. Makarova, E. V. Koonin, P. A. Sharp and F. Zhang (2015). "In vivo genome editing using Staphylococcus aureus Cas9." Nature **520**(7546): 186-191.
Ran, F. A., P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang and F. Zhang (2013). "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity." Cell **154**(6): 1380-1389.
Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang (2013). "Genome engineering using the CRISPR-Cas9 system." Nat Protoc **8**(11): 2281-2308.
Reiman, A., V. Srinivasan, G. Barone, J. I. Last, L. L. Wootton, E. G. Davies, M. M. Verhagen, M. A. Willemsen, C. M. Weemaes, P. J. Byrd, L. Izatt, D. F. Easton, D. J. Thompson and A. M. Taylor (2011). "Lymphoid tumours and breast cancer in ataxia telangiectasia; substantial protective effect of residual ATM kinase activity against childhood tumours." Br J Cancer **105**(4): 586-591.
Ribeil, J. A., S. Hacein-Bey-Abina, E. Payen, A. Magnani, M. Semeraro, E. Magrin, L. Caccavelli, B. Neven, P. Bourget, W. El Nemer, P. Bartolucci, L. Weber, H. Puy, J. F. Meritet, D. Grevent, Y. Beuzard, S. Chretien, T. Lefebvre, R. W. Ross, O. Negre, G. Veres, L. Sandler, S. Soni, M. de Montalembert, S. Blanche, P. Leboulch and M. Cavazzana (2017). "Gene Therapy in a Patient with Sickle Cell Disease." N Engl J Med **376**(9): 848-855.
Rio, P., R. Banos, A. Lombardo, O. Quintana-Bustamante, L. Alvarez, Z. Garate, P. Genovese, E. Almarza, A. Valeri, B. Diez, S. Navarro, Y. Torres, J. P. Trujillo, R. Murillas, J. C. Segovia, E. Samper, J. Surralles, P. D. Gregory, M. C. Holmes, L. Naldini and J. A. Bueren (2014). "Targeted gene therapy and cell reprogramming in Fanconi anemia." EMBO Mol Med **6**(6): 835-848.

Robbins, P. D. and S. C. Ghivizzani (1998). "Viral vectors for gene therapy." Pharmacol Ther **80**(1): 35-47.
Rodwell C, A. S. (July 2014). 2014 Report on the State of the Art of Rare Disease Activities in Europe., EUCERD.
Ruis, B. L., K. R. Fattah and E. A. Hendrickson (2008). "The catalytic subunit of DNA-dependent protein kinase regulates proliferation, telomere length, and genomic stability in human somatic cells." Mol Cell Biol **28**(20): 6182-6195.
Sadelain, M., E. P. Papapetrou and F. D. Bushman (2011). "Safe harbours for the integration of new DNA in the human genome." Nat Rev Cancer **12**(1): 51-58.
Sanjana, N. E., O. Shalem and F. Zhang (2014). "Improved vectors and genome-wide libraries for CRISPR screening." Nat Methods **11**(8): 783-784.
Sapranauskas, R., G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath and V. Siksnys (2011). "The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli." Nucleic Acids Res **39**(21): 9275-9282.
Sauer, A. V., B. Di Lorenzo, N. Carriglio and A. Aiuti (2014). "Progress in gene therapy for primary immunodeficiencies using lentiviral vectors." Curr Opin Allergy Clin Immunol **14**(6): 527-534.
Sauer, B. (1998). "Inducible gene targeting in mice using the Cre/lox system." Methods **14**(4): 381-392.
Sauer, B. and N. Henderson (1988). "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1." Proc Natl Acad Sci U S A **85**(14): 5166-5170.
Schatz, D. G. and Y. Ji (2011). "Recombination centres and the orchestration of V(D)J recombination." Nat Rev Immunol **11**(4): 251-263.
Schatz, D. G. and P. C. Swanson (2011). "V(D)J recombination: mechanisms of initiation." Annu Rev Genet **45**: 167-202.
Secretan, M. B., Z. Scuric, J. Oshima, A. J. Bishop, N. G. Howlett, D. Yau and R. H. Schiestl (2004). "Effect of Ku86 and DNA-PKcs deficiency on non-homologous end-joining and homologous recombination using a transient transfection assay." Mutat Res **554**(1-2): 351-364.
Seymour, L. W. and A. J. Thrasher (2012). "Gene therapy matures in the clinic." Nat Biotechnol **30**(7): 588-593.
Shi, W., G. S. Arnold and J. S. Bartlett (2001). "Insertional mutagenesis of the adeno-associated virus type 2 (AAV2) capsid gene and generation of AAV2 vectors targeted to alternative cell-surface receptors." Hum Gene Ther **12**(14): 1697-1711.
Shibata, A. and P. A. Jeggo (2014). "DNA double-strand break repair in a cellular context." Clin Oncol (R Coll Radiol) **26**(5): 243-249.
Sibanda, B. L., D. Y. Chirgadze, D. B. Ascher and T. L. Blundell (2017). "DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair." Science **355**(6324): 520-524.
Sibanda, B. L., D. Y. Chirgadze and T. L. Blundell (2010). "Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats." Nature **463**(7277): 118-121.
Silva, G., L. Poirot, R. Galetto, J. Smith, G. Montoya, P. Duchateau and F. Paques (2011). "Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy." Curr Gene Ther **11**(1): 11-27.
Sonoda, E., H. Hochegger, A. Saberi, Y. Taniguchi and S. Takeda (2006). "Differential usage of non-homologous end-joining and homologous recombination in double strand break repair." DNA Repair (Amst) **5**(9-10): 1021-1029.
Stankovic, T., A. M. Kidd, A. Sutcliffe, G. M. McGuire, P. Robinson, P. Weber, T. Bedenham, A. R. Bradwell, D. F. Easton, G. G. Lennox, N. Haites, P. J. Byrd and A. M. Taylor (1998). "ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer." Am J Hum Genet **62**(2): 334-345.

