ADDRESSING BARRIERS TO GENE THERAPY:
SAFETY AND EFFICACY

by

Kateryna Veronika Kratzer

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2022

This thesis is dedicated to Jane.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

The development of the CRISPR/Cas9 gene editing system has driven renewed interest in the field of gene therapy. However, despite recent advancements, barriers to clinical gene therapies remain. My research aims to address some of these barriers.

The first topic of this thesis is the identification of gene editing loci (GELs), genomic regions where genetic information can be safely incorporated. Targeting these sites will improve gene editing safety and efficacy. Here, I demonstrate the functionality of targeting a proto-GEL on human chromosome 15 and discuss a script to identify novel GELs.

The second topic focuses on the efficiency of gene insertion by DNA repair. I show that translesion DNA polymerase eta is important in the homology-directed repair pathway, which could help improve gene editing outcomes. I also reveal a potentially novel role of this enzyme in another aspect of genome maintenance, highlighting the dual nature of HDR-affecting enzymes.

LIST OF ABBREVIATIONS USED

AAV(S1)	Adeno-associated virus (integration site 1)
ADA	Adenosine deaminase
AID	Activation-induced cytidine deaminase
AiO	All-in-one (plasmid)
ALT	Alternative lengthening of telomeres
APB	ALT-associated PML nuclear bodies
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ATM	Ataxia telangiectasia mutated kinase
ATR	ATM Rad3-related kinase
BED	Browser extensible data
BER	Base excision repair
BIR	Break induced replication
BLM	Blood Syndrome helicase
BRCA1/2	Breast cancer 1/2, early onset
BSA	Bovine serum albumin
CAR-T	Chimeric antigen receptor T-cell
Cas	CRISPR-associated protein
Cas9n	CRISPR-associated protein 9 nickase
CCR5	C-C motif chemokine receptor 5
cDNA	Complementary DNA
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane regulator
Chr	Chromosome
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CtIP	Carboxy-terminal binding protein-interacting protein
DAPI	4'6-diamidino-2-phenylindole
DDR	DNA damage response
DEPC	Diethyl pyrocarbonate
dH_2O	Distilled water
D-loop	Displacement loop
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit

DNA2	DNA replication ATP-dependent helicase-like homolog
DSB	Double-stranded break
DSBR	Double-stranded break repair
EGFP	Enhanced green fluorescent protein
EXO1	Exonuclease 1
FA	Fanconi anemia
FACS	Fluorescence-activated single cell sorting
FANCD2	Fanconi anemia group D2 protein
FBS	Fetal bovine serum
GEL	Gene editing locus
GFP	Green fluorescent protein
G1/2 (phase)	Growth 1/2 (phase)
gRNA	Guide RNA
HA	Homology arm
HDR	Homology-directed repair
HIV	Human immunodeficiency virus
HNH	Histidine-Asparagine-Histidine [protein]
HR	Homologous recombination
hROSA26	Human homolog reverse orientation splice acceptor 26
HSC	Hematopoietic stem cell
IGR	Intergenic region
iRFP	Near-infrared fluorescent protein
KD	Knockdown
KI	Knock-in
KO	Knockout
LCA	Leber congenital amaurosis
LIG4	DNA ligase 4 (V)
LINE-1 / L-1	Long interspersed nuclear element, class 1 retrotransposon
LMNA	Nuclear lamin A gene
lncRNA	Long non-coding RNA
LPL(D)	Lipoprotein lipase (deficiency)
MRE11	Meiotic recombination 11 protein
MRN	MRE11, RAD50, NBS1 complex
NBS1	Nijmegen breakage syndrome 1 protein
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
OTC	Ornithine transcarbamylase
PALB2	Partner and localizer of BRCA2

PAM	Protospacer-adjacent motif
PAXX	Paralog of XRCC4 and XLF
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pegRNA	Prime editing gRNA
PFA	Paraformaldehyde
PML	Promyelocytic leukemia
PuroR	Puromycin resistance
RAD51	Recombinase protein homolog of bacterial RecA
RNA	Ribonucleic acid
RNase H2	Ribonuclease H2
RPA	Replication protein A
RT	Reverse transcriptase
RT-qPCR	Reverse transcription quantitative real-time PCR
RuvC	Crossover junction endodeoxyribonuclease
SCID	Severe combined immunodeficiency
SDSA	Synthesis-dependent strand annealing
SHS	Safe harbour site
S (phase)	Synthesis (phase)
SPV	Shope papilloma virus
ssDNA	Single-stranded DNA
TAL(EN)	Transcription activator-like (effector nuclease)
TLS	Translesion synthesis
tracrRNA	Trans-activating crRNA
Ub	Ubiquitin
UV	Ultraviolet
WB	Western blot
WT	Wildtype
XLF	XRCC4-like factor
XP(V)	Xeroderma pigmentosum (variant)
XRCC4	X-ray repair cross-complementing protein 4
ZFN	Zinc finger nuclease
53BP1	p53-binding protein 1
αMEM	Alpha modified Eagle's minimum essential medium

ACKNOWLEDGEMENTS

This chapter contains material (sections 1.1, 1.3.1, 1.3.2, 1.4) originally published in:

1.1 History of gene therapy

The foundation of modern gene therapy was laid in the early to mid-twentieth century, when researchers discovered the phenomena of bacterial transformation and transduction, demonstrating that cell phenotypes can be altered by the introduction of exogenous genetic material [1-4]. Scientists later learned to employ enzymes to recombine DNA into novel configurations. However, recombinant DNA research was temporarily put on hold in the 1970s after safety concerns were raised. These were summarily mitigated by review of the technology in 1975 [5-8], reopening the door for research on recombinant DNA that would beget future gene therapy attempts.
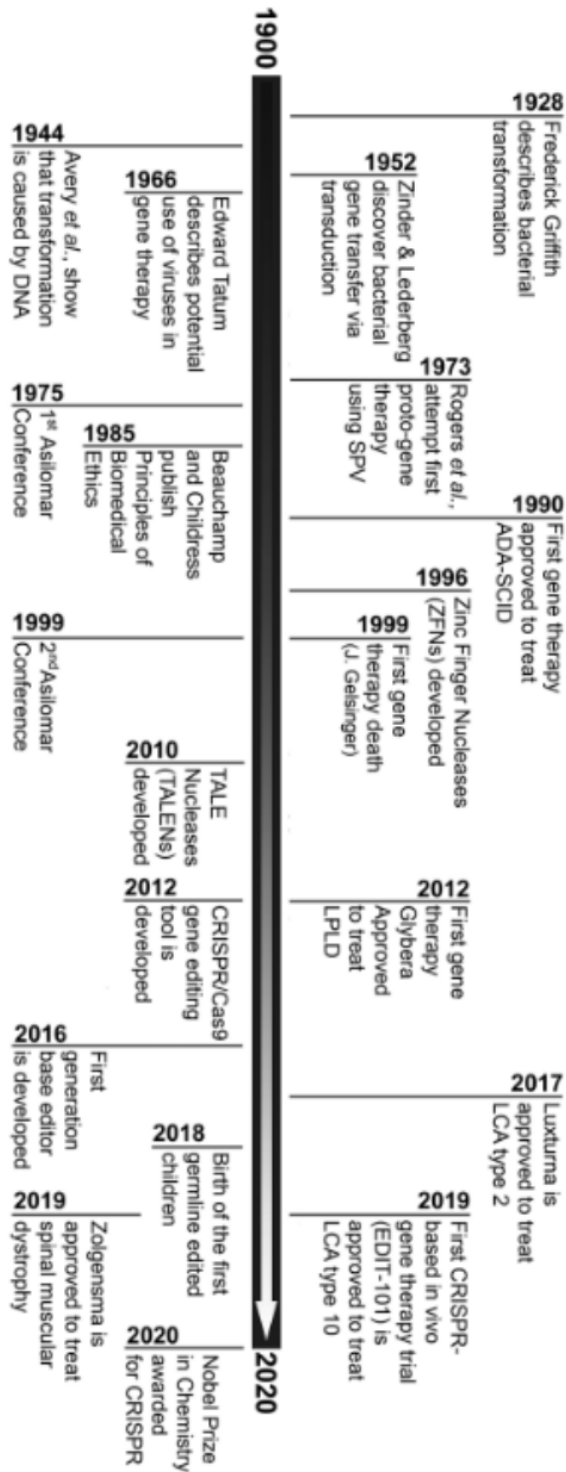
Figure 1. Timeline of important discoveries and events in the field of gene therapy.
Key events in the history of gene therapy are presented between 1900 and 2020. ADA-SCID adenosine deaminase-severe combined immunodeficiency, LCA Leber congenital amaurosis, LPLD lipoprotein lipase deficiency, SPV Shope papilloma virus. From Kratzer et al. 2021.

By this time, the first proto-gene therapy approaches were being tested. One was the use of infection with the Shope papilloma virus (SPV) to treat hyperargininemia, a liver urea cycle disorder characterized by elevated serum arginine, as the virus was believed to encode an arginase enzyme [3,9]. Despite the negative outcome of this trial – sequencing later revealed that the SPV genome did not encode an arginase – early discoveries and experiments paved the way for virus-based delivery of gene replacement and the gene therapy boom of the 1990s [3]. In 1990, the first gene therapy trial was launched to treat adenosine deaminase (ADA) deficiency, a form of severe combined immunodeficiency (SCID) [3]. This gene replacement approach involved removing diseased T-cells, correcting them *ex vivo* using a retrovirus to deliver a copy of the functional *ADA* gene, and infusing them back into the patient [10]. Although the therapy was effective at partially restoring the patient's immune system and allowed her to develop normally, this effect was temporary; today, she requires regular enzyme replacement therapy [3,11]. Despite the incomplete success, this remains a landmark event in gene therapy by demonstrating the potential efficacy of gene replacement and paving the way for future trials [11].

Soon after, an infamous trial was conducted to treat a young man named Jesse Gelsinger who had a mild form of the rare disease ornithine transcarbamylase (OTC) deficiency. This genetic disorder prevents the breakdown of ammonia, which can build up and cause brain damage, organ failure, and death in severe cases [3,12,13]. The controversy of this trial arises from the recruitment, treatment, and subsequent death of Gelsinger, as it was the first death directly attributed to gene therapy [3,10,12–14]. The backlash to the OTC gene replacement trial was swift. Rather than the cautious optimism

which had surrounded the field since 1975, the future of gene therapy was in doubt. A second meeting was held in 2000 to discuss the future of gene therapies [12]. There were calls for a complete halt to gene therapy, and human trials were suspended in the United States [12,13]. Ultimately, stricter policies were set, including a decade-long self-imposed global moratorium [12].

A major requirement for safe and effective gene therapy is the availability of precision gene editing tools. These tools began to emerge in the late 1990s with the development of programmable endonucleases designed to generate site-specific DNA double-strand breaks (DSBs) such as zinc finger nucleases (ZFNs) [15–18] and transcription activator-like (TAL) effector nucleases (TALENs) [19,20] (Fig. 2). ZFNs, which fuse an adaptable zinc finger DNA binding protein with a *Flavobacterium okeanokoites* (FokI) endonuclease catalytic domain (Fig. 2), were one of the first systems used for targeted gene editing [21,22]. Similarly to ZFNs, TALENs consist of FokI fused to multiple repetitive TAL effector domains to facilitate site-specific DNA cleavage and subsequent gene editing via DSB repair (Fig. 2). However, the size and repetitive nature of TALENs makes them less suitable for viral vector delivery, and both ZFNs and TALENs are more labour-intensive and expensive to design and build, hampering their versatility compared to the clustered regularly interspaced short palindromic repeats (CRISPR) gene editing system [22–25]. These programmable endonucleases bind as dimers to DNA to facilitate site-specific DNA recognition and induction of a DSB, which is then resolved by the cell's own DNA repair machinery, resulting in targeted gene editing (Fig. 2).

Following induction of a DSB, one of two endogenous DNA repair mechanisms may occur, producing different gene editing outcomes: non-homologous end joining

(NHEJ) or homology-directed repair (HDR) (Fig. 2). The predominant pathway throughout the majority of the cell cycle, NHEJ repairs broken DNA strands with no requirement for homology, causing random insertion/deletion (indel) mutations that are capable of disrupting a gene [26]. In brief, the DSB is rapidly recognized and bound by the Ku70/80 heterodimer, which stabilizes the broken ends and recruits a variety of NHEJ factors including the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) [27–30]. Binding of DNA-PKcs causes rotation of Ku70/80 away from the DSB [31], enabling loading of other repair factors promoted by p53-binding protein 1 (53BP1) [32], as required by the severity of the DNA damage [33–38]. DNA polymerases of the X family including μ and λ fill any gaps left by end processing [39]. Regardless of the factors recruited in the preceding steps, the broken strands are resected to create blunt ends which are finally ligated by ligase IV [40], inevitably altering genetic information at the break site (Fig. 2).

When a targeted break is repaired by NHEJ, this deliberate disruption can knockout (KO) a gene of interest. Alternatively, HDR relies on a homologous template normally sourced from undamaged sister chromatids for error-free repair, and thus is primarily restricted to late S and G2 stages of the cell cycle when a homologous sequence is present [41,42]. Less efficient but more accurate than NHEJ, the HDR pathway can be co-opted for gene editing when a repair template is provided that includes a new sequence of interest in addition to homology to the break site. Incorporation of this template through the mechanism of homologous recombination (HR) results in precise insertion of novel genetic material [43] (Fig. 2).

Just as in NHEJ, a variety of repair proteins and factors are involved in HR. The DSB is initially recognized by Ataxia telangiectasia mutated (ATM) kinase and a

complex of meiotic recombination 11 (MRE11), RAD50, and Nijmegen breakage syndrome 1 (NBS1) proteins, termed the MRN complex. The MRN complex tethers the ends of the DSB and initiates 5' end processing [44–46]. Additionally, MRN recruits the carboxy-terminal binding protein-interacting protein (CtIP) responsible for end resection, resulting in 3' overhangs [47]. These single stranded sections of DNA (ssDNA) seem to be the defining feature of DSB repair pathway choice since factors necessary for NHEJ are unable to assemble on ssDNA [48]. Once pathway choice is determined, the initial end resection is extended by DNA replication ATP-dependent helicase-like homolog (DNA2), exonuclease 1 (EXO1), and Bloom Syndrome helicase (BLM) [47]. 53BP1 and the Ku70/80 heterodimer inhibit the HR pathway by preventing accumulation of breast cancer 1, early onset (BRCA1) and stabilizing the break site against the formation of unstable ssDNA structures, respectively [47,48]. In HR, replication protein A (RPA) coats the ssDNA to promote stability and protect against degradation. ATM Rad3-related (ATR) kinase is subsequently activated by RPA and, together with ATM, phosphorylates BRCA1 [49]. BRCA1 likely serves many important roles in HR. It seems to promote the most high-fidelity pathway available, acting antagonistically with 53BP1 at DNA ends of a DSB when a homologous sequence is available for HR [50]. BRCA1 facilitates the recruitment of BRCA2 through direct interaction with the partner and localizer of BRCA2 (PALB2) [47,51], and in turn, BRCA2 recruits the recombinase protein homolog of bacterial RecA (RAD51). RAD51 replaces RPA to coat and protect the ssDNA filaments, and BRCA1 promotes RAD51-mediated strand invasion and homologous pairing with the donor strand, resulting in displacement (D-) loop formation [47,48,51]. The D-loop is resolved and DNA repair is achieved through one of three sub-HR pathways: double-

strand break repair (DSBR), synthesis-dependent strand annealing (SDSA), or break-induced replication (BIR) [48].

Reliance on these endogenous mechanisms for gene editing necessitates a deeper understanding of the repair factors and processes. However, much remains to be elucidated, including the broad role of BRCA1, and the function of Fanconi anemia (FA) and translesion synthesis (TLS) proteins in HR (further discussed in Section 4).
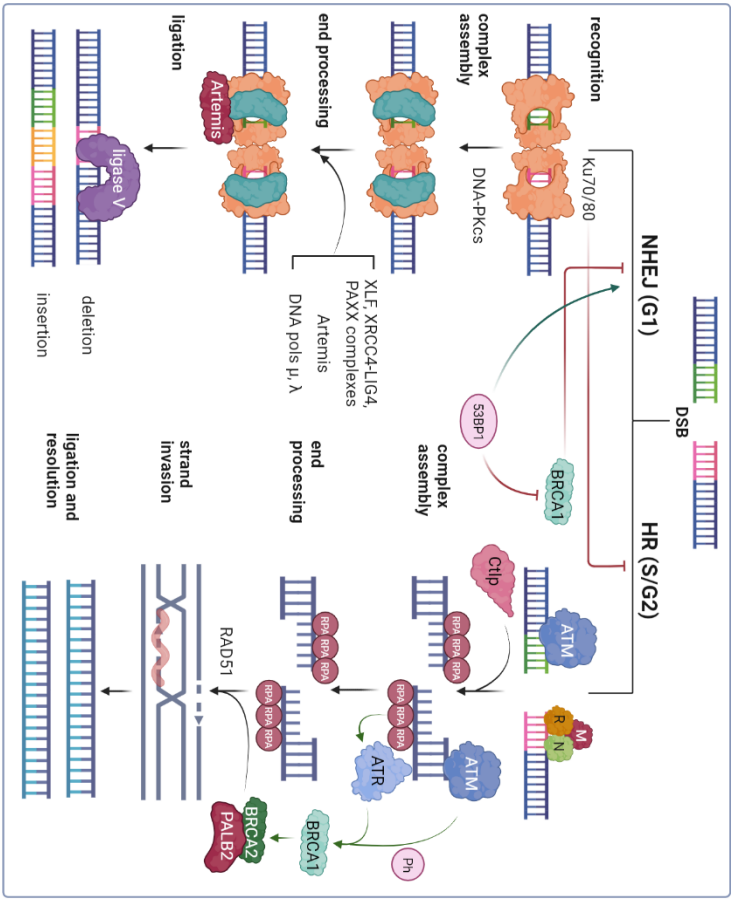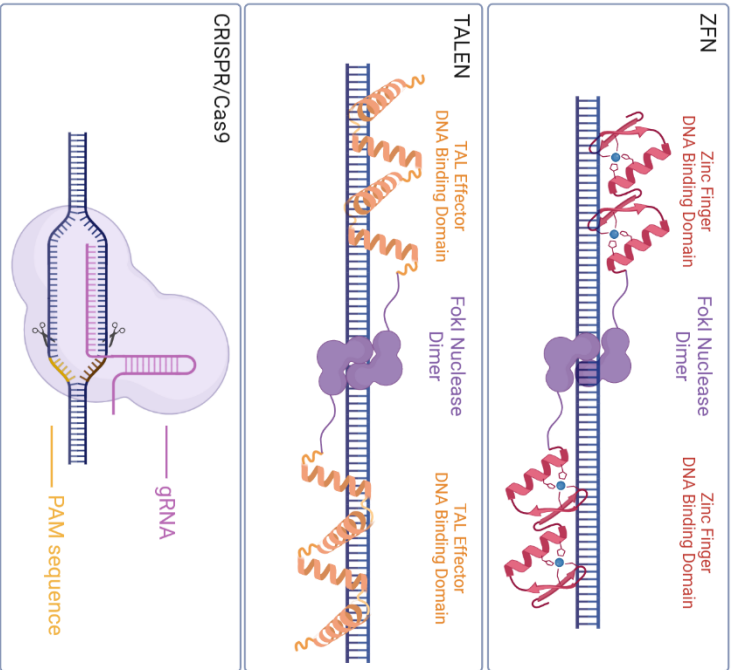
Figure 2. Gene editing platforms harness programmable nucleases and endogenous DNA repair. The three major gene editing systems (ZFNs, TALENs, CRISPR/Cas9) employ nucleases coupled with DNA-recognizing motifs to create targeted DSBs. ZFNs and TALENs are based on the fusion of a nonspecific cleavage FokI domain to a site-specific DNA binding protein domain, while CRISPR/Cas9 pairs the bacterial enzyme Cas9 with gRNA complementary to the target site. After site recognition, DSBs are induced at targeted sites and repaired by one of two major cellular repair pathways, resulting in gene editing.

In G1 phase of the cell cycle, the cell undergoes the highly efficient but more mutagenic NHEJ pathway. The Ku70/80 heterodimer binds the broken ends of DNA, followed by DNA-PKcs. Various end processing enzymes are recruited to the site, followed by final ligation by LIG4. In S/G2 phases when homologous sister chromatids are available, or when a homologous donor sequence is provided, BRCA1 restricts NHEJ in favour of HR. ATM and the MRN complex initially bind, sensing the break site. End resection occurs, resulting in long stretches of ssDNA that are repaired using an undamaged template strand. Green arrows denote promotion while red lines indicate inhibition.

ZFN zinc finger nuclease, TALEN transcription activator-like effector nucleases, CRISPR clustered regularly interspaced short palindromic repeats, Cas9 CRISPR-associated protein 9, gRNA guide RNA, PAM protospacer-adjacent motif, DSB double-stranded break, NHEJ non-homologous end-joining, HR homologous recombination, G1/2 growth phase 1/2, S synthesis phase, BRCA1/2 breast cancer 1/2, early onset, DNA-PKcs DNA-dependent protein kinase catalytic subunit, 53BP1 p53 binding protein 1, XLF XRCC4-like factor, XRCC4 X-ray repair cross-complementing protein 4, LIG4 DNA ligase 4 (V), PAXX paralog of XRCC4 and XLF, ATM Ataxia telangiectasia mutated, MRN meiotic recombination 11 (MRE11), RAD50, and Nijmegen breakage syndrome 1 (NBS1) complex, CtIP carboxy-terminal binding protein-interacting protein, RPA, ATR ATM Rad3-related kinase, Ph phosphorylation, PALB2 partner and localizer of BRCA2, RAD51 recombinase protein homolog of bacterial RecA.

Created with BioRender.com.

1.2 CRISPR/Cas9 and the resurgence of gene editing

A resurgence of interest in the field of gene editing occurred in 2012 with the application of CRISPR-associated protein 9 (Cas9). By isolating this protein from *Streptococcus pyogenes* and providing "programmable" RNAs to direct the endonuclease activity, a simple and precise genome editing system was created [52]. While this system was quickly named Science's Breakthrough of the Year in 2015 and went on to spur numerous gene therapy trials, CRISPR sequences were discovered serendipitously in 1987, though their function was not yet known [53]. CRISPR/Cas systems are a form of prokaryotic adaptive immunity whereby the short palindromic repeats in a host genome flank the bacteriophage sequences acquired from a previous infection [25]. These foreign sequences are processed by Cas enzymes into CRISPR RNA (crRNA), and the resulting complex recognizes and cleaves nucleic acids complementary to the crRNA, termed "protospacers" [52,54]. Therefore, upon re-infection by a recognized bacteriophage, the crRNA-Cas protein complex pairs to the complementary viral sequence for targeted destruction.

In the specific CRISPR system harnessed for gene editing in 2012, Cas9 proteins require a double-stranded RNA (dsRNA) structure of crRNA bound to a complementary approximately 100 nucleotide long trans-activating crRNA (tracrRNA) [52]. Cleavage occurs at sites determined by (a) base pairing between crRNA and the target sequence, and (b) a trinucleotide (NGG) protospacer adjacent motif (PAM) near the target region, identified by the PAM-interacting (PI) domain of the Cas enzyme (Fig. 2) [43,52]. The application of this system came from the discovery that an easily customizable gRNA molecule with partially double-stranded secondary structure could be used in place of the

crRNA-tracrRNA duplex [52]. The approximately 20 nucleotide long targeting segment of the gRNA can pair with complementary DNA not only in viral genomes, but in any target organism, and together with provided Cas9 enzymes, trigger cutting of the DNA at a site adjacent to a PAM sequence [43]. Subsequent repair of the DSB results in gene editing as with previous engineered nucleases (Fig. 2).

The CRISPR/Cas9 system has been further expanded to address additional challenges. In these augmented systems, one or both DSB-causing active sites on the Cas nuclease are deactivated, resulting in a broader Cas-based toolkit. Deactivation is caused by engineering point mutations to substitute an alanine amino acid into the active sites of the RuvC (D10A) or HNH (H840A) domains (Fig. 3A) [55]. When only one catalytic site remains active, Cas becomes a nickase (Cas9n), an enzyme with the same DNA-targeting abilities as the wildtype, but the restricted ability to cleave only a single strand of DNA (ssDNA). The Cas9n system has improved specificity due to the requirement for two binding-and-cutting events to co-occur, 40 – 130 nucleotides apart, on each strand of DNA before a DSB is created (Fig. 3B). While the gRNA for WT Cas9 ideally targets a single unique site in the genome, many sequences will have partial homology to additional sites, leading to the risk of off-target effects (see Sections 1.4.2 and 1.4.3 for further discussion). Requiring two distinct gRNAs to successfully target their associated Cas9n greatly reduces off-target effects by 50- to 1500-fold, though at the expense of efficiency [56].
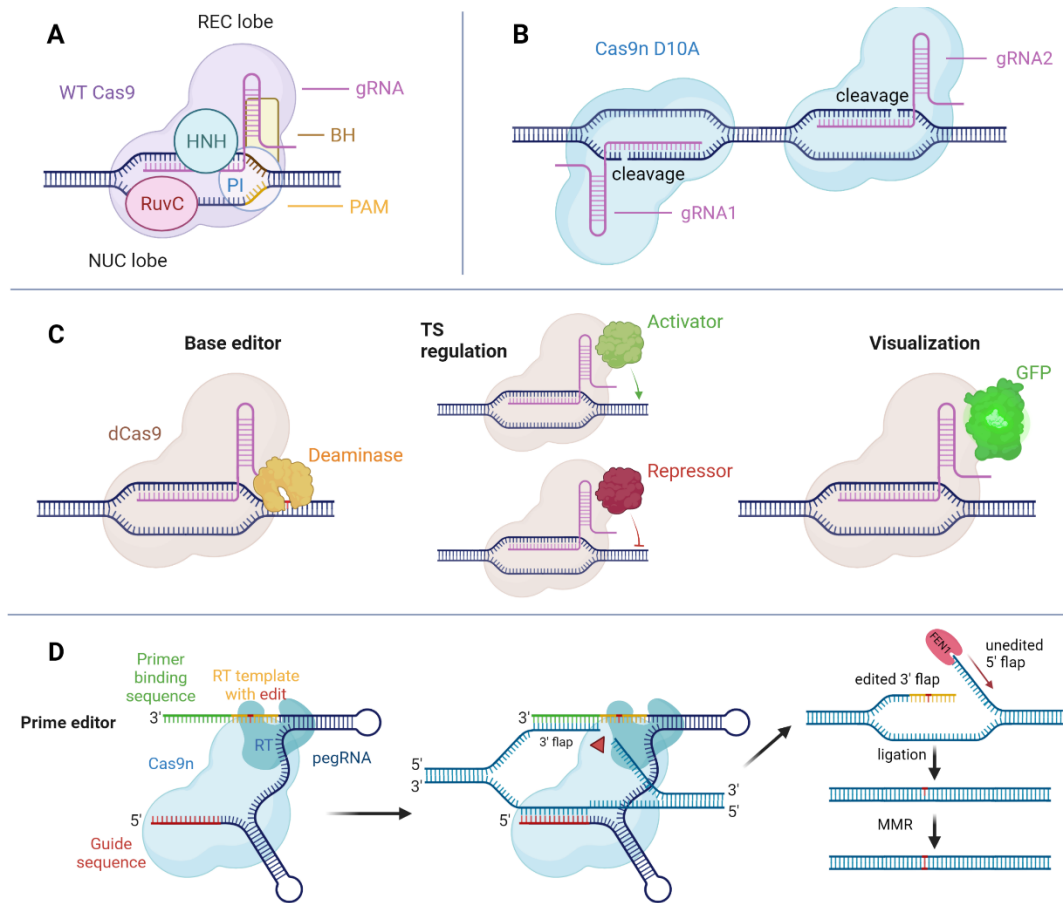
Figure 3. Variants of Cas9 for an expanded cell biology toolkit. A. The domains of *S. pyogenes* Cas9. Note that the active site of the HNH domain is responsible for cleavage of the target strand, while that of the RuvC domain cleaves the non-target strand. These domains are flexible and adopt transient conformations; the static gRNA-bound phase is shown here. B. Mechanism of paired nickases with two gRNAs to initiate a DSB. Two D10A nickases with mutated RuvC domains are shown. C. Tools based on dCas9 for base editing, transcription regulation, and nucleic acid visualization. D. Cas9n fused to a reverse transcriptase to form a prime editing system. The enzyme creates a nick, allowing RT to occur using the template with a single base mutation (red). The resulting unedited 5' flap of endogenous DNA is degraded, and the mutation is incorporated via cellular ligation and repair mechanisms. Cas9 CRISPR-associated protein 9, dCas9 catalytically dead Cas9 protein, REC recognition lobe, NUC nuclease lobe, HNH histidine-asparagine-histidine endonuclease domain, RuvC endonuclease domain named for homology to *Escherichia coli* RuvC crossover junction protein, PI PAM-interacting domain, BH bridge helix, DSB double-stranded break, gRNA guide RNA, PAM protospacer-adjacent motif, TS transcription, GFP green fluorescent protein, RT reverse transcriptase, pegRNA prime editing guide RNA, FEN1 flap endonuclease 1, MMR mismatch repair. Created with BioRender.com.

Mutating both RuvC and HNH active domains of WT Cas9 removes all catalytic activity, resulting in a catalytically dead protein (dCas9) [56]. Rather than cutting DNA, these systems take advantage of the targetable nature of the ribonucleoprotein complex to fuse additional proteins to dCas9 for homology-mediated delivery to a sequence of interest (Fig. 3C). A fused deaminase enzyme results in a base editor system, which allows for targeted deamination of a single DNA base – effectively changing the sequence of DNA without requiring induction of a potentially pathogenic DSB [25]. Cytidine deaminases coupled to dCas9 mutate cytosine to uridine in DNA, which is converted to thymidine by DNA repair mechanisms such as mismatch repair (MMR) or base excision repair (BER) [57,58]. Another method of achieving transition mutations is using an adenine base editor, which employs an adenine deaminase to achieve A-T to G-C conversion through the deamination of adenine to inosine, which is read as guanine by DNA polymerases [59]. Base editors therefore allow for targeted, precision editing without inducing potentially mutagenic DSBs, opening the door to corrective gene editing for point-mutation caused diseases.

A similar adaptation of the CRISPR system uses Cas9n rather than dCas9, resulting in the prime editing (PE) system. For PE, a reverse transcriptase (RT) enzyme is fused to the nickase which carries a multifunctional prime editing guide RNA (pegRNA) (Fig. 3D). This complex binds the target and creates a nick upstream of the PAM sequence, releasing a 3' flap which binds to the primer binding site (PBS) of the pegRNA [60–62]. The RT extends this flap by copying the pegRNA template, and the edited 3' flap ultimately displaces the unedited 5' flap, which is removed by the endogenous flap endonuclease 1 (Fen1) [60–62]. After ligation, the resulting double stranded DNA contains a

mismatch, which is repaired by MMR to either reflect the desired mutation or return to the original sequence. If the latter occurs, the PE complex can bind and start the process again. The advantage of this system is the ability to introduce desired base pair changes, unlimited by the action of available deaminases, and without causing potentially pathogenic DSBs.

Variants of dCas9 have even been employed for approaches beyond gene editing (Fig. 3C). In these methods, the CRISPR system is used as a modifiable nucleic acid binding platform. The CRISPR interference (CRISPRi) system causes a targetable steric block to transcription (TS), repressing a gene of interest [63]. By fusing TS activator (e.g., VP64) or repressor domains (e.g., KRAB) to dCas9, the system can further enhance or reduce gene expression, respectively [64]. The addition of acetyltransferases and demethylases facilitates epigenome modifications [65–67], paving the way for the treatment of epigenetic conditions [25,68]. An RNA editing platform was developed from another member of the Cas family, the protein Cas13, which can be further diversified with the fusion of additional modification proteins [69]. Finally, the fusion of fluorescent proteins to Cas9 enables visualization of 3D DNA and chromatin structure, including the labelling of up to six distinct loci [70,71].

In addition to the development of novel tools based on Cas9, synthetic biology has further refined the CRISPR/Cas9 system by modifying recognition and PAM-interacting sites to remove bulk, create high-fidelity enzymes, alter PAM recognition, or facilitate editing of multiple genomic sites [55,72]. The rapid advancement of a diverse range of genetic engineering tools has therefore expedited gene therapy and clinical trial progression, with some already available to patients.

