Method development for the generation and selection of precise genomic edits in human cells

Alice Lightowlers

Exeter College

University of Oxford

A thesis submitted for the degree of:

Doctor of Philosophy

January 2020

Word Count:

41413

Supervisors:

Professor Tudor Fulga

Dr David Knapp

Contents

Abstract

Gene editing is the process of generating targeted changes in the genome of an organism or cell. Changes can be imprecise insertions or deletions (indels) which generate null alleles and permit loss of function studies. Alternatively, an exogenous donor can be provided which allows precise changes to be copied into the genome. Precise editing can be used to advance scientific understanding by generating cell lines with specific mutations. Furthermore, precise editing has therapeutic potential: pathogenic mutations in patient's cells can be reversed or compensatory mutations can be introduced.

The relatively low efficiency of precise editing limits applicability of genome editing and increases the burden of labour on the researcher. In this thesis, I describe my efforts to improve the efficiency of editing and enrich for correctly edited cells. I study the impact of published methods shown to improve efficiency of editing. I then develop two potential FACS-based methods to enrich for correctly edited cells. Firstly, I attempt to use molecular beacons to directly recognise sequence changes in mRNA, and find that, despite promising *in vitro* results, the beacons are highly non-specific within cells. I then develop a transient Cas9-based circuit to detect the presence of a genomically-integrated crRNA. This crRNA-expression system can be incorporated beside the intended edit, using an ssODN donor of less than 200 base pairs. Finally, I find that using the transient Cas9-based circuit, I can enrich for precisely edited HEK293T cells by 20-40 fold, achieving final editing efficiencies of 20-60% alleles.

Acknowledgements

I would like to acknowledge the contribution of all past and present members of the Fulga lab to my growth and development as a scientist. In particular, I would like to thank Max Jamilly, Andrew Ramos, and Hector Barbosa for their companionship and support; Yale Michaels for his mentorship and wisdom; and Markus Toegel for his experimental support and selflessness. I am beyond grateful for my supervisor, Dr David Knapp, who's incredible insight and ideas are the basis of my thesis and who always goes above and beyond to help and support others in the lab. A special thanks goes to my supervisor, Tudor Fulga, who inspired me to come to Oxford, and who's infectious enthusiasm for Science and creativity I will always envy.

I am very grateful to the support and kindness provided by Marella De Brujin, without which I am sure my thesis would not have been finished. I thank my thesis advisory committee, Deborah Gill and Jim Hughes, for their wise words and sound advice, and the members of the WIMM FACS facility for all their help.

I am thankful for the love and support of my friends and family: for Fred Collman, for giving me a place to live, and good company; for Bethan Page for providing a space to work; for Hele Francis, Blane Scott and Martin Engelcke, for their shared experiences and warmth; for Andy, Sophie, Sara and Hatem for their love, commiserations and advice. Above all I am thankful for my parents: for my Mum, for her limitless love and support; and for my Dad, who I know would be proud of me.

List of Abbreviations

ALT-EJ	Alternate End-Joining
AML5	Acute Myeloid Leukaemia 5
BE	Base Editor
BIR	Break Induced Replication
bp	Base Pair
BRCA	Breast Cancer Associated
Cas	CRISPR Associated Protein
CCR5	Chemokine Receptor 5
cDNA	Complementary DNA
Cpf1	CRISPR from Prevotella and Francisella 1
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
Ct	Cycle of Threshold
CTS1/2	CRISPR Targeting Sequence ½
D32	Delta 32
dCas9	Dead Cas9
d/E/G/Y/C/iRFP	Destabilised/Enhanced/Green/Yellow/Cyan/Near-InfraRed Fluorescent Protein
DHJ	Double Holliday Junction
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethylsulfoxide
DNA	Deoxyribonucleic Acid
DSB	Double Strand Break
DSBR	Double Strand Repair
dsDNA	Double Stranded DNA
ES	Embryonic Stem

FA	Fanconi Anaemia
FACS	Fluorescent Activated Cell Sorting
FBS	Fetal Bovine Serum
FIAU	1-(2'-deoxy-2'-fluoro-1-β-D-arabinofuranosyl)-5-iodouracil
gDNA	Genomic DNA
HBA/HBB	Haemoglobin A/B
HDR	Homology Directed Repair
HR	Homologous Recombination
HSF1	Heat shock factor 1
HUDEP	Human Umbilical Cord Blood-Derived Erythroid Progenitor
ICL	Inter-strand Cross Link
iPSC	Induced Pluripotent Stem Cell
IVT	*In Vitro* Transcription
LMB	Landing Pad Molecular Beacon
LP	Landing Pad
LWy	Landing Pad Wyvern Beacon
MCP	Methyl-accepting chemotaxis protein
MCS-R2	Multi-Species Conserved Site – Regulatory 2
MFI	Mean Fluorescence Intensity
miRNA	Micro RNA
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NPRL3	Nitrogen permease regulator-like 3
ORF	Open Reading Frame
PAM	Protospacer Adjacent Motif
PE	Prime Editor