Staples, E. R., E. M. McDermott, A. Reiman, P. J. Byrd, S. Ritchie, A. M. Taylor and E. G. Davies (2008). "Immunodeficiency in ataxia telangiectasia is correlated strongly with the presence of two null mutations in the ataxia telangiectasia mutated gene." Clin Exp Immunol **153**(2): 214-220.
Stiff, T., M. O'Driscoll, N. Rief, K. Iwabuchi, M. Lobrich and P. A. Jeggo (2004). "ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation." Cancer Res **64**(7): 2390-2396.
Stray-Pedersen, A., T. Jonsson, A. Heiberg, C. R. Lindman, E. Widing, I. S. Aaberge, A. L. Borresen-Dale and T. G. Abrahamsen (2004). "The impact of an early truncating founder ATM mutation on immunoglobulins, specific antibodies and lymphocyte populations in ataxia-telangiectasia patients and their parents." Clin Exp Immunol **137**(1): 179-186.
Szczepek, M., V. Brondani, J. Buchel, L. Serrano, D. J. Segal and T. Cathomen (2007). "Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases." Nat Biotechnol **25**(7): 786-793.
Tang, L., Y. Zeng, H. Du, M. Gong, J. Peng, B. Zhang, M. Lei, F. Zhao, W. Wang, X. Li and J. Liu (2017). "CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein." Mol Genet Genomics.
Taylor, A. M. and P. J. Byrd (2005). "Molecular pathology of ataxia telangiectasia." J Clin Pathol **58**(10): 1009-1015.
Taylor, A. M., D. G. Harnden, C. F. Arlett, S. A. Harcourt, A. R. Lehmann, S. Stevens and B. A. Bridges (1975). "Ataxia telangiectasia: a human mutation with abnormal radiation sensitivity." Nature **258**(5534): 427-429.
Taylor, A. M., Z. Lam, J. I. Last and P. J. Byrd (2015). "Ataxia telangiectasia: more variation at clinical and cellular levels." Clin Genet **87**(3): 199-208.
Tebas, P., D. Stein, W. W. Tang, I. Frank, S. Q. Wang, G. Lee, S. K. Spratt, R. T. Surosky, M. A. Giedlin, G. Nichol, M. C. Holmes, P. D. Gregory, D. G. Ando, M. Kalos, R. G. Collman, G. Binder-Scholl, G. Plesa, W. T. Hwang, B. L. Levine and C. H. June (2014). "Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV." N Engl J Med **370**(10): 901-910.
Teraoka, S. N., M. Telatar, S. Becker-Catania, T. Liang, S. Onengut, A. Tolun, L. Chessa, O. Sanal, E. Bernatowska, R. A. Gatti and P. Concannon (1999). "Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences." Am J Hum Genet **64**(6): 1617-1631.
Thomas, K. R. and M. R. Capecchi (1986). "Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene." Nature **324**(6092): 34-38.
Tsai, S. Q., N. Wyvekens, C. Khayter, J. A. Foden, V. Thapar, D. Reyon, M. J. Goodwin, M. J. Aryee and J. K. Joung (2014). "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing." Nat Biotechnol **32**(6): 569-576.
Tsai, S. Q., Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee and J. K. Joung (2015). "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases." Nat Biotechnol **33**(2): 187-197.
Uematsu, N., E. Weterings, K. Yano, K. Morotomi-Yano, B. Jakob, G. Taucher-Scholz, P. O. Mari, D. C. van Gent, B. P. Chen and D. J. Chen (2007). "Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks." J Cell Biol **177**(2): 219-229.
Urnov, F. D., J. C. Miller, Y. L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory and M. C. Holmes (2005). "Highly efficient endogenous human gene correction using designed zinc-finger nucleases." Nature **435**(7042): 646-651.
Urnov, F. D., E. J. Rebar, M. C. Holmes, H. S. Zhang and P. D. Gregory (2010). "Genome editing with engineered zinc finger nucleases." Nat Rev Genet **11**(9): 636-646.
Ussowicz, M., J. Musial, E. Duszenko, O. Haus and K. Kalwak (2013). "Long-term survival after allogeneic-matched sibling PBSC transplantation with conditioning