1.3 Current gene therapies

1.3.1 Ex vivo gene therapy

In *ex vivo* gene therapy, cells are extracted, modified *in vitro*, and reintroduced to the patient (Fig. 4). This approach enables verification of on-target edits and thus has a more acceptable safety profile than *in vivo* gene therapy. Heritable blood disorders that benefit from gene replacement are ideal candidates for gene therapy. Hemophilia, a heritable blood clotting defect, can be treated by viral-mediated supplementation of the deficient clotting protein factor VIII, although there is also interest in direct editing strategies to correct the mutated gene [73]. As of the time of writing, over 45 gene therapy treatments are being developed for hemophilia [74], with several in late-stage clinical trials (see the 2022 pipeline report for detailed profiles: https://www.delveinsight.com/sample-request/hemophilia-a-pipeline-insight-2020?utm_source=cision&utm_medium=pressrelease&utm_campaign=kupr). Hemoglobinopathies such as beta-thalassemia and sickle cell anemia are defects in mature hemoglobin caused by various mutations on the β-globin chain-encoding gene *HBB* [75]. Disruption of the *BCL11A* locus, which encodes a transcription factor (TF) that suppresses fetal hemoglobin, has been shown to promote hemoglobin production by hematopoietic stem cells (HSCs), potentially providing an effective treatment for heritable hemoglobinopathies [76,77]. Editing the *HBB* locus itself can permanently correct the mutation; promising results were seen in preliminary *in vitro* experiments [78–81], and clinical trials are currently underway [82,83].
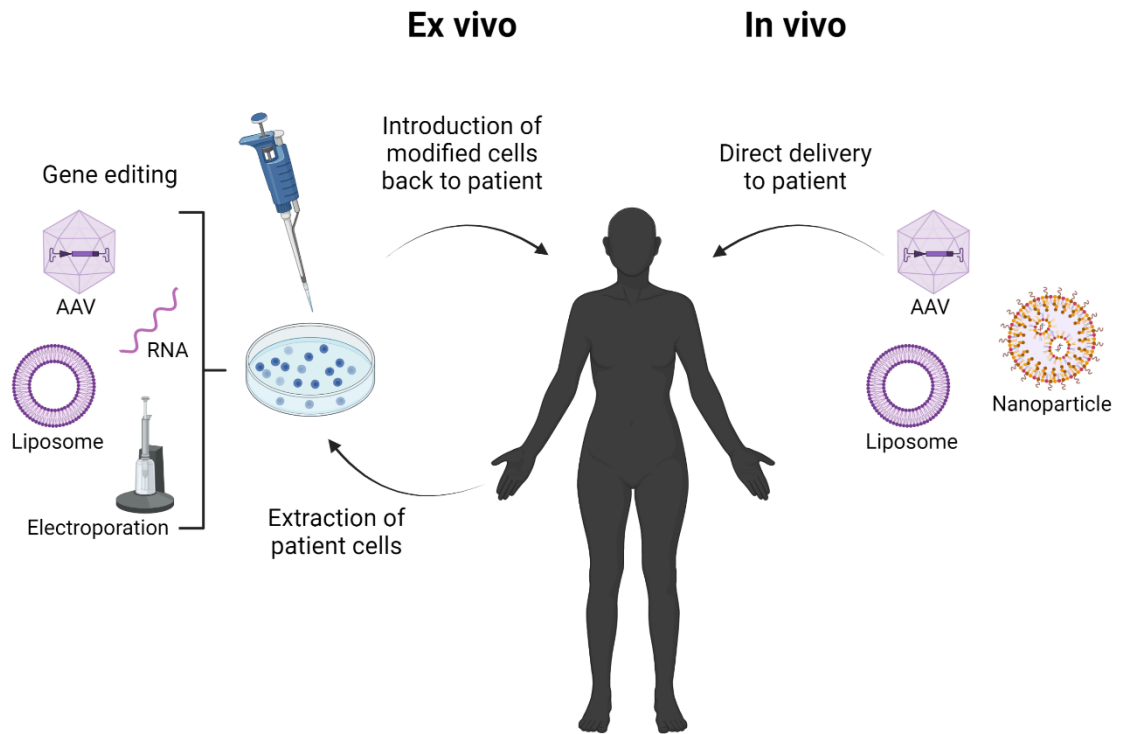
Figure 4. Overview of gene therapy approaches. In *ex vivo* gene therapy (left), patient cells (target tissue, stem cells, or immune cells) are extracted, modified by a gene editing tool which can be delivered to the cells by various transfection methods, and ultimately reintroduced to the patient. Alternatively, therapeutic agents may be directly delivered *in vivo* using viral vectors such as AAV, liposomes, or nanoparticles (right). *In vivo* gene therapy may involve direct introduction of a functional gene copy, but gene editing approaches pose greater potential risk to the patient from unintended on- and off-target effects or the initiation of an immune response. *Ex vivo* approaches allow gene editing side effects to be addressed prior to reintroduction, can prevent rejection or pathogenic immune response, and can employ methods such as modified RNAs or electroporation that would be less effective or impractical *in vivo*. AAV adeno-associated virus, RNA ribonucleic acid. Adapted from Kratzer *et al.* 2021. Created with BioRender.com.

Other types of diseases well-suited to *ex vivo* gene therapy are those that fall under the umbrella of severe combined immunodeficiency (SCID), a group of disorders caused by various mutations that affect the function of immune cells including thymus lymphocytes (T cells), bursa lymphocytes (B cells), and natural killer (NK) cells, and can lead to life-threatening infections. The potential for a genetic-based treatment for these diseases was recognized prior to 1990 (Fig. 1) and the first attempt was partially successful (see Section 1.1 for further detail). In a following clinical trial for the treatment of X-linked SCID (X-SCID), which was otherwise successful, four of ten patients developed leukemia due to insertional mutagenesis by the retroviral vector used for *ex vivo* gene delivery to endogenous HSCs which were then infused back into the patients [84,85]. However, modern approaches have improved the safety of delivery systems, and various medical products for the gene-based treatment of SCID disorders are in clinical trials with promising preliminary results or have been approved [86]. With the development of CRISPR and an ever-expanding toolbox of CRISPR-based editing tools (see Section 1.2), there is evidence that base editing can revert mutation status to rescue normal T cell activity [87]. Though more trials are required, promising results hint at a future where SCID patients no longer require expensive and painful bone marrow transplants for treatment.

A third branch of candidate diseases that benefit from *ex vivo* gene editing are blood cancers such as lymphomas, leukemias, and multiple myelomas that can be treated by chimeric antigen receptor (CAR) -T cell therapy. Using this method, gene editing is employed not to correct a pathogenic mutation, but as an immunotherapy to engineer patients' immune cells to target their cancers. Follow-up studies of patients who went

into remission show long-term stability and persistence of the CAR-T cell population [88], with six CAR-T cell therapies currently approved and in use for cancer treatment [89]. *Ex vivo* gene therapy approaches (Fig. 4) are therefore becoming well established as alternative means of disease treatment.

1.3.2 In vivo gene therapy

In vivo gene therapy directly introduces genetic material or modifies patient DNA. This approach carries significantly more risk due to unintended off-target effects. However, with improvement in the efficiency and efficacy of current gene editing tools, there are certain disease scenarios where *in vivo* gene therapy may be the preferred treatment method. These include monogenic, multiorgan diseases, or those with a self-contained scope of effect; namely, heritable disorders of the eye.

Mutation on any of 18 different genes causes Leber congenital amaurosis (LCA), a group of inherited retinal disorders that causes vision loss and is collectively the most common cause of inherited blindness. Mutations in the *RPE65* gene cause LCA type 2 (LCA2), as the RPE65 enzyme is required to regenerate the visual pigment necessary for rod and cone vision [90]. In 2017, Luxturna, a one-time injection to deliver a functional copy of *RPE65* to treat biallelic LCA type 2, was approved as the first gene therapy treatment for an inherited retinal disease [91]. Another promising *in vivo* gene therapy, EDIT-101, is undergoing Phase I/II clinical trials in patients with LCA10, another form of heritable vision loss caused by mutations in the *CEP290* gene [92]. In contrast to previously established supplementation forms of gene delivery, EDIT-101 is the first to use CRISPR to edit the genetic code directly inside the human body by targeting the

disease-causing point mutation in IVS26, the intron of *CEP290* that creates a cryptic splice site resulting in a truncated non-functional CEP290 protein [92,93].

Genetic engineering using CRISPR or related tools *in vivo* is still rare in clinical trials due to a greater degree of uncertainty. Instead, direct gene replacement treatments via viral vectors have become increasingly available in the last decade for various monogenic disorders. The first gene therapy was made commercially available in 2013, when Glybera was approved to treat the rare genetic disorder lipoprotein lipase deficiency (LPLD) through viral vector delivery of the lipoprotein lipase (*LPL*) gene [75]. However, after treatment of only 31 people worldwide, the drug was withdrawn due to high cost and commercial failure [94–96]. In 2019, Zolgensma debuted to treat the previously incurable disease spinal muscular atrophy, and quickly gained notoriety as the world's most expensive drug [97].

A disease for which *in vivo* gene therapy has been attempted, but with limited success in patients, is cystic fibrosis (CF). CF is caused by mutations in the *CFTR* gene which encodes the cystic fibrosis transmembrane regulator (CFTR) protein, a chloride channel that affects multiple organ systems, most notably the lungs [98]. Impaired CFTR transport of chloride ions in the lungs dysregulates salt and fluid homeostasis and promotes a pro-inflammatory environment through recurring infections and chronic lung damage, leading to organ failure and death [43,98]. Due to the multiorgan nature of CF, and as the most common fatal genetic disorder, this monogenic disease has been an intriguing candidate for gene therapy [99]. However, clinical trials have thus far been disappointing. Due to a variety of factors including the difficulty in targeting the affected organs and that a full-size functional copy of the *CFTR* gene is too large for most delivery systems for gene

replacement to be successful, only one Phase I/II clinical trial was in place and has since been discontinued [99,100]. However, a potential breakthrough emerged with the expansion of the CRISPR toolbox, as prime editing has been shown to successfully correct mutated *CFTR* in patient-derived organoids [101]. Despite the promise of this technology, further optimization and technological development is needed in order to achieve success in future *in vivo* clinical gene therapy for CF, as well as many other diseases.

1.3.3 Breaking a faulty gene

In many disease scenarios, it may be simpler to disable the mutated gene and prevent the production of faulty protein altogether rather than attempting to edit or correct the mutation. This approach is especially attractive for diseases with a single genetic cause such as Huntington's disease; a rare, fatal, autosomal dominant neurodegenerative disorder caused by mutations in the cytosine-adenine-guanine (CAG) trinucleotide repeat region of the *HTT* gene [102,103]. *HTT* encodes a protein which plays an important yet undefined role in neurons [104,105]. When the CAG repeat extends past a certain length, the pathogenically elongated protein is degraded into fragments which form toxic aggregates on neurons [103]. The resulting neural dysfunction leads to total physical and mental deterioration prior to death. With no treatment for Huntington's, gene therapy is an attractive method to potentially halt the devastating effects of the disease. In 2021, two clinical trials that employed antisense oligonucleotides targeting *HTT* were halted early due to negative results [106]. Data from ongoing 2022 Phase I/II trials of AMT-130, a microRNA delivered via an adeno-associated viral (AAV) vector directly to the brain to prevent the translation of huntington protein, are promising [107,108].

However, AMT-130 non-selectively targets both mutated and normal copies of *HTT*, as did the two previous therapies. Ongoing clinical trials are needed to ensure the safety and functionality of an *HTT*-knockdown therapy for Huntington's disease.

In another example of breaking a gene that poses a problem, two independent studies have recently shown that *in vitro* CRISPR-mediated disruption of human immunodeficiency virus (HIV) sequences can eliminate the virus from infected cells [109,110]. As a retrovirus, HIV integrates into the genome of immune cells, causing progressive immune failure termed acquired immunodeficiency syndrome (AIDS) [111]. Though AIDS was once seen as a terminal disease, it must be noted that modern antiretroviral therapy can allow individuals with HIV to live normal, healthy lives with virtually zero chance of transmitting the infection [112–114]. However, the potential to eliminate the virus from the genome using gene therapy remains intriguing. By using the CRISPR system paired with gRNAs targeting either multiple exons [110] or long terminal repeat (LTR) regions [109] of HIV, researchers were able to excise the viral sequence from the genome of infected cells with no detected off-target effects [110], though the viral sequences were able to persist and potentially display residual activity unless multiple targets were cleaved [109]. These proof-of-concept studies demonstrate the feasibility of a CRISPR-mediated cure for HIV, though work remains to ensure safety and adequate delivery of the therapy to patients.

1.3.4 Harnessing new therapeutic applications of human gene editing beyond genetic diseases

The broad appeal of gene therapy approaches is even being explored in the context of non-genetic diseases. Type 1 diabetes (T1D) is an autoimmune disorder which can be caused by various factors and is characterized by the inability of pancreatic β-cells to produce insulin [115,116]. Insulin is a hormone required for the entry of glucose into cells to provide energy; without insulin, consumed glucose accumulates, and high blood glucose leads to serious long-term complications [116,117]. Current treatment for T1D involves continuous blood glucose monitoring and daily insulin shots, therefore a potential cure mediated by genetic engineering of the defective β-cells is appealing. Pre-clinical trials have shown promising results for gene therapies that correct insulin deficiency [118,119], promote β-cell replacement outcomes [120,121], modulate the immune response to β-cells transplant [122,123], or promote insulin production by non-β-cells [124–126].

Genetic engineering to combat disease is not limited to editing human cells. Pathogens such as bacteria can be targeted by CRISPR/Cas9 gene editing approaches; indeed, the CRISPR system naturally occurred in prokaryotes prior to being harnessed as a gene editing tool (see Section 1.2). With the growing concern of antibiotic resistance, the use of bacteria-targeting viruses to combat bacterial infections – phage therapy – is increasing in popularity as an alternative to standard antibiotic treatment [127]. Bacteriophages can be genetically modified *ex vivo* prior to introduction to the patient to improve their effectivity against infectious bacteria. For example, this approach may prove superior in treating chronic urinary tract infections [128], the current treatment of which is typically a continuous antibiotic regiment [129].

The pre-clinical and clinical trials discussed here are merely a sample of the gene therapies currently in development. There is immense promise in personalized genetic cures for a plethora of diseases, many of which currently have inadequate treatment options; however, there are still barriers to address before gene therapy becomes a widely accessible clinical option.

1.4 Addressing barriers to clinical gene editing

1.4.1 Delivery systems

A number of technical hurdles still stand in the way of the widespread application of gene therapy and medical gene editing. One common obstacle is the delivery of the therapeutic components to target cells. Viral vectors remain at the forefront of the field for their relative ease of manufacturing and innate ability to target cells [75,130,131], although due to potential immunologic and oncogenic adverse effects, other delivery methods and vectors are being developed [130–133]. Synthetic systems overcome the potential problems associated with modified viruses as delivery vehicles such as immune response, undesired integration into the genome, and small carrying capacity [130–132,134–136] (Table 1). Work is also being done to improve viral vector design [137], with improvements achieved by capsid optimization [138], oversized adeno-associated virus (AAV) gene delivery [135], and dual vector systems [130,135,139]. Ongoing efforts to improve vector design will enhance the safety and efficacy of gene therapy delivery.

Table 1. Advantages and disadvantages of gene delivery systems.

		Advantages	Disadvantages
Viral vectors	**Adenovirus**	High efficiency *in vitro* and *in vivo*	Episomal
			Transient expression
		Transduces different tissues and cell types	Requires packaging cell line
		Effective in proliferating and non-proliferating cells	Induces immune/inflammatory response
		Stable in various solvents and simple to purify	No genomic targeting
		Large insert size (36kb)	
		High titres	
	Adeno-associated virus	Integration into host genome at **AAVS1 SHS**	Possible insertional mutagenesis
		Robust transcription and sustained expression	Requires packaging cell line
			No directed genomic targeting
		Effective in proliferating and non-proliferating cells	Tedious to scale to therapeutically relevant titre
			Small insert size (5kb)
	Herpes simplex virus	Efficient transduction *in vivo*	Transient expression
		Neuronal tropism	Cytotoxic
		Latent, long-term expression	Requires packaging cell line
		Large insert size (40-50kb)	No genomic targeting
			Tedious to scale to therapeutically relevant titre
	Retrovirus (e.g., lentivirus)	Integration into host genome via RT	Low efficiency
		Sustained expression	Possible insertional mutagenesis
		Transduces different tissues and cell types	Requires packaging cell line
		Medium insert size (8-12kb)	No genomic targeting
		High titres	Tedious to manufacture stock
			Safety concerns (HIV)

		Advantages	**Disadvantages**
Non-viral physical methods	**Electroporation**	Improved uptake of naked DNA *ex vivo* and *in vivo*	Limited *in vivo* application to accessible organs
		Low immunogenic profile	High electric density at needle tip of electrode may cause tissue damage
		Optimized transfection efficiency and cell survival for various cell types	
		No gene size limit	Expression is not targetable
		Gene expression is constrained to treatment area	
	Sonoporation	Improved uptake of naked DNA *ex vivo* and *in vivo*	Limited load capacity
		Non-invasive	Short circulation half-life of microbubbles
		Low immunogenic profile	
		Able to reach internal organs without surgical intervention	May damage surrounding tissue
		Improves permeability of blood-brain barrier	Expression is not targetable
		Microbubbles can be further engineered	
	Particle bombardment	No gene size limit	Limited *in vivo* application to accessible organs
		Simple production of metal particles	Higher immune response than microinjection
		Nontoxic components of inert gas and heavy metal particles	Potential tissue damage
			Gene expression is short term and low
			Expression is not targetable
	Microinjection	May be combined with other methods for improved gene uptake	Requires expensive equipment and high degree of skill
		Gene delivery may be physically targeted to cellular region	Low transfection rate
		No gene size limit	Inefficient procedure injects one cell at a time
		Low toxicity and immunogenic response	

		Advantages	Disadvantages
Non-viral physical methods	**Hydrodynamics**	Improved uptake of naked DNA *ex vivo* and *in vivo*	Overload of fluid can cause tissue damage and cardiac failure
		Enhanced transduction efficiency when coupled with viral vectors	Invasive procedure
			Effectiveness and tolerability depend on animal physiology
		Targetable to any organ via vein/capillary or local injection	Requires precise control of injection volume
Non-viral chemical methods and synthetic vectors	**Liposome**	Effective *ex vivo* and *in vivo*	Short-term gene expression
		Effective at low doses	Less effective than viral vectors *in vivo*
		Easily formulated and produced	
		No biological damage or immunological response	
		No risk of insertional mutagenesis	
		Easily modifiable	
		No gene size limit	
	Nanoparticle	Effective *ex vivo* and *in vivo*	Short-term gene expression
		No biological damage or immunological response	Less effective than viral vectors *in vivo*
		No risk of insertional mutagenesis	
		Easily modifiable	
		No gene size limit	
		Deliverable directly to the nucleus, avoiding cellular degradation	
	Polymer	Wide array of natural and synthetic polymers available for various applications	Short-term gene expression
			Less effective than viral vectors *in vivo*
		Biodegradable polymers improve safety compared to viral vectors	Tedious design and selection process
		No risk of insertional mutagenesis	Small carrying capacity

Due to their innate ability to infect and deliver encapsulated material to cells, viruses were the first vectors employed for gene therapy delivery vehicles. They are simple to administer via injection or inhalation methods but due to diffusion, high titres are required for an effective treatment. Additional obstacles, such as potential patient reactions to viral vectors, prompted the development of alternative gene delivery methods. Physical approaches apply a force to increase cell membrane permeability and promote gene transfer; however, these forces can cause tissue damage. Chemical methods utilize synthetic and naturally occurring non-viral vectors, drastically reducing the risk of initiating a dangerous immune response to the therapy. HIV human immunodeficiency virus; SHS safe harbour site; RT reverse transcription.

1.4.2 Genomic targets for safe gene insertion

While much progress has been made improving the targeting precision of Cas-based systems to reduce the risk of off-target effects (Fig. 3), an often-overlooked potential problem is unintended *on*-target effects. Cas-induced mutations may be harnessed therapeutically but undesirable mutations at the target site can be detrimental [133]. In addition, targeting a gene for insertion of a therapeutic segment of DNA, even if done with minimal or no potential for off-target cleavage, can potentially affect the transcription of that gene or genes in the vicinity of the edited locus. Our current lack of knowledge about the function and regulation of many genes and intergenic regions, some of which may be considered 'unimportant' and therefore 'safe' to target (e.g., *CCR5*), could lead to unintended consequences in a particular cell or tissue type that relies on the function(s) of these and co-regulated loci. This is concerning, potentially when considering genetic interventions that alter DNA in multiple tissues or the germline. Two possible solutions are (a) increasing the comprehensiveness of current genome annotation data and (b) restricting targeting efforts to safe harbour sites (SHS), regions in the genome where novel genetic information can be predictably incorporated without altering host genome function.

1.4.3 Overcoming inefficiency of homologous recombination for precision gene insertion

A final technical hurdle to widely available gene therapy is the relatively low efficiency of HR compared to NHEJ and the potential effect of homeology, the degree of mismatches between two similar DNA sequences [140,141], on the efficiency of HDR. Mismatches within four to 13 base pairs of the PAM sequence [142,143], required for precise

localization of CRISPR-mediated gene editing (Fig. 3A), can reduce Cas target

specificity and promote error during HDR. Natural genetic variation between individuals

could therefore alter the specificity of CRISPR/Cas-based therapies [144]. Other factors

such as cell cycle stage, overexpression of BRCA1 mutants, the presence of

oligonucleotides homologous to the insert sequence, and increased ratios of gRNA to

donor sequence can also affect HDR efficiency and thus both the application and

precision of gene editing [26,145–147]. The NHEJ repair pathway is active throughout the cell

cycle; conversely, as a repair mechanism that requires the presence of a template

sequence, HR is primarily restricted to S/G2 phase [42,148]. HR is therefore typically limited

to cycling cells, complicating gene editing strategies in tissues with little to no cellular

turnover such as the brain [42,43,149,150]. Much remains to be understood regarding safely

improving the rate of HDR for gene editing purposes.

1.5 Overview

In this thesis, I explain the barriers still present to widespread clinical gene therapy and present the results of my studies aimed at improving the safety and efficacy of CRISPR/Cas-mediated gene editing. As discussed above, sites used for gene insertion have been classically termed "safe harbour sites" but most, if not all, are found within genes and have not been thoroughly tested for safety in all developmental and tissue contexts. Therefore, I discuss the use of the term gene editing loci (GELs) as an alternative name for these sites and explore these regions as permissible sites for gene insertion by developing a toolkit designed to target genes of interest to a proto-GEL on human chromosome 15 (GEL15). I show how available human genome data can be searched within criteria frameworks for novel GELs. Additionally, I explore the issue of efficiency in CRISPR-mediated HDR for gene insertion and show a novel role for translesion polymerase eta (η) in this process. I created cell lines to study this enzyme and further examined the diversity of its functions, providing evidence of a role for Polη in other aspects of genome maintenance beyond HDR. Collectively, these findings explore aspects of genomic targeting and DNA repair that may improve gene editing safety and efficacy.

2.1 Identification of genomic sites for gene insertion

We previously identified and targeted an intergenic region on human chromosome 15 (GRCh37.p.13 primary assembly chr15:74696007-74698086) for gene insertion based upon visual scanning of the genetic structure according to the UCSC Genome Browser 146,151. This site was chosen for its proximity to the study gene of interest, the promyelocytic leukemia (PML) gene, and apparent open chromatin structure based upon methylation annotation tracks. At the time of discovery of the GEL15 site, there were no known nearby genes. Further interest in the use of safe harbour sites as targets for gene insertion prompted more detailed characterization of the GEL15 site. An updated scan was performed using criteria based on 152 (Table 2). External searches were performed using COSMIC (Tate et al., 2018; https://cancer.sanger.ac.uk) and NCBI 154.

Table 2. Criteria and corresponding data sources for identification of genomic regions amenable to gene insertion.

	Criterion	Data Source
Safety	>300kb from any cancer-related mutation	*COSMIC database*
	>150kb from any 5' gene end	Genes and gene predictions: UCSC Genes
	Intergenic region (not located within gene) *	Genes and gene predictions: UCSC Genes
	>300kb from functional small RNA	Genes and gene predictions: sno/miRNA
Functional silence	>50kb from any site of origin	Regulation: UW Repli-seq: Peaks
	>50kb from any ultra-conserved element	Regulation: VISTA Enhancers
	Low transcriptional activity (no mRNA ± 25kb)	mRNA and EST: Human mRNAs
Accessibility	Not in copy number variable region	Repeats: Segmental Dups
	In open chromatin (DHS signal ± 1kb)	Regulation: ENC Dnase/FAIRE: Open Chrom Synth
	Targetable by CRISPR/Cas9 (<30 Efficiency)	Genes and gene predictions: CRISPR Targets
Uniqueness	Single copy in human genome	*NCBI BLAST search output*

UCSC Genome Browser assembly hg19. Italic data sources are not tracks in the UCSC Genome Browser. *Not included as a criterion for the GEL identification code (Section 3.2.2). Adapted from [152].

2.2 Plasmid reagents and construction

All plasmids used and/or generated in this study are listed in Table 3.

Table 3. Plasmids used and/or generated in this study.

Plasmid	Description	Resistance	Source
pBlueScript II SK (+)	Empty cloning vector	Amp	Addgene #1946
Clover-LMNA donor	Clover-tagged LMNA CRISPR donor for HDR assay	Amp	[155]*
px330-LMNA gRNA	Lamin Cas9/gRNA for HDR assay	Amp	[155]*
iRFP	Expression of far-red protein for transfection control	Kan/Neo	Addgene #31857
Clover-J1	Clover expression for flow cytometry single colour control	Amp	Addgene #40259
AiO-Puro	All-in-one expression of Cas9n(D10A) and a gRNA pair with puromycin resistance gene	Amp	Addgene #74630
pGIPZ	Empty lentiviral vector with puromycin resistance gene	Amp	Addgene
GEL15-Cas9n-gRNA_1	Intermediate GEL15 Cas9n/gRNA1 in AiO-Puro	Amp	This study
GEL15-Cas9n-gRNA	Final Cas9n and gRNA pair to target GEL15	Amp	This study
pEGFP-Puro	Intermediate GEL15 Puro payload inserted into pEGFP-N1	Kan/Neo	This study
pEGFP-2A-Puro	Intermediate GEL15 EGFP-2A-Puro payload inserted into pEGFP-N1	Kan/Neo	This study
GEL15-MCS	Final GEL15 HAs flanking MCS for programmable targeting into GEL15	Amp	This study
GEL15-Puro	Final GEL15 HAs flanking puromycin resistance gene	Amp	This study – *in progress*
GEL15-EGFP-N1	Final GEL15 HAs flanking EGFP with C-terminal MCS	Amp	This study
GEL15-EGFP-C1	Final GEL15 HAs flanking EGFP with N-terminal MCS	Amp	This study

Plasmid	Description	Resistance	Source
GEL15-EGFP-2A-Puro	Final GEL15 HAs flanking EGFP-2A-Puro construct	Amp	This study
GEL15-EGFP-N1-POLH	Final GEL15 HAs flanking EGFP-N1 tagged polymerase eta	Amp	This study
pEGFP-N1	C-terminal GFP tagged protein expression	Kan/Neo	Addgene #2491
pEGFP-C1	N-terminal GFP tagged protein expression	Kan/Neo	Addgene #2487
pEGFP-N1-POLH[1]	N-terminal GFP tagged polymerase eta expression	Kan/Neo	J-Y. Masson[‡]
AiO-Puro-POLH_1	Intermediate POLH Cas9n/gRNA1 in AiO-Puro	Amp	This study
AiO-Puro-POLH2	Final Cas9n and gRNA pair designed to target exon 3 of POLH for puromycin-selectable knockout	Amp	This study
pBABE-Puro	Empty vector with selectable puromycin resistance	Amp	Addgene #1764
L1-neo-TET	Expression of hL1.3 with intron disrupting neo gene	Amp	Addgene #51284

Intermediate as well as final constructs are noted. [1]See Appendix B Figure 1 for a full plasmid map. *Dalhousie University, [‡]Université Laval. Amp ampicillin, Neo/Kan neomycin/kanamycin, HAs homology arms, MCS multiple cloning site, gRNA guide RNA, Cas9n nickase, hL1 human LINE-1.

2.2.1 Generation of GEL15 donor construct

The Ruby-LMNA donor vector constructed by Pinder et al. (2015) in the pCR2.1-TOPO backbone (Thermo Fisher Scientific) was digested at HindIII and XhoI restriction sites to remove the insert and allow for later insertion of the chr15 homology arm – multiple cloning site – homology arm (HA-MCS-HA) construct (Fig. 5A).

The two chr15 homology arm (HA) sequences were previously designed by Attwood et al. (2020) from whole-genome U2OS DNA in the pCR2.1-TOPO backbone (Thermo Fisher Scientific) (Fig. 5B) and were amplified for the purposes of this study using polymerase chain reaction (PCR) (Fig. 5C). The 3' and 5' HAs were amplified with the addition of a restriction site (XhoI or HindIII, respectively) and the MCS sequence which was: PacI-BlpI-MluI-SacI-Eco53kI encoded by 5'- TTA ATT AAG CTC AGC ACG CGT GAG CTC ATT AAT. The PCR master mix was: 34uL dH$_2$O, 10uL 5x HF buffer (Thermo Fisher Scientific), 1uL 10mM dNTPs, 1uL each forward and reverse primers (Table 6), 2uL backbone, and 1uL Phusion polymerase (Thermo Fisher Scientific). The PCR protocol was performed as follows:

Table 4. Step 1 PCR thermocycler profile.

	Temp (C)	Time	Cycles
Initial Denaturation	98	1min	1
Denaturation	98	10s	
Annealing	58	30s	35
Extension	72	25s	
Final Extension	72	3min	1
Hold	12	∞	1

The resulting intermediates were two HA-MCS fragments with restriction enzyme sites compatible with the linearized Ruby-LMNA donor described above. The fragments were gel isolated and extracted (QIAGEN, Hilden, Germany) prior to overlap extension PCR to create the HA-MCS-HA fragment. The master mix was: 33uL dH$_2$O, 10uL 5x HF buffer, 1uL 10mM dNTPs, 1uL 5'HA_F primer, 1uL 3'HA_R primer, 2uL 5'HA-MCS fragment (16ng/uL), 1uL 3'HA-MCS fragment (30ng/uL), and 1uL Phusion polymerase (Thermo Fisher Scientific). The PCR protocol was as follows:

Table 5. Step 2 overlap extension PCR thermocycler profile.

	Temp (C)	Time	Cycles
Initial Denaturation	98	1min	1
Denaturation	98	10s	
Annealing	60	30s	25
Extension	72	20s	
Final Extension	72	3min	1
Hold	12	∞	1

The resulting intermediate was the HA-MCS-HA fragment with compatible restriction enzyme sites on either side. This fragment was digested with XhoI and HindIII, PCR purified (QIAGEN, Hilden, Germany), and cloned into XhoI/HindIII-digested Ruby-LMNA donor backbone to create the GEL15 MCS donor plasmid. Clones confirmed to have the correct insert size by diagnostic restriction enzyme digest were sent for sequencing by Plasmidsaurus.com.

A

TTAATTAAGCTCAGCACGCGTGAGCTCATTAAT

PacI-BlpI-MluI-SacI-Eco53kI

M13_R   M13_F

5'HA   MCS   3'HA

Amp   Neo/Kan

B

Legend:

Blue text = sequence used for 5' homology template (5'HA)
Red text = sequence used for 3' homology template (3'HA)
Green text = PAM sequence
Bold = GEL15 target sequence (51bp)
Bold italic = Cas9n gRNAs
Underlined = sequenced PCR amplicon from wild-type U2OS genomic DNA

```
GGACAAAGATTAGATGCATCTTTTTGTTGTGAAATTTTGGTAAACTACTTACTCTTTCTCAGCCTCAGTTC
TTTGAACTATGAAATGGAGACAATAGTAATAGTATCAACCTCAAAGGGCTGGTGAGAAAACTAACCAAGAC
AATTCACACAAAGTTTCAGAACAGTCCCGGGCAGGTGCGTGACTTACGCATTTGGCTTTTGTTCACCACTA
TTACTACTTGCTTATTGAAGCTGAGCCAGGTGGCTGAGCAATGTCCCTTTCTGGCAAGGAGGGTGGAGGCA
AGTGTCACATTTCAGAAAGATCCCTGGGCTCCTGTTGAATCTGTCCCCACACCAAAACATGCCCGTGGGGG
CTCTAGCTCCTGCCCCAGCAGCGAAAGGGGCTCCTCCTTCTCTGGTGCCCAGTCCTAGCTCACCTGCTCTT
CACAAGCCTCCTCCCAGCACAGCCTGCACCACACAGGTGCTCAGTAAGGGGGTTGCTCTGAACAGGGAGCA
GGCCCACATGCTGAGAGTGGTGGAGGCGTGCAGGCTCTGATGACCACCTGAACCCAGACTCCAGGGCAGGA
CAGGTCCTTACAGACCGTAAGGGCTTCCTGCCTACCCTCTTCCCATGCAGTTACCTCAACCCTGGGACACC
AGGAAGGTTGAGCATTCCAGGAAAGCCCCGACCCACCCAGGAAACCCAAGTACCTGGTGGGGCCTCATCTG
GCCCATTGTCCTGAAAGGACAGGAATAAAGGCTCTCTCCTCATGGCCACACACCCTAGGGACGTCCAGCCT
CCAGCTTTCTGTGGGTGGGGCAAATGACTGAGGAGGCTTCCCTCCCTGCAGGCGCCTCCCTCTGGGGCCTC
CTGTGGTCCAGTCGTCCCTGGTCCCCAGCTCTCAGGTGGGTCATAGGGCATCTCAGCAGGACCACCAGGAA
GAGCATCAGGAGGGGAAGGAGCCAGAGAGCCCTCCAGGCCAGTGCACTGCTAACTCAACATCTCATCTTCC
TATCTCGTCCCTTCTGCTCACCTCCGGAGACTCCCCCTCCCCTCTCATGCTCTTCACTGGCAGGTTATGAT
GGCCCAGAGATCCCCACCAGAGTCCCAGATATCAGTGGGGATCCCAGCCTATTCCTCCAAGCTCCCAGTTG
CCAGTGACAGGGAGGCAGAAACAGCCTGGAAATTAACTGTGGGTTCTCCCTGAGCCTGTGGATACATGGAA
AGAGAGCTTCACCACCACCCCCTTCTCTGAATATACATACAGGATCTGAGGAGGGAGGATTCTGTGTGGAT
GCATGTGTGCACAGATGTGCATGTTTGTACGTGTGTATGGGTGAGCAGGTCATACGTGTGGTCCCATGTGC
CCACCTGTTGTATATATGCATGCTTGTTTTGTGTAAGTGAGCATAAGAGGTCAAGCCAGAGAGAGAAATAG
GAACAAGGAGAGACAGAGAGAGAGAGAATAGAAACAGAGACCAACAGGGAAACAGAATGAGACTGAGACAG
AGAACAACAGAAGGCCAGACAGGTAGACAAGCCACAGTTCCATCTTCCCTGGTGCTGGCTGCCAAGGCAGA
CAACTCTCCTCTCACTTCATTGATGTTCTGCCTGTGTTTATACCCATGATTGGGGTGGAGTTCAGGAAAGG
ATGTATCAGGGCTAGGACGCCCCCACCTCCACACTCACAGGGCCAGTCAGGGCCCTCTGCAGGGGCTGCCA
TTTCCTCAAGCTGAGGAATGGGCTTGGATCCCTCGGATGCCTTCCAGGGTGCAGGGGCGCTCCCAGTGACT
GGGCTCTTGTGGTGGCTGGGGAGGGGGAGCTCCTCTGACAGGCCCCATGCTGGGGGAAACGTCCTGAGACT
GACCTATCACGGGCTGACATGCACACCAAGGAGGTCTCACATCTGCTGGGAAATGAGGGGCTGGTGGAGGT
GAGAAAACCTCAAGTTCTAATGCCTTCCCAACACCTTCACCTCACCTGAGAAGCTTCTAGATTCCACCCAC
CCCTGTTTCCTTGGGGGAGGGTGTTGTGTGGAGGTGGGAAGAGGCAGGAGTCCAGGTTCTAGGCCAGGCTC
TGAAGGGGCAGGTAGCTCAG
```

C

5'HA_F
5'HA   MCS
5'HA_R

Step 1.