pegRNA	Prime Editor Guide RNA
Pol II/Pol III	RNA Polymerase II/III
RD	Restriction Digest
RFLP	Restriction Fragment Length Polymorphism
RGEN	RNA-Guided Endonuclease
RNA	Ribonucleic Acid
RNP	RiboNucleoprotein
RPMI	Roswell Park Memorial Institute
RT-qPCR	Real Time quantitative Polymerase Chain Reaction
SCD	Sickle Cell Disease
SDN	Site Directed Nuclease
SDSA	Synthesis-Dependent Strand Annealing
sgRNA	Single Guide RNA
SNP	Single Nucleotide Polymorphism
SSA	Single Strand Annealing
SSB	Single Strand Break
ssDNA	Single Stranded DNA
ssODN	Single stranded Oligonucleotide Donor
SSTR	Single Stranded Template Repair
T7E1	T7 Endonuclease I
TALEN	Transcription Activator Like Effector Nucleases
TIDE	Tracking of Indels by Decomposition
TK	Thymidine Kinase
tracrRNA	Trans-Activating CRISPR RNA
tRNA	Transfer RNA
UTR	Untranslated Region
WT	Wild Type
ZFN	Zinc Finger Nuclease

Declaration of Authorship

This thesis, and the work presented within it is my own unless otherwise specified. This work was carried out wholly at the Weatherall Institute of Molecular Medicine, University of Oxford. I am submitting this thesis for the degree of DPhil in Medical Sciences. This thesis has not been submitted for any other degree.

The following experiments were conducted in collaboration with others:

1) Design of all molecular beacons (excepting GFP-7_21_0) was performed by David Knapp.
2) I performed the flow Cytometry-based assays for experiments in Chapter 4 and Chapter 5 in collaboration with David Knapp.
3) FACS of cells using BD FACSAria™ Fusion was performed by Kevin Clark.
4) One replicate of the bulk sorting and enrichment experiment shown in Figure 5.7 was performed by David Knapp and Markus Toegel.
5) S2 cell transfections were completed in collaboration with Markus Toegel
6) The concept and design of both molecular beacon-based enrichment and circuit-based enrichment was devised by David Knapp.

Chapter 1: Introduction

1.1 Motivation

After recent advances in gene editing technology there has been a rush of interest in using genome engineering to introduce precise edits into cells, both for therapeutic endeavours and basic science. However, the efficiency of this process is limiting. The aim of the research presented in this thesis is to develop solutions to this problem by either improving the probability of incorporating correct edits or enriching for correctly edited cells within a heterogenous population.

1.2 Genome Engineering

Genetic Engineering is the process of adding new elements or removing or altering existing elements to the genomes of cells. Scientists have been attempting these processes since the 1970s. Originally, this process involved delivering donor DNA in the form of plasmids, containing transgenes flanked by homologous sequences to the intended target region, into cells[1–4]. In a small proportion of cells (typically 1 in 10^3 to 10^9), transgenes are spontaneously incorporated into the host's genome by the DNA repair process, homologous recombination (HR). This phenomenon was useful as it allowed the creation of transgenic organisms as disease models, for example the creation of transgenic mice through genetic engineering of mouse ES cells[5]. However, the process of creating a transgenic cell line or organism was inefficient and laborious. The genetic engineering process was improved massively on the discovery that site-directed nucleases could be used to create a double strand break in the targeted region, and native repair machinery co-opted to incorporate the foreign DNA as part of the repair process[6–8]. This process, referred to as homology directed repair or HDR, is distinct from the natural repair process, HR. Site-directed nucleases increase the efficiency of targeted payload integration by 10^2-10^6 fold.

In addition to improving the efficiency of sequence knock-ins, the use of site-specific nucleases for genome engineering presented an alternative method of gene disruption: gene knockout via indel

(insertion and deletion) formation[9]. Indels are formed when random errors occur in the native repair process, non-homologous end joining (NHEJ). In roughly $^{2}/_{3}$ of cases, indels cause frameshift mutations which cause nonsense-mediated decay of mRNA transcripts and gives a loss of function phenotype[10]. Mistakes in the NHEJ repair process occur at approximately 1 in 1000 DSBs (provided breaks are blunt ended) but eventually transpire during multiple cycles of cutting and re-ligation at the nuclease target site[11].

Despite the utility of NHEJ in loss of function studies, it presents a problem when attempting to incorporate precise edits into the genomes of mammalian cells. In the presence of a site directed nuclease and homologous donor DNA, repair processes are far more likely to produce NHEJ-based indels than precise HDR-mediated edits[12]. Therefore, a clear understanding of these complex repair processes is important to enable their manipulation for distinct gene editing purposes.