consisting of low-dose busilvex and fludarabine in a 3-year-old boy with ataxia-telangiectasia syndrome and ALL." Bone Marrow Transplant **48**(5): 740-741.
van der Burg, M., H. Ijspeert, N. S. Verkaik, T. Turul, W. W. Wiegant, K. Morotomi-Yano, P. O. Mari, I. Tezcan, D. J. Chen, M. Z. Zdzienicka, J. J. van Dongen and D. C. van Gent (2009). "A DNA-PKcs mutation in a radiosensitive T-B- SCID patient inhibits Artemis activation and nonhomologous end-joining." J Clin Invest **119**(1): 91-98.
van der Burg, M., J. J. van Dongen and D. C. van Gent (2009). "DNA-PKcs deficiency in human: long predicted, finally found." Curr Opin Allergy Clin Immunol **9**(6): 503-509.
van Meerloo, J., G. J. Kaspers and J. Cloos (2011). "Cell sensitivity assays: the MTT assay." Methods Mol Biol **731**: 237-245.
Verma, I. M. and M. D. Weitzman (2005). "Gene therapy: twenty-first century medicine." Annu Rev Biochem **74**: 711-738.
Vigna, E. and L. Naldini (2000). "Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy." J Gene Med **2**(5): 308-316.
Vladutiu, A. O. (1993). "The severe combined immunodeficient (SCID) mouse as a model for the study of autoimmune diseases." Clin Exp Immunol **93**(1): 1-8.
Vogt, V. M. and M. N. Simon (1999). "Mass determination of rous sarcoma virus virions by scanning transmission electron microscopy." J Virol **73**(8): 7050-7055.
Wang, H., A. R. Perrault, Y. Takeda, W. Qin, H. Wang and G. Iliakis (2003). "Biochemical evidence for Ku-independent backup pathways of NHEJ." Nucleic Acids Res **31**(18): 5377-5388.
Wang, H. and X. Xu (2017). "Microhomology-mediated end joining: new players join the team." Cell Biosci **7**: 6.
Wanisch, K. and R. J. Yanez-Munoz (2009). "Integration-deficient lentiviral vectors: a slow coming of age." Mol Ther **17**(8): 1316-1332.
Weiss, R. A. (1998). "Retroviral zoonoses." Nat Med **4**(4): 391-392.
Wiedenheft, B., K. Zhou, M. Jinek, S. M. Coyle, W. Ma and J. A. Doudna (2009). "Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense." Structure **17**(6): 904-912.
Wolff, J. A. and J. Lederberg (1994). "An early history of gene transfer and therapy." Hum Gene Ther **5**(4): 469-480.
Woodbine, L., J. A. Neal, N. K. Sasi, M. Shimada, K. Deem, H. Coleman, W. B. Dobyns, T. Ogi, K. Meek, E. G. Davies and P. A. Jeggo (2013). "PRKDC mutations in a SCID patient with profound neurological abnormalities." J Clin Invest **123**(7): 2969-2980.
Wu, Z., A. Asokan and R. J. Samulski (2006). "Adeno-associated virus serotypes: vector toolkit for human gene therapy." Mol Ther **14**(3): 316-327.
Yanez-Munoz, R. J., K. S. Balaggan, A. MacNeil, S. J. Howe, M. Schmidt, A. J. Smith, P. Buch, R. E. MacLaren, P. N. Anderson, S. E. Barker, Y. Duran, C. Bartholomae, C. von Kalle, J. R. Heckenlively, C. Kinnon, R. R. Ali and A. J. Thrasher (2006). "Effective gene therapy with nonintegrating lentiviral vectors." Nat Med **12**(3): 348-353.
Yanez, R. J. and A. C. Porter (1998). "Therapeutic gene targeting." Gene Ther **5**(2): 149-159.
Yang, Z., C. Steentoft, C. Hauge, L. Hansen, A. L. Thomsen, F. Niola, M. B. Vester-Christensen, M. Frodin, H. Clausen, H. H. Wandall and E. P. Bennett (2015). "Fast and sensitive detection of indels induced by precise gene targeting." Nucleic Acids Res **43**(9): e59.
Yoder, K. E. and R. Fishel (2008). "Real-time quantitative PCR and fast QPCR have similar sensitivity and accuracy with HIV cDNA late reverse transcripts and 2-LTR circles." J Virol Methods **153**(2): 253-256.
Zetsche, B., J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin and F. Zhang (2015). "Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system." Cell **163**(3): 759-771.