3'HA_F
MCS   3'HA
3'HA_R

Step 2.

5'HA_F
HindIII — 5'HA   MCS   3'HA — XhoI
3'HA_R

Figure 5. Design of the GEL15 donor construct. A. Representative map of GEL15 MCS donor plasmid. M13_R and M13_F are universal primers. B. Reference sequence of an intergenic region on chromosome 15 (hg19 chr15:74696007-74698086). Underlined text denotes the region of the PCR amplicon from wild-type U2OS genomic DNA that was verified through Sanger sequencing by Attwood et al. (2020). Adapted from [146]. C. PCR and overlap extension PCR strategy to create two HA-MCS intermediates (Step 1), then a HA-MCS-HA fragment (Step 2) with compatible restriction sites for digestion and insertion into backbone. Coloured regions of primers denote homology, black denotes additional sequence. HA homology arm, MCS multiple cloning site. Created with BioRender.com.

2.2.2 Generation of chromosome 15 targeted reporter constructs in the GEL15-MCS donor

Five reporter gene payloads were created for insertion into and verification of the GEL15-MCS vector (Table 3). All primers are detailed in Table 6. For GEL15-Puro, the puromycin resistance cassette was first amplified from pGIPZ (Addgene #2918) using the Puro F/R primers to add sites for EcoRI and NotI, PCR purified, and subsequently ligated into the pEGFP-N1 vector that had been linearized with these restriction enzymes. A similar two-step process was employed to create the dual reporter GEL15-EGFP-2A-Puro vector. The polycistronic expression cassette 2A-Puro was amplified from AiO-Puro (Addgene #74630) using 2A-Puro F/R to add sites for EcoRI and KpnI, which was then purified and ligated into the MCS of pEGFP-N1. Clones were selected for further processing based on correct insert size band dropout from diagnostic restriction enzyme digests.

All payload intermediates in pEGFP vectors were amplified using EGFP F/R primers to insert PacI and SacI sites. After digestion and purification using the QIAquick PCR purification kit (QIAGEN, Hilden, Germany), payloads were ligated into the complementarily linearized GEL15-MCS construct. Ligation reactions were prepared as follows: 2 uL 10x T4 DNA ligase buffer (Thermo Fisher Scientific), 50 ng linearized vector DNA, insert DNA at a 3:1 molar ratio (http://nebiocalculator.neb.com/#!/ligation), 1 uL T4 DNA ligase (Thermo Fisher Scientific), and dH$_2$O up to 20 uL. The mixture was incubated for 10 min at room temp to allow ligation of sticky ends, then heat inactivated at 65° for 10 min. Half of the ligation reaction (10 uL) was chilled on ice and transformed into 90 uL MAX Efficiency DH5α competent cells (Thermo Fisher Scientific).

The mixture was stirred gently with a sterile pipette tip and incubated on ice for 30 min prior to a 42° heat shock for 60 sec. Following ≥2 min incubation on ice, transformed competent cells were incubated in 500 uL LB media, shaking, at 37° for 45 – 60 min. Cultures were then plated on LB agar plates supplemented with the appropriate antibiotic for selection and incubated at 37° overnight. The next day, individual colonies were picked with a sterile toothpick and inoculated into LB media supplemented with the appropriate antibiotic to be incubated shaking at 37° overnight. Cultures were then mini- or midi-prepped using the applicable kit (QIAGEN, Hilden, Germany) for DNA extraction and purification. Constructs were sequence verified and mapped by Plasmidsaurus.com.

2.3 Primer and oligonucleotide reagents

All primers and oligonucleotides were ordered from Integrated DNA Technologies (IDT; Coralville, USA) and resuspended to 100 uM in nuclease-free water. Primers were designed, as described above, to PCR amplify intermediates and reporter constructs for GEL15 targeting (Table 6), and for reverse transcription quantitative-PCR (RT-qPCR) validation of POLH clonal cell lines (Table 7, see Section 2.6).

Table 6. Primers used for cloning in this study.

Primer	Function	Sequence (5'–3')
5'HA F 5'HA R	Amplify GEL15 5' HA with added MCS and HindIII site	<u>CATTAG</u>**AAGCTT**GGCTCTGATGACCACC <u>ATTAATGAGCTCACGCGTGCTGAGCTTAATTAA</u>AGGAATAGGCTGGGATC
3'HA F 3'HA R	Amplify GEL15 3' HA with added MCS and XhoI site	<u>TTAATTAAGCTCAGCACGCGTGAGCTCATTAAT</u>GTTCTCCCTGAGCCTG <u>TTGCAC</u>**CTCGAG**GACTGGCCCTGTGAGT
EGFP F EGFP R_SacI EGFP R_MluI	Amplify CMV-[protein/MCS]-EGFP-[protein/MCS]-PolyA out of pEGFP-N1 or -C1	<u>CATTAG</u>**TTAATTAA**ATAGTAATCAATTACGGGGTC <u>TTGCAC</u>**GAGCTC**TAAGATACATTGATGAGTTTGGAC <u>TTGCAC</u>**ACGCGT**TAAGATACATTGATGAGTTTGGAC
Puro F Puro R	Amplify Puromycin resistance cassette out of pGIPZ	<u>CATTAG</u>**GAATTC***ACCATGG*CCACCGAGTACAAG <u>GTGCAA</u>**CGCCGGCG**TCAGGCACCGGGCTTG
2A-Puro F_EcoRI 2A-Puro R	Amplify 2A-Puro out of AiO-Puro	<u>CATTAG</u>**GAATTC**TTAAAAGGCCGGCGGC <u>CATTAG</u>**GGTACC**TCAGGCACCGGGCTTG

5' primer extensions that do not match template are underlined, added restriction sites are bolded, Kozak consensus sequences are italicized. Multiple cloning site (MCS) indicated in blue.

Table 7. Primers used for RT-qPCR analysis.

Primer	Target	Sequence (5'–3')
POLH 1 F	POLH E2-3	TCATGGAAGGGTGGTGGAATA
POLH 1 R		TGAGGTTAGCTTTCCCACGG
POLH 5 F	POLH E11	ATCCAGACAGAATGGTCTCCT
POLH 5 R		GGTAACTGGCACCTTTGGCA
POLH 6 F		AATCCAGACAGAATGGTCTCCT
POLH 6 R		TGCTCAAGAAGCTGGTGATGT

E2-3 POLH exon 2 – exon 3 junction. E11 POLH exon 11.

Oligonucleotides were designed to create sgRNAs for targeting Cas9n(D10A) to either the GEL15 site or exon 3 of POLH (Table 8). Single-stranded oligonucleotides were annealed to form double-stranded gRNAs for insertion into AiO-Puro (Addgene #74630) as follows: 2 ug each oligonucleotide in 48 uL annealing buffer (10 mM Tris, pH 7.5-8.0, 50 mM NaCl, 1 mM EDTA) were incubated at 100°C for 5 min. Oligonucleotides were allowed to gradually cool to room temperature for 45 min to allow for annealing to occur, after which they were stored at -20°C.

Table 8. CRISPR gRNA oligonucleotide sequences.

CRISPR gRNA	Oligonucleotide sequence (5'–3')
AiO Chr15 gRNA1AX	ACCGGTCACTGGCAACTGGGAGCT
AiO Chr15 gRNA1AY	AAACAGCTCCCAGTTGCCAGTGAC
AiO Chr15 gRNA1BX	ACCGAACAGCCTGGAAATTAACTG
AiO Chr15 gRNA1BY	AAACCAGTTAATTTCCAGGCTGTT
AiO koPOLH gRNA1 m2AX	ACCGAACTCACTGCAATTATTCTA
AiO koPOLH gRNA1 m2AY	AAACTAGAATAATTGCAGTGAGTT
AiO koPOLH gRNA2 m2BX	ACCGTTGGAGTCACTAGAAGTATG
AiO koPOLH gRNA2 m2BY	AAACCATACTTCTAGTGACTCCAA

5' primer extensions that do not match template are underlined.

2.4 Cell culture, cell line generation, and transfection

2.4.1 Cell culture

All cell lines used and generated in this study are described in Table 9. Human normal fibroblast NHDF (ATCC) and XPV patient-derived fibroblast GM03617 (Coriell Institute for Medical Research) were cultured in Alpha modified Eagle's minimum essential medium (αMEM) (Sigma Aldrich) supplemented with 15% fetal bovine serum (FBS) (Thermo Fisher Scientific), 1% GlutaMAX^{TM} (Thermo Fisher Scientific), and 1% penicillin/streptomycin (Thermo Fisher Scientific) at 37°C with 5% CO_2. Human cancer cell lines used in this study, and those derived from them, were cultured in Dulbecco's modified Eagle's minimum essential medium (DMEM) (Sigma Aldrich) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) and 1% penicillin/ streptomycin (Thermo Fisher Scientific) at 37°C with 5% CO_2.

For ultraviolet (UV) irradiation, media was removed from plates and replaced with PBS to keep cells moist. Plates were placed inside a decontaminated (70% EtOH) Stratagene UV Stratalinker 2400 (Agilent Genomics, Santa Clara, CA, USA) and lids were removed to avoid blocking radiation. Cells were irradiated (uJ x 100) or left in the Stratalinker for 10 sec for 0 J controls. The plates were then covered and removed from the Stratalinker, growth media was replaced, and cells were allowed to incubate in the conditions described above until fixation.

2.4.2 Generation of polymerase eta knockout cell lines

Clonal polymerase eta knockout (ΔPOLH) cell lines were generated from human osteosarcoma U2OS cells (American Type Culture Collection HTB-96) using AiO-Puro (Addgene #74630), modified to contain a gRNA pair targeting Cas9n(D10A) to exon 3 of POLH (Table 8), and linearized pBABE-Puro (Addgene #1764), which were gifts from Steve Jackson and Hartmut Land *et al.*, respectively. These plasmids were used in a two-step selection-based knockout process to insert a puromycin resistance cassette into exon 3 of POLH, disrupting the gene with a selectable marker (Fig. 6). Cells were grown at low density such that single cells could be selected and grown into clonal populations. U2OS ΔPOLH clonal cell lines were cultured as described above for wildtype U2OS and maintained in 4 ug/mL puromycin (Table 9).
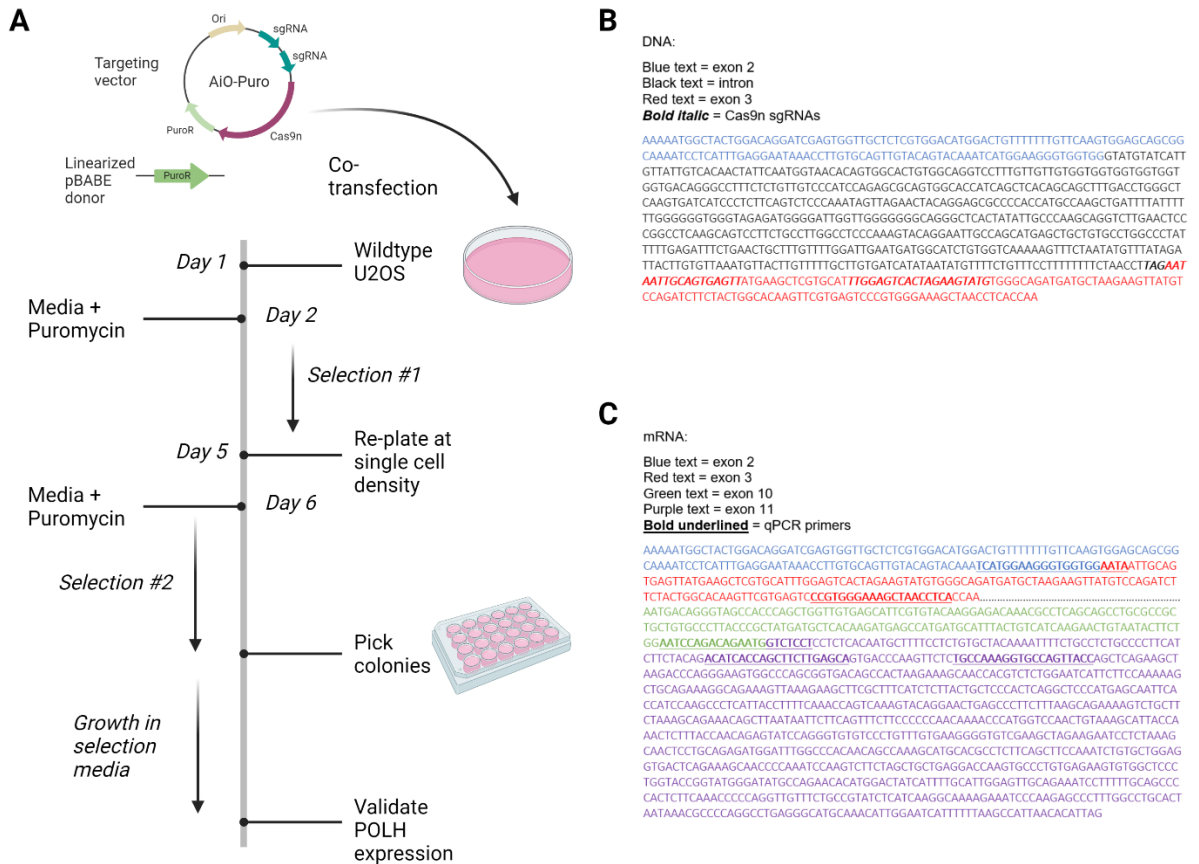
Figure 6. Generation of selectable POLH clonal cell lines. A. Transfection and selection method for cell line construction. B. DNA sequence of POLH exon 2 – exon 3 with Cas9n sgRNAs targeting exon 3 noted (bold italic). C. mRNA sequence of POLH exon 2 – exon 3 and exon 10 – exon 11 junctions with qPCR primers noted (bold underlined). Created with BioRender.com.

Table 9. Cell lines used and/or generated in this study.

Cell line	Selection	Source
NHDF	None	ATCC: PCS-201-010
GM03617	None	CIMR
HeLa	None	ATCC: CCL-2
U2OS	None	ATCC: HTB-96
U2OS CRISPR ΔPOLH clone 1	4ug/mL puromycin	This study
U2OS CRISPR ΔPOLH clone 2	4ug/mL puromycin	This study
U2OS CRISPR ΔPOLH clone 3	4ug/mL puromycin	This study
U2OS CRISPR ΔPOLH clone 4	4ug/mL puromycin	This study
U2OS CRISPR ΔPOLH clone 5	4ug/mL puromycin	This study
U2OS CRISPR ΔPOLH clone 6	4ug/mL puromycin	This study

2.4.3 Transfection

For lipofection, wildtype (WT) and ΔPOLH U2OS cells were transfected with DNA using Lipofectamine 2000 (L2K) (Life Technologies) at a ratio of 4 μl L2K : 2 μg DNA. L2K and DNA were diluted separately in 200 μl serum-free DMEM and combined dropwise. After 20 min incubation at room temperature (RT), the transfection mixture was pipetted dropwise onto cells and allowed to incubate at 37°C with 5% CO_2 overnight. Media was changed after 24 hours to prevent excess cell death. Fibroblast cell lines including NHDF and GM03617 cells were transfected with an optimized ratio of 2 μl L2K : 2 μg DNA, followed by incubation as described above. HeLa cells were transfected using Lipofectamine 3000 (L3K) (Life Technologies) at a ratio of 3.75 μl L3K : 2 μg DNA + 5 μl P3000 (P3K) reagent (Life Technologies). L3K and DNA/P3K were diluted separately in 125 μl Opti-MEM reduced serum media (Thermo Fisher Scientific) and combined dropwise. After 5 min incubation at RT, the transfection mixture was added to cells and allowed to incubate as described above.

For electroporation, cells were transfected using the Neon transfection system (Life Technologies) with 100 μl pipette tips and the recommended settings for each cell type (Table 10) according to the manufacturer's instructions. All cells were trypsinized, washed twice with PBS, and resuspended in Buffer R (Thermo Fisher). After electroporation, cells were plated in pre-warmed media supplemented with serum (10% or 15% FBS, as described above) but without antibiotics.

Table 10. Neon electroporation settings according to cell type.

Cell line	Pulse voltage (v)	Pulse width (ms)	Pulse number
NHDF	1400	20	2
GM03617	1400	20	2
HeLa	1005	35	2
U2OS	1230	10	4
U2OS CRISPR ΔPOLH clones	1230	10	4

2.5 Quantitative reverse-transcription PCR

To extract RNA, confluent cells were rinsed with PBS and lysed using dropwise application of Trizol Reagent (Invitrogen). After a 10 min incubation at room temperature (RT), cell lysate was scraped into a sterile tube and incubated for a further 5 min at RT. Lysates were mixed vigorously with chloroform to promote phase separation. After a second 5 min incubation at RT, lysates were spun for 15 min 12,000 xg at 4°C. The clear upper phase was removed without disturbing the white middle phase, combined with equal volume 70% EtOH, and loaded onto RNeasy MinElute spin columns (Qiagen, Hilden, Germany). The manufacturer's instructions were followed, in brief: 1x Wash Buffer I, 15 min incubation with DNase to remove any contaminating DNA, 1x Wash Buffer I, 2x Wash Buffer II, elution with DEPC-treated H_2O.

cDNA was synthesized by creating a mixture of 4 uL 5x iScript RT Supermix (BioRad) and 1 ug RNA in a total volume of 20 uL with DEPC-treated H_2O and running the following thermocycler profile: 25°C for 5 min, 46°C for 20 min, 95°C for 1 min. Primers were designed to target exon junctions at the insert site (exon 2 – exon 3) and downstream of the insert site (exon 10 – exon 11) (Fig. 6, Table 7). 4 uL of diluted cDNA (1:10 with RNase free dH_2O) was combined with 5 uL SYBR Green Master Mix (BioRad) and 1 uL of diluted (1:25) F/R primer mix to a total of 20 uL. Each sample was run in triplicate and gene expression was determined by normalization to three reference genes: *B2M*, *RAC1*, and *TBP*. Samples were run on a CFX Connect Real-Time System thermocycler (Bio-Rad) with the following profile: 95°C for 30 sec, [95°C for 10 sec, 60°C for 30 sec] 40x, melt curve 65°C to 95°C in 0.5°C increments for 5 sec. Expression and corrected expression standard error of mean (SEM), as calculated by Bio-Rad CFX

Manager (v.3.1) using WT as the relative control, were plotted using GraphPad Prism (v.9.3.1 471, San Diego, California USA, www.graphpad.com).

2.6 Microscopy

2.6.1 Immunofluorescence microscopy

Cells were grown on glass coverslips then washed briefly with PBS, fixed in 2% PFA for 20 min, washed 2 x PBS, permeabilized with 0.5% Triton X-100 in PBS for 5 min, and washed again 2 x PBS. Slips were blocked in 4% BSA in PBS for 20 min then incubated for 1 h with antibodies specific to POLH (Thermo Fisher Scientific, rabbit polyclonal; Santa Cruz, sc-17770, mouse monoclonal; both 1:300) and γH2AX (Thermo Fisher Scientific, 1:1000) diluted in 4% BSA. Following primary antibody incubation, cells were washed 3 x in PBS and incubated for 45 min with Alexa-Fluor 555 donkey anti-mouse and Alexa-Fluor 488 donkey anti-rabbit (Thermo Fisher Scientific) secondary antibodies each diluted to 1:2000 in 4% BSA. Cells were washed again in PBS for 5 min, incubated with 1 ug/mL of 4',6-diamidino-2-phenylindole (DAPI) for 10 min to visualize DNA, and washed for a final 5 min in PBS. Coverslips were mounted on glass microscope slides using Dako fluorescent mounting medium (Dako Agilent Pathology Solutions). Fluorescent images were captured with a BSI scientific complementary metal-oxide-semiconductor (sCMOS) camera (Photometrics) on a custom-built Zeiss Cell Observer Microscope using a 1.3 NA 40X immersion oil objective lens and LED illumination via a Spectra light engine (Lumencor). Images were processed using only linear adjustments (e.g., brightness/contrast) and analyzed using Slidebook (Intelligent Imaging Innovations, Boulder, CO) and Adobe Photoshop CS (v.6.2).

2.6.2 Brightfield microscopy

To visualize fluorescence, cells in 6 well plates were imaged using Brightfield and FITC filters on an AxioZoom fluorescent macroscope equipped with ZEN Pro 2 acquisition software (v.2.0.0.0) (Carl Zeiss Microscopy GmbH).

2.7 Flow cytometry

To fix cells for fluorescence-activated single cell sorting (FACS), cells were trypsinized from 9 cm plates and resuspended in PBS. After washing twice with PBS (0.3 x g for 5min, resuspend, repeat), cells were fixed in 2% PFA for 15 minutes at room temperature with constant agitation. Cells were washed twice with PBS, pelleted, and resuspended in 300 uL PBS for analysis.

Flow cytometry was performed using a FACSCalibur flow cytometer (BD Biosciences) and data analyzed using FCS Express (De Novo Software, CA, USA; v7.12.0007) and GraphPad Prism for Windows (v.9.3.1 471, San Diego, California USA, www.graphpad.com) software. Cells were first gated for intact cell population using forward scatter *vs.* side scatter plots, then gated for transfected cells that were Clover-positive based on expression of the transfection control Far-Red protein iRFP and CRISPR-mediated Clover-LMNA expression, respectively. Single colour controls (iRFP only, Clover-J1 only) were used to determine the threshold of positive *vs.* negative populations. Data were normalized to a donor-only negative control to account for background signal from the Clover-LMNA donor plasmid.

2.8 CRISPR/Cas9-mediated homology-directed repair assay

A visual transfection-based reporter assay designed by Pinder et al. (2015) was performed as previously described. Briefly, $2x10^6$ cells were co-transfected with Clover-LMNA CRISPR donor and gRNA : Cas9 vector in a 1 : 3 ratio (Fig. 7), along with a plasmid expressing Far-Red protein as a transfection control using either Lipofectamine reagent or the Neon transfection system, as described above. In transfections without donor or gRNA : Cas9, pBlueScript II SK (pBSKII) (AddGene #212205) was used as a negative control. Cells were plated into 9 cm plates or 6-well plates containing a sterile coverslip and harvested for flow cytometry and/or IF after 72 h.

Quantification of cells which had undergone productive HDR (HDR+) by IF, as observed by Clover green fluorescent nuclear lamin, was determined by manually counting the number of HDR+ cells in 100 transfected (iRFP+) cells per experiment. The means and standard errors from independent experiments were calculated and statistical significance was determined using unpaired nonparametric Mann-Whitney U tests or paired parametric t-test in GraphPad Prism (v.9.3.1 471, San Diego, California USA, www.graphpad.com).
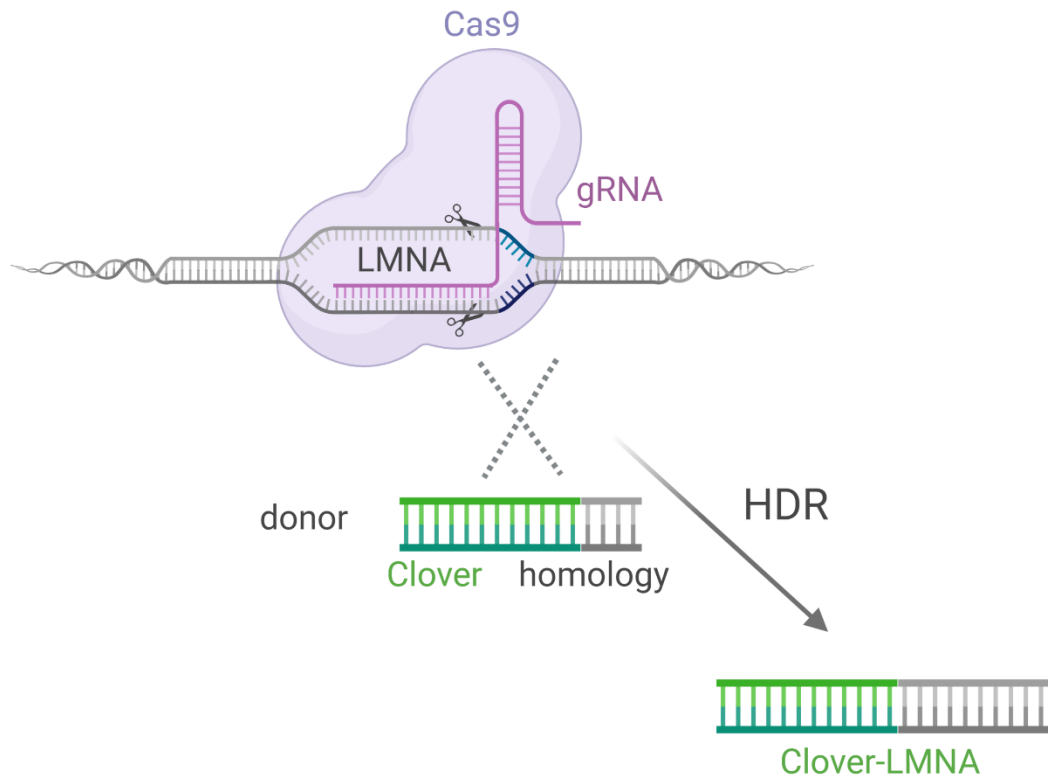
Figure 7. Cas9-directed knock-in of Clover in the nuclear LMNA gene for fluorescent detection of HDR+ cells. Adapted from Pinder et al. (2015). Created with BioRender.com.

2.9 LINE-1 retrotransposition selection-based assay

On Day 1, cells were seeded in 6-well plates such that they would be approximately 70% confluent the following day. After 24 h to allow for cell adherence, cells were co-transfected with the LINE-1 retrotransposition plasmid pJM101 [155] (Fig. 8) and the plasmid of interest (or pBSKII as control) in a 1:1 ratio using Lipofectamine reagent as described in Section 2.4.3. Three days post transfection, cells were trypsinized, counted on a hemocytometer, and plated at a density of $2x10^4$ cells per 9 cm dish. Cells were incubated for two days in growth media prior to selection with G418 (500 ug/mL). After 12 – 14 days of selection, media was removed and G418 resistant (G418R) cells were rinsed with PBS, fixed with methanol for 15 min at room temperature, then stained with crystal violet for 30 min at room temperature. Excess stain was removed with distilled water and plates were allowed to dry overnight before fixed G418R colonies were counted.
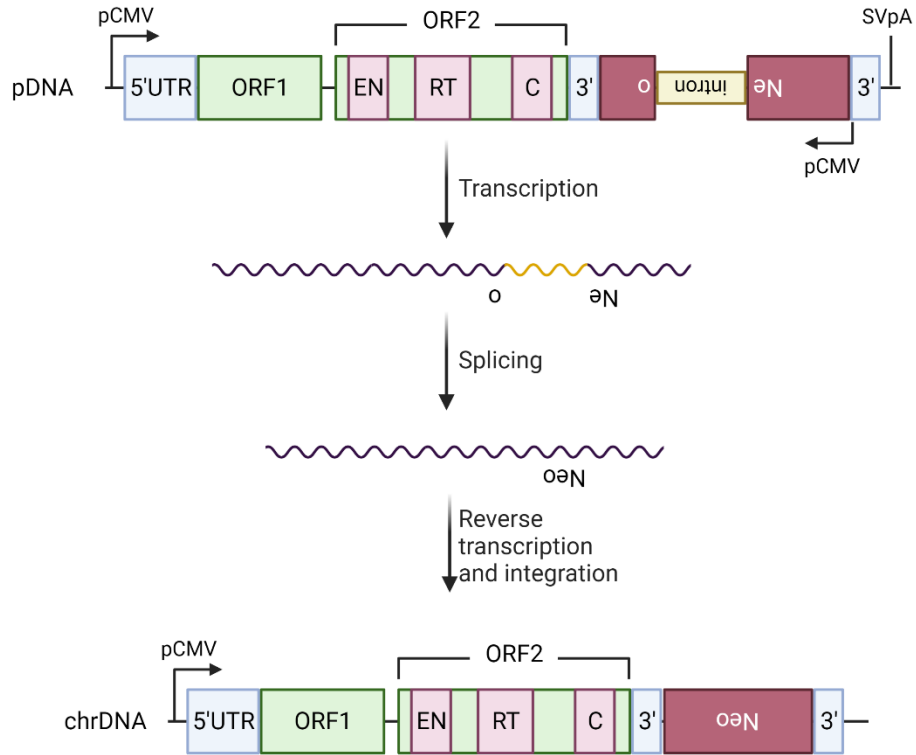
Figure 8. Function of LINE-1 assay for quantification of retrotransposition activity.
A neomycin resistance cassette (Neo) conferring antibiotic resistance to G418 was
inserted within the 3' UTR of human L1.2 in the antisense orientation and interrupted by
intron 2 of the human ɣ-globin gene. Transcription of the retrotransposon driven by the
promoter, splicing of the ɣ-globin intron, reverse transcription, and integration into the
genome are required for successful expression of the Neo gene. After G418 selection,
only cells that underwent successful retrotransposition will express Neo. The number of
surviving colonies is therefore proportional to the number of retrotransposition events.
The resulting L1.3mneol construct was subcloned to create the vector L1-neo-TET
(Moran et al. 1996). ORF open reading frame, pDNA plasmid DNA, chrDNA
chromosomal DNA, pCMV cytomegalovirus promoter, SVpA SV40 poly adenylation.
Adapted from [155,156]. Created with BioRender.com.

Colonies were counted manually within a random 2 x 2 in square, then extrapolated to estimate the total number of colonies per plate. The average of two counts was taken. The means and standard errors from independent experiments were calculated and statistical significance was determined using the paired t-test in GraphPad Prism software (v.9.3.1 471, San Diego, California USA, www.graphpad.com). Dry plates were scanned using a Canon CanoScan LiDE 220 photo scanner and the IJ Scan Utility app v.14.0 with the following settings: Photo Scan; Color Mode, Color; Paper Size, Custom Full Plate; Resolution, 300 dpi; Sharpen outline selected.

3.1 Background

The ability of CRISPR systems to insert sequences into the genome is contrasted with the potential for unintended consequences (see Section 1.4.2). Altering genetic sequence in this manner, regardless of whether the intention is to replace a non-functional pathogenic gene or introduce entirely novel information, can result in unintended mutation at the target site. The most commonly documented Cas9-mediated mutations are single-base and small indels [157,158]. However, larger-scale inversions [159,160] and translocations [161,162] have been recorded, as well as rare subsequent events such as chromosome loss [163] and chromothripsis [164], the catastrophic clustering of extensive chromosomal rearrangements. While certain mutations may be beneficial for the purposes of gene knockout via NHEJ (Section 1.1), when these changes unintentionally affect the function of the target or nearby regions, they can also induce undesirable "on-target effects".

An approach that mitigates on-target effects is the use of safe harbour sites (SHS) as targets for gene insertion [75,152,165]. These genomic regions, which may be genes or non-coding loci, are permissive of gene insertion and expression. Importantly, such regions are free from on-target effects that might disrupt the expression of genes or regulatory elements in their vicinity (or their disruption does not have a pathological effect) and currently include the *AAVS1*, *ROSA26*, *SHS231,* and *CCR5* loci [152]. Of these, *AAVS1* is the most commonly targeted site in basic and clinical studies, based on an integration site of adeno-associated virus serotype 2 in the intron between exons 1 and 2 of the *PPP1R12C* gene of chromosome 19, and thus far no adverse effects have been reported

[152,165,166]. The *ROSA26* site was initially discovered in mouse gene targeting experiments and an equivalent human locus has been mapped between exons 1 and 2 [152,167]. *SHS231*, the most recent addition to the SHS catalogue, was identified by scanning human genome data for nuclease accessibility and adherence to safety criteria [152]. The most controversial of these is *CCR5*. This gene encodes the C-C motif chemokine receptor 5, a key receptor for both immune cells and the entry of HIV [168,169]. As mentioned above (Section 1.3.3), the potential use of gene therapy to cure HIV remains appealing. A now-infamous experiment attempted to establish HIV resistance by disrupting the *CCR5* locus [170]; however, this locus' immune functions are implicated in defense against other viral infections such as influenza and West Nile which are likely to be compromised by mutation of this gene [171,172], throwing into question the safety of targeting this site [173]. The additional problematic nature of this experiment arose from the *in vivo* germline editing nature of the experiment conducted in human embryos, two of which were eventually brought to term. The *CCR5* locus therefore does not adhere to the definition of a safe harbour, as disruption does not have a negligible effect. Further, while "safe harbour site" implies that targeting a particular region is inherently non-pathogenic and free of unintended side effects, the term SHS may be misleading due to the constantly evolving understanding of the human genome. Therefore, despite the common use of SHS to describe such a location, the term "gene editing locus" (GEL) will be used for the remainder of this thesis to avoid making presumptions about the safety of targeting various genomic sites.