1.3 DNA Repair Pathways

Eukaryotic cells invest a great deal of resources in maintaining genome integrity and faithful genetic replication[13]. Perhaps the greatest threat to genome integrity are double strand breaks in DNA, which if unresolved can cause severe chromosomal rearrangements[14]. When repairing DSBs, mammalian cells have two broad pathways to choose from. The major DNA repair pathway, NHEJ, is active through all phases of the cell cycle and results in quick, efficient religation of the two blunt DNA ends with no need for template-based DNA synthesis[15]. However, as previously discussed, repetitive cycles of targeted nuclease-induced DNA cleavage, are eventually terminated by mutation of the target sequence due to indel formation[11].

DSBs are rapidly recognised due to conformational changes in DNA, initiating a cascade of interactions which results in chromatin relaxation and recruitment of Mre11, a protein in the MRN complex (Mre11, Rad50, Nbs1)[16,17]. The MRN complex has multiple functions in DSB repair. It is thought to act as a scaffold to recruit other factors involved in DNA repair and is important in tethering the broken ends of DNA[17]. This recognition and repair initiation process is common to both the main pathways of

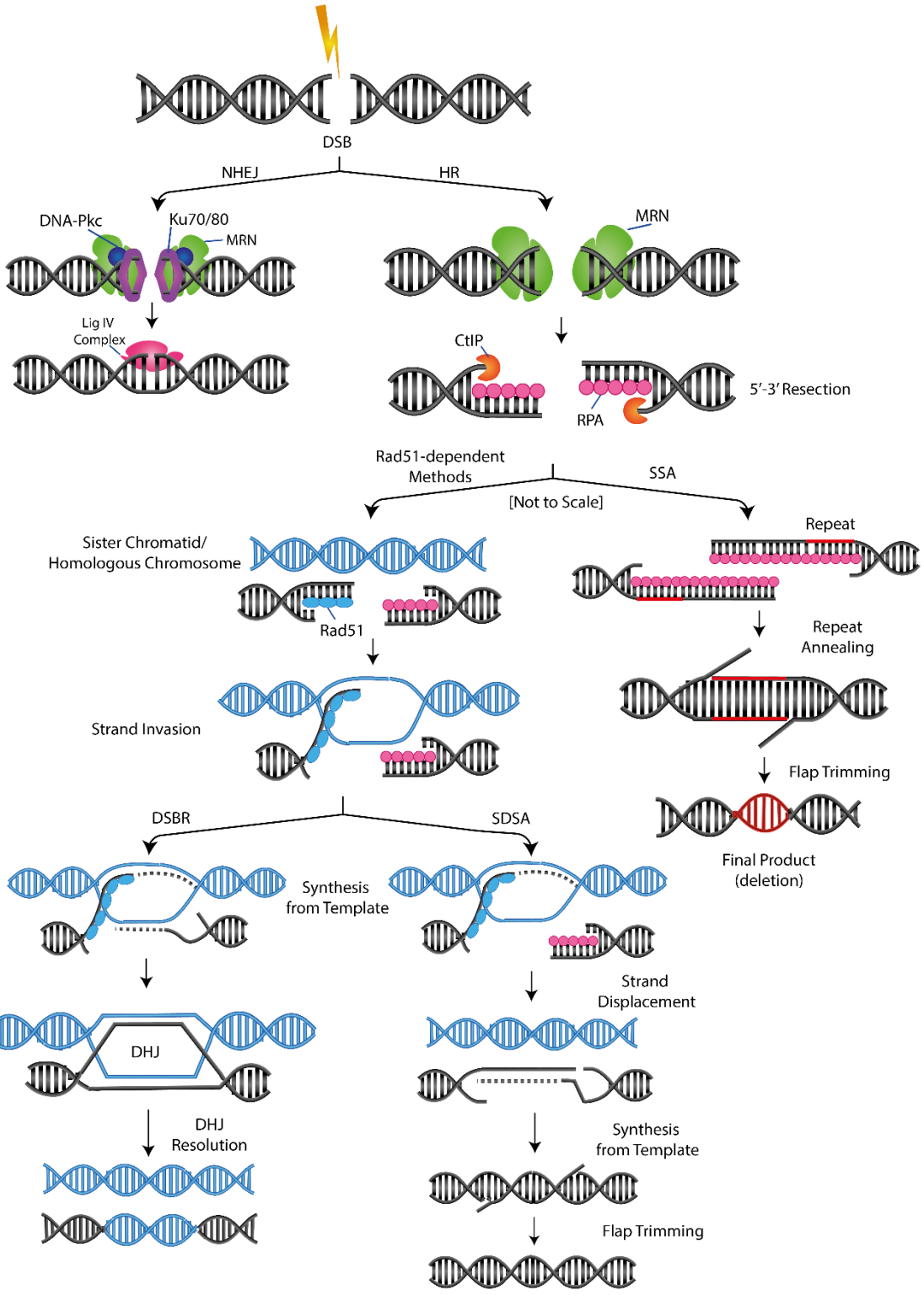


Figure 1.1 DSB Repair Pathways. DSBs are repaired through fast acting NHEJ or slower HR. In NHEJ blunt ends are recognised by Ku70/800 and DNA-Pkcs. The LigIV complex is recruited and blunt ends are religated. Alternatively, in HR, ends resection by CtIP prevents NHEJ and results in homologous recombination. This can be achieved through faithful, Rad51 dependent methods, or deleterious, Rad51-independent methods (SSA). Rad51 mediates strand invasion of the 3′ overhang into the sister chromatid (or other homologous sequence), and synthesis proceeds using this template. The strand is then either displaced, to be used as a template for the second strand, SDSA, or the second strand is also synthesized using the sister chromatid (DSBR). In SSA, homologous repeat sequences are identified and anneal through an unknown mechanism. Flaps are trimmed and nicks sealed. This process results in large deletions.

repair, however, depending on which factors are first recruited to by the MRN complex, the pathways diverge (fig 1.1)[18]. In NHEJ, the Ku70/80 complex, which binds broken DNA ends, is the secondary responder[19]. Ku70/80 then recruits the Ligase IV complex and broken ends are rapidly and efficiently re-ligated[20].