Zhang, N., P. Chen, K. K. Khanna, S. Scott, M. Gatei, S. Kozlov, D. Watters, K. Spring, T. Yen and M. F. Lavin (1997). "Isolation of full-length ATM cDNA and correction of the ataxia-telangiectasia cellular phenotype." Proc Natl Acad Sci U S A **94**(15): 8021-8026.
Zufferey, R., T. Dull, R. J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini and D. Trono (1998). "Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery." J Virol **72**(12): 9873-9880.
Zufferey, R., D. Nagy, R. J. Mandel, L. Naldini and D. Trono (1997). "Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo." Nat Biotechnol **15**(9): 871-875.

APPENDIX I - List of *ATM* Mutations

Exon and intervening sequence (IVS)	Length (bp)	CDS position *(1 is A in ATG; 9,171 is A in TGA)*	Mutations from Human Gene Mutation Database® (updated 26 March 2016)				
			Missense and nonsense (370)	***Splicing (112)***	***Small deletions (160)***	***Small insertions (51)***	***Indels (22)***
3	355						
IVS 3	4408						
4	102	1-72	c.1A>G		c.27delT		
			c.2T>C		c.97delC		
			c.3G>A				
			c.37C>T				
			c.67C>T				
IVS 4				c.72+2T>C			
				c.73-2A>G			
5	113	73-185	c.103C>T		c.138_141delTTCA		
			c.128T>C		c.193delC		
			c.146C>G				
			c.170G>A				
IVS 5							
6	146	186-331	c.271C>T		c.217_218delGA		
			c.295A>G				
			c.299T>A				
			c.299T>A				
IVS 6				c.331+2T>G			
				c.331+5G>A			
				c.332-1G>A			
7	165	332-496	c.309C>G	c.487C>T	c.364_368delAATTA	c.397_398insT	
			c.320G>A		c.368delA	c.432dupA	

			c.331A>T		c.370_374delATCAT		
			c.362T>A		c.381delA		
			c.378T>A		c.450_453delTTCT		
			c.434T>G		c.480_484delTCAGC		
			c.452C>T				
			c.458G>T				
			c.484C>T				
IVS 7				c.496+5G>A			
8	166	497-662	c.513C>G	c.513C>T	c.510_511delGT		
			c.538C>T	c.662G>T	c.640delT		
			c.544G>C				
			c.572T>A				
			c.590G>A				
			c.597T>A				
			c.609C>A				
			c.610G>T				
IVS 8							
9	239	663-901	c.664C>T	c.748C>T	c.689delA		c.738_739delC TinsA
			c.667G>T	c.802C>T	c.717_720delCCTC		
			c.670A>G		c.741delT		
			c.692A>G		c.756_757delTG		
			c.691C>T		c.782delC		
			c.743G>T		c.790delT		
			c.742C>T		c.824delT		
			c.749G>A				
			c.748C>T				
			c.802C>T				
			c.875C>T				
IVS 9				c.901+1G>A			