3.2 Results

3.2.1 A plasmid toolkit for targeted gene insertion at a human chromosome 15 locus

3.2.1.1 Characterization of GEL15 site

To expand the number of potentially useful human GELs available for targeted

gene insertion, a previously utilized intergenic region on human chromosome 15

(GRCh37.p.13 primary assembly chr15:74696007-74698086; Figure 5; Attwood et al.,

2020) was characterized with updated (at time of printing) genome annotation data from

the UCSC Genome Browser [151]. According to more stringent criteria based upon those

originally outlined by Pellenz et al. (2019), the GEL15 target site (52bp at

chr15:74697131-74697182) scores higher than the current SHS described above and

meets the same number of criteria as the recently identified SHS231 (Table 11).

However, updates to genome annotation since the identification of GEL15 revealed that

this site does not adhere to two criteria: that it be > 150 kb away from a 5' gene end, and

that there is low transcriptional activity (no mRNA within 25 kb). Visual inspection of

GEL15 through the UCSC Genome Browser [151] revealed that a long noncoding RNA as

well as three protein-coding genes are located within 50 kb of GEL15: *CYP11A*, which

encodes a cytochrome; *SEMA7A*, a multi-functional membrane-bound semaphorin; and

UBL7, which is predicted to enable polyubiquitination-dependent protein binding activity

(GeneCards GC15M074337, GC15M074409, GC15M090219).

Table 11. Criteria and data sources for the identification and functionality scoring of intergenic loci for gene editing.

	Criterion	Data Source	AAVS1 19:55,625,241-55,629,351	hROSA26 3:9,415,082-9,414,043	CCR5 3:46,414,443-46,414,942	SHS231 4:58,976,613-58,976,632	GEL15 15:74,697,131-74,697,182
Safety	>300kb from any cancer-related mutation	*COSMIC database*	-	-	-	+	+
	>150kb from any 5' gene end	Genes and gene predictions: UCSC Genes	-	-	-	+	-
	Intergenic region (not located within gene) *	Genes and gene predictions: UCSC Genes	-	-	-	+	+
	>300kb from functional small RNA	Genes and gene predictions: sno/miRNA	+	+	+	+	+
Functional silence	>50kb from any site of origin	Regulation: UW Repli-seq: Peaks	+	-	+	+	+
	>50kb from any ultra-conserved element	Regulation: VISTA Enhancers	+	+	+	+	+
	Low transcriptional activity (no mRNA ± 25kb)	mRNA and EST: Human mRNAs	-	-	-	+	-
Accessibility	Not in copy number variable region	Repeats: Segmental Dups	+	+	+	+	+
	In open chromatin (DHS signal ± 1kb)	Regulation: ENC Dnase/FAIRE: Open Chrom Synth	+	-	+	-	+
	Targetable by CRISPR/Cas9 (<30 Efficiency)	Genes and gene predictions: CRISPR Targets	+	+	+	-	+
Uniqueness	Single copy in human genome	*NCBI BLAST search output*	+	+	+	+	+

Genomic sites were assessed for each criterion by visual inspection of the corresponding data source or track on the UCSC Genome Browser (assembly hg19) to the noted region size on either side of the coordinate. Where a buffer region is not noted, only the region within the coordinate was analyzed. Green + adheres to the criterion, red - does not adhere to the criterion. Italic data sources are not tracks in the UCSC Genome Browser. *Not included as a criterion for the GEL identification code (Section 3.2.2). Adapted from Pellenz et al. (2019).

3.2.1.2 Construction of GEL15 donor plasmids

Multiple donor plasmids were constructed, including a flexible cloning site and various reporter constructs located between GEL15 homology arms, to characterize the stability, expression, and functionality of GEL15-integrated genes (see Section 2.2). Donor constructs based on GEL15-MCS including GEL15-EGFP-N1, -EGFP-C1, EGFP-2A-Puro, and -EGFP-N1-POLH were validated by whole plasmid sequencing (Plasmidsaurus.com). Construction of GEL15-Puro is ongoing. See Appendix B Figures B2-7 for plasmid maps of all GEL15 constructs.

3.2.1.3 Validation of GEL15 donor plasmids

GEL15 donor constructs were additionally validated *in vitro* by Lipofectamine-mediated transfection. Three donors containing fluorescent reporter genes (EGFP-N1, EGFP-C1, EGFP-N1-POLH) were transfected into different cell types, including WT U2OS, HeLa, and NHDF (Section 2.4.3). EGFP-N1 and -C1 were visualized in all three cell types, while EGFP-tagged Polη was only seen in U2OS and HeLa cells (Fig. 9). The number of EGFP+ cells was generally higher for the EGFP only constructs than EGFP-N1-POLH (Fig. 10), as expected based on previous transfection experiments with the original pEGFP-N1 and EGFP-N1-POLH plasmids (see Section 4, Fig. 17). Contrary to expectations that U2OS would be most amenable to GEL15-targeted expression of all donors due to U2OS-based HA design (Fig. 5), the highest percentage of EGFP-N1-POLH+ cells was observed in HeLa cells, both 24 h after transfection with donor only (Fig. 10A) and post HDR-mediated insertion (Fig. 10B). However, fluorescent gene expression was maintained in both U2OS and HeLa cells for twelve days (Fig. 11).
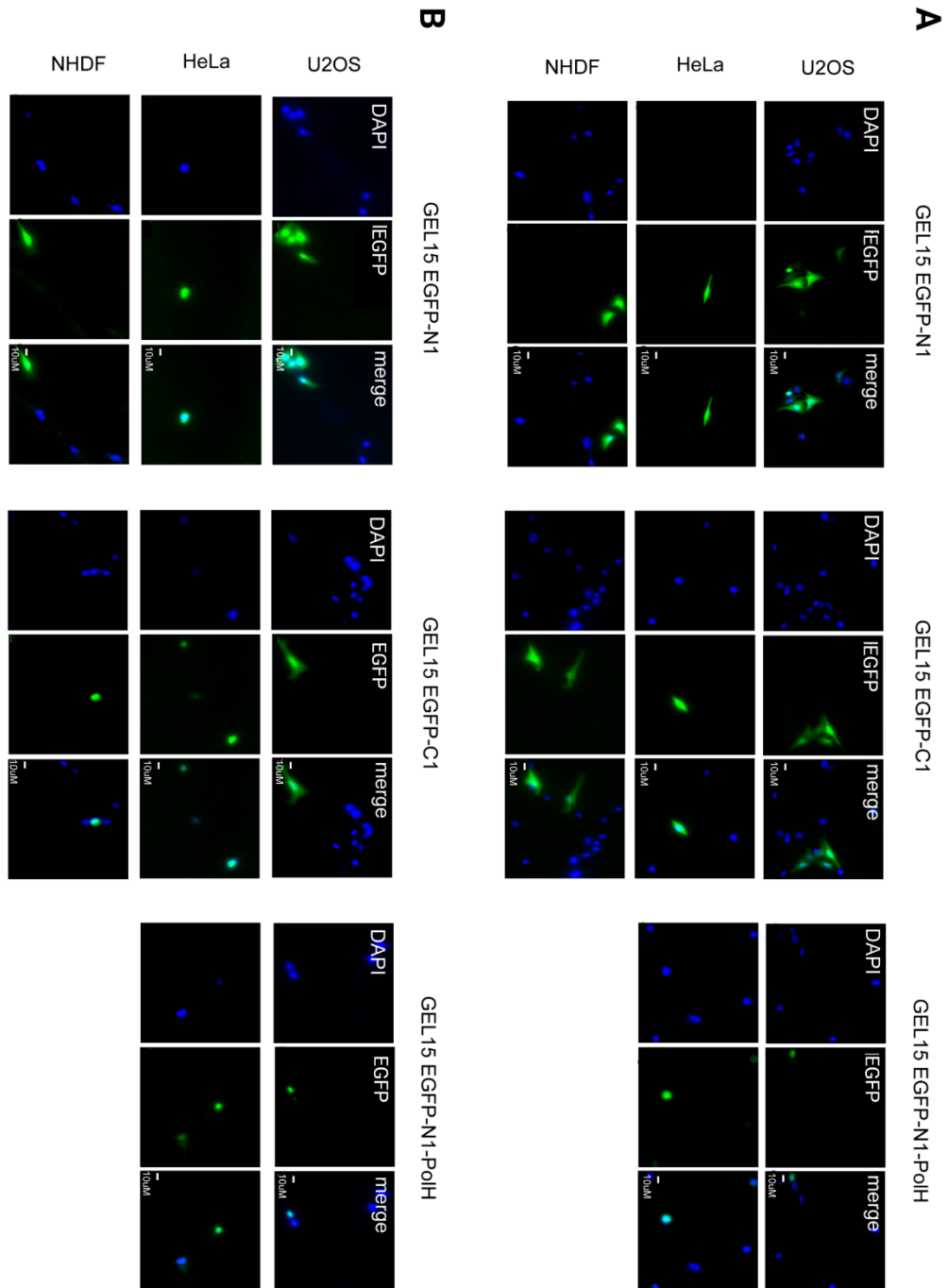
Figure 9. GEL15 donor constructs produce detectable green fluorescence in multiple cell types. Expression of GEL15-EGFP-N1, -C1, -POLH (left, middle, right) in WT U2OS, HeLa, and NHDF cells (top, middle, bottom). n = 1. A. Fluorescent expression 24 h post transfection with donor only. B. Fluorescent expression 3 days post transfection with donor + gRNA. Created with Slidebook (Intelligent Imaging Innovations, Boulder, CO), Adobe Photoshop CS (v.6.2), and BioRender.com.
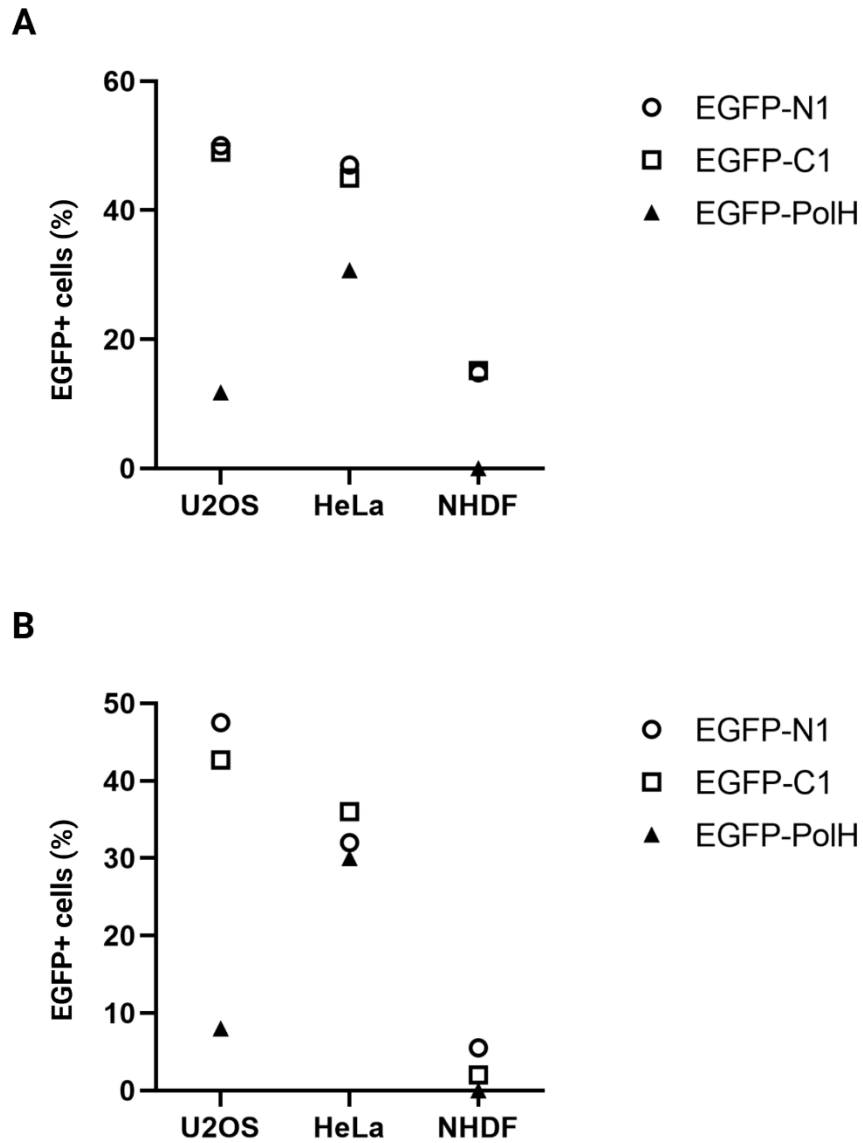
Figure 10. Percentage of EGFP+ cells from transfection of EGFP GEL15 donor plasmids + gRNA in different cell types. Percentage determined as number of EGFP+ cells in 100 DAPI-stained cells, visualized by IF. n = 1. A. Cells fixed 24 hrs post-transfection. B. Cells fixed 3 days post-transfection with gRNA. Created with GraphPad Prism (v.9.3.1 471, San Diego, California USA, www.graphpad.com) and BioRender.com.

Fluorescent expression of EGFP-N1 and -C1 donor constructs were highest in

U2OS cells, with 47.5% expressing EGFP-N1 and 42.7% expressing EGFP-C1 three

days post-transfection (Fig. 10). However, only 8.0% of U2OS cells were positive for

EGFP-N1-POLH fluorescence after three days, compared to 30.0% in HeLa cells. HeLa

cells had generally consistent levels of expression across the three fluorescent donors

(~30%) while NHDFs had the lowest, with < 10% expressing either EGFP donor, and no

cells detected expressing EGFP-N1-POLH.
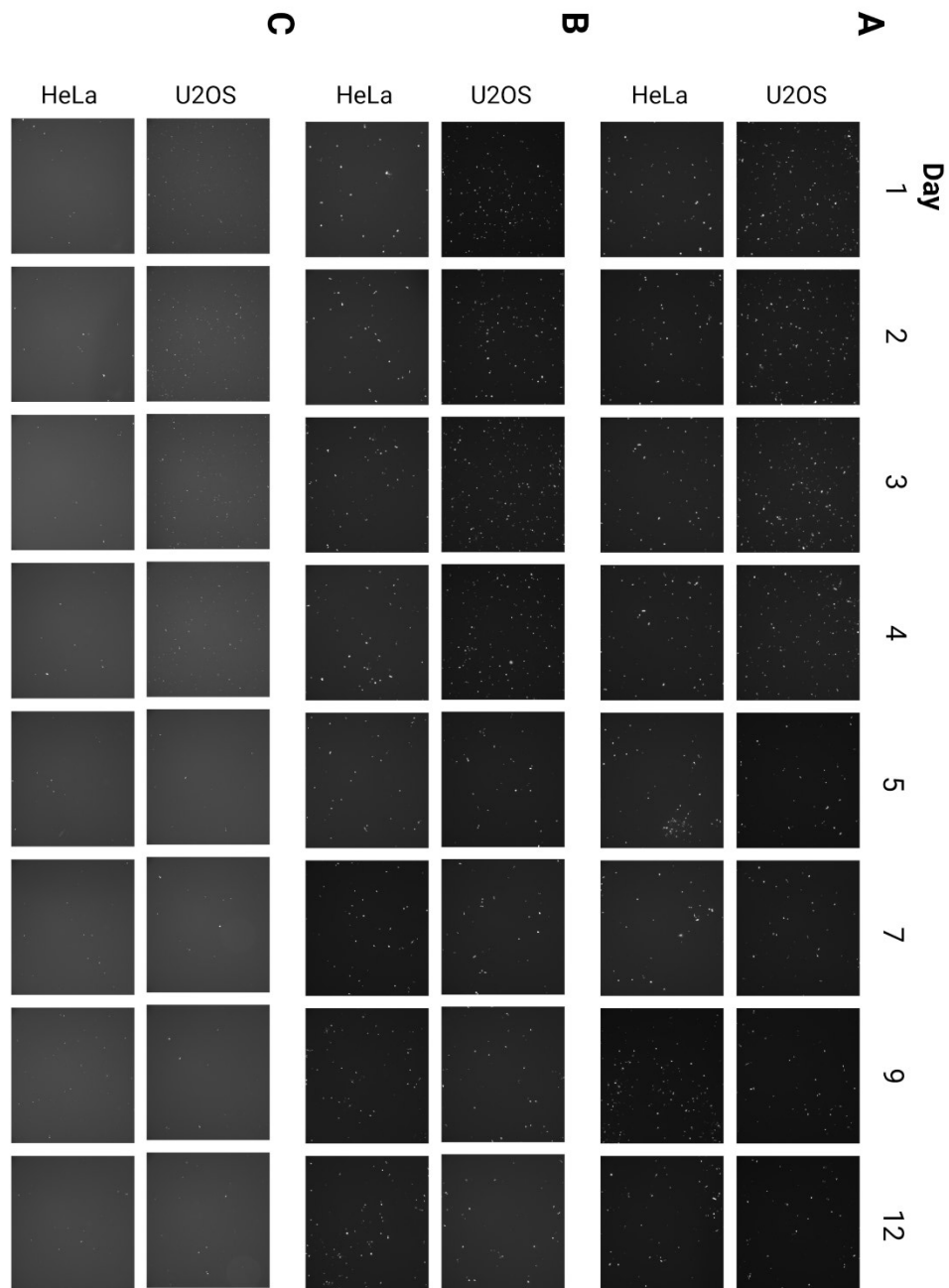
Figure 11. GEL15 EGFP expression in U2OS and HeLa WT cells maintained in growth media for 12 days post co-transfection with GEl15 gRNA. Images captured at 10x magnification with a FITC filter, so green fluorescent cells appear white. A. GEL15 EGFP-N1. B. GEL15 EGFP-C1. C. GEL15 EGFP-N1-POLH. Images processed with Zen Pro (v.2.0.0.0) software and ImageJ 1.53s. n = 1. Created with BioRender.com.

3.2.2 A virtual toolkit for identification of novel GELs

To ameliorate future scenarios in which a putative GEL must be manually scanned, we created a versatile and modular script in collaboration with Ariel Lisogorsky (Waterloo University) to output genome regions that adhere to our criteria (Table 2). Additional criteria may be modified or added as required by the user, and buffer regions may be edited to increase or decrease stringency. The workflow of the script is described in Figure 12 and the code is accessible through GitHub (https://github.com/kkratzer/gene-editing-loci/).
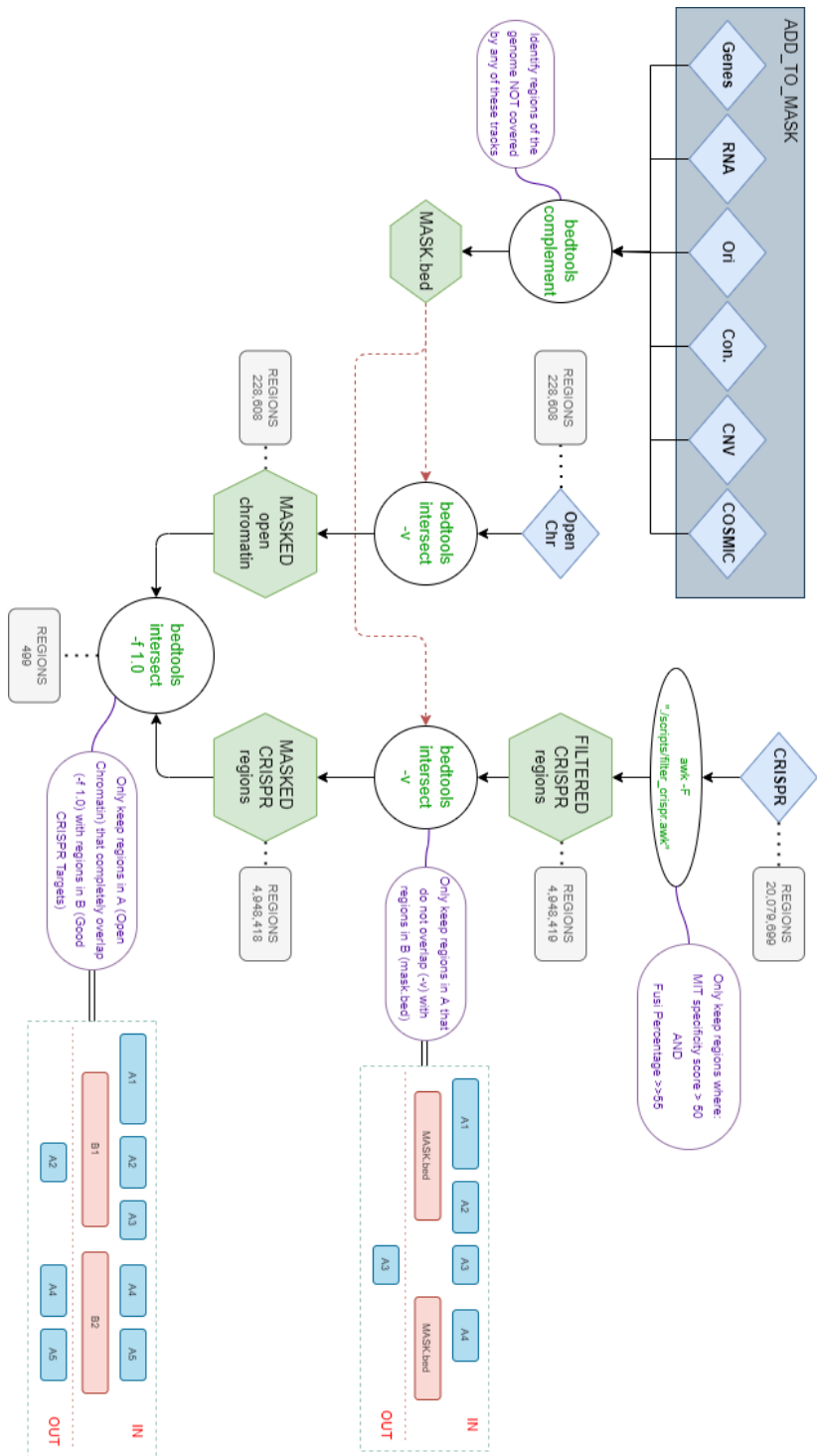
Figure 12. Workflow of GEL identification code. Blue diamonds, UCSC tracks; green hexagons, output files. Adapted from Ariel Lisogorsky. Created with diagrams.net (JGraph 2021, https://www.diagrams.net v.15.5.2).

In brief, a Masterfile titled "MASK.bed" was created using the BEDTools suite function 'complement' with the corresponding UCSC Genome Browser tracks to identify suitable regions of the human genome according to the criteria listed in Table 2 [174]. This mask was applied to both the Open Chromatin track and a filtered CRISPR Target track using BEDTools 'intersect', resulting in masked BED files of both tracks. Finally, 'intersect' was applied to these intermediate output files such that only regions of masked open chromatin that completely overlapped with high-scoring CRISPR targets remained as an output of 499 potential GEL regions. The regions are displayed visually by chromosome in Figure 13.

CRISPR-accessible regions were determined by limiting GEL output to regions with an MIT specificity score > 50 and a Doench/Fusi efficiency > 55 [176,177]. The specificity score summarizes off-targets into a number from 0 – 100, with lower scores indicating less unique sequences, while the efficiency score predicts the efficiency of Cas-mediated cleavage at the target site as determined by gRNA screens performed by Doench et al. (2014) and Sullender et al. (2015). These scores were calculated using the tool CRISPOR [176] and mapped to potential CRISPR PAM -NGG motif targets in the genome to create the CRISPR target track [176,177].

Figure 13. Map of current and predicted GEL sites. Magenta lines denote approximate locations of predicted GEL sites based on the identification script. Thick black bands are cytobands. Arrows indicate locations of current SHS and GELs: hROSA26 on chromosome 3 (black), CCR5 on chromosome 3 (white), SHS231 on chromosome 4, GEL15 on chromosome 15, AAVS1 on chromosome 19. Created with BioRender.com.

3.3 Discussion

Genome engineering techniques often employ the strategy of targeting novel genetic information to one of the currently known 'safe harbour' sites (SHS). These sites should, at minimum, allow transgene expression without pathogenically disrupting other regions. Identifying suitable sites for sequence insertion is an increasingly important part of developing safe gene therapies as well as in basic synthetic biology approaches [148,152]. Gene editing loci (GELs) are regions in the genome where new genetic information can be reliably incorporated without losing host function, as per the definition of a genomic SHS. In addition, GELs encompass further safety criteria. A very small pool of so-called SHS for gene editing are currently available and in use [152]. These were discovered by various methods including viral insertion sites (AAVS1), homology (hROSA26), and more recently, a genome-wide scan for areas that fit a set of first-pass criteria (SHS231). Here we characterized and validated a proto-GEL (GEL15) previously shown to be amenable to gene editing [146], and additionally identified nearly 500 novel putative sites in the human genome through the application of a modifiable auto-search script.

An intergenic region on human chromosome 15 was previously identified by visual analysis of UCSC Genome Browser tracks and chosen for its proximity to the study gene of interest [146]. After successful targeting of this region by Attwood et al. (2020), further validation revealed the utility of the novel GEL15. We first cloned the Cas9n targeting vector from scratch, based upon work done by Attwood et al. (2020). Next, we built an all-purpose donor vector with an empty MCS between the HAs. Then, we inserted various payloads that could act as reporter genes into the MCS by digestion/ligation. The reporter donor vectors were validated by visualization of EGFP by

IF (Fig. 9) and by whole-plasmid sequencing from Plasmidsaurus (Appendix B Figs. 2-7). As the homology arms of the donor vector were designed from genomic U2OS DNA (Fig. 5) we determined whether targeting could occur to this site in other cell types. The percentage of cells expressing EGFP-N1 and -C1 from GEL15 donors were highest in U2OS cells, as expected (Fig. 10). However, fewer U2OS cells were positive for EGFP-N1-POLH fluorescence after three days compared to HeLa cells which had generally consistent levels of expression across the three fluorescent donors (~30%). NHDFs had the lowest, with < 10% expressing either EGFP donor and no cells detected expressing EGFP-N1-POLH.

Generally lower EGFP expression efficiencies observed in non-target cells cannot be assumed to be due to homeology between the genome and the donor HA. While the rate of HDR may depend on the degree of sequence similarity between target and donor, initial expression of the EGFP GEL15 donor plasmids was lower in non-target cell types compared to U2OS after only 24 h (Fig. 10), when EGFP should be optimally expressed post-L2K transfection (as seen in Fig. 17B), and when no site-directed insertion via HDR had occurred. To definitively determine if homeology plays a role in targeting GEL15 and other sites, the GEL15 locus should be sequenced in a variety of different cell types to identify any variation at the site. Donor vectors may then be redesigned with HAs homologous to the region of another cell type, as determined by genomic DNA isolation. Additionally, an interesting route of exploration would be to target a reporter gene to GEL15 in induced pluripotent stem cells (iPSCs) and follow gene expression post-differentiation. Such an experiment would reveal not only the amount of time that GEL15 expression persists (theoretically indefinitely through cellular passages), but also the

efficiency of targeting GEL15 in many different cell types with an identical HA sequence. If homeology is found to be a compounding factor in CRISPR-mediated HDR, this will help further refine the criteria for putative GELs. This would also have important implications for clinical gene editing, as well as highlight the importance of personalized medicine. When performing insertional mutagenesis, the target site should be sequenced to design the optimal donor vector for maximum insertion efficiency, and presumably, maximum therapeutic benefit. Future directions include comparing the targeting efficiency of GEL15 with other in-use SHS, though one limitation is that transfection efficiency depends heavily upon method used (Table 1). In addition, it may be useful to attempt dual targeting to multiple GELs concurrently or in a stepwise fashion, to further explore the flexibility and utility of these gene editing tools.

Researchers may desire novel regions for gene insertion in addition to the currently accepted sites for a plethora of reasons beyond improved safety in clinical gene editing, including the flexibility to co-target sites for multiple subsequent or simultaneous edits; the ability to tag cells in a novel way; and the option to select transfected cells, therefore improving transfection efficiency. A selectable reporter construct, such as Puro or EGFP-2A-Puro, may be used as part of a co-transfection cocktail to allow for selection of transfected cells, thereby increasing the efficacy of a transfection-based assay.

We therefore showed the utility of GEL15 as the newest option available in the CRISPR-mediated HDR gene editing toolkit. However, GEL15 does not adhere to all criteria (Table 2). As such, we designed a bash script code using human genome data available from UCSC Genome Browser tracks and the COSMIC cancer database to mask regions of the genome assembly that do not adhere to a set of safety, accessibility, and

uniqueness criteria. These criteria were built upon those proposed by Pellenz et al. (2019). The first group of principals that cumulatively fall under 'safety' was adhered to by placing buffer regions of exclusion around known genes, functional small RNA, and cancer-related mutations (as determined by the COSMIC database). In Table 2, we added an additional criterion to differentiate between intergenic sites that are within 150 kb of a gene (GEL15) and sites that are directly within a gene itself (AAVS1, hROSA26, CCR5). However, for the script, we only included a criterion to mask all regions within 150 kb of known genes. This served to increase stringency and therefore safety, and also to restrict the output to a manageable number of putative GELs.

The following criteria, which adhere to the concept of functional silence, ensure that the inserted sequence does not disrupt the expression of other genes by avoiding replication origins, conserved elements, and regions of mRNA expression. Despite avoiding known functional regions, the GELs had to be accessible. This included avoiding copy number variable regions, which typically increase uncertainty of Cas9 targeting [175], as well as restricting sites to regions of open chromatin that are also targetable by Cas9.

The final group of criteria falls under the category of structural accessibility. For ease of targeting, a putative GEL must not be located in a copy number variable region and should be unique in the entire genome (automatically determined by the CRISPR target track in the script). To allow expression of the inserted genetic material, the GEL must likely be in a region of open chromatin. However, SHS231 'fails' this criterion, yet high targeting efficiency was observed [152]. This paradox highlights the duality of the 'open chromatin' criterion. Despite the fact that euchromatin is more permissible to gene

expression, there may be yet unknown reasons why intergenic regions are endogenously

maintained in this state. As human genome data coverage improves, we could discover

that regions previously thought of as safe contain an encoded factor or are otherwise

important in regulatory or structural contexts. Targeting heterochromatin may therefore

further improve safety, and expression of exogenous sequences can be maintained by

including insulating DNA sequences in targeting constructs [180] or epigenetically altering

chromatin at the target site, as shown by [181]. For this reason, the GEL code is modifiable

and can be run with any desired version of the human genome assembly; however, we

included tracks in our analysis that are only available for assembly hg19. This modality

allows for a customizable search that can be continuously updated as datasets improve.

Here, we offer an additional tool to achieve the current optimized outcomes for safe,

successful gene editing.

4.1 Background

4.1.1 Polymerase eta plays dual roles in translesion synthesis and homologous recombination repair pathways

Gene editing relies on the endogenous ability of cells to repair damaged DNA. Due to the numerous sources and types of DNA damage, living organisms have evolved a plethora of DNA repair proteins, mechanisms, and pathways. For example, within humans, there are at least 15 known DNA polymerases [182,183], the enzymes responsible for synthesizing DNA polymers from nucleic acids. The Y-family of DNA polymerases consists of enzymes involved in a still relatively undefined type of repair called translesion synthesis (TLS). The four Y-family TLS polymerases can each bypass specific types of DNA lesions, allowing DNA synthesis or repair to proceed when replicative polymerases fail to bypass the lesions. Specifically, polymerase eta (Polη, encoded by POLH) is involved in the resolution of stalled replication forks and provides high fidelity bypass of UV damage (Fig. 14) [184–186]. Polκ, which normally bypasses guanine lesions, and Polι, which indiscriminately incorporates thymine bases, can act together to partially fill the role of non-functional Polη (see Section 4.1.2), albeit with lower fidelity in repairing UV dimers [185,187,188]. Finally, Rev1 inserts cytosine bases opposite lesions, although its major role seems to be indirect; it recruits other TLS polymerases to DNA damage sites and is involved in the inhibition of the p53 tumor suppressor [189], which in turn regulates expression of Polη and Polκ [183,190].
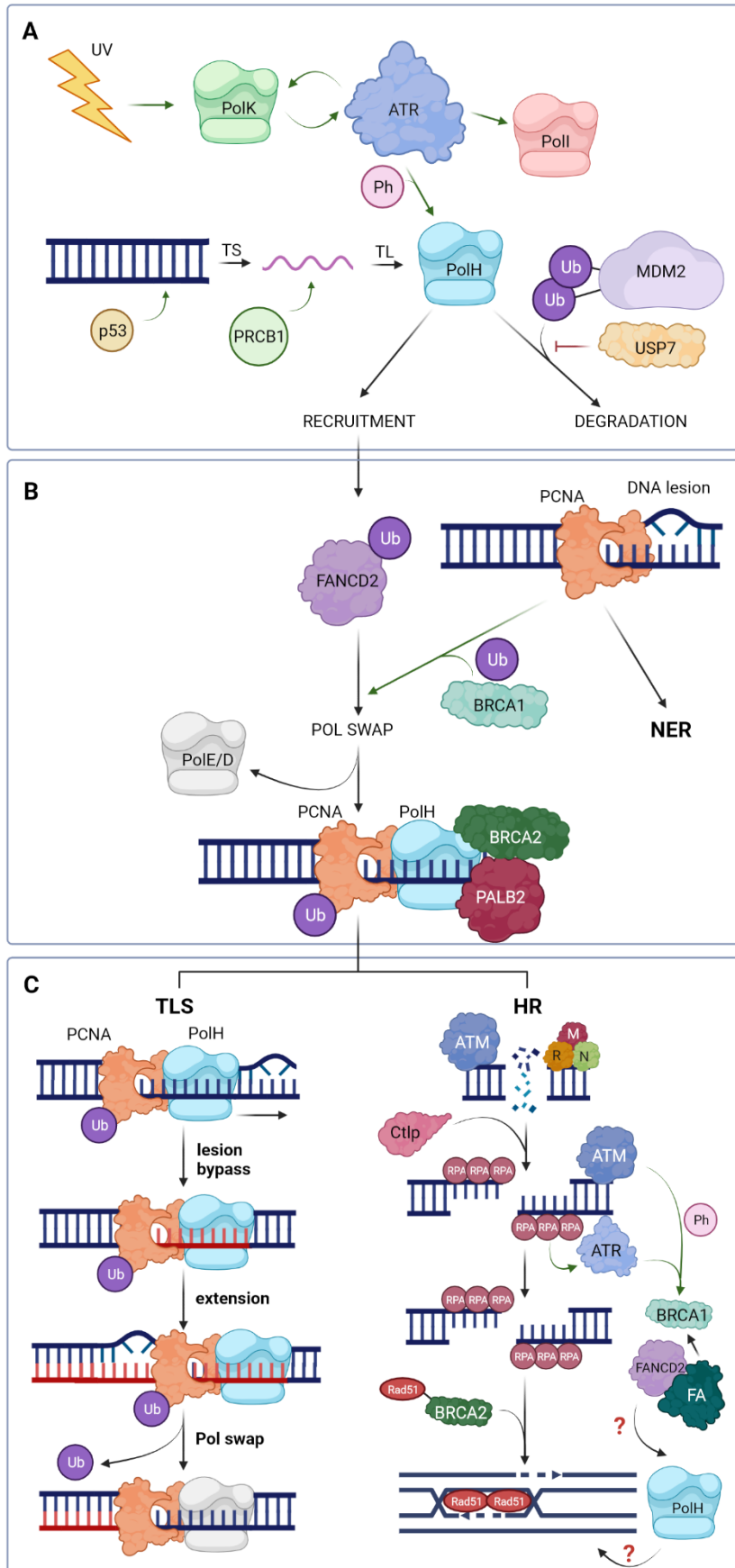
Figure 14. Summary of Polη regulation and function. A. Polη regulation occurs at various stages of expression. The p53 tumor suppressor acts as a transcription factor to promote POLH transcription. The stability of the POLH mRNA transcript is positively regulated by PRCB1. In response to UV exposure, Polη expression is promoted through Polκ and ATR activation, though Polη is rapidly degraded by the proteosome. MDM2 tags Polη for Poly-Ub-dependent degradation and PIRH2 targets Polη for mono-Ub-dependent degradation, while USP7 promotes Polη stability through competitive action with MDM2. Mono-Ub of Polη by RCHY1/PIRH2 is promoted by TRAIP and inhibits Polη interaction with PCNA and its ability to bypass UV lesions. B. Multiple DNA repair proteins interact to recruit Polη to DNA lesions. Mono-Ub FANCD2 recruits Polη to the mono-Ub PCNA clamp at stalled replication forks, where it co-localizes with PALB2 and BRCA2. BRCA1 promotes Ub-PCNA; without it, repair pathway choice proceeds to NER. C. The DNA replication fork stalls at DNA lesions. Ub-PCNA has increased affinity for the Ub-binding domain of Polη, causing a polymerase swap. DNA repair proceeds through one of two Polη-dependent pathways: TLS or HR.