The second DSB repair pathway, HR, is a much slower process involving coordinated sequences of complex events, which are limited by the cell cycle, presence of homologous templates, and other factors, some of which remain unclear[21]. During HR the damaged strands are resected and resynthesized using a homologous DNA template. HR is dependent on the presence of this homologous donor template, which is often the sister chromatid but can also be from a homologous chromosome or homologous repeat region. Use of the homologous chromosome or homologous repeat regions for repair of DSBs can lead to loss of heterozygosity and spontaneous mutations, respectively[22]. Thus, during G2 and S phase, when sister chromatids are synthesised, HR is permitted, via the targeted cyclin dependent kinase (CDK)-dependent phosphorylation and activation of the 5'-3' endonuclease CtIP[23,24]. CtIP is recruited to DSBs by the MRN complex and initiates the pivotal step of HR, 5'-3' end resection, a process that creates long 3' single stranded tails which cannot be substrates for NHEJ-mediated ligation[25].

CtIP-mediated, 5'-3' resection is common to all forms of HR. In precise, Rad51-dependent HR, the exposed single stranded regions are bound and protected by RPA, before being sequentially replaced: first by Rad52, and then by Rad51[26,27]. Rad51 mediates strand invasion of the 3' single stranded tails into a homologous template[28,29]. Canonical HR then proceeds through one of three pathways: Double strand break repair (DSBR), Synthesis Dependent Strand Annealing (SDSA), and Break Induced Replication (BIR), which occurs specifically at replication forks. For all three cases, template-dependent synthesis of the damaged strand occurs. In DSBR, synthesis of the second 3' tail using the invaded chromatid is initiated (fig 1.1). This results in a double Holliday junction intermediate between the damaged and invaded chromatids, which can be resolved with or without crossing over[4]. In

contrast, in SDSA, only one strand is synthesised using the invaded chromatid as a template[30]. The second strand is then synthesized using the first strand as a template, and resulting flaps are trimmed prior to ligation. The third pathway, BIR occurs only as a result of a stalled replication fork and thus involves only one blunt end[31]. Here, after 3' strand invasion, synthesis of both the leading (3' tail) and lagging (resected 5' end) strands continues until either it re-joins a second replication fork or it reaches the end of the chromosome.

Interstrand crosslinks (ICLs) are also repaired through a DSB intermediate, HR-based pathway, known as the Fanconi Anaemia (FA) pathway[32]. ICLs typically cause stalling of replication forks and thus lead to genetic instability; mutations in the FA pathway are commonly associated with cancer[33]. The FA pathway model begins with recognition of the ICL site with two stalled convergent replication forks by FANCM and recruitment of the FA complex. The FA pathway then feeds into the above HR pathways with controlled DSB formation, followed by repair of the DSB via Rad51-dependent HR.