				c.901+2T>A			
				c.901+3A>T			
				c.902-1G>T			
10	164	902-1065	c.902G>A		c.1027_1030delGAAA	c.1055dupT	
			c.967A>G		c.1028_1031delAAAA		
			c.995A>G		c.1037_1040delTTGA		
			c.998C>T				
			c.1009C>T				
IVS 10				c.1066-6T>G			
11	170	1066-1235	c.1110C>G	c.1235G>C	c.1110delC	c.1139_1142dupACAG	
			c.1120C>T		c.1158delG		
			c.1132A>G		c.1179_1180delGG		
			c.1153A>G		c.1215delT		
			c.1156A>G				
			c.1229T>C				
IVS 11				c.1236-404C>T			
				c.1236-2A>G			
				c.1236-2A>T			
12	372	1236-1607	c.1240C>T	c.1236G>T	c.1290_1291delTG	c.1332_1333dupCC	
			c.1316T>C		c.1348delG		
			c.1339C>T		c.1355delC		
			c.1346G>C		c.1358delC		
			c.1369C>T		c.1402_1403delAA		
			c.1396C>T		c.1514_1515delTT		
			c.1440A>C		c.1524delT		
			c.1464G>T		c.1564_1565delGA		
			c.1463G>A				
			c.1464G>A				
			c.1514T>C				

			c.1520T>A				
			c.1541G>A				
			c.1547T>C				
IVS 12				c.1607+1G>T			
13	195	1608-1802	c.1636C>G	c.1802G>T	c.1753_1756delTTAG		
			c.1648A>G		c.1782delA		
			c.1709T>C				
			c.1744T>C				
			c.1784T>C				
IVS 13							
14	96	1803-1898	c.1810C>T		c.1847_1849delCTA	c.1863dupA	
			c.1837G>T				
			c.1880T>G				
IVS 14				c.1898+1G>T	c.1898+3_1898+4delAT		
				c.1898+2T>G			
15	226	1899-2124	c.1924G>T		c.2062delG	c.1930_1931ins16	c.2114_2119delACTCATinsTCATAC
			c.1931C>A				
			c.1931C>G				
			c.1960C>A				
			c.2066C>G				
			c.2104C>G				
			c.2119T>C				
IVS 15							
16	126	2125-2250	c.2135C>A			c.2192dupA	c.2152_2154delTGTinsAAAC
			c.2158C>T				
IVS 16				c.2250+2T>C			
				c.2251-10T>G			

				c.2251-1G>A			
17	126	2251-2376	c.2276G>A	c.2376G>A	c.2284_2285delCT		
			c.2302A>G				
			c.2336T>C				
			c.2353C>T				
			c.2362A>C				
IVS 17				c.2376+1G>A			
18	90	2377-2466	c.2383C>A				
			c.2414G>A		c.2466+1delG		
			c.2413C>T				
			c.2465T>G				
			c.2512A>G				
			c.2548G>T				
			c.2554C>T				
			c.2572T>C				
			c.2614C>T				
IVS 18				c.2466+1G>A	c.2464_2466+2delTTAGT		
					c.2466+1delG		
19	172	2467-2638			c.2483delA	c.2502dupA	c.2547_2548delGGinsTT
IVS 19				c.2638+2T>G	c.2521delG		
				c.2838+1G>A			
				c.2639-384A>G			
20	200	2639-2838	c.2650C>T		c.2680delG	c.2806_2809dupCTAG	
			c.2662G>T		c.2689_2693delTTCTT		
			c.2734C>T		c.2720_2723delGTGT		
			c.2771G>A				
			c.2789T>G				

			c.2804C>G				
IVS 20							
21	83	2839-2921	c.2849T>G		c.2880delC		c.2839-3_2839delTAGTinsGATACTA
			c.2877C>G				
			c.2909T>G				
IVS 21				c.2921+1G>A		c.2921+2dupT	
				c.2921+3A>T			
				c.2921+5G>A			
				c.2922-2A>G			
				c.2922-1G>A			
22	156	2922-3077	c.2924A>G			c.3058dupA	
			c.2932T>C				
			c.2941C>T				
			c.3002T>A				
IVS 22				c.3078-1G>A			
23	76	3078-3153	c.3102T>A		c.3085delA		
			c.3102T>G		c.3111delT		
			c.3115A>T		c.3110delC		
			c.3136C>T				
IVS 23				c.3154-2A>C			
				c.3154-2A>G			
24	131	3154-3284	c.3161C>G	c.3284G>C	c.3219delA		c.3245_3247delATCinsTGAT
			c.3174G>A		c.3273_3284+4del16		
			c.3214G>T				
			c.3245A>T				
			c.3248A>G				
			c.3273G>T				
			c.3275C>A				