In addition to its well-established role in TLS, Polη has been implicated as a key player in the HR pathway. In TLS, Polη is recruited to the lesion by monoubiquitination of the PCNA clamp where it inserts potentially mispaired bases opposite the lesion to allow replication to proceed [191–193]. Polη quickly dissociates and another polymerase swap occurs, with a high fidelity polymerase, often Polδ, taking over to resume replication fork progression [185,193,195–197]. Any mismatches can be repaired by NER. Alternatively, a stalled replication fork may regress to form a structure that can undergo nucleolytic cleavage [185,193]. NER repairs the lesion, and the resulting DSB triggers recruitment of HR repair factors. Rad51, recruited to the DSB by BRCA2, promotes strand invasion and the formation of a D-loop intermediate structure that can be extended by Polη [185,192]. Rad51 may directly recruit Polη, although Polη has also been shown to associate with BRCA1-dependent Fanconi anemia (FA) protein FANCD2, particularly in the context of recruitment to Ub-PCNA [192,194]. PALB2 and BRCA2 have also been shown to interact directly with Polη and are necessary for increased D-loop extension [198]. Polη extends the D-loop until crossover cutting can occur to re-establish a replication fork [185]. Polymerase swap back to Polδ allows high fidelity DNA replication to proceed. Despite this additional role of Polη in HR, opening the door for its involvement in other genome maintenance processes (see Section 4.1.3), no investigation has been done on the role of Polη in the process of CRISPR-mediated HDR.

4.1.2 Lack of functional translesion polymerase eta causes xeroderma pigmentosum variant disease

Xeroderma pigmentosum (XP) is a group of rare autosomal recessive disorders characterized by hypersensitivity to UV, solar-induced deterioration of the eyes and exposed skin, and a predisposition to the development of skin cancer caused by deficiencies in DNA repair [199–203]. Progressive neurodegeneration, including acquired microcephaly, occurs in approximately 25% of cases [204]. Though classified as a rare genetic disease, XP affects 1/22,000 people in Japan, North Africa, and the Middle East, and typically results in a shortened lifespan [204]. Seven subtypes of XP disease are due to mutations in various XP genes involved in the nucleotide excision repair (NER) pathway, an alternative UV repair mechanism to TLS (Fig. 14). The eighth subtype, termed xeroderma pigmentosum variant (XPV), displays the same symptoms but is due to decreased or dysfunctional Polη-mediated TLS.

Pathogenic mutations in POLH typically cause a truncated Polη protein, though frameshift and missense mutations are also well documented [205–208]. These mutations vary in their affect on protein function and are not obviously related to clinical severity [205]. The N-terminal of Polη, which is usually intact in XPV, forms the active site and is highly conserved among Y-family polymerases; meanwhile, the C-terminal is responsible for nuclear localization and accumulation of Polη into replication foci [205,209]. A truncated protein would therefore not be able to localize to damage sites if it escaped nonsense-mediated decay of mRNA. In cell cultures derived from XPV patients, rates of both TLS and HR-mediated repair of DSBs are decreased [185,210,211], while the accumulation of stalled replication forks is increased compared to normal fibroblasts [212]. The sensitivity of these cells to UV irradiation is exacerbated by low doses of caffeine [213,214]. In addition,

tumor cells from XPV patients have altered expression levels of other TLS polymerases, seemingly in an effort to compensate for the lack of functional Polη, despite their reduced ability to replicate faithfully over UV lesions: Polι levels are increased, while expression of Polκ and Rev1 are decreased [215].

4.1.3 Polymerase eta strikes a balancing act within its many roles in genome maintenance

With its established roles in two distinct DNA repair pathways, Polη has been implicated in a wide variety of genome maintenance mechanisms. Due to its innate ability to replicate through UV-induced crosslinks, overexpression of Polη increases chemoresistance in various cancers by allowing DNA replication to proceed through chemotherapy-induced crosslinks such as those created by the platinum-based drug cisplatin [216–220]. Indeed, upregulation of TLS polymerases seen in some cancers may promote cancer cell survival and replication, with no normal checks on or consequences for the mutagenic nature of these enzymes (Table 12) [184,221,222]. More work remains to be done to determine whether TLS polymerase upregulation is a consequence of cancer, or a potential cause of it.

Table 12. Overexpression of TLS polymerases positively correlates with breast cancer biomarkers.

	Luminal A and B			HER2-enriched
	ESR1	*ESR2*	*PGR*	*HER2*
POLH		*		
POLK	*	*	*	
POLI	*	*	*	
REV1	*			

Green * indicates significant (q-value < 0.05) co-occurrence of gene overexpression (EXP > 1 Std Dev) with markers of Luminal A/B (ESR1, ESR2, PGR) and HER2-enriched (HER2) breast cancers. Data source: Breast Cancer Invasive Carcinoma (TCGA, PanCancer Atlas) accessed on Sept. 24 2020 at https://www.cbioportal.org/.

Another cancer-related process that involves Polη is the Alternative Lengthening of Telomeres (ALT) pathway which maintains telomere length in cancers that lack the enzyme telomerase. Polη is efficient when replicating through repetitive regions [184,223], potentially partially explaining its recruitment to telomeres [224]. However, Polη has been implicated directly in the HR-dependent ALT mechanism, which is predicted to use a sister chromatid telomere as a template for repair [224,225].

Finally, Polη has also been shown to play a role in the diversification of immunoglobulin (Ig) genes mediated by the activation-induced cytidine deaminase (AID)/Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of enzymes [226]. We have seen significant co-occurrence of Polη expression and some APOBEC3 members (Table 13), which are specifically known as retroelement inhibitors [227,228]. The function of Polη in an APOBEC-associated hypermutation mechanism is even less well characterized than its role in ALT, and thus merits further study. However, it is known that these cytidine deaminases deaminate cytosine to uracil

in response to pro-inflammatory cytokine signalling, which can then function beneficially to (a) induce base-pair mismatches in Ig genes, thereby furthering genetic diversity and B cell evolution; (b) cause mutations in retroelement sequences, preventing their potentially pathogenic genomic integration; or, negatively, by (c) increasing genome instability and promoting apoptosis or cancer (Fig. 15) [227–237]. Presumably, Polη may be involved in the mutagenic repair of these APOBEC-induced mismatches or resulting DSBs, further diversifying Ig genes or increasing carcinogenesis in other contexts [226,230,231,238]. We therefore sought to investigate whether Polη may also play a role in the restriction of retrotransposition, specifically of long interspersed nuclear element type 1 (LINE-1) retroelements using a selection-based assay. LINEs are transposable elements in the human genome that have the ability to self replicate. They have been pinpointed as a major source of evolution-driving diversity [239], making up 17 – 20% of the human genome [240–242], although LINE overactivity has been shown to cause genome instability and cancer [243–245]. To keep this activity in check, cells have evolved numerous retroelement-restricting defense mechanisms, including AID/APOBEC3 enzymes, P53, and HR repair factors such as BRCA1 and FA factors [157,235,237,241,242,246–254]. Due to its associations with both APOBECs and HR repair factors, is likely that Polη is also involved in LINE-1 repression, though at what stage of the mechanism is unknown.

Table 13. Expression of Polη significantly co-occurs with some APOBEC3 enzymes in breast cancer.

Gene A	Gene B	q-Value	Tendency
POLH	**A3A**	**0.012**	**Co-occurrence**
POLH	**A3B**	**<0.001**	**Co-occurrence**
POLH	A3C	0.580	Co-occurrence
POLH	A3D	0.214	Co-occurrence
POLH	**A3F**	**0.003**	**Co-occurrence**
POLH	A3G	0.624	Mutual exclusivity
POLH	A3H	0.442	Mutual exclusivity

Bold indicates significant q-value (< 0.05). Default expression values used. Data source: Breast Cancer Invasive Carcinoma (TCGA, PanCancer Atlas, samples with mRNA data (RNA Seq V2) (n = 1082) accessed on Sept. 24 2020 at https://www.cbioportal.org/.
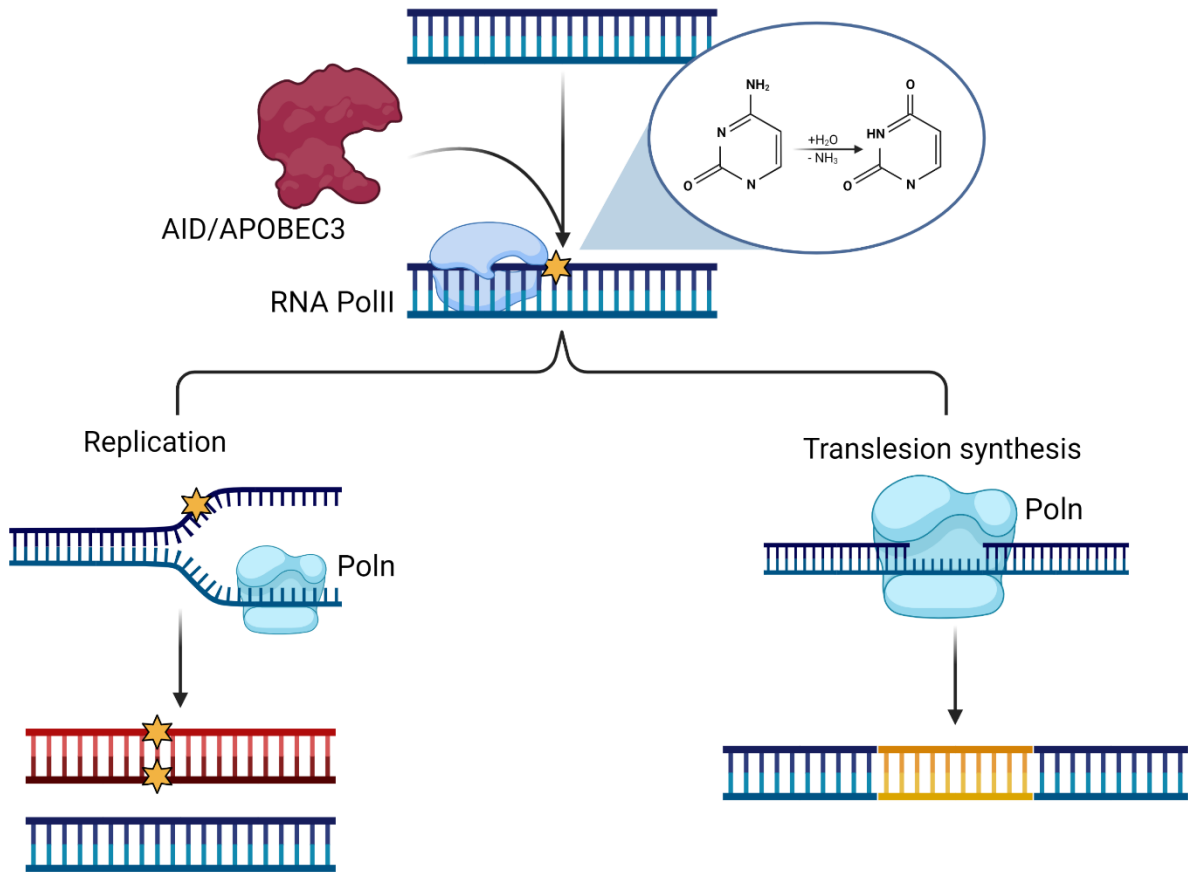
Figure 15. Mechanism of hypermutation initiated by AID/APOBEC and repaired by Polη. Transcription by RNA polymerase II (RNA PolII) exposes a single strand DNA substrate for AID/APOBEC enzymatic action. After deamination of cytosine to uracil (callout), the mismatch is processed by one of two pathways. It can undergo replication, perpetuating the mutation through half of the daughter strands and potentially requiring Polη to resolve a replication fork stalled at a mutation. Alternatively, the mutated base and surrounding sequence are removed by base excision repair (BER) and the resulting gap is filled in by error-prone TLS Polη, resulting in a mutated patch. Created with BioRender.com.
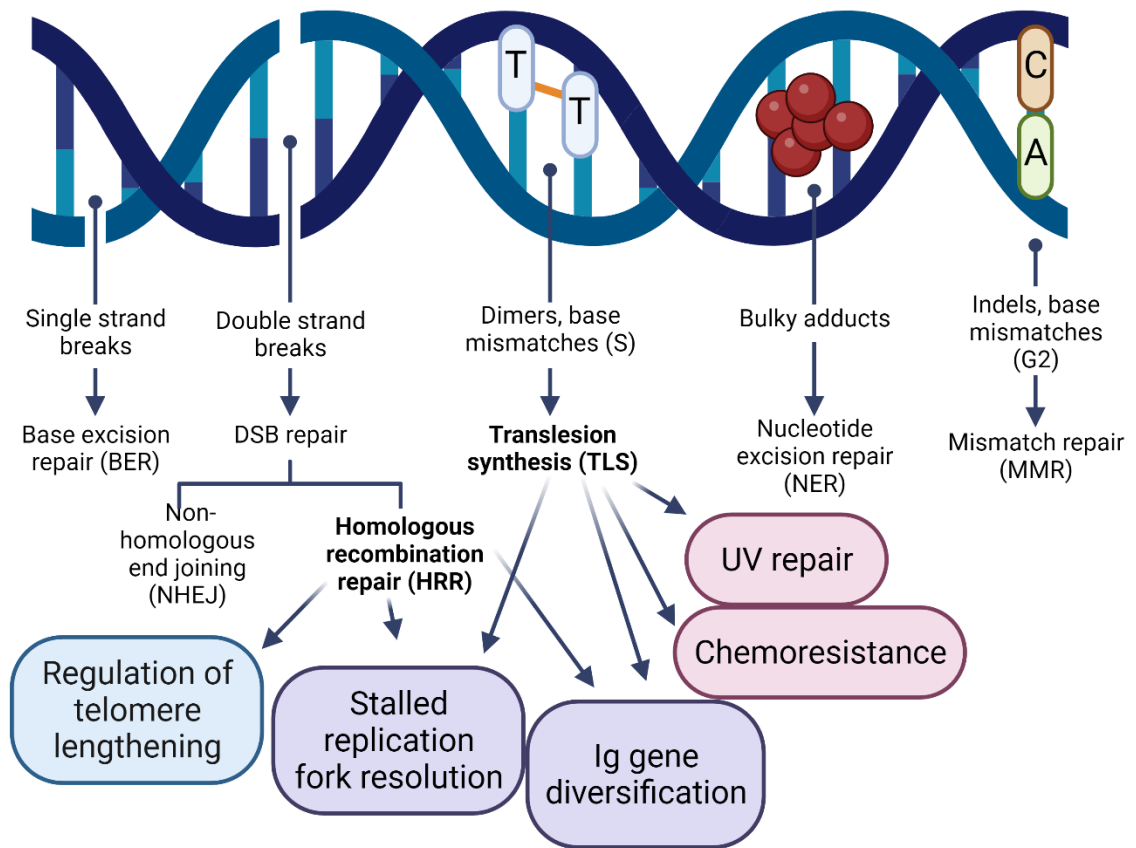
Figure 16. Summary of multiple roles of Polη in genome maintenance through its dual action in the TLS and HR pathways. Created with BioRender.com.

4.2 Results

4.2.1 Overexpression of EGFP-tagged polymerase eta results in low expression

One of the main tools used here for the study of Polη was the EGFP-N1-POLH plasmid construct gifted by Jean-Yves Masson (Université Laval). This plasmid encodes green fluorescent C-terminal tagged Polη (Appendix B Fig. 1) for visualization of the protein post transfection. Overexpression of Polη in all cell types resulted in two transfection phenotypes, as seen in Fig. 17A: diffuse EGFP signal throughout the nucleus and distinct nuclear foci. Additionally, I observed low fluorescence intensity compared to EGFP alone (Fig. 17A) and relatively low transfection efficiency across multiple cell types (Fig. 17B, C). Despite the short turnover time of endogenous Polη (Section 4.1.1) [185,193], EGFP-N1-POLH-transfected cells were visualized 24 h post-transfection for optimal Polη expression, unless otherwise noted (Fig. 17B, C).
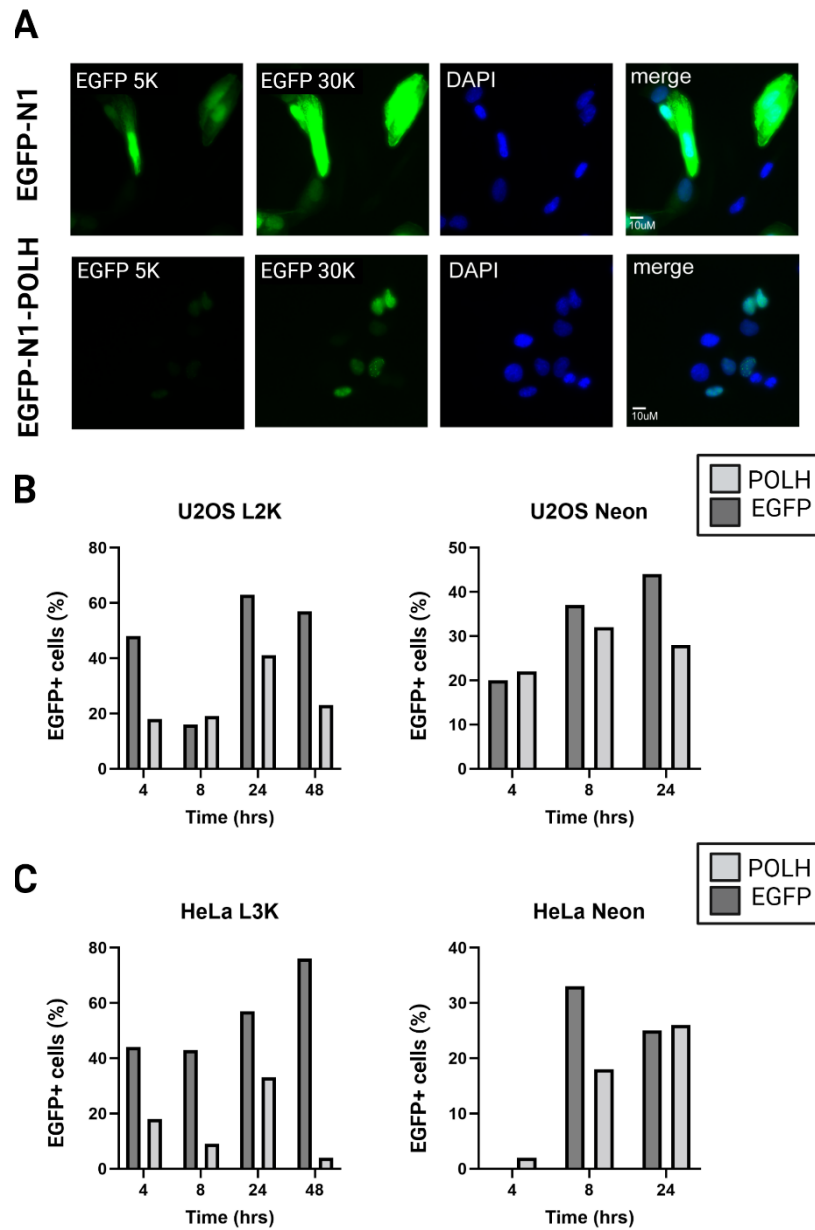
Figure 17. Transfection profile of EGFP-tagged Polη. A. Representative IF images of WT U2OS cells with Lipofectamine-mediated transfection of pEGFP-N1 (upper) and the EGFP-N1-POLH construct (lower). Maximum FITC fluorescence was set to 5000 (left) and 30000 (right). B. Transfection efficiency of FITC+ expression in U2OS WT cells transfected with Lipofectamine 2000 (left) and Neon electroporation (right) over various timepoints. C. Transfection efficiency of FITC+ expression in HeLa WT cells transfected with Lipofectamine 3000 (left) and Neon electroporation (right) over various timepoints. n = 1. Created with GraphPad Prism (v.9.3.1 471, San Diego, California USA, www.graphpad.com) and BioRender.com.

4.2.2 Homology-directed repair is decreased in cells deficient in polymerase eta

4.2.2.1 Construction and validation of stable puromycin-resistant Polη-deficient cell lines

To examine novel roles of Polη in genome maintenance, U2OS ΔPOLH clonal cell lines were constructed as described in Section 2.4.2. In brief, the 5' end of POLH exon 3 (the first exon common between the three non-pathogenic isoforms of POLH) was targeted by Cas9n-mediated DSB formation. The insertion of a co-transfected linearized PuroR-SV40 sequence by endogenous NHEJ at the break site was predicted to result in cells with selectable puromycin resistance and severely truncated POLH. These clonal cells, grown from single cell colonies and maintained in media supplemented with 4 ug/uL puromycin, exhibited no growth defects (data not shown; as seen in [211]). However, validating POLH KO and a lack of functional Polη protein production was not straightforward.

A rabbit polyclonal anti-POLH antibody (Thermo Fisher Scientific) produced high background when analyzed by IF microscopy (data not shown). A mouse monoclonal anti-POLH antibody (Santa Cruz, sc-17770) was able to bind EGFP-tagged Polη, as shown by focal colocalization in U2OS and HeLa WT cells transfected with EGFP-N1-POLH and subsequently stained with anti-POLH (Fig. 18A, B). However, no nuclear staining of endogenous Polη was seen, even upon UV treatment to induce POLH expression (Fig. 19A) [191,195,196]. There was a slightly higher intensity of anti-POLH stain in WT U2OS compared to ΔPOLH cells (Fig. 19B), but as this signal was extranuclear, it was likely background and not the nuclear TLS polymerase.
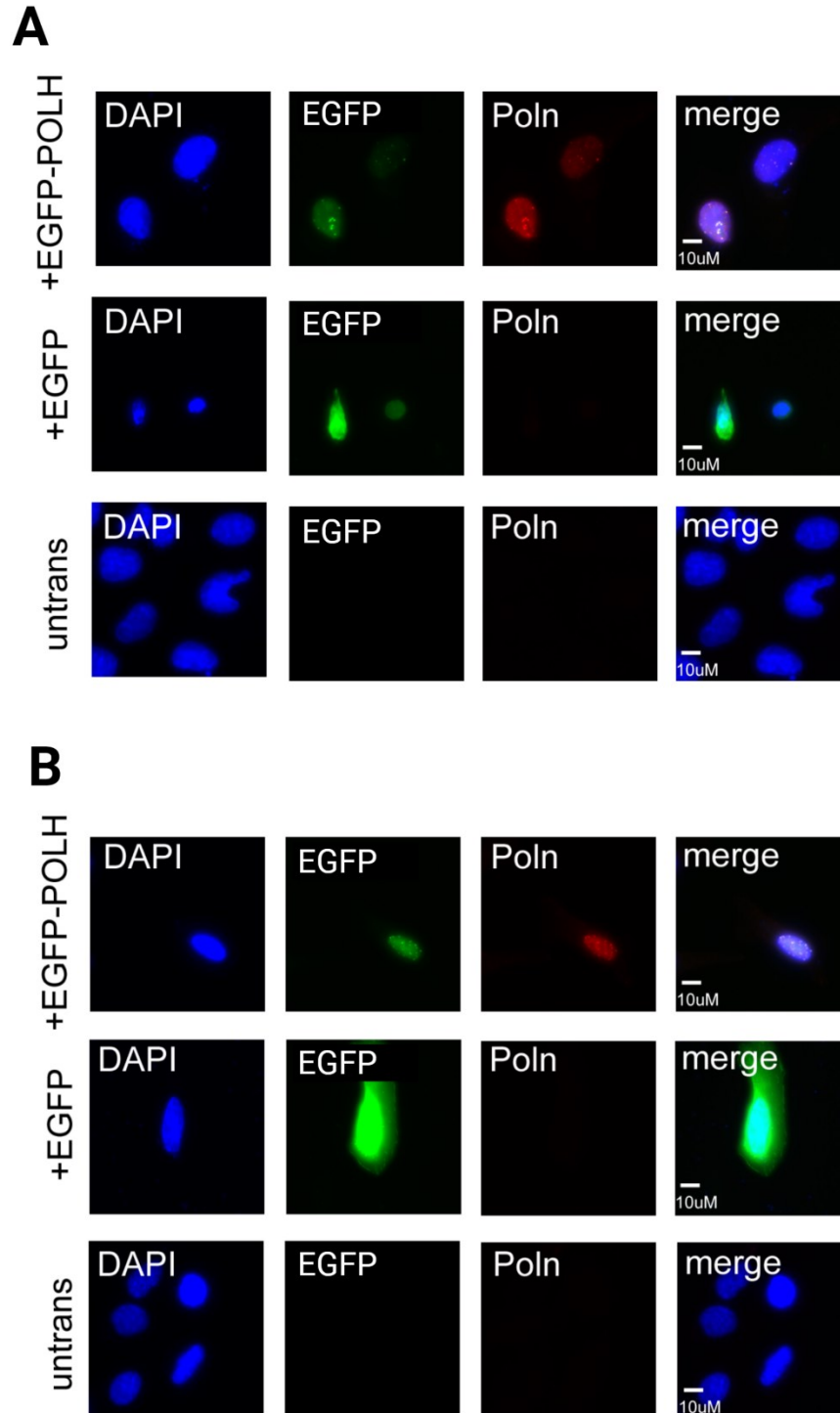
Figure 18. Validation of mouse anti-POLH antibody staining of overexpressed EGFP-tagged Polη. Anti-POLH is stained red (1:300) and nuclei are stained blue with DAPI, while overexpressed Polη is tagged green. A. U2OS WT. B. HeLa WT. Created with Slidebook (Intelligent Imaging Innovations, Boulder, CO), Adobe Photoshop CS (v.6.2), and BioRender.com.
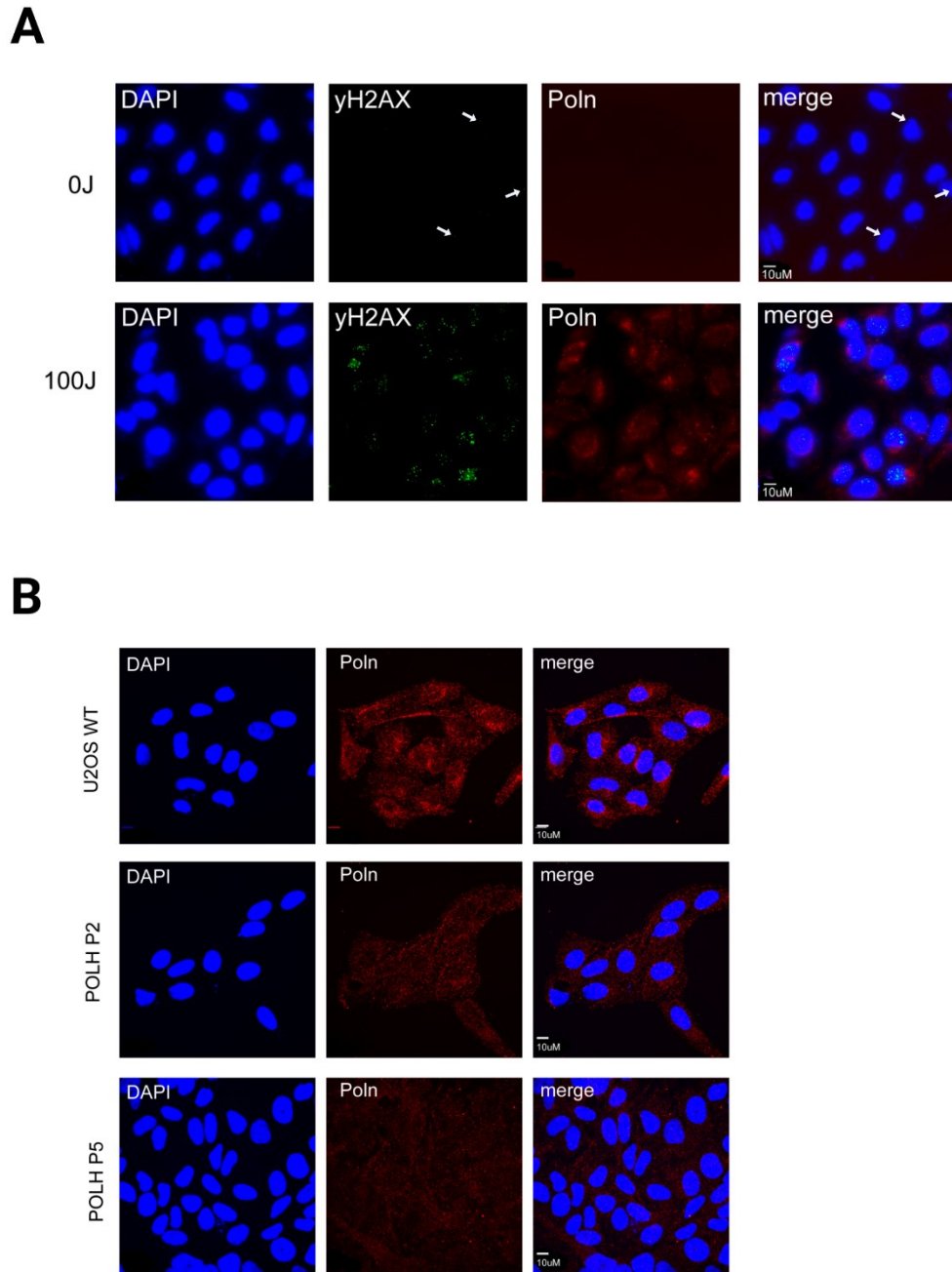
Figure 19. Validation of mouse anti-POLH antibody staining of endogenous Polη.
Anti-POLH is stained red (1:300) and nuclei are stained blue with DAPI. A. U2OS WT
cells were irradiated with 0J (upper) and 100J (lower) of UV. After 30 min incubation,
slips were fixed and stained as previously described. Double stranded DNA breaks are
visualized with anti-γH2AX antibody (green). B. Untreated U2OS WT and ΔPOLH cells
were stained for anti-POLH (red). Created with Slidebook (Intelligent Imaging
Innovations, Boulder, CO), Adobe Photoshop CS (v.6.2), and BioRender.com.

Following failure to visualize endogenous Polη by IF, whole cell lysates were analyzed by SDS-PAGE and Western blot. All cell types (Table 9) were, through multiple trials, lysed using two different methods, run on both polyvinylidene difluoride and nitrocellulose membranes, stained with multiple anti-Polη antibodies, and blocked with both BSA and milk. No experiments produced clean, definitive bands at the expected molecular weight (79 kDa) (Santa Cruz, sc-17770).

Finally, multiple quantitative PCR (qPCR) primer pairs were designed to target various regions of the POLH gene that are common between all three isoforms (Table 7, Fig. 6). Three reference genes were used to normalize expression data across all samples: *B2M*, *RAC1*, and *TBP*. Target stability values were calculated automatically by Bio-Rad CFX Manager (v.3.1) to be mean coefficient of variation (CV) = 0.3176 and mean M value = 0.8366, which are below the threshold values required for heterogenous samples [255,256]. Samples included were: U2OS WT, XPV3617, and U2OS ΔPOLH P2-6 (sample P1 was excluded due to lack of cDNA).