The previously mentioned HR mechanisms are faithful and error free, provided a sister chromatid is used as a template. However, initiation of HR does not require DNA to have been replicated, only the presence of CtIP for 5'-3' resection[34]. Other intrachromosomal DSB repair pathways which do not require the sister chromatid often lead to loss of genetic information. These pathways include single strand annealing (SSA) and an alternative to NHEJ, alternate end joining (ALT-EJ, also referred to as microhomology mediated end joining or MMEJ). SSA and ALT-EJ are similar in that they both involve repair of DSBs using homologous repeat regions, 5'-3' DNA resection, non-homologous 3' flap trimming, single-stranded annealing of homologous regions, DNA-polymerase mediated gap filling, and finally re-ligation of the DNA backbone[35,36]. Thus, both SSA and ALT-EJ are likely to lead to loss of some genetic information caused by deletion of the non-homologous sequence between repeats. Both processes, unlike non-deleterious HR mechanisms, are independent of Rad51, suggesting they do not involve strand invasion. In SSA, Rad52 has been found to be required for annealing of the complementary ssDNA regions[37]. The major differences between SSA and ALT-EJ lie in the size of the

homologous regions required, the length of resected 3' tails, and the key proteins involved. In SSA, the size of homologous repeat and length of 3' tails are typically much longer than in ALT-EJ. In *S. Cerevisiae* deletions caused by SSA can be over 25kb[38].

In contrast, single strand breaks (SSBs) are far less deleterious than DSBs, as damaged DNA does not become physically separated, thus danger of translocations and loss of genetic information is much reduced. However, repair of SSBs is important as they are the most common lesions arising in cells[39]. SSBs occur as results of direct damage from oxidation and from mistakes at replication forks, as well as indirectly as intermediates in other repair processes. SSBs are gaps in the phosphodiester backbone, usually accompanied by loss of a nucleotide and sometimes accompanied by loss of several nucleotides. Once detected SSBs undergo end processing to produce 3'P and 5'OH groups. Single nucleotide gaps are sealed by Pol β and longer nucleotide gaps (2-12nt) are filled by Pol δ and/or Pol ε. Flaps are then removed and the broken backbone resealed by Lig3[40].

Alternatively the 3' end of an SSB may also become a substrate for HR[41]. This process has been shown to occur in vertebrates at the IgG variable (V) regions in chicken cells, and aids hypermutation[42]. The free 3' end invades the duplex containing the inactive pseudo-V regions and primes DNA synthesis against this template. The resulting complex is thought to form a single Holliday junction but evidence is still inconclusive[43–45].

Overall, repair of both SSBs and DSBs is complex and interwoven. Prior to their use in gene editing, site-directed nucleases were instrumental in the dissection of the mechanics and dynamics of these repair processes and they remain invaluable tools.

1.4 Site directed Nucleases

Site-directed nucleases have evolved over time to better suit the needs of researchers. In the following section I introduce the most commonly used nucleases and discuss the advantages and disadvantages they provide for as gene editors.

1.4.1 Protein-based site-directed nucleases

The first site-directed nucleases used for targeted genome engineering consisted of proteins with DNA binding domains and nuclease domains. The first of these pioneer nucleases to be used in this manner were the naturally occurring meganucleases. Meganucleases are a family of restriction enzyme with long 14-40 bp recognition sequences, the most well-known of which is I-SceI, originally found in *S Cerevisiae*[6–8] (fig 1.2.A). The long recognition sequences of meganucleases imparts a high degree of