IVS 24				c.3285-2A>G			
25	118	3285-3402	c.3295G>A		c.3381_3384delTCAG		c.3754_3756delTATinsGA
			c.3334C>G		c.3388_3391delGGAA		
			c.3335C>T				
			c.3349C>T				
			c.3372C>G				
			c.3383A>G				
			c.3382C>T				
IVS 25				c.3403-12T>A			
26	174	3403-3576	c.3467C>T	c.3576G>A			
			c.3485T>G				
			c.3511C>T				
			c.3541A>T				
IVS 26				c.3576+687C>T			
27	170	3577-3746	c.3630G>A		c.3602_3603delTT		
			c.3663G>A		c.3626_3627delTT		
			c.3668A>G		c.3668delA		
			c.3673C>T		c.3693_3697delATCTT		
			c.3763T>G		c.3695_3713del19		
IVS 27				c.3746+1G>A			
				c.3747-1G>C			
28	247	3747-3993	c.3836G>A		c.3802delG	c.3880dupA	
			c.3848T>C		c.3850delA	c.3894dupT	
			c.3886C>T		c.3856delT		
			c.3931C>T		c.3905_3911delGTACCAG		
					c.3910_3916delAGAGACA		
IVS 28	3,026			c.3993+1G>A			

				c.3993+5G>T			
				c.3994-159A>G			
				c.3994-2A>G			
29	116	3994-4109	c.4066A>G		c.4002_4005delCTTA	c.4007_4008insA	
					c.4052delT		
					c.4105delT		
IVS 29				c.4110-9C>G			
				c.4110-1G>A			
30	127	4110-4236	c.4138C>T		c.4111delG	c.4143dupT	
			c.4148C>T				
			c.4198A>T				
IVS 30				c.4237-2A>G			
31	200	4237-4436	c.4258C>T		c.4358_4359delTA	c.4344dupA	
			c.4279G>A		c.4373delG	c.4415dupT	
			c.4303A>T		c.4389delT		
			c.4310G>A				
			c.4318A>T				
			c.4324T>C				
			c.4362A>C				
			c.4370T>G				
			c.4388T>G				
			c.4394T>C				
			c.4396C>T				
			c.4400A>G				
IVS 31				c.4437-1G>C			
32	175	4437-4611	c.4477C>G				
			c.4507C>T				
			c.4565G>A				
			c.4574T>C				
			c.4587T>G				

			c.4597G>T				
IVS 32				c.4612-12A>G	c.4611_4611+9del10		
33	165	4612-4776	c.4664T>A		c.4632_4635delCTTA	c.4661dupA	
			c.4697C>G		c.4642_4645delGATA		
			c.4709T>C		c.4649delA		
			c.4722G>T		c.4683_4689delTTTAGAT		
			c.4724G>A				
			c.4732C>T				
IVS 33				c.4776+2T>A			
				c.4776+2T>C			
				c.4777-2A>C			
34	133	4777-4909	c.4852C>T	c.4777G>T	c.4844delA	c.4840_4841dupCT	
			c.4858C>T				
IVS 34				c.4909+1G>A			
35	96	4910-5005			c.4989_4990delTG		c.4956_4957delGCinsTT
IVS 35							
36	172	5006-5177	c.5044G>T				
			c.5071A>C				
IVS 36				c.5177+1G>A			
				c.5177+5G>A			
37	142	5178-5319	c.5089A>G	c.5319G>A	c.5290delC		
			c.5186T>C		c.5309delC		
			c.5189G>T				
			c.5188C>T				
			c.5197G>C				
			c.5228C>T				
			c.5262G>T				
			c.5293C>T				

IVS 37				c.5319+2T>C	c.5320-5_5320-2delTCTA		
				c.5319+9A>G			
38	177	5320-5496	c.5326G>T		c.5324delT	c.5405dupA	
			c.5354C>A		c.5366delT	c.5441dupT	
			c.5410A>T		c.5435_5437delCTT		
			c.5414G>A				
			c.5415G>A				
			c.5419A>G				
			c.5441T>A				
			c.5471T>G				
IVS 38				c.5496+2T>C			
				c.5497-2A>C			
39	178	5497-5674	c.5515C>T		c.5539_5549del11	c.5554dupC	c.5631_5639delCTCGCAAACinsAAACA
			c.5554C>T		c.5549delT	c.5664dupC	
			c.5557G>A		c.5559_5560delTA		
			c.5558A>T		c.5650_5657delACAACCCC		
			c.5566G>T		c.5653delA		
			c.5574G>A				
			c.5623C>T				
			c.5644C>T				
IVS 39				c.5675-2A>C			
40	88	5675-5762	c.5692C>T			c.5712dupA	
			c.5738T>G				
			c.5741A>G				
			c.5749A>T				
IVS 40	2,175			c.5763-1050A>G			
				c.5763-2A>T			