Samples were initially amplified with primer pairs POLH5 and POLH6 (Table 7). Both primer pairs target the last and largest exon of POLH, exon 11, which is far downstream of the PuroR insert site in exon 3 (Fig. 6). ΔPOLH mutants would therefore be expected to have virtually no transcript at this site due to nonsense mutation downstream of the insertion; however, while relative gene expression was decreased in XPV and three ΔPOLH mutants relative to WT, two mutants (P2 and P5) had increased expression (Fig. 20A, B). This unexpected result is likely due to the presence of the SV40 promoter in the linearized PuroR sequence and the unpredictability of NHEJ repair. Therefore, an additional primer pair was designed to target the junction of POLH exons 2

and 3 (Fig. 6, Table 7). Expression of this amplicon was decreased in all mutants and XPV relative to U2OS WT (Fig. 20C), as expected for successful interruption of gene sequence at this site.
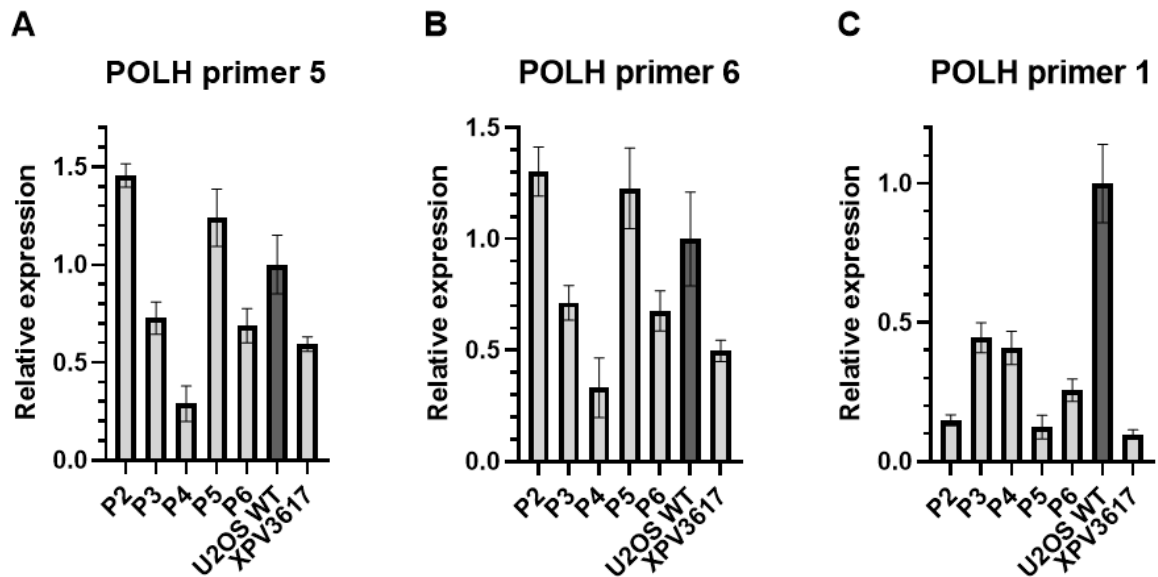
Figure 20. Validation of ΔPOLH cell lines by qPCR. Gene expression values relative to U2OS WT (dark gray) are plotted from n = 3 technical replicates. A. Primer pair 5, downstream of the edit site. B. Primer pair 6, downstream of the edit site. C. Primer pair 1, at the edit site. Created with GraphPad Prism (v.9.3.1 471, San Diego, California USA, www.graphpad.com) and BioRender.com.

4.2.2.2 HDR assay

Since Polη has been shown to play a role in HR-mediated repair of DSBs (Section 4.1.1), we sought to determine whether this mutagenic TLS polymerase may also be involved in CRISPR-mediated HDR-dependent gene editing. To examine the role of Polη in HDR, we employed an HDR assay developed by Pinder et al. (2015) which allows quantification of cells that have undergone productive HDR by targeted knock-in of the fluorescent reporter gene Clover to the nuclear lamin A locus (Fig. 7, Section 2.8). HDR+ cells therefore have fluorescent nuclear lamin, allowing their quantification by IF or flow cytometry (Fig. 21A).

We first observed decreased HDR efficiency in Polη-deficient XPV cells compared to WT U2OS ($p = 0.0556$) and WT HeLa ($p = 0.3333$), as determined by the percentage of transfected (iRFP+) cells with green fluorescent nuclear lamin, though this was not statistically significant (Fig. 21B). However, conducting the assay in the six ΔPOLH U2OS cell lines compared to WT consistently showed significantly lower HDR efficiency by both IF and FACS (Fig. 21C, 22B), which was restored to near-WT levels by the addition of EGFP-N1-POLH (co-stained with red anti-POLH to differentiate HDR+ alone, which were not counted, and HDR+ POLH+ cells) (Fig. 21D).
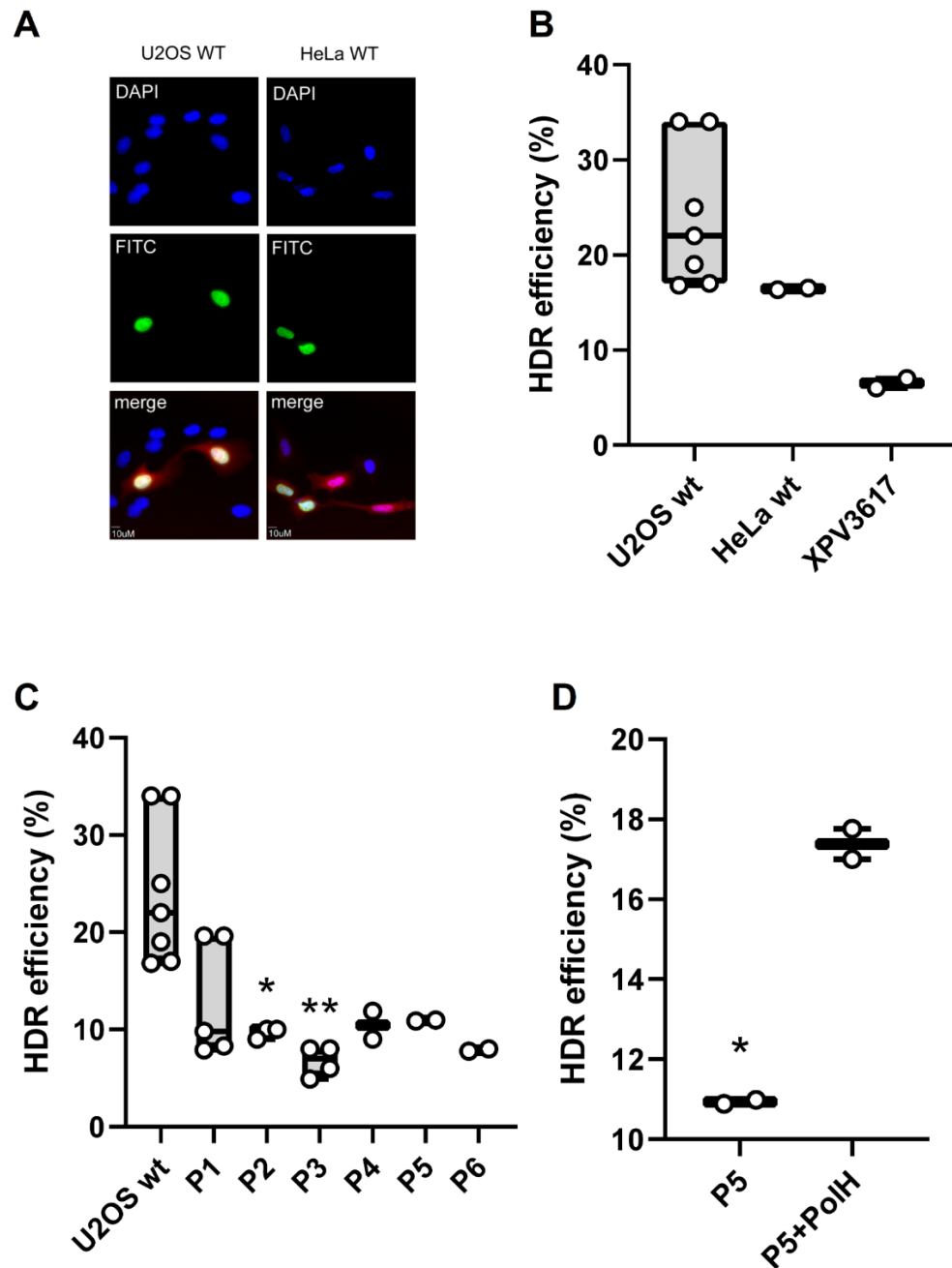
Figure 21. Clover LMNA HDR assay by IF microscopy. HDR efficiencies are represented as percentage of transfected (iRFP+) cells. * p < 0.05. ** p < 0.01. n = 2 – 7, as indicated by white circles. A. Representative images of HDR+ cells in U2OS WT (left) and HeLa WT (right). iRFP (red) acted as a transfection control. Nuclei are stained blue with DAPI. B. HDR efficiency of different WT cell types. C. HDR efficiencies of U2OS WT compared to ΔPOLH clonal lines P1-6. D. HDR efficiency of ΔPOLH P5 with and without addback of EGFP-N1-POLH construct, stained with 555-anti-Polη. Created with GraphPad Prism (v.9.3.1 471, San Diego, California USA, www.graphpad.com), Slidebook (Intelligent Imaging Innovations, Boulder, CO), Adobe Photoshop CS (v.6.2), and BioRender.com.

93
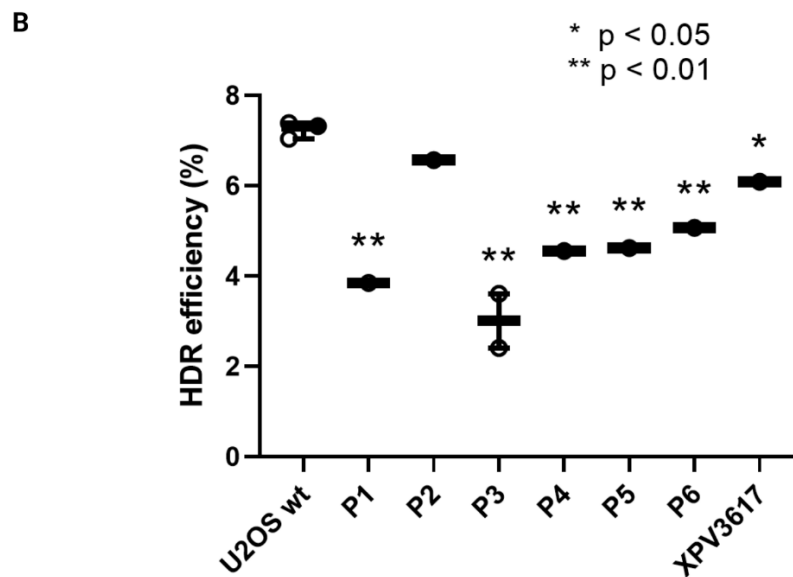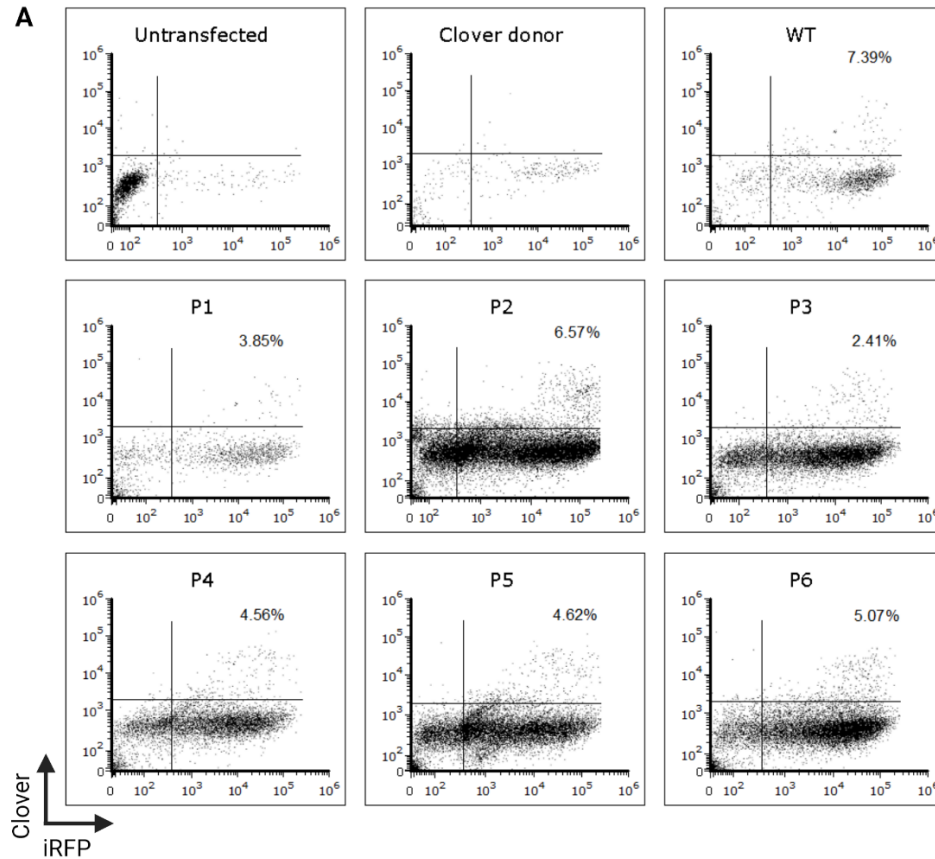
Figure 22. Clover LMNA HDR assay by FACS. A. Representative dot plots of FACS data. Created with FCS Express (De Novo Software, CA, USA; v7.12.0007). B. HDR efficiency as a percentage of total cell population analyzed by FACS. n varied by experiment: U2OS wt (n = 3), ΔPOLH clone P2 (n = 2), ΔPOLH clones P1, P2, P4 – P6, XPV3617 (n = 1). Created with GraphPad Prism for Windows (v.9.3.1 471, San Diego, California USA, www.graphpad.com).

4.2.3 A novel role for translesion polymerase eta in the repression of retrotransposition activity

Another mechanism in which Polη may be involved is the repression of retrotransposition activity, as seen with APOBEC enzymes, which are known to promote Ig gene diversity in cooperation with the mutagenic activity of Polη. To investigate this possibility, we used a G418 selection-based LINE-1 assay, which enables quantification of cells that have undergone LINE-1 retrotransposition in order to express the neomycin resistance G418R gene. Colonies grown from surviving cells are visualized with crystal violet staining (Fig. 23A). When either the pEGFP-N1 or EGFP-N1-POLH plasmids were co-transfected into HeLa cells with the L1-neo plasmid, approximately half as many colonies grew on the +POLH plates (Fig. 23A, B), indicating potential LINE-1 repression by Polη.

To investigate the effect of loss of Polη function on LINE-1 activity, the assay was repeated using U2OS WT and ΔPOLH cells. No significant difference was observed between cell types when examining only retrotransposition activity (Fig. 23C). However, when these cells were co-transfected with pEGFP-N1 or EGFP-N1-POLH, results were inconsistent between ΔPOLH cell lines, with some showing increased survival with EGFP (P1, P4) and others showing increased survival with POLH (P5, P6) (Fig. 23D). WT U2OS cells behaved similarly to WT HeLa when transfected with EGFP *vs.* POLH (Fig. 23D). Further repetition of this assay, as well as study of Polη in the context of other retro elements and APOBECs, will help elucidate the role of Polη in this factor of genome maintenance (Fig. 24).
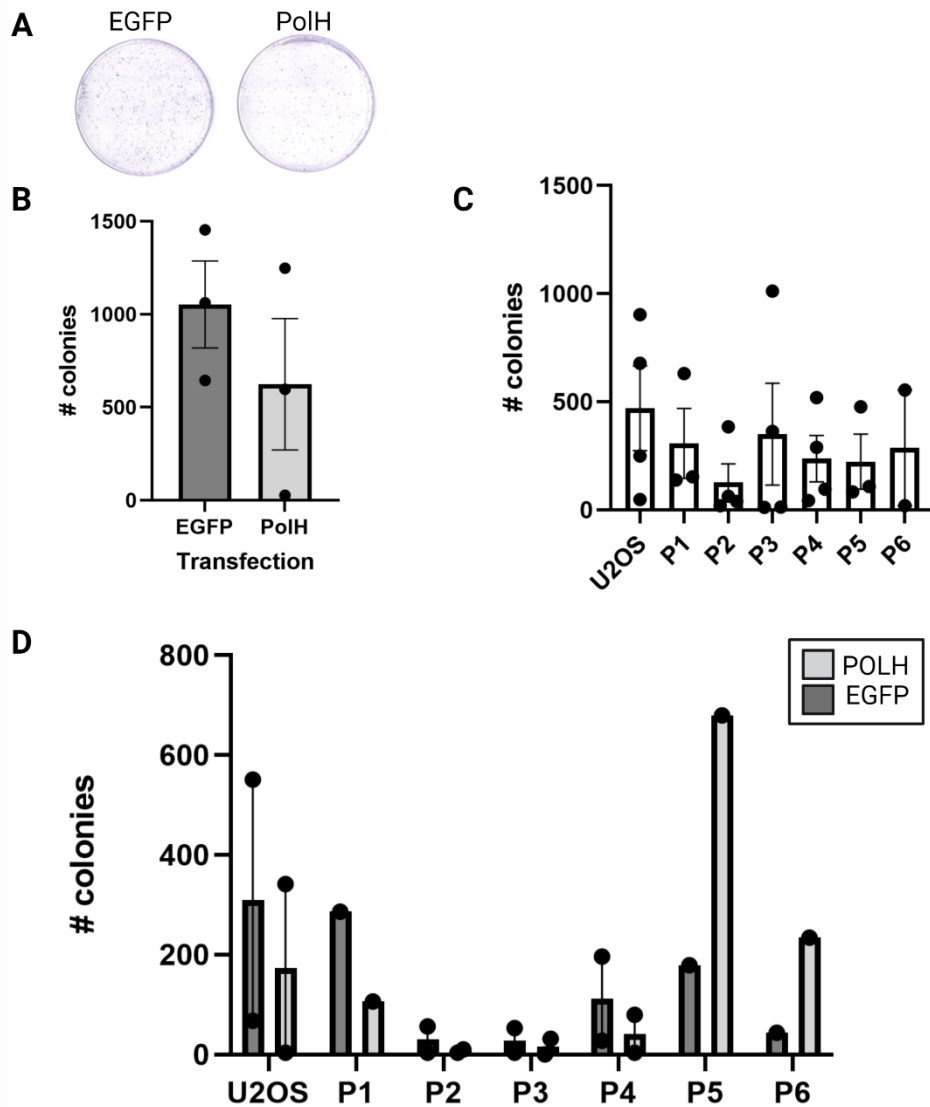
Figure 23. The effect of Polη on the LINE-1 neomycin retrotransposition assay.
EGFP dark gray, PolH light gray, no co-transfection white. Means and standard errors of means (calculated automatically by GraphPad Prism) are shown. A. Representative images of HeLa WT cells co-transfected with L1-neo and either EGFP-N1 (left) or EGFP-N1-POLH (right). B. HeLa WT cells +L1-neo +EGFP-N1 (left) or +EGFP-N1-POLH (right). n = 3. C. U2OS WT and ΔPOLH cells co-transfected with L1-neo and pBSKII (empty vector). n varied by experiment: U2OS wt, ΔPOLH clones P2 – P4 (n = 4), ΔPOLH clones P1, P5 (n = 3), ΔPOLH clone P6 (n = 2). D. U2OS WT and ΔPOLH cells co-transfected with L1-neo and pEGFP-N1 (dark gray, left) or EGFP-N1-POLH (light gray, right). n = 1 – 2. Created with GraphPad Prism for Windows (v.9.3.1 471, San Diego, California USA, www.graphpad.com) and BioRender.com.
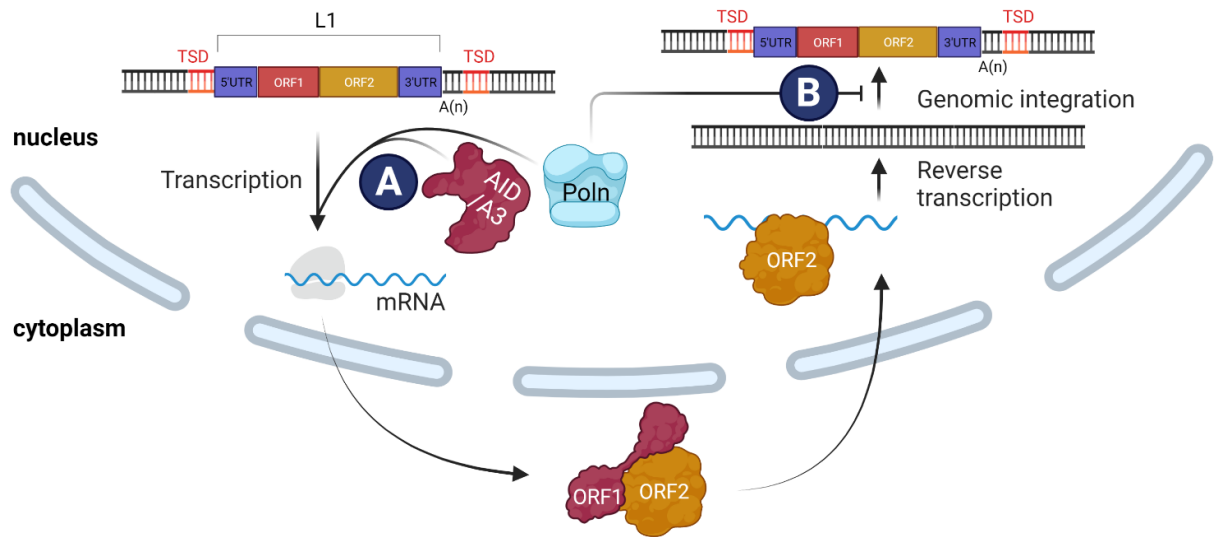
Figure 24. Potential roles of Polη in the LINE-1 retrotransposition mechanism.
Potential modes of action are noted in navy circles. A. Polη acts in a similar cooperative mechanism as in Ig diversification, by erroneously repairing AID/APOBEC-mediated mutations. In this scenario, the mutated transcript would be unable to produce functional ORF proteins. B. Polη has a non-enzymatic function in sterically blocking genomic integration, either by competition with RNaseH2 (not shown) or by stabilizing the ssDNA : dsDNA intermediate structure. ORF open reading frame, UTR untranslated region, TSD target site duplications. Created with BioRender.com.

4.3 Discussion

EGFP-tagged Polη had two distinct phenotypes after expression: diffuse and focal, both constrained to the nucleus (Fig. 17A). The diffuse pattern likely represents overexpression to the point that it is impossible to discern distinct foci, while the foci may be a more accurate pattern of endogenous Polη behaviour [224,257]. However, I was unable to determine what these foci were localizing to in multiple experiments, as they did not co-localize to DNA damage markers such as 53BP1 (as expected), nor to H2AX or γH2AX. Additionally, EGFP-Polη foci did not localize to PML or TRF2 in U2OS cells, as would be expected of ALT bodies [224,258,259]. They may colocalize to other known protein interactors such as ubiquitinated PCNA, PALB2, BRCA2, and/or FANCD2 (Section 4.1.1), which merits further investigation. Alternatively, these foci may not colocalize to any cellular structure as they could simply be due to overexpression and represent Polη nuclear bodies that have not yet been degraded by endogenous cellular machinery. Finally, overexpressed Polη foci could be at a location the endogenous protein does not normally localize to, such as DNA that does not contain lesions or DSBs. In this case, our EGFP-tagged Polη would be expected to have increased mutagenic properties compared to the endogenous TLS pol, which is tightly regulated (Section 4.1.1, Fig. 14).

This last scenario could explain the other notable phenotypes seen when transfecting this plasmid, namely, the low intensity of fluorescence compared to EGFP alone (Fig. 17A) and the relatively low transfection efficiency across multiple cell types (Fig. 17B, C). If overexpression of Polη is excessively mutagenic and potentially toxic, cells may be actively repressing its expression, degrading the protein, or simply

undergoing apoptosis if the mutagenic load becomes too high [260]. However, cell cultures transfected with EGFP-N1-POLH did not exhibit growth defects or lower cell density compared to those transfected with EGFP alone, and this does not explain the fact that optimal expression of EGFP-tagged Polη is seen at 24 h post transfection (Fig. 17B, C). Polη-induced toxicity also seems unlikely, as cells were able to express EGFP-tagged Polη for an extended period of time when the construct was inserted into the GEL15 site (Section 3.2.1.3, Fig. 11). Lower transfection efficiency may simply be due to difference in size between the two plasmids, with smaller plasmids taken up more efficiently by cells in culture [261,262].

An additional confounding factor when working with this protein was the fact that we were not able to validate our ΔPOLH cells with an anti-POLH antibody. The mouse monoclonal antibody had perfect overlap with overexpressed EGFP-tagged Polη by IF but did not seem able to pick up endogenous Polη by either IF (Fig. 19) or WB. This may be due to relatively low levels of endogenous expression under normal conditions, though Polη should be highly expressed 30 min post UV exposure [191,195,196], which we did not observe here (Fig. 19A). Alternatively, the C-terminal EGFP tag may have altered Polη protein structure sufficiently to allow detection by the monoclonal anti-POLH antibody, which is designed to bind near the C-terminal end of the protein (Santa Cruz, sc-17770). In addition to our collaborators who created the EGFP-N1-POLH construct [198], others (e.g., [224]) have used fluorescent-tagged POLH to study this protein.

We therefore validated the U2OS POLH cell lines by qPCR. Initial targeting of exon 11 of POLH with two independent primer pairs revealed unexpected results (Fig. 20A, B). Due to PuroR insertion at exon 3 leading to nonsense-mediated decay of mRNA

or protein truncation, as in XPVs, virtually no transcript was expected this far downstream of the insert site. However, while gene expression was decreased in XPV and three ΔPOLH mutants relative to WT, two mutants (P2 and P5) had increased expression (Fig. 20A, B). Primers targeting the junction of POLH exon 2 and 3 (Fig. 6, Table 7) resulted in decreased expression of all ΔPOLH cell lines and XPV compared to WT U2OS (Fig. 20C). These results are likely due to the relatively unpredictable nature of DNA repair by NHEJ, which can insert blunt-end sequences, as the linearized PuroR was, in either orientation at the break site [263]. The increased expression of exon 11 in ΔPOLH lines P2 and P5 was likely due to PuroR insertion in the sense orientation, as the SV40 promoter included in the linear construct would have driven not only PuroR expression, but any sequence downstream of it. The remaining ΔPOLH clones (P3, P4, P6) presumably had PuroR insertion in the antisense orientation, leading to SV40-driven expression of the PuroR gene, but not downstream POLH exons. Regardless of the genetic structure of these cell lines, POLH expression is functionally similar to that in naturally Polη-deficient XPV cells, as evidenced by decreased expression of the POLH exon 2 – 3 junction as well as their ability to survive in high doses of puromycin (Table 9).

The ΔPOLH cell lines were used in two different assays to examine potential roles of Polη in novel genome maintenance mechanisms. First, we showed that Polη is involved in CRISPR-mediated gene editing by HDR. Previous work has implicated Polη in the HR mechanism, and it seems able to independently extend D-loop intermediates [185]. Here, a loss of functional Polη reduced HDR efficiency, which was restored to near-WT levels with addback of Polη (Fig. 21, 22). An effect on HDR could have implications

for gene editing, with increased gene insertion efficiency potentially improving gene editing outcomes [144,148,149]. However, caution must be taken when attempting to modify gene editing efficiencies, especially in this context. Overexpression of TLS polymerases, including Polη, are seen in many different types of cancers and are associated with increased resistance to DNA-targeting chemotherapies [264,265]. Therefore, while potential promotion of the HDR mechanism by Polη is interesting, further study is required to determine the safety of manipulating this and other effectors of HR to improve gene editing outcomes. For example, oncologic research is required to determine whether TLS polymerase overexpression is a consequence of unmitigated growth and mutation in cancer cells, or a driving force of it.

A limitation of this section is that while no growth difference was noticed in the ΔPOLH cells compared to WT U2OS, we did not formally explore cell cycle distribution [146]. Future work should involve investigating whether loss of Polη alters or impairs cell cycling, which could explain it as an apparent requirement for higher efficiency HDR, as HR is restricted to S/G2 phases [42,43]. Additionally, increased experimental repetition and higher cell numbers collected for FACS will strengthen these results.

An additional genome maintenance role for Polη was studied; its involvement in the repression of LINE-1 retrotransposition. Polη is known to be involved in Ig hypermutation, following initial G-to-A mutagenesis by AID/APOBEC enzymes, which have been shown to repress LINE-1 activity (see Section 4.1.3). Regulation of retro elements, including viruses and retrotransposons, has even been proposed as the original function of the APOBEC3 family [157]. However, exactly which APOBEC3 enzymes can repress LINE-1 retrotransposition and with which mechanisms is still in question, with

contradictory results found in the literature. All APOBEC3 enzymes have been found to differentially inhibit LINE-1 activity, though some studies did not find any retrotransposition restriction activity of A3G [157,249,250]. A3A, a nuclear APOBEC, functions in LINE-1 repression through its deaminase activity [250,251]; other nuclear APOBECs including A3B and A3H, as well as cytoplasmic A3F and A3G, display deamination-independent mechanisms [237,250,266]. Interestingly, we observed significant co-occurrence of Polη with both nuclear and cytoplasmic A3s that have been shown to strongly suppress LINE-1 activity, namely A3A, A3B, and A3F (Table 13) [250,251,266]. Here I attempted to discern whether Polη also plays a role in retrotransposition regulation, hypothesizing that it may do so in cooperation with APOBECs, similar to the mechanism of Ig diversification.

In this scenario, Polη would perform mutagenic repair of APOBEC-induced deamination mutations during retroelement transcription (Fig. 24A). This would prevent transcription of functional LINE-1 protein, and therefore further work should involve monitoring LINE-1 ORF1p and ORF2p expression in the context of Polη overexpression and knockdown or knockout, as in Kinomoto et al. (2007). Alternatively, (Fig. 24B) Polη could be acting in cooperation with other HR repair factors to repress LINE-1 integration in a post-translational mechanism [242]. This repression activity would be independent of APOBEC3 function and may or may not depend on the enzymatic function of Polη itself. Polη is recruited during HR by PALB2, BRCA2, and Rad51, the latter which functions to stabilize HR intermediate structures [192,198,267] (see Section 4.1.1). Polη could be acting to stabilize the intermediate ssDNA : dsDNA structure prior to LINE-1 integration, thereby delaying or preventing the final step of the LINE-1 pathway (Fig. 24B).

Limitations in this study include the lack of inclusion of additional LINE-1 repression elements such as APOBECs and other HR repair factors, preventing the elucidation of the mechanism of Polη-induced LINE-1 repression, and the contradictory results seen when the LINE-1 assay was applied to ΔPOLH cells (Fig. 23C, D). One would expect that if Polη plays a substantial role in LINE-1 repression, then ΔPOLH cells should show increased colony growth compared to WT. However, that was not observed in this study; potentially due to compensation by other mechanisms (as in XPV disease, see Section 4.1.2), or perhaps due to an as-of-yet undiscovered growth defect in the novel cell lines. Future work must seek to resolve these issues and further investigate potential modes of cooperation between Polη, APOBECs, and HR proteins, in order to resolve the functions of these mutagenic genome maintenance factors.

CHAPTER 5: CONCLUSION

Despite recent rapid advancements in the field of gene therapy, concerns remain about the safety and efficacy of various techniques, particularly when inserting novel genetic material. Here, using genetic engineering systems, I examined two potential solutions to these issues: the lack of safe genomic sites amenable to gene insertion, and the uncertainty still surrounding the mechanism of homologous recombination generally necessary for gene insertion.

First, I characterized a gene editing locus previously identified on human chromosome 15. Although this locus does not adhere to our stringent criteria of a GEL site, it has previously proven useful in gene editing, and I further demonstrated the feasibility of targeting this site using fluorescent reporter genes. This adds another tool to the 'molecular toolkit' of available sites for safe gene editing. Further work should examine the safety and long-term efficacy of gene insertion at GEL15. To further expand the available database of GELs, providing increased flexibility in basic research and potentially reinforcing the importance of considering safety during clinical gene therapy studies, we created a preliminary open-access GEL identification code such that other researchers can identify novel putative GELs for their own needs.

Next, I examined the relatively recently discovered widespread role of the TLS polymerase η in various mechanisms of genome maintenance. I showed that Polη affects CRISPR-mediated HDR, identifying it as a potential factor in gene editing efficiency. However, balance of this and other mutagenic proteins such as APOBECs must be stressed; while they may help repress excessive retrotransposition and genome instability, they have been implicated in cancer initiation and progression. Care must therefore be

taken when attempting to modify cellular mechanisms to increase the efficiency of HDR

for gene editing purposes, as we are still likely unaware of many unintended

consequences in this field.

REFERENCES

1. Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine* **79**, (1944).

2. Griffith, F. The Significance of Pneumococcal Types. *Journal of Hygiene* **27**, (1928).

3. Wirth, T., Parker, N. & Ylä-Herttuala, S. History of gene therapy. *Gene* **525**, 162–169 (2013).

4. Zinder, N. D. & Lederberg, J. Genetic exchange in Salmonella. *J Bacteriol* **64**, 679–699 (1952).

5. Berg, P., Baltimore, D., Brenner, S., Roblin, R. O. & Singer, M. F. Summary statement of the Asilomar conference on recombinant DNA molecules. *Proceedings of the National Academy of Sciences* **72**, (1975).

6. Berg, P. & Singer, M. F. The recombinant DNA controversy: twenty years later. *Proceedings of the National Academy of Sciences* **92**, (1995).

7. Getz, L. J. & Dellaire, G. Angels and Devils: Dilemmas in Dual-Use Biotechnology. *Trends Biotechnol* **36**, 1202–1205 (2018).

8. Meagher, K. M., Allyse, M. A., Master, Z. & Sharp, R. R. Reexamining the Ethics of Human Germline Editing in the Wake of Scandal. *Mayo Clin Proc* **95**, 330–338 (2020).

9. Terheggen, H. G., Lowenthal, A., Lavinha, F. & Colombo, J. P. Familial hyperargininaemia. *Arch Dis Child* **50**, (1975).