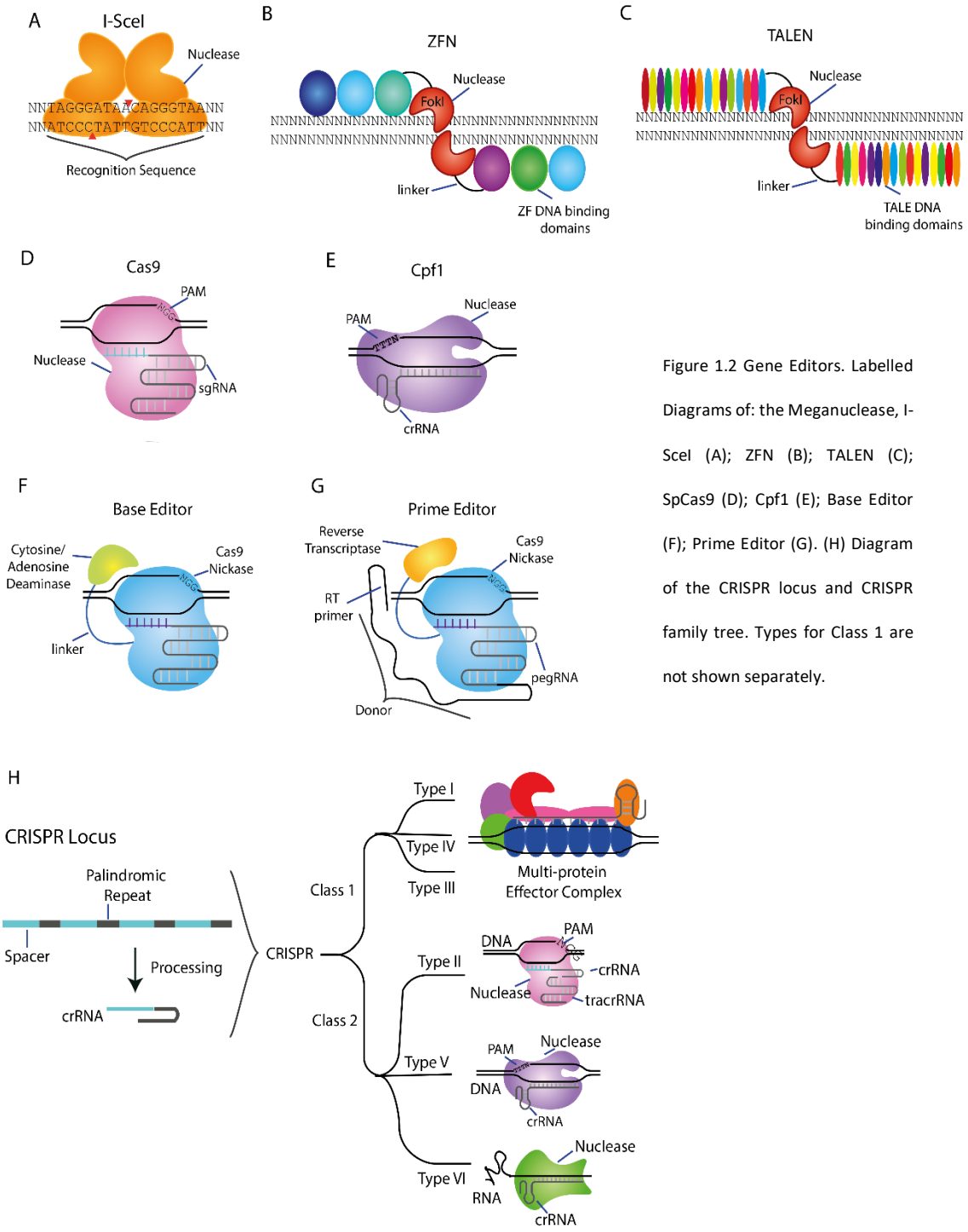


Figure 1.2 Gene Editors. Labelled Diagrams of: the Meganuclease, I-SceI (A); ZFN (B); TALEN (C); SpCas9 (D); Cpf1 (E); Base Editor (F); Prime Editor (G). (H) Diagram of the CRISPR locus and CRISPR family tree. Types for Class 1 are not shown separately.

specificity and therefore precision to these site-directed nucleases, but it also limits their applicability. The probability of a gene of interest containing a meganuclease recognition sequence is extremely low. Meganucleases can be artificially evolved through a process of randomised mutation and selection, to increase the diversity of target DNA sequences available[46]. However, this process is laborious and expensive.

Therefore, a more versatile alternative: the zinc finger nucleases (ZFNs), were developed. ZFNs are synthetic site directed nucleases formed by the fusion of zinc finger DNA binding domains to endonucleases (fig 1.2.B). Zinc finger domains are the most common DNA binding domains in eukaryotic cells[47]. Their DNA-binding specificity is imparted by finger region amino acids in each domain which make contact with DNA bases through the major groove of the double helix[48]. Each individual zinc finger targets a 3 bp sequence of DNA. Within a single ZFN, 3-6 zinc finger DNA binding domains are combined into arrays, targeting DNA stretches of 9-18 bp. The nuclease domain of ZFNs is typically the type II restriction enzyme FokI, which functions as a homodimer. Thus, ZFNs usually function as pairs, with three zinc finger DNA binding domains fused to each individual FokI domain, which function to bring the nucleases into proximity at the target site for cleavage.

This modular approach to target specificity means ZFNs are initially simpler to engineer than the meganucleases. However, specificity of each zinc finger is also subject to context specific variations from the surrounding DNA sequence and neighbouring zinc fingers. Thus, modularly designed ZFNs often achieve imperfect targeting and require further protein engineering and selection mechanisms to improve specificity[49].

Transcription activator like effector nucleases (TALENs) are a similar form of synthetic site directed nucleases to ZFNs, formed from fusion of the TAL effector DNA binding domains to the FokI nuclease domain (fig 1.2.C). TAL effector (TALE) domains were originally found in *Xanthomonas* bacteria and consist of homologous sequences 33-34 amino acids in length[50]. Each TALE has divergent amino acids at its 12th and 13th position, the repeat variable dinucleotide (RVD). Choice of this dinucleotide imparts

nucleotide specificity to each TALE[51]. Like ZFNs, 3 or more individual TALEs combine with a single FokI, and each TALEN acts as part of a pair, nicking opposing strands of the double helix to cleave DNA. Unlike ZFNs, the specificity of each TALE is much less liable to be influenced by the surrounding sequence context, thus they are more reliable to design and require less engineering. However, TALENs are highly repetitive due to the homology of individual TALEs. This makes these sequences difficult to clone, as they are likely to undergo recombination in the bacteria used for plasmid propagation[52,53].