41	156	5763-5918	c.5771C>A		c.5809_5813delAATTA		c.5791delGinsCCT
			c.5821G>C		c.5878delA		
			c.5825C>T		c.5910delA		
			c.5858C>T				
			c.5882A>G				
			c.5890A>G				
			c.5890A>T				
			c.5908C>T				
IVS 41							
42	88	5919-6006	c.5908C>T	c.5932G>T	c.5979_5983delTAAAG	c.5977dupA	
			c.5932G>T	c.5971G>T			
			c.5971G>T				
			c.6004C>T				
			c.6040G>T				
			c.6047A>G				
			c.6056A>G				
			c.6067G>A				
			c.6082C>T				
			c.6088A>G				
			c.6095G>A				
IVS 42							
43	89	6007-6095		c.6095G>A	c.6056delA	c.6015dupC	
IVS 43				c.6095+1G>A	c.6096-9_6096-5delTTCTT		
44	103	6096-6198	c.6100C>T				
			c.6108T>G				
			c.6115G>A				
			c.6145T>G				
			c.6188G>A				

IVS 44				c.6198+1G>A			
				c.6199-1G>T			
45	149	6199-6347	c.6200C>A	c.6154G>A	c.6215delG		
			c.6203T>C		c.6228delT		
			c.6315G>C				
			c.6314G>C				
			c.6325T>G				
IVS 45				c.6347+1G>A			
46	105	6348-6452	c.6326G>A		c.6444delA	c.6371_6372insG	
			c.6332A>G			c.6404_6405insTT	
			c.6374A>G				
IVS 46				c.6452+1G>A	c.6453-1_6457delGAGTAA		
47	120	6453-6572	c.6482G>C				
			c.6491A>C				
IVS 47				c.6572+1G>A			
				c.6572+1G>T			
				c.6573-9G>A			
48	235	6573-6807	c.6575C>G		c.6573_6577delATCAG	c.6776_6777dupCT	c.6672_6680delGGCTCTACGinsCTC
			c.6679C>T		c.6629delA		c.6736_6755delins20
			c.6685G>T		c.6710delA		
			c.6761A>C		c.6754delA		
					c.6779_6780delTA		
IVS 48							
49	168	6808-6975	c.6814G>A		c.6812delC	c.6867dupT	c.6910_6911delGAinsTG
			c.6820G>A		c.6916_6917delAG	c.6908dupA	
			c.6829C>G				
			c.6838C>T				

			c.6860G>C				
			c.6873G>A				
			c.6890A>C				
			c.6913C>T				
			c.6919C>T				
IVS 49				c.6976-3C>G			
50	114	6976-7089	c.6988C>G		c.7010_7011delGT	c.6987_6988insT	
			c.6995T>C			c.6997dupA	
			c.7013T>C			c.7037dupC	
			c.7028A>G				
IVS 50				c.7090-2A>C			
51	218	7090-7307	c.7096G>T		c.7150delA	c.7159_7160insAGCC	
			c.7166C>G		c.7241_7244delAAGC		
			c.7181C>T		c.7278_7283delCCTTAG		
			c.7187C>G		c.7279_7284delCTTAGG		
			c.7202T>C				
			c.7271T>G				
			c.7291A>G				
			c.7305C>G				
IVS 51				c.7307+1G>A			
52	208	7308-7515	c.7311C>A	c.7449G>A	c.7519_7520delGA	c.7481dupA	
			c.7316T>C		c.7517_7520delGAGA		
			c.7327C>T				
			c.7355T>C				
			c.7375C>G				
			c.7390T>C				
			c.7397C>T				
			c.7408T>G				

			c.7446G>A				
			c.7449G>A				
			c.7456C>T				
			c.7462T>C				
			c.7463G>T				
			c.7463G>A				
			c.7471T>C				
IVS 52							
53	114	7516-7629	c.7542T>A				c.7793_7798delGAACAGinsCC
			c.7542T>G				
			c.7570G>C				
			c.7592T>C				
			c.7606G>T				
			c.7622T>G				
IVS 53				c.7630-2A>C	c.7629_7629+4delTGTAA		
				c.7630-2A>T			
54	159	7630-7788	c.7660C>G	c.7653T>C	c.7638_7646delTAGAATTTC		
			c.7709A>G		c.7671_7674delGTTT		
			c.7775C>G		c.7705_7718del14		
			c.7777C>T				
IVS 54				c.7788+8G>T			
				c.7789-3T>G			
55	139	7789-7927	c.7792C>T	c.7789G>T	c.7829_7830delGA	c.7876_7877dupGC	c.7875_7876delTGinsGC
			c.7840C>T	c.7865C>T	c.7886_7890delTATTA		
			c.7865C>T	c.7788G>A	c.7925_7926delGA		
			c.7874A>G	c.7926A>C			