10. Dellaire, G. Editorial: CRISPR Medicine: From Bench to Bedside. *Curr Gene Ther* **17**, 261–262 (2018).

11. Blaese, R. M. *et al.* T lymphocyte-directed gene therapy for ADA-SCID: Initial trial results after 4 years. *Science (1979)* **270**, 475–480 (1995).

12. Shreenivas, S. Who killed Jesse Gelsinger? Ethical issues in human gene therapy. *Monash Bioeth Rev* **19**, 35–43 (2000).

13. Sibbald, B. Death but one unintended consequence of gene-therapy trial. *CMAJ* **164**, (2001).

14. Stolberg, S. G. The Biotech Death of Jesse Gelsinger - The New York Times. *The New York Times Magazine* https://www.nytimes.com/1999/11/28/magazine/the-biotech-death-of-jesse-gelsinger.html (1999).

15. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156–1160 (1996).

16. Klug, A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* **43**, 1–21 (2010).

17. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A* **104**, 3055–3060 (2007).

18. Urnov, F. D., Rebar, E. J., Holmes, M. C., Steve Zhang, H. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. (2010) doi:10.1038/nrg2842.

19.   Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 756–761 (2010).

20.   Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143–150 (2011).

21.   Singh, J. K. & van Attikum, H. DNA double-strand break repair: Putting zinc fingers on the sore spot. *Semin Cell Dev Biol* (2020) doi:10.1016/j.semcdb.2020.09.003.

22.   Li, H. *et al.* Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Nature* **5**, (2019).

23.   Salsman, J. & Dellaire, G. Precision genome editing in the CRISPR era. *Biochemistry and Cell Biology* **95**, 187–201 (2017).

24.   Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol* **21**, 1468–1478 (2019).

25.   Broeders, M., Herrero-Hernandez, P., Ernst, M. P. T., van der Ploeg, A. T. & Pijnappel, W. W. M. P. Sharpening the Molecular Scissors: Advances in Gene-Editing Technology. *iScience* **23**, 100789 (2020).

26.   Pinder, J., Salsman, J. & Dellaire, G. Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res* **43**, 9379–9392 (2015).

27.   Uematsu, N. *et al.* Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biol* **177**, 219 (2007).

28.   Andrin, C. *et al.* A requirement for polymerized actin in DNA double-strand break repair. *Nucleus* **3**, 384–395.

29.   Mari, P. O. *et al.* Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A* **103**, 18597–18602 (2006).

30.   Xu, X. *et al.* DNA-PKcs: A Multi-Faceted Player in DNA Damage Response. *Front. Genet* **11**, 607428 (2020).

31.   Yoo, S. & Dynan, W. S. *Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein*. *Nucleic Acids Research* vol. 27 https://academic.oup.com/nar/article/27/24/4679/2901950 (1999).

32.   Gupta, A. *et al.* Role of 53BP1 in the regulation of DNA double-strand break repair pathway choice. *Radiat Res* **181**, 1–8 (2014).

33.   Reynolds, P. *et al.* The dynamics of Ku70/80 and DNA-PKcs at DSBs induced by ionizing radiation is dependent on the complexity of damage. doi:10.1093/nar/gks879.

34.   Li, S. *et al.* Polynucleotide kinase and aprataxin-like forkhead-associated protein (PALF) acts as both a single-stranded DNA endonuclease and a single-stranded DNA 3′ exonuclease and can participate in DNA end joining in a biochemical system. *Journal of Biological Chemistry* **286**, 36368–36377 (2011).

35.   Jefferson, J. *et al.* WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nat Struct Mol Biol* **13**, (2006).

36. Povirk, L. F., Zhou, T., Zhou, R., Cowan, M. J. & Yannone, S. M. Processing of 3′-phosphoglycolate-terminated DNA double strand breaks by artemis nuclease. *Journal of Biological Chemistry* **282**, 3547–3558 (2007).

37. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794 (2002).

38. Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res* **2**, (2013).

39. Nick McElhinny, S. A. & Ramsden, D. A. Polymerase Mu Is a DNA-Directed DNA/RNA Polymerase. *Mol Cell Biol* **23**, 2309–2315 (2003).

40. Wyman, C. & Kanaar, R. DNA Double-Strand Break Repair: All's Well that Ends Well DSB: DNA double-strand break. (2006) doi:10.1146/annurev.genet.40.110405.090451.

41. Karanam, K., Kafri, R., Loewer, A. & Lahav, G. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. *Mol Cell* **47**, 320–329 (2012).

42. Rothkamm, K., Krüger, I., Thompson, L. H. & Löbrich, M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Mol Cell Biol* **23**, (2003).

43. Kratzer, K., Getz, L. J., Peterlini, T., Masson, J.-Y. & Dellaire, G. Addressing the dark matter of gene therapy: technical and ethical barriers to clinical application. *Hum Genet* 1–19 (2021) doi:10.1007/s00439-021-02272-5.

44. Kim, J.-S. *et al.* Independent and sequential recruitment of NHEJ and HR factors to DNA damage sites in mammalian cells. *J Cell Biol* **170**, 341–347 (2005).

45. Williams, R. S. *et al.* Nbs1 Flexibly Tethers Ctp1 and Mre11-Rad50 to Coordinate DNA Double-Strand Break Processing and Repair. *Cell* **139**, 87–99 (2009).

46. de Jager, M. *et al.* Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol Cell* **8**, 1129–1135 (2001).

47. Gupta, A. *et al.* Role of 53BP1 in the regulation of DNA double-strand break repair pathway choice. *Radiat Res* **181**, 1–8 (2014).

48. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* **5**, (2013).

49. Saha, J. & Davis, A. J. Unsolved mystery: The role of BRCA1 in DNA end-joining. *J Radiat Res* **57**, i18–i24 (2016).

50. Saha, J. & Davis, A. J. Unsolved mystery: The role of BRCA1 in DNA end-joining. *J Radiat Res* **57**, i18–i24 (2016).

51. Liu, Y. & Lu, L. Y. BRCA1 and homologous recombination: Implications from mouse embryonic development. *Cell Biosci* **10**, 1–10 (2020).

52. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (1979)* **337**, (2012).

53. Ishino, Y., Krupovic, M. & Forterre, P. History of CRISPR-Cas from Encounter with a Mysterious. *J Bacteriol* **200**, e00580-17 (2018).

54. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* E2579–E2586 (2012) doi:10.1073/pnas.1208507109.

55. Meaker, G. A., Hair, E. J. & Gorochowski, T. E. Advances in engineering CRISPR-Cas9 as a molecular Swiss Army knife. *Synth Biol* **5**, ysaa021 (2020).

56. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell* **154**, (2013).

57. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science (1979)* **353**, (2016).

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

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

60. Anzalone, A. v. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

61. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–4 (2016).

62. Scholefield, J. & Harrison, P. T. Prime editing – an update on the field. *Gene Therapy 2021 28:7* **28**, 396–401 (2021).

63. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. (2013) doi:10.1038/nprot.2013.132.

64. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442 (2013).

65. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510–519 (2015).

66. Kearns, N. A. *et al.* Functional annotation of native enhancers with a cas9-histone demethylase fusion. *Nat Methods* **12**, (2015).

67. Nuñez, J. K. *et al.* Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* (2021) doi:10.1016/j.cell.2021.03.025.

68. Hampson, D. R., Hooper, A. W. M. & Niibori, Y. The Application of Adeno-Associated Viral Vector Gene Therapy to the Treatment of Fragile X Syndrome. *Brain Sci* **9**, 32 (2019).

69. Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science (1979)* **358**, 1019–1027 (2017).

70. Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature Bio* **34**, 528–530 (2016).

71. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).

72.   Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems HHS Public Access. *Science (1979)* **339**, 819–823 (2013).

73.   Huai, C. *et al.* CRISPR/Cas9-mediated somatic and germline gene correction to restore hemostasis in hemophilia B mice. *Hum Genet* **136**, (2017).

74.   Batty, P. & Lillicrap, D. Hemophilia Gene Therapy: Approaching the First Licensed Product. *Hemasphere* (2021) doi:10.1097/HS9.0000000000000540.

75.   Prakash, V., Moore, M. & Yáñez-Muñoz, R. J. Current Progress in Therapeutic Gene Editing for Monogenic Diseases. *Molecular Therapy* **24**, (2016).

76.   Chang, K. H. *et al.* Long-Term Engraftment and Fetal Globin Induction upon BCL11A Gene Editing in Bone-Marrow-Derived CD34+ Hematopoietic Stem and Progenitor Cells. *Mol Ther Methods Clin Dev* **4**, 137–148 (2017).

77.   Zipkin, M. CRISPR's "magnificent moment" in the clinic. *Nat Biotechnol* (2019) doi:10.1038/d41587-019-00035-2.

78.   Hoban, M. D. *et al.* Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells Key Points. *Blood* **125**, 2597–2604 (2015).

79.   Sebastiano, V. *et al.* In Situ Genetic Correction of the Sickle Cell Anemia Mutation in Human Induced Pluripotent Stem Cells Using Engineered Zinc Finger Nucleases. *Stem Cells* **29**, 1717–1726 (2011).

80.   Sun, N. & Zhao, H. Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnol Bioeng* **111**, 1048–1053 (2014).

81.   Voit, R. A., Hendel, A., Pruett-Miller, S. M. & Porteus, M. H. Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res* **42**, 1365–1378 (2014).

82.   Intondi, A. Gene Correction in Autologous CD34+ Hematopoietic Stem Cells (HbS to HbA) to Treat Severe Sickle Cell Disease - Full Text View - ClinicalTrials.gov. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT04819841.

83.   Locatelli, F. *et al.* Betibeglogene Autotemcel Gene Therapy for Non–β 0 /β 0 Genotype β-Thalassemia . *New England Journal of Medicine* **386**, 415–427 (2022).

84.   Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* **118**, 3143–3150 (2008).

85.   Pike-Overzet, K., van der Burg, M., Wagemaker, G., van Dongen, J. J. M. & Staal, F. J. T. New insights and unresolved issues regarding insertional mutagenesis in X-linked SCID gene therapy. *Molecular Therapy* **15**, 1910–1916 (2007).

86.   Ferrua, F. & Aiuti, A. Twenty-Five Years of Gene Therapy for ADA-SCID: From Bubble Babies to an Approved Drug. doi:10.1089/hum.2017.175.

87.   Abstract Details | ASGCT 25th Annual Meeting. https://annualmeeting.asgct.org/abstracts/abstract-details?abstractId=6370.

88.   Joseph Melenhorst, J. *et al.* Decade-long leukaemia remissions with persistence of CD4 + CAR T cells. *Nature* **602**, 503 (2022).

89.   CAR T Cells: Engineering Immune Cells to Treat Cancer - NCI.
        https://www.cancer.gov/about-cancer/treatment/research/car-t-cells.

90.   Cai, X., Conley, S. M. & Naash, M. I. RPE65: Role in the visual cycle, human retinal
        disease, and gene therapy. *Ophthalmic Genet* **30**, 57 (2009).

91.   FDA Approves Spark Therapeutics' LUXTURNA™ (voretigene neparvovec-rzyl), a One-
        time Gene Therapy for Patients with Confirmed Biallelic RPE65 Mutation-associated
        Retinal Dystrophy – Spark Therapeutics. https://sparktx.com/press_releases/fda-approves-
        spark-therapeutics-luxturna-voretigene-neparvovec-rzyl-a-one-time-gene-therapy-for-
        patients-with-confirmed-biallelic-rpe65-mutation-associated-retinal-dystrophy/.

92.   Ramirez, V. B. First Human CRISPR Trial in the US Aims to Cure Inherited Blindness.
        Singularity Hub https://singularityhub.com/2019/07/28/first-human-crispr-trial-in-the-us-
        aims-to-cure-inherited-blindness/ (2019).

93.   Albright, C. Discovery of EDIT-101 for the Treatment of Leber's Congenital Amaurosis
        Type 10. *Editas Medicine* Preprint at (2018).

94.   Crowe, K. Canadian breakthrough that became the world's most expensive drug, then
        vanished, gets second chance | CBC News. *CBC News*
        https://www.cbc.ca/news/health/glybera-lpld-rare-drug-orphan-disease-nrc-cbc-price-
        1.5312177 (2019).

95.   Crowe, K. The million-dollar drug | CBC News. *CBC News*
        https://newsinteractives.cbc.ca/longform/glybera (2018).

96.   McConaghie, A. Glybera, the most expensive drug in the world, to be withdrawn after
        commercial flop -. *PharmaPhorum* https://pharmaphorum.com/news/glybera-expensive-
        drug-world-withdrawn-commercial-flop/ (2017).

97.   Chase, L. 10 Most Expensive Drugs in the U.S., Period . *GoodRx*
        https://www.goodrx.com/blog/most-expensive-drugs-period/ (2020).

98.   Stoltz, D. A., Meyerholz, D. K. & Welsh, M. J. Origins of cystic fibrosis lung disease. *N
        Engl J Med* **372**, 351–362 (2015).

99.   Lee, J. A. *et al.* Gene therapy for cystic fibrosis: new tools for precision medicine. *Journal
        of Translational Medicine 2021 19:1* **19**, 1–15 (2021).

100.   MRT5005 | CFF Clinical Trials Tool.
        https://apps.cff.org/trials/pipeline/details/10157/MRT5005.

101.   Geurts, M. H. *et al.* Evaluating CRISPR-based prime editing for cancer modeling and
        CFTR repair in organoids. (2021) doi:10.26508/lsa.202000940.

102.   The Huntington's Disease Collaborative Research Group. A novel gene containing a
        trinucleotide repeat that is expanded and unstable in Huntington's disease chromosomes.
        Cell **72**, 971–983 (1993).

103.   Walker, F. O. Huntington's disease. *The Lancet* **369**, 218–228 (2007).

104.   Kadakkuzha, Y.-B. M., Liu, B. M., Akhmedov, X.-A. & Kandel, K. R. Huntingtin Is
        Critical Both Pre-and Postsynaptically for Long-Term Learning-Related Synaptic Plasticity
        in Aplysia. *PLoS One* **9**, 103004 (2014).

105. Vitet, H., Brandt, V. & Saudou, F. Traffic signaling: new functions of huntingtin and axonal transport in neurological disease. *Curr Opin Neurobiol* **63**, 122–130 (2020).

106. Kwon, D. Failure of genetic therapies for Huntington's devastates community. *Nature* **593**, 180 (2021).

107. uniQure. Huntington's Disease - Gene Therapy. https://www.uniqure.com/gene-therapy/huntingtons-disease.php.

108. uniQure. Phase I/II Clinical Trial of AMT-130 - Patients. https://www.uniqure.com/patients/phase-1-2-clinical-trial-of-amt-130.php (2021).

109. Lai, M. *et al.* CRISPR/Cas9 Ablation of Integrated HIV-1 Accumulates Proviral DNA Circles with Reformed Long Terminal Repeats. *J Virol* **95**, (2021).

110. Herskovitz, J. *et al.* CRISPR-Cas9 Mediated Exonic Disruption for HIV-1 Elimination. *EBioMedicine* **73**, 103678 (2021).

111. Powell, M. K. *et al.* Opportunistic Infections in HIV-Infected Patients Differ Strongly in Frequencies and Spectra between Patients with Low CD4+ Cell Counts Examined Postmortem and Compensated Patients Examined Antemortem Irrespective of the HAART Era. *PLoS One* **11**, (2016).

112. Rodger, A. J. *et al.* Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. *JAMA* **316**, 171–181 (2016).

113. Bavinton, B. R. *et al.* Viral suppression and HIV transmission in serodiscordant male couples: an international, prospective, observational, cohort study. *Lancet HIV* **5**, e438–e447 (2018).

114. Eisinger, R. W., Dieffenbach, C. W. & Fauci, A. S. HIV viral load and transmissibility of HIV infection undetectable equals untransmittable. *JAMA - Journal of the American Medical Association* **321**, 451–452 (2019).

115. Srinivasan, M., Raja Thangaraj, S. & Arzoun, H. Gene Therapy-Can it Cure Type 1 Diabetes? (2021) doi:10.7759/cureus.20516.

116. Sonksen, P. & Sonksen, J. Insulin: understanding its action in health and disease. *Br J Anaesth* **85**, 69–79 (2000).

117. Akil, A. A. S. *et al.* Diagnosis and treatment of type 1 diabetes at the dawn of the personalized medicine era. *Journal of Translational Medicine 2021 19:1* **19**, 1–19 (2021).

118. Xia, F. *et al.* Reg3g overexpression promotes β cell regeneration and induces immune tolerance in nonobese-diabetic mouse model. *J Leukoc Biol* **99**, 1131–1140 (2016).

119. Mallol, C. *et al.* AAV-mediated pancreatic overexpression of Igf1 counteracts progression to autoimmune diabetes in mice. *Mol Metab* **6**, 664–680 (2017).

120. Rao, P., Cozar-Castellano, I., Roccisana, J., Vasavada, R. C. & Garcia-Ocaña, A. Hepatocyte growth factor gene therapy for islet transplantation. *Expert Opin Biol Ther* **4**, 507–518 (2005).

121. Lin, Y. & Sun, Z. Antiaging gene klotho attenuates pancreatic β-cell apoptosis in type 1 diabetes. *Diabetes* **64**, 4298–4311 (2015).

122. Li, D. *et al.* Transplantation of Aire-overexpressing bone marrow-derived dendritic cells delays the onset of type 1 diabetes. *Int Immunopharmacol* **49**, 13–20 (2017).

123. Finckh, A. & Gabay, C. At the horizon of innovative therapy in rheumatology: new biologic agents. *Curr Opin Rheumatol* **20**, 269–275 (2008).

124. Banga, A., Greder, L. v, Dutton, J. R. & Slack, J. Stable insulin-secreting ducts formed by reprogramming of cells in the liver using a three-gene cocktail and a PPAR agonist. *Gene Ther* **21**, 19–27 (2014).

125. Cheung, A. T. *et al.* Glucose-dependent insulin release from genetically engineered K cells. *Science* **290**, 1959–1962 (2000).

126. Chellappan, D. K. *et al.* Gene therapy and type 1 diabetes mellitus. *Biomed Pharmacother* **108**, 1188–1200 (2018).

127. Tóthová, Ľ. *et al.* Phage therapy of Cronobacter-induced urinary tract infection in mice. *Med Sci Monit* **17**, BR173 (2011).

128. Zalewska-Piatek, B. & Piatek, R. Phage Therapy as a Novel Strategy in the Treatment of Urinary Tract Infections Caused by E. Coli. *Antibiotics* **304**, (2020).

129. Recurrent Urinary Tract Infections in Women: Diagnosis and Management. https://www.aafp.org/pubs/afp/issues/2010/0915/p638.html.

130. Dhooge, P. P. A., Valkenburg, D. & Hoyng, C. B. Gene therapy for inherited retinal diseases. in *Genetics and Genomics of Eye Disease* (ed. Gao, X. R.) 279–295 (2020). doi:10.1016/B978-0-12-816222-4.00017-4.

131. Rodrigues, G. A. *et al.* Pharmaceutical Development of AAV-Based Gene Therapy Products for the Eye. *Pharm Res* **36**, (2019).

132. Conboy, I., Murthy, N., Etienne, J. & Robinson, Z. Making gene editing a therapeutic reality [version 1; referees: 2 approved]. *F1000Res* **7**, 1–10 (2018).

133. Lee, M. & Kim, H. Therapeutic application of the CRISPR system: current issues and new prospects. *Hum Genet* **138**, 563–590 (2019).

134. Alapati, D. *et al.* In utero gene editing for monogenic lung disease. *Sci Transl Med* **11**, (2019).

135. Patel, A., Zhao, J., Duan, D. & Lai, Y. Design of AAV vectors for delivery of large or multiple transgenes. in *Methods in Molecular Biology* (ed. Castle, M. J.) vol. 1950 19–33 (Springer Nature, 2019).

136. Sanjurjo-Soriano, C. & Kalatzis, V. Guiding Lights in Genome Editing for Inherited Retinal Disorders: Implications for Gene and Cell Therapy. *Neural Plast* **2018**, 5056279 (2018).

137. Xue, K., Groppe, M., Salvetti, A. P. & MacLaren, R. E. Technique of retinal gene therapy: Delivery of viral vector into the subretinal space. *Eye (Basingstoke)* **31**, 1308–1316 (2017).

138. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov* **18**, 358–378 (2019).

139. Sandoval, I. M., Collier, T. J. & Manfredsson, F. P. Design and assembly of CRISPR/Cas9 lentiviral and rAAV vectors for targeted genome editing. in *Methods in Molecular Biology* (eds. Manfredsson, F. P. & Benskey, M. J.) vol. 1937 29–45 (Springer Nature, 2019).

140. Nelson, H. H., Sweetser, D. B. & Nickoloff, J. A. Effects of terminal nonhomology and homeology on double-strand-break-induced gene conversion tract directionality. *Mol Cell Biol* **16**, 2951–2957 (1996).

141. Li, S., Wehrenberg, B., Waldman, B. C. & Waldman, A. S. Mismatch tolerance during homologous recombination in mammalian cells. *DNA Repair (Amst)* **70**, 25–36 (2018).

142. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat Biotechnol* **32**, (2014).

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

144. Hewes, A. M., Sansbury, B. M., Barth, S., Tarcic, G. & Kmiec, E. B. gRNA Sequence Heterology Tolerance Catalyzed by CRISPR/Cas in an In Vitro Homology-Directed Repair Reaction. *Mol Ther Nucleic Acids* **20**, (2020).

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

146. Attwood, K. M. *et al.* PML isoform expression and DNA break location relative to PML nuclear bodies impacts the efficiency of homologous recombination. *Biochemistry and Cell Biology* **98**, 314–326 (2020).

147. Kuwayama, H., Yanagida, T. & Ueda, M. DNA oligonucleotide-assisted genetic manipulation increases transformation and homologous recombination efficiencies: Evidence from gene targeting of Dictyostelium discoideum. *J Biotechnol* **133**, 418–423 (2008).

148. Kratzer, K., Getz, L. J., Peterlini, T., Masson, J. Y. & Dellaire, G. Addressing the dark matter of gene therapy: technical and ethical barriers to clinical application. *Human Genetics* Preprint at https://doi.org/10.1007/s00439-021-02272-5 (2021).

149. Salsman, J., Masson, J.-Y., Orthwein, A. & Dellaire, G. CRISPR/Cas9 Gene Editing: From Basic Mechanisms to Improved Strategies for Enhanced Genome Engineering In Vivo. *Curr Gene Ther* **17**, (2017).

150. Orthwein, A. *et al.* A mechanism for the suppression of homologous recombination in G1 cells. *Nature* **528**, 422–426 (2015).

151. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res* **12**, 996–1006 (2002).

152. Pellenz, S. *et al.* New Human Chromosomal Sites with 'Safe Harbor' Potential for Targeted Transgene Insertion. *Hum Gene Ther* **30**, 814–828 (2019).

153. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, 941–947 (2018).

154. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).

155. Moran, J. v. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).

156. Lindič, N. *et al.* Differential inhibition of LINE1 and LINE2 retrotransposition by vertebrate AID/APOBEC proteins. *Retrovirology* **10**, 156 (2013).

157. Shan, S., Soltis, P. S., Soltis, D. E. & Yang, B. Considerations in adapting CRISPR/Cas9 in nongenetic model plant systems. *Appl Plant Sci* **8**, (2020).

158. Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol* **37**, 64–82 (2019).

159. Zhang, C. *et al.* Creation of targeted inversion mutations in plants using an RNA-guided endonuclease. *Crop Journal* **5**, 83–88 (2017).

160. Binda, C. S., Klaver, B., Berkhout, B. & Das, A. T. CRISPR-Cas9 Dual-gRNA Attack Causes Mutation, Excision and Inversion of the HIV-1 Proviral DNA. *Viruses* **12**, (2020).

161. Brunet, E. & Jasin, M. Induction of Chromosomal Translocations with CRISPR-Cas9 and Other Nucleases: Understanding the Repair Mechanisms That Give Rise to Translocations. *Adv Exp Med Biol* **1044**, 15–25 (2018).

162. Jeong, J. *et al.* High-efficiency CRISPR induction of t(9;11) chromosomal translocations and acute leukemias in human blood stem cells. *Blood Adv* **3**, 2825–2835 (2019).

163. Zuccaro, M. v. *et al.* Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell* **183**, 1650-1664.e15 (2020).

164. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat Genet* **53**, 895–905 (2021).

165. Hong, S. G. *et al.* Rhesus iPSC Safe Harbor Gene-Editing Platform for Stable Expression of Transgenes in Differentiated Cells of All Germ Layers. *Molecular Therapy* **25**, 44–53 (2017).

166. Ward, P. & Walsh, C. E. Targeted integration of a rAAV vector into the AAVS1 region. *Virology* **433**, (2012).

167. Irion, S. *et al.* Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol* **25**, 1477–1482 (2007).

168. Acharya, P., Lusvarghi, S., Bewley, C. A. & Kwong, P. D. HIV-1 gp120 as a therapeutic target: navigating a moving labyrinth. *Expert Opin Ther Targets* **19**, (2015).

169. Oliveira, C. E. C. de *et al.* CC Chemokine Receptor 5: The Interface of Host Immunity and Cancer. *Dis Markers* **2014**, (2014).

170. Gu, W.-G. & Chen, X.-Q. Targeting CCR5 for anti-HIV research. *European Journal of Clinical Microbiology & Infectious Diseases* **33**, (2014).

171. Li, L. *et al.* Genomic Editing of the HIV-1 Coreceptor CCR5 in Adult Hematopoietic Stem and Progenitor Cells Using Zinc Finger Nucleases. *Molecular Therapy* **21**, (2013).

172. Styer, K. L., Click, E. M., Hopkins, G. W., Frothingham, R. & Aballay, A. Study of the role of CCR5 in a mouse model of intranasal challenge with Yersinia pestis. *Microbes Infect* **9**, (2007).

173. Cohen, J. Did CRISPR help—or harm—the first-ever gene-edited babies? *Science (1979)* (2019) doi:10.1126/science.aay9569.

174. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *BIOINFORMATICS APPLICATIONS NOTE* **26**, 841–842 (2010).

175.	Radford, F., Elliott, S. D., Schepartz, A., Farren, & & Isaacs, J. Targeted editing and evolution of engineered ribosomes in vivo by filtered editing. doi:10.1038/s41467-021-27836-x.

176.	Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Web Server issue Published online* **46**, (2018).

177.	Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 1–12 (2016).

178.	Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**, 1262–1267 (2014).

179.	Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191 (2016).

180.	Emery, D. W. Methods Review The Use of Chromatin Insulators to Improve the Expression and Safety of Integrating Gene Transfer Vectors. doi:10.1089/hum.2010.233.

181.	Chen, E., Lin-Shiao, E., Doost, M. S. & Doudna, J. A. Decorating chromatin for enhanced genome editing using CRISPR-Cas9. *bioRxiv* 2022.03.15.484540 (2022) doi:10.1101/2022.03.15.484540.

182.	Loeb, L. A. & Monnat, R. J. DNA polymerases and human disease. *Nature Reviews Genetics* vol. 9 594–604 Preprint at https://doi.org/10.1038/nrg2345 (2008).

183.	Barnes, R. & Eckert, K. Maintenance of genome integrity: How mammalian cells orchestrate genome duplication by coordinating replicative and specialized DNA polymerases. *Genes (Basel)* **8**, (2017).

184.	Mcllwraith, M. J. *et al.* Human DNA polymerase η promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol Cell* **20**, 783–792 (2005).

185.	Roerink, S. F., Koole, W., Stapel, L. C., Romeijn, R. J. & Tijsterman, M. A Broad Requirement for TLS Polymerases η and κ, and Interacting Sumoylation and Nuclear Pore Proteins, in Lesion Bypass during C. elegans Embryogenesis. *PLoS Genet* **8**, e1002800 (2012).

186.	Tissier, A. *et al.* Co-localization in replication foci and interaction of human Y-family members, DNA polymerase polη and REVl protein. *DNA Repair (Amst)* **3**, 1503–1514 (2004).

187.	Wang, Y. *et al.* Evidence that in xeroderma pigmentosum variant cells, which lack DNA polymerase η, DNA polymerase ι causes the very high frequency and unique spectrum of UV-induced mutations. *Cancer Res* **67**, 3018–3026 (2007).

188.	Ziv, O., Geacintov, N., Nakajima, S., Yasui, A. & Livneh, Z. DNA polymerase ζ cooperates with polymerases κ and ι in translesion DNA synthesis across pyrimidine photodimers in cells from XPV patients. *Proc Natl Acad Sci U S A* **106**, 11552–11557 (2009).

189.	Yoon, J. H. *et al.* Rev1 promotes replication through UV lesions in conjunction with DNA polymerases η, ι, and κ but not DNA polymerase ζ. *Genes Dev* **29**, 2588–2602 (2015).

190.	Jung, Y. S., Qian, Y. & Chen, X. DNA polymerase eta is targeted by Mdm2 for polyubiquitination and proteasomal degradation in response to ultraviolet irradiation. *DNA Repair (Amst)* **11**, 177–184 (2012).

191. Li, J., Holzschu, D. L. & Sugiyama, T. PCNA is efficiently loaded on the DNA recombination intermediate to modulate polymerase δ, η, and ζ activities. *Proc Natl Acad Sci U S A* **110**, 7672–7677 (2013).

192. Kannouche, P. L., Wing, J. & Lehmann, A. R. Interaction of human DNA polymerase η with monoubiquitinated PCNA: A possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell* **14**, 491–500 (2004).

193. Fu, D. *et al.* Cell Cycle Recruitment of DNA polymerase eta by FANCD2 in the early response to DNA damage) Recruitment of DNA polymerase eta by FANCD2 in the early response to DNA damage. *www.landesbioscience.com Cell Cycle 803 Cell Cycle* **12**, 803–809 (2013).

194. Cipolla, L. *et al.* UBR5 interacts with the replication fork and protects DNA replication from DNA polymerase η toxicity. *Nucleic Acids Res* **47**, 11268–11283 (2019).

195. Bertoletti, F. *et al.* Phosphorylation regulates human polη stability and damage bypass throughout the cell cycle. *Nucleic Acids Res* **45**, 9441–9454 (2017).

196. Jung, Y.-S., Qian, Y. & Chen, X. DNA polymerase eta is targeted by Mdm2 for polyubiquitination and proteasomal degradation in response to ultraviolet irradiation. *DNA Repair (Amst)* **11**, 177–184 (2012).

197. Buisson, R. *et al.* Breast cancer proteins PALB2 and BRCA2 stimulate polymerase η in recombination-associated DNA Synthesis At Blocked Replication Forks. *Cell Rep* **6**, 553–564 (2014).

198. Thakur, M. *et al.* DNA polymerase η undergoes alternative splicing, protects against UV sensitivity and apoptosis, and suppresses Mre11-dependent recombination. *Genes Chromosomes Cancer* **32**, 222–235 (2001).

199. Limoli, C. L., Giedzinski, E. & Cleaver, J. E. Alternative recombination pathways in UV-irradiated XP variant cells. *Oncogene* **24**, 3708–3714 (2005).

200. Cleaver, J. E. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218**, 652–656 (1968).

201. Cleaver, J. E. *et al.* Increased Ultraviolet Sensitivity and Chromosomal Instability Related to P53 Function in the Xeroderma Pigmentosum Variant. *Cancer Res* **59**, 1102–1108 (1999).

202. Patton, J. D. *et al.* Xeroderma Pigmentosum Fibroblasts Including Cells From Xp Variants Are Abnormally Sensitive To the Mutagenic and Cytotoxic Action of Broad Spectrum Simulated Sunlight. *Photochem Photobiol* **39**, 37–42 (1984).

203. Xeroderma Pigmentosum - NORD (National Organization for Rare Disorders). https://rarediseases.org/rare-diseases/xeroderma-pigmentosum/.

204. Broughton, B. C. *et al.* Molecular analysis of mutations in DNA polymerase η in xeroderma pigmentosum-variant patients. *Proc Natl Acad Sci U S A* **99**, 815–820 (2002).

205. Masutani, C. *et al.* The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature* **399**, 700–704 (1999).

206. Johnson, R. E., Kondratick, C. M., Prakash, S. & Prakash, L. hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science (1979)* **285**, 263–265 (1999).

207. Zamani, G. Y., Khan, R., Karim, N., Ahmed, Z. M. & Naeem, M. Identification of Frameshift Variants in POLH Gene Causing Xeroderma Pigmentosum in Two Consanguineous Pakistani Families. *Genes (Basel)* **13**, (2022).

208. Kannouche, P. *et al.* Domain structure, localization, and function of DNA polymerase η, defective in xeroderma pigmentosum variant cells. *Genes Dev* **15**, 158–172 (2001).

209. Cordonnier, A. M., Lehmann, A. R. & Fuchs, R. P. P. Impaired Translesion Synthesis in Xeroderma Pigmentosum Variant Extracts. *Mol Cell Biol* **19**, 2206–2211 (1999).

210. Kawamoto, T. *et al.* Dual roles for DNA polymerase η in homologous DNA recombination and translesion DNA synthesis. *Mol Cell* **20**, 793–799 (2005).

211. Despras, E., Daboussi, F., Hyrien, O., Marheineke, K. & Kannouche, P. L. ATR/Chk1 pathway is essential for resumption of DNA synthesis and cell survival in UV-irradiated XP variant cells. doi:10.1093/hmg/ddq046.

212. Itoh, T. *et al.* Xeroderma Pigmentosum Variant Heterozygotes Show Reduced Levels of Recovery of Replicative DNA Synthesis in the Presence of Caffeine after Ultraviolet Irradiation. *Journal of Investigative Dermatology* **115**, 981–985 (2000).

213. Cleaver, J. E. Caffeine toxicity is inversely related to DNA repair in simian virus 40-transformed xeroderma pigmentosum cells irradiated with ultraviolet light. *Teratog Carcinog Mutagen* **9**, 147–155 (1989).

214. Guo, J., Zhou, G., Zhang, W., Song, Y. & Bian, Z. A novel POLH mutation causes XP-V disease and XP-V tumor proneness may involve imbalance of numerous DNA polymerases. *Oncol Lett* **6**, 1583–1590 (2013).