In conclusion, protein guided targeted nucleases are incredibly valuable and have made huge advancements in the field, however, their use is limited as they require intensive protein engineering.

1.4.2 CRISPR/Cas9

The discovery of RNA-guided nucleases, i.e. the CRISPR (clustered regularly interspaced short palindromic repeats) system, and their adaptation for genome editing, provided an alternative to the purely protein guided targeted nucleases[54–56].The CRISPR system removed the need for protein engineering to give the nucleases specific-targeting capabilities, instead relying on complementary base-pairing of the target sequence with a guide RNA. This massively improved the accessibility of genome editing, resulting in a wave of interest in the area.

1.4.2.1 Discovery of the CRISPR system

CRISPR elements were first discovered in 1987 as a sequence of interest in the E. Coli genome[54]. They were initially characterised as an array of short (30-40 bp) partially palindromic repeats separated by variant 'spacer' sequences and were hence named, cluster of regularly interspaced palindromic repeats (CRISPR). These sequences were not homologous to known sequences, but similar arrays were found to be common across many bacterial and archaeal genomes[57]. Further investigation determined the spacer sequences were comprised of sequences found in viral genomes and plasmid DNA, leading to the hypothesis that the CRISPR arrays were involved in an adaptive immune response utilising Watson-Crick base pairing, similar to RNAi in eukaryotes[58].

The CRISPR system was further elucidated with the discovery of CRISPR-associated (CAS) proteins, which were used to classify the system into two classes, subdivided into six types(fig 1.2.H)[59,60]. These six types all utilise a guide RNA known as the CRISPR RNA (crRNA) transcribed from a single palindromic repeat and spacer sequence to target invading viral or plasmid material. However, they differ in the protein effectors which complex with the crRNA and enact the immune response. The rich diversity of mechanisms of action and effector proteins in the CRISPR immune system is likely due to an evolutionary arms race with pathogenic invading viruses[61].

Class 1, comprised of the subtypes I, III, and IV, includes immune systems with multiprotein effector complexes or cascades[60]. This complexity makes the Class 1 systems less useful for genome engineering. Whereas, Class 2 systems contain a single effector protein[62]. Class 2 systems are further subdivided into Type II, V, and VI, based on the nuclease domains of their effector protein. Type II effector proteins, i.e. Cas9, contain two single stranded DNA cutting domains: RuvC and HNH, allowing efficient production of DSBs in DNA[63–65]. Type V effector proteins also produce DSBs in DNA, albeit using two RuvC domains and with a single RNA component (crRNA)[66]. In contrast, Type VI effector proteins cleave single stranded RNA (ssRNA) using two HEPN domains[67].

1.4.2.2 The Type II CRISPR System

The Type II CRISPR system is naturally comprised of: the Cas9 effector protein; the DNA-targeting crRNA; a trans-activating CRISPR RNA (tracrRNA) which hybridises with the crRNA and Cas9 to produce a targeted nuclease[68]; the Cas1, Cas2 and Cas4 proteins involved in integrating foreign spacer elements into the CRISPR locus of the bacterial genome[69].

Type II CRISPR systems, along with the majority of other CRISPR systems, require a 2-6 bp protospacer adjacent motif (PAM) within the DNA sequence which interacts with the Cas9 protein[58,70]. The PAM sequence both increases specificity and, most importantly, prevents self-cleavage of the CRISPR locus in the bacterial genome by its own immune response. However, the PAM also imposes a limitation on which sequences are eligible for Cas9-targeting.

Different Cas orthologues from different species differ in size, PAM sequence, and cleavage location in the spacer. For example, the 1368 residue Streptococcus Pyogenes (Sp)Cas9 has a relatively simple PAM, NGG, found roughly every 8-12bp[55]. Therefore, despite its large size SpCas9 is a popular genome editing tool. The smaller, 1053 residue Staphylococcus aureus (Sa)Cas9 is also commonly used in situations where size is limiting, despite it's more restrictive PAM, NNGRRT[71]. Cas9 variants have also been engineered to recognise simpler PAMs to increase the range of loci accessible for editing[72,73]. The diversity of the CRISPR family provides a wealth of proteins which can be co-opted for different purposes in synthetic biology and genome editing.

1.4.2 Developments of the CRISPR system for gene editing

The first step towards producing the Cas9 system commonly used today was the development of the crRNA-tracrRNA fusion for SpCas9 by Doudna and Charpentier, known as the single guide RNA (sgRNA), which they then used to edit the bacterial genome[55]. Soon after this development, a codon-optimised SpCas9 was successfully developed which was used to edit the genome of mammalian cells[56,74]. A two-component ribonucleoprotein (RNP) comprised of the SpCas9 nuclease and the targeting single guide RNA (sgRNA) is now the most commonly used gene editing tool used by researchers (fig1.2.D).

These papers also showed how Cas9 could be multiplexed with more than one sgRNA to edit different loci simultaneously. The separation of the nuclease and targeting functions into components that are able to assemble within cells allows multiple RNPs to form, each guided by different sgRNAs[74,75]. Therefore, several sgRNAs can be used: either to investigate phenotypes caused by multiple mutations or to produce layered effects in synthetic circuits[56,76–79]. Additionally, the simplicity of sgRNA production means this system can be used very effectively in genome wide screens[80–82].

Beyond the Cas9 proteins utilised in these pioneering studies, researchers have also taken advantage of other natural and engineered effector variants. In 2015, two variants of the type V effector nuclease, Cpf1, also became co-opted for gene editing in human cells (fig 1.2.E)[83]. Cpf1, also known as

Cas12a, which requires no tracrRNA, exhibits T rich 5' PAMs and produces DSBs with staggered ends[83]. Cpf1 has a few potential advantages compared to Cas9. Firstly, the smaller size of its crRNA (~42 bp) compared to the longer sgRNA (~100 bp) may be preferable for delivery. Secondly, when editing AT rich genomes it may be preferable to use a T rich PAM than a G rich PAM. Finally, a direct comparison of SpCas9 with Cpf1 suggests that Cpf1 generally gives higher rates of HDR when using non-target strand ssODNs[84].

One of the key issues limiting the application of CRISPR systems, is their relative non-specificity when compared with ZFNs and TALENs[85]. Off-target cleavage activity has been found to occur in sequences with up to 5 bp mismatches to the sgRNA in the PAM-distal region of the spacer[86–89]. Multiple techniques have been used and additional techniques developed to accurately identify probable off target sites[71,86,90–92]. One simple technique found to improve the specificity of Cas9 is truncation of the spacer sequence to 17-19 nt[93]. This reduces the energy in the sgRNA:DNA duplex, restricting the interaction of Cas9 with mismatched sequences. Limiting the time that the active Cas9 and sgRNA complex spend within the nucleus also reduces non-specific cutting. This principle has been demonstrated by exchanging plasmid delivered Cas9 for the more rapidly degrading protein Cas9; and by using exogenous small molecules or light to activate Cas9 for specific lengths of time[94–96]. Researchers also rationally designed Cas9 to reduce the strength of non-specific electrostatic contacts between the protein and DNA, and found that this significantly reduced the occurrence of off-target editing whilst maintaining on-target editing efficiency[97,98]. The probability of off target indels is massively reduced if two sgRNAs are required to introduce DSBs. Cas9 nickases, where either the RuvC or HNH nuclease domain is inactive, used with two sgRNAs which target opposite strands in close proximity have been used to prove this principle[88,99]. Additionally, catalytically inactive dead Cas9 (dCas9) fused to FokI can be used in a similar manner to reduce the probability of off-target editing[100]. However, using paired guides requires that the locus of interest contains two suitably positioned PAMs with efficient sgRNAs.

The CRISPR toolset has revolutionised gene editing as a tool for biochemical analysis. However, as with other genome editing nucleases, use of Cas9 to produce donor-specific precise edits is inefficient, and more often results in NHEJ-induced indels[101,102]. Despite this, the increased convenience of CRISPR has meant technology in this area has developed rapidly with innovative solutions to this problem. In Section 1.6, I discuss solutions to this problem which involve shifting the balance of the cellular repair processes towards HDR. However, I will briefly discuss here developments from the Liu lab, which provide an alternative strategy to increase the efficiency of Cas9-based precise edits. Base-editors, a fusion of dCas9 and either cytosine or adenosine deaminase, achieve precise editing by modifying DNA bases without inducing DSBs(fig 1.2.F)[103,104]. Base editors massively reduce the probability of indels; however, they can still cause off-target base changes and unintended changes to bases within a 5 bp window of the targeted edit. Additionally, base editors are unable to make precise insertions or deletions.

Furthermore, a more recent development from the Liu lab can generate both precise modifications and small insertions. Prime editors are composed of RNPs with a Nickase:reverse transcriptase fusion and an sgRNA with an appended donor template(fig 1.2.G)[105]. Following Cas9 nicking, the RNA donor anneals to the broken single stranded end, which primes extension of the DNA against the RNA template by the reverse transcriptase. The single strand opposite the break is then displaced by the newly synthesized strand and flaps are trimmed. For short inserts, prime editors have a higher editing efficiency and reduced off-target effects compared to Cas9. However, the efficacy of prime editing for larger insertions is currently unknown.

These new gene editors are particularly advantageous as they remove the need for deleterious DSBs. However, the (currently) more commonly used gene editors require DSBs to edit efficiently, thus, it is important for researchers to understand the mechanisms underlying the appropriation of DSB repair for gene editing.

1.5 Repair Pathways and Targeted Nucleases