			c.7879T>C				
			c.7913G>A				
			c.7926A>C				
IVS 55				c.7928-1G>A	c.7927+5delG		
56	83	7928-8010	c.7967T>C		c.7989_7991delTGT		
			c.7985T>A				
			c.8000T>C				
			c.8003A>G				
IVS 56				c.8010+186C>T			
				c.8011-2A>C			
57	141	8011-8151	c.8030A>G	c.8150A>G	c.8030delA		
			c.8045C>G		c.8070delT		
IVS 57							
58	117	8152-8268	c.8071C>T	c.8156G>A	c.8206_8207delAA		c.8201_8211delins11
			c.8089A>G		c.8251_8254delACTA		c.8236_8239delAGGAinsTCCCT
			c.8096C>T		c.8264_8268delATAAG		
			c.8100A>T				
			c.8098A>T				
			c.8105T>G				
			c.8110T>C				
			c.8110T>G				
			c.8122G>A				
			c.8124T>A				
			c.8140C>T				
			c.8147T>C				
			c.8155C>T				
			c.8156G>A				

			c.8161G>A				
			c.8177C>T				
			c.8187A>C				
			c.8185C>T				
			c.8189A>C				
			c.8264A>G				
			c.8265T>G				
			c.8266A>T				
IVS 58				c.8269-1G>A			
59	150	8269-8418	c.8287C>T		c.8283_8284delTC		
			c.8293G>A		c.8368delA		
			c.8307G>A		c.8395_8404del10		
			c.8314G>A		c.8408delA		
			c.8373C>A				
IVS 59					c.8418+5_8418+8delGTGA		
60	166	8419-8584	c.8471G>A		c.8422delG	c.8476_8477dupAA	c.8565_8566delTGinsAA
			c.8473C>T		c.8432delA		
			c.8480T>G		c.8440delG		
			c.8479T>A		c.8484delA		
			c.8486C>T		c.8504_8518del15		
			c.8489T>G		c.8578_8580delTCT		
			c.8494C>T				
			c.8535G>A				
			c.8546G>C				
			c.8545C>T				
			c.8558C>G				
			c.8560C>T				
			c.8565T>A				
			c.8565T>G				

IVS 60				c.8585-2A>C			
				c.8585-1G>C			
61	87	8585-8671	c.8599G>C		c.8624delA		
			c.8620C>T				
			c.8624A>G				
			c.8633T>G				
			c.8659C>G				
IVS 61							
62	115	8672-8786	c.8672G>A		c.8711_8716delAGACAG	c.8714_8715dupCA	c.8720delCinsTT
			c.8711A>G			c.8766dupT	
			c.8725A>G			c.8769dupT	
			c.8726G>C				
			c.8734A>G				
			c.8737G>T				
			c.8741T>C				
			c.8756G>A				
			c.8774G>T				
IVS 62				c.8786+1G>A			
				c.8787-1G>T			
63	64	8787-8850	c.8793T>A	c.8850G>T	c.8802delC	c.8818_8821dupAACT	
			c.8824C>T		c.8814_8824del11		
			c.8850G>T		c.8833_8834delCT		
IVS 63				c.8851-973A>C			
				c.8851-1G>T			
64	137	8851-8987	c.8879G>A		c.8876_8879delACTG	c.8947dupA	c.8978_8981delGAAAinsAT
			c.8968G>T				
			c.8977C>T				
IVS 64				c.8987+1G>T			

				c.8988-1G>C			
65	3,775	8988-9171 (+UTR)	c.9008A>G		c.9001_9002delAG	c.9021dupA	c.9003_9004delTTinsC
			c.9019G>T		c.9145_9146delTT	c.9064dupG	
			c.9022C>T		c.9170_9171delGA	c.9079dupA	
			c.9023G>A				
			c.9029T>G				
			c.9031A>G				
			c.9068G>T				
			c.9103C>T				
			c.9139C>T				
			c.9169T>G				
			c.9170G>C				