215. Laporte, G. A. *et al.* The role of double-strand break repair, translesion synthesis, and interstrand crosslinks in colorectal cancer progression—clinicopathological data and survival. *J Surg Oncol* (2020) doi:10.1002/jso.25737.

216. Dasari, S. & Bernard Tchounwou, P. Cisplatin in cancer therapy: Molecular mechanisms of action. *Eur J Pharmacol* **740**, 364–378 (2014).

217. Enoiu, M., Jiricny, J. & Schärer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res* **40**, 8953–8964 (2012).

218. Mentegari, E. *et al.* Ribonucleotide incorporation by human DNA polymerase η impacts translesion synthesis and RNase H2 activity. *Nucleic Acids Res* **45**, 2600–2614 (2017).

219. Gomes, L. R. *et al.* ATR mediates cisplatin resistance in 3D-cultured breast cancer cells via translesion DNA synthesis modulation. *Cell Death Dis* **10**, (2019).

220. Zhang, J. *et al.* A POLH transcript with a short 30UTR enhances POLH expression and mediates cisplatin resistance. *Cancer Res* **79**, 3714–3724 (2019).

221. Flanagan, A. M. *et al.* The human POLH gene is not mutated, and is expressed in a cohort of patients with basal or squamous cell carcinoma of the skin. *Int J Mol Med* **19**, 589–596 (2007).

222. Twayana, S. *et al.* Translesion polymerase eta both facilitates DNA replication and promotes increased human genetic variation at common fragile sites. **118**, (2021).

223. Garcia-Exposito, L. *et al.* Proteomic Profiling Reveals a Specific Role for Translesion DNA Polymerase η in the Alternative Lengthening of Telomeres. *Cell Rep* **17**, 1858–1871 (2016).

224. Nabetani, A. & Ishikawa, F. Alternative lengthening of telomeres pathway: Recombination-mediated telomere maintenance mechanism in human cells. *J Biochem* **149**, 5–14 (2011).

225. Supek, F. & Lehner, B. Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. *Cell* **170**, 534-547.e23 (2017).

226. Cullen, B. R. Role and Mechanism of Action of the APOBEC3 Family of Antiretroviral Resistance Factors. *J Virol* **80**, 1067–1076 (2006).

227. Refsland, E. W., Harris, R. S. & Cullen, B. R. The APOBEC3 Family of Retroelement Restriction Factors. *Curr Top Microbiol Immunol* **371**,.

228. Siriwardena, S., Chen, K. & Bhagwat, A. S. The Functions and Malfunctions of AID/APOBEC Family Deaminases: the known knowns and the known unknowns. *Chem Rev* **116**, 12688–12710 (2016).

229. Roberts, S. A. *et al.* An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* (2013) doi:10.1038/ng.2702.

230. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov* **5**, 704 (2015).

231. Rebhandl, S., Huemer, M., Greil, R. & Geisberger, R. AID/APOBEC deaminases and cancer. *Oncoscience* **2**, 320–333 (2015).

232. Chen, J. & Furano, A. v. Breaking bad: The mutagenic effect of DNA repair. *DNA Repair (Amst)* (2015) doi:10.1016/j.dnarep.2015.04.012.

233. Xu, W. K., Byun, H. & Dudley, J. P. The Role of APOBECs in Viral Replication. *Microorganisms* **8**, 1–46 (2020).

234. Orecchini, E., Frassinelli, L., Galardi, S., Ciafrè, S. & Michienzi, A. Post-transcriptional regulation of LINE-1 retrotransposition by AID/APOBEC and ADAR deaminases. *Chromosome Res* **26**, 45–59 (2018).

235. Halemano, K. *et al.* Immunoglobulin somatic hypermutation by APOBEC3/Rfv3 during retroviral infection. *Proc Natl Acad Sci U S A* **111**, 7759–7764 (2014).

236. Feng, Y., Goubran, M. H., Follack, T. B. & Chelico, L. Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. *Scientific Reports 2017 7:1* **7**, 1–11 (2017).

237. Teng, G. & Papavasiliou, F. N. Immunoglobulin Somatic Hypermutation. (2007) doi:10.1146/annurev.genet.41.110306.130340.

238. Cervantes-Ayalc, A., Ruiz Esparza-Garrido, R. & Velázquez-Flores, M. Á. Long Interspersed Nuclear Elements 1 (LINE1): The chimeric transcript L1-MET and its involvement in cancer. *Cancer Genetics* Preprint at https://doi.org/10.1016/j.cancergen.2019.11.004 (2020).

239. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280–5285 (2003).

240. Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 8780–8785 (2006).

241. Mita, P. *et al.* BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat Struct Mol Biol* **27**, 179–191 (2020).

242. McKerrow, W. *et al.* LINE-1 expression in cancer correlates with DNA damage response, copy number variation, and cell cycle progression. *bioRxiv* (2020) doi:10.1101/2020.06.26.174052.

243. Scott, E. C. & Devine, S. E. The role of somatic L1 retrotransposition in human cancers. *Viruses* **9**, (2017).

244. Iwagami, S. *et al.* LINE-1 hypomethylation is associated with a poor prognosis among patients with curatively resected esophageal squamous cell carcinoma. *Ann Surg* **257**, 449–455 (2013).

245. Liang, W. *et al.* APOBEC3DE Inhibits LINE-1 Retrotransposition by Interacting with ORF1p and Influencing LINE Reverse Transcriptase Activity. *PLoS One* **11**, (2016).

246. Bogerd, H. P., Wiegand, H. L., Doehle, B. P., Lueders, K. K. & Cullen, B. R. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res* **34**, 89–95 (2006).

247. Muckenfuss, H. *et al.* APOBEC3 Proteins Inhibit Human LINE-1 Retrotransposition. *Journal of Biological Chemistry* **281**, 22161–22172 (2006).

248. Arias, J. F., Koyama, T., Kinomoto, M. & Tokunaga, K. Retroelements versus APOBEC3 family members: No great escape from the magnificent seven. *Front Microbiol* **3**, 275 (2012).

249. Kinomoto, M. *et al.* All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res* **35**, 2955 (2007).

250. Richardson, S. R., Narvaiza, I., Planegger, R. A., Weitzman, M. D. & Moran, J. v. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* **3**, (2014).

251. Koito, A. & Ikeda, T. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Front Microbiol* **4**, 28 (2013).

252. Hulme, A. E., Bogerd, H. P., Cullen, B. R. & Moran, J. v. Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* **390**, 199–205 (2007).

253. Tiwari, B. *et al.* P53 directly represses human LINE1 transposons. *Genes Dev* (2020) doi:10.1101/gad.343186.120.

254. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, 1–12 (2002).

255. Nygard, A. B., Jørgensen, C. B., Cirera, S. & Fredholm, M. Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. *BMC Mol Biol* **8**, 1–6 (2007).

256. Barnes, R. P., Tsao, W. C., Moldovan, G. L. & Eckert, K. A. DNA polymerase eta prevents tumor cell-cycle arrest and cell death during recovery from replication stress. *Cancer Res* **78**, 6549–6560 (2018).

257. Chung, I., Osterwald, S., Deeg, K. I. & Rippe, K. PML body meets telomere: The beginning of an ALTernate ending? *Nucleus (United States)* vol. 3 Preprint at https://doi.org/10.4161/nucl.20326 (2012).

258. Bilaud, T. *et al.* Telomeric localization of TRF2, a novel human telobox protein. *Nat Genet* **17**, 239 (1997).

259. Roos, W. P. & Kaina, B. DNA damage-induced cell death by apoptosis. *Trends Mol Med* **12**, 440–450 (2006).

260. Pezzoli, D., Giupponi, E., Mantovani, D. & Candiani, G. Size matters for in vitro gene delivery: investigating the relationships among complexation protocol, transfection medium, size and sedimentation. *Nature Publishing Group* (2017) doi:10.1038/srep44134.

261. Lesueur, L. L., Mir, L. M. & André, F. M. Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. *Mol Ther Nucleic Acids* **5**, e291 (2016).

262. Pannunzio, N. R., Watanabe, G. & Lieber, M. R. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry* **293**, 10512–10523 (2018).

263. Zhao, Y. *et al.* Structural basis of human DNA polymerase η-mediated chemoresistance to cisplatin. *Proc Natl Acad Sci U S A* **109**, 7269–7274 (2012).

264. Bassett, E. *et al.* The Role of DNA Polymerase η in Translesion Synthesis Past Platinum-DNA Adducts in Human Fibroblasts. *Cancer Res* **64**, 6469–6475 (2004).

265. Stenglein, M. D. & Harris, R. S. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* **281**, 16837–16841 (2006).

266. Mcllwraith, M. J. *et al.* Human DNA polymerase η promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol Cell* **20**, 783–792 (2005).

267. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192 (2012).

268. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

APPENDIX A: COPYRIGHT PERMISSIONS

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 26, 2022

_____

_____

This Agreement between Kateryna Kratzer ("You") and Springer Nature ("Springer Nature")consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number   5296550447018

License Date   Apr 26, 2022

Licensed Content Publisher   Springer Nature

Licensed Content Publication   Human Genetics

Licensed Content Title   Addressing the dark matter of gene therapy: technical and ethical barriers to clinical application

Licensed Content Author   Kateryna Kratzer et al

Licensed Content Date   Apr 8, 2021

Type of Use   Thesis/Dissertation

Requestor type   academic/university or research institute

Format print and electronic

Portion   full article/chapter

Will you be translating?   no

Circulation/distribution	1000 - 1999
Author of this Springer Nature content	yes
Title	MSc student
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Requester Location	Kateryna Kratzer, 3501 Dutch Village Rd Halifax, NS R3N2S8 Canada
Total	0.00 CAD

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between youand **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1.1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1.2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the materialyou are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1.3. If the credit line on any part of the material you have requested indicates that itwas reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

123

2. Scope of Licence

2.1. You may only use the Licensed Content in the manner and to the extent permittedby these Ts&Cs and any applicable laws.

2.2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid inthe language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2.3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com forthese rights.

2.4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided thatthe material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2.5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3.1. A licence for is valid from the date of purchase ('Licence Date') at the end of therelevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4.1. The Licensor's permission must be acknowledged next to the Licenced Material inprint. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5.1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Anyother changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5.2. You must not use any Licensed Material as part of any design or trademark.

5.3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAPsites prior to final publication.

6. Ownership of Rights

6.1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8.1. ___BOOKS ONLY:___ Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity,NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8.2. For content reuse requests that qualify for permission under the STM PermissionsGuidelines, which may be updated from time to time, the STM Permissions Guidelinessupersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9.1. Licences will expire after the period shown in Clause 3 (above).

9.2. Licensee reserves the right to terminate the Licence in the event that payment is notreceived in full or if there has been a breach of this agreement by you.

Appendix 1 – Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication).

For Advance Online Publication papers:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advanceonline publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/ Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g.Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any replication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of CancerResearch UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK:[**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (yearof publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
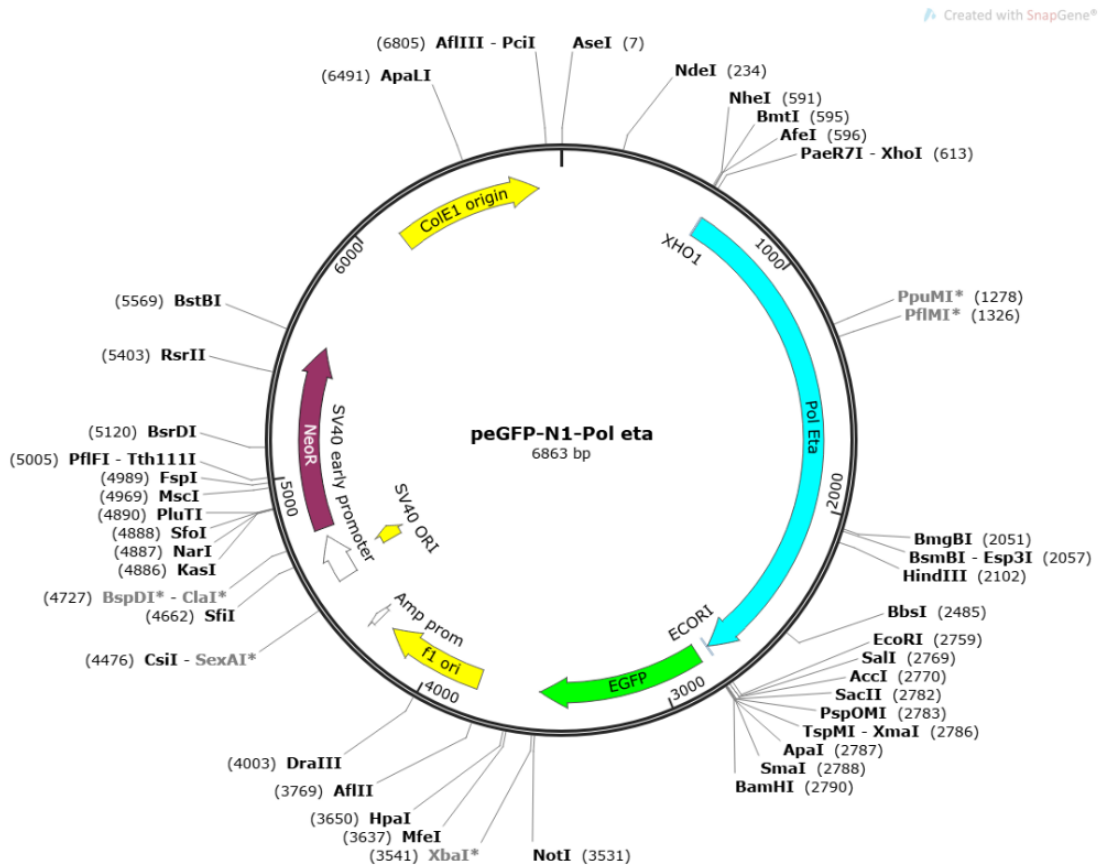Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

**Questions?** customercare@copyright.com **or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.**

Figure B1. Annotated plasmid map of pEGFP-N1-POLH. Unique 6-mer restriction enzyme sites are shown. Ori, site of origin, prom promoter; NeoR kanamycin/neomycin resistance. Created by Amélie Rodrigue using DNA Strider, edited by Kateryna Kratzer in SnapGene. From [198].
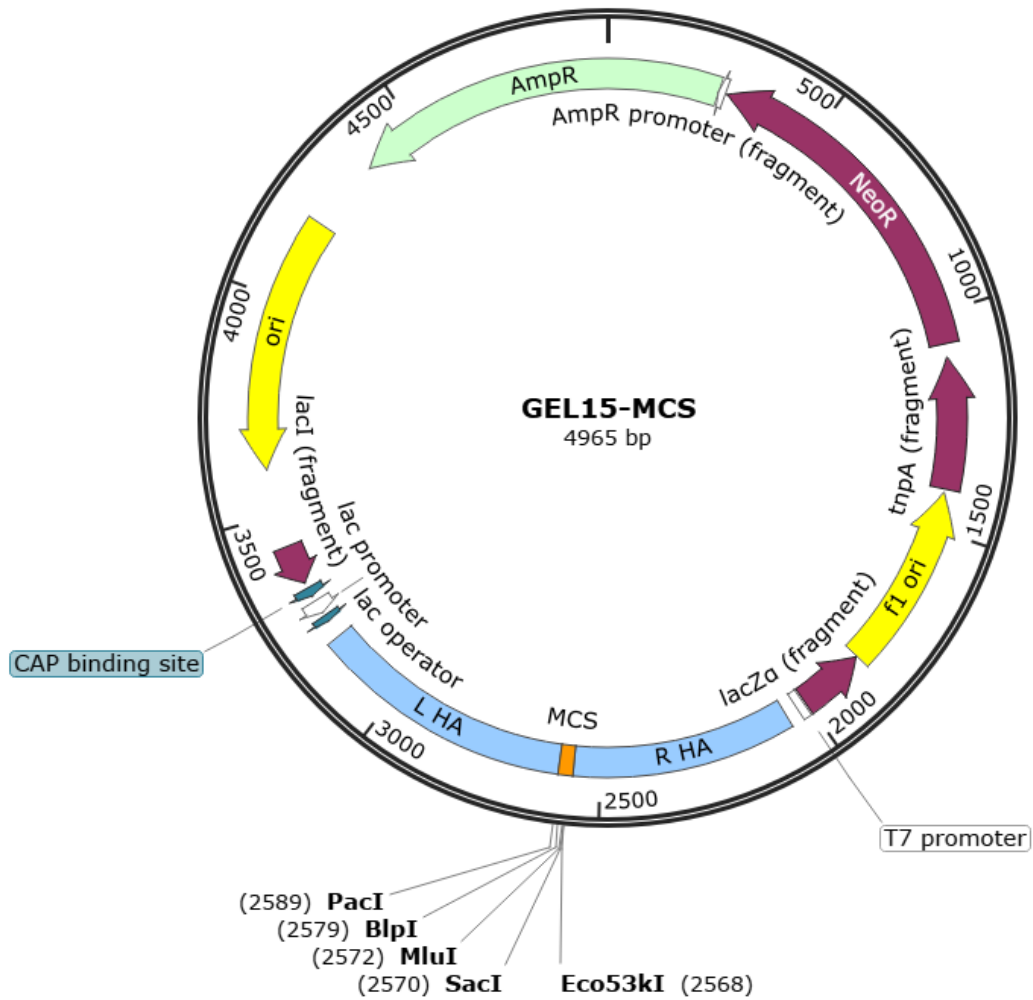
Figure B2. Annotated plasmid map of GEL15-MCS donor construct. Only MCS restriction sites are shown (orange). AmpR, ampicillin resistance; (L/R) HA, left/right chr15 homology arms; MCS, multiple cloning site; NeoR, neomycin/kanamycin resistance; ori, site of origin. Created by Plasmidsaurus.com and edited using SnapGene.
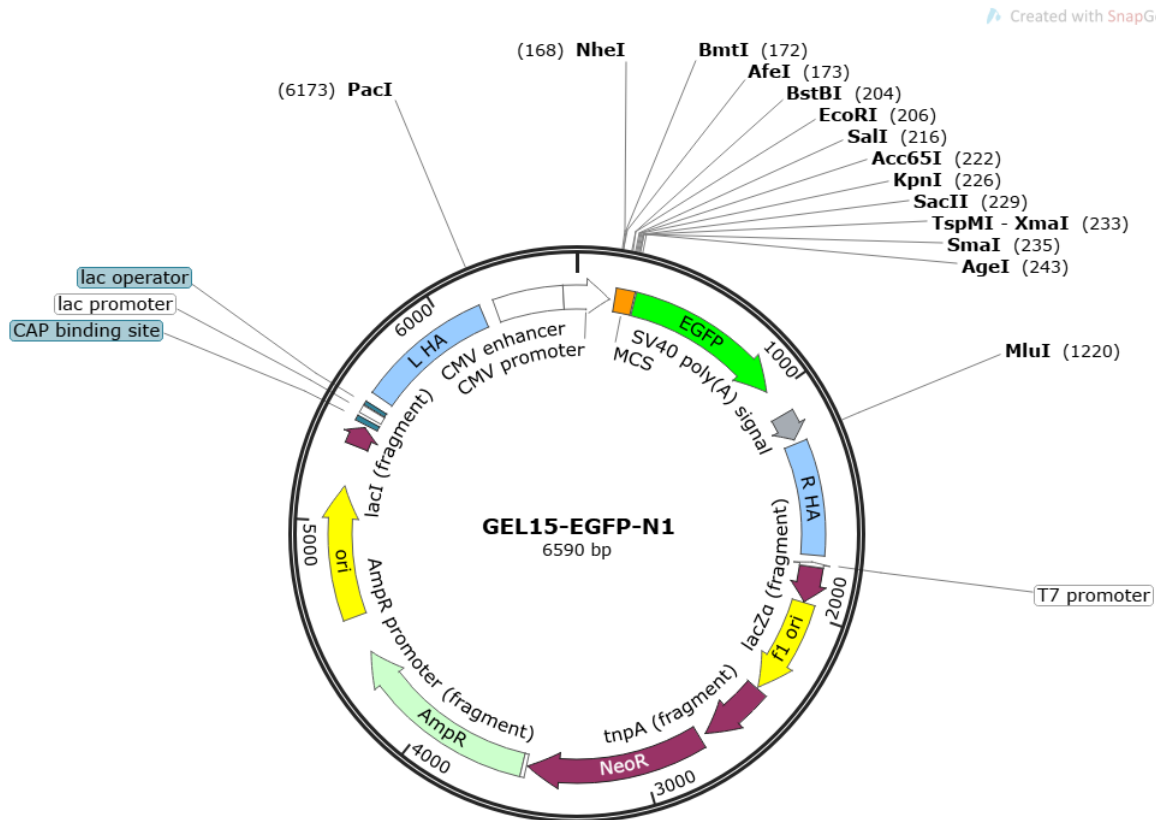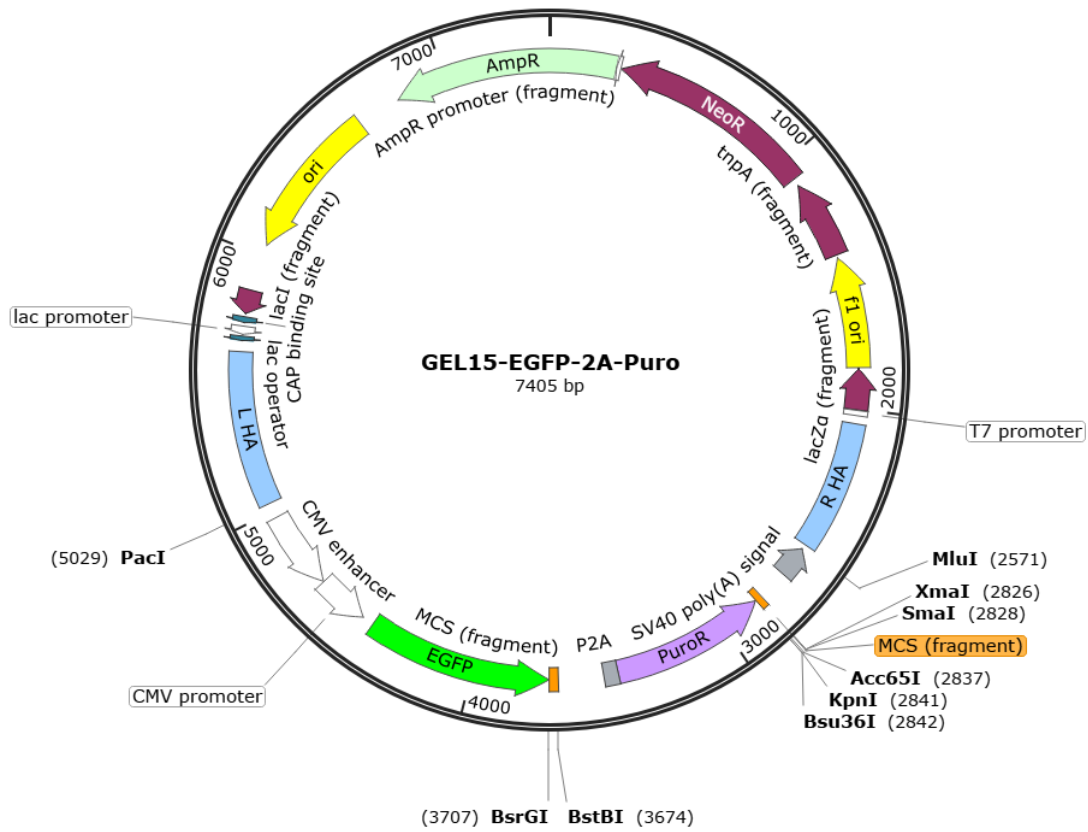
Figure B3. Annotated plasmid map of GEL15-EGFP-N1. The intact pEGFP-N1 MCS is shown (orange), as are the remaining restriction sites from the GEL15-MCS (PacI, MluI). AmpR, ampicillin resistance; (L/R) HA, left/right chr15 homology arms; MCS, multiple cloning site; NeoR, neomycin/kanamycin resistance; ori, site of origin. Created by Plasmidsaurus.com and edited using SnapGene.
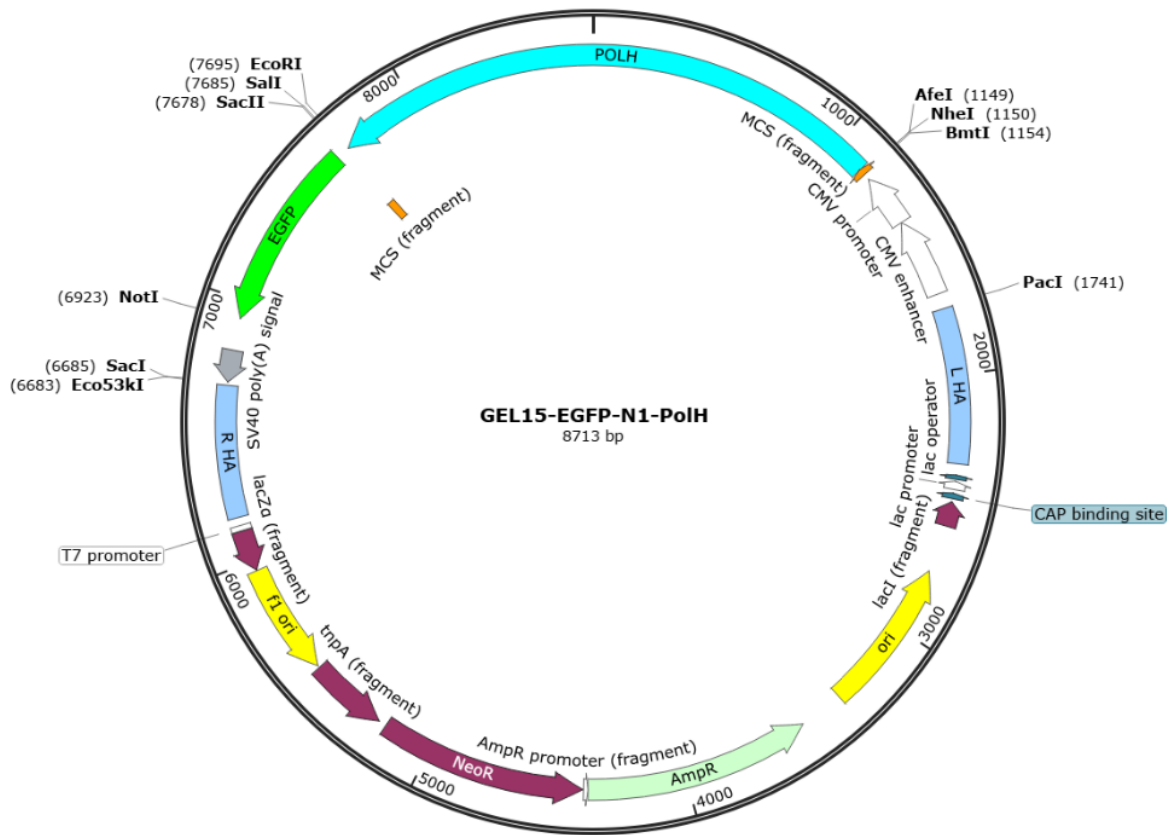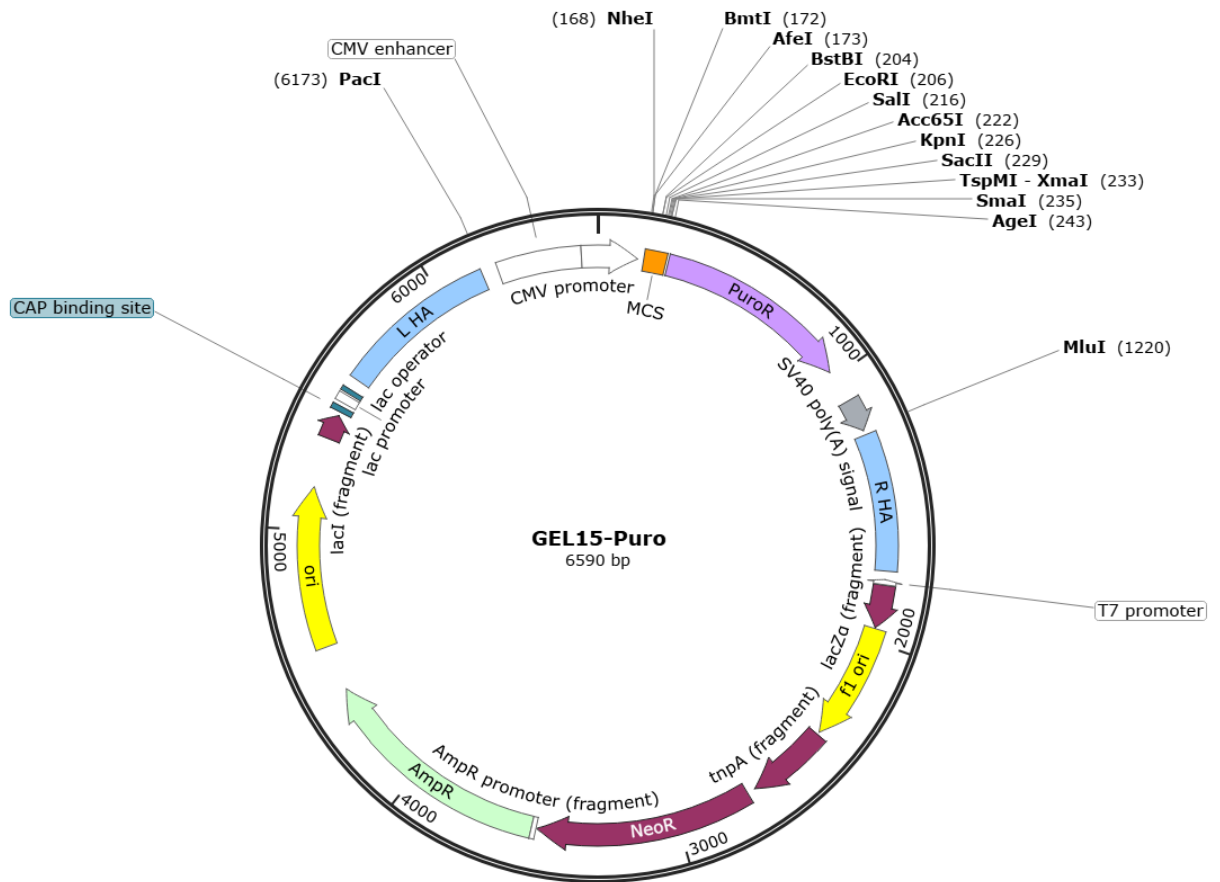
Figure B4. Annotated plasmid map of GEL15-EGFP-C1. The intact pEGFP-C1 MCS is shown (orange), as are the remaining restriction sites from the GEL15-MCS (PacI, MluI). AmpR, ampicillin resistance; (L/R) HA, left/right chr15 homology arms; MCS, multiple cloning site; NeoR, neomycin/kanamycin resistance; ori, site of origin. Created by Plasmidsaurus.com and edited using SnapGene.

Figure B5. Annotated plasmid map of GEL15-EGFP-2A-Puro. The remaining pEGFP-N1 MCS is shown (orange), as are the remaining restriction sites from the GEL15-MCS (PacI, MluI). AmpR, ampicillin resistance; (L/R) HA, left/right chr15 homology arms; MCS, multiple cloning site; NeoR, neomycin/kanamycin resistance; ori, site of origin; PuroR, puromycin resistance. Created by Plasmidsaurus.com and edited using SnapGene.

Figure B6. Annotated plasmid map of GEL15-EGFP-N1-POLH. The remaining pEGFP-N1 MCS is shown (orange). AmpR, ampicillin resistance; (L/R) HA, left/right chr15 homology arms; MCS, multiple cloning site; NeoR, neomycin/kanamycin resistance; ori, site of origin. Created by Plasmidsaurus.com and edited using SnapGene.

Figure B7. Annotated plasmid map of GEL15-Puro (predicted construct). The remaining pEGFP-N1 MCS is shown (orange), as are the remaining restriction sites from the GEL15-MCS (PacI, MluI). AmpR, ampicillin resistance; (L/R) HA, left/right chr15 homology arms; MCS, multiple cloning site; NeoR, neomycin/kanamycin resistance; PuroR, puromycin resistance; ori, site of origin. Created by Plasmidsaurus.com and edited using SnapGene.

APPENDIX C: GEL IDENTIFICATION CODE

The entire editable and executable bash script was written by Ariel Lisogorsky with input from Kateryna Kratzer and is located at (https://github.com/kkratzer/gene-editing-loci/).

UCSC Genome Browser files and COSMIC data must be downloaded (automated). Once the downloads are complete, the mask is created. "file_sequence.tsv" describes how the mask was built:

```
action   method   source   arg_one arg_two buffer   description file_out
add_to_mask script  ./scripts/get_cosmic.sh ""   ""   ""   ""
add_to_mask mysql   knownGene   chrom,txStart,txEnd,name   ""   150 any
gene    BED
add_to_mask mysql   wgRna   defaultBED   ""   300 any miRNA/other functional
small RNA    BED
add_to_mask script  ./scripts/get_repliseqpeaks.sh   ""   ""   50   any
replication origin  BED
add_to_mask mysql   vistaEnhancers  defaultBED   ""   50   any ultra-
conserved element BED
skip    mysql   all_mrna   NONE   ""   25   low transcriptional activity
(no mRNA)   BED
add_to_mask mysql   genomicSuperDups   defaultBED   ""   0   copy number
variable region BED
apply_mask  mysql   wgEncodeOpenChromSynthGm12878Pk
defaultBED   ""   0   regions in OPEN chromatin   BED
apply_mask  rsync   rsync://hgdownload.soe.ucsc.edu/gbdb/hg19/crispr/crisp
r.bb  ""   ""   0   regions targetable by CRISPR/Cas9   BB
```

Once the mask has been built, it is applied to regions of open chromatin and CRISPR targets. The regions that intersect between these two masked tracks are outputted as a comprehensive BED file that can be viewed using genome visualization software (e.g., UCSC Genome Browser, [151]; Integrative Genomics Viewer,[268,269]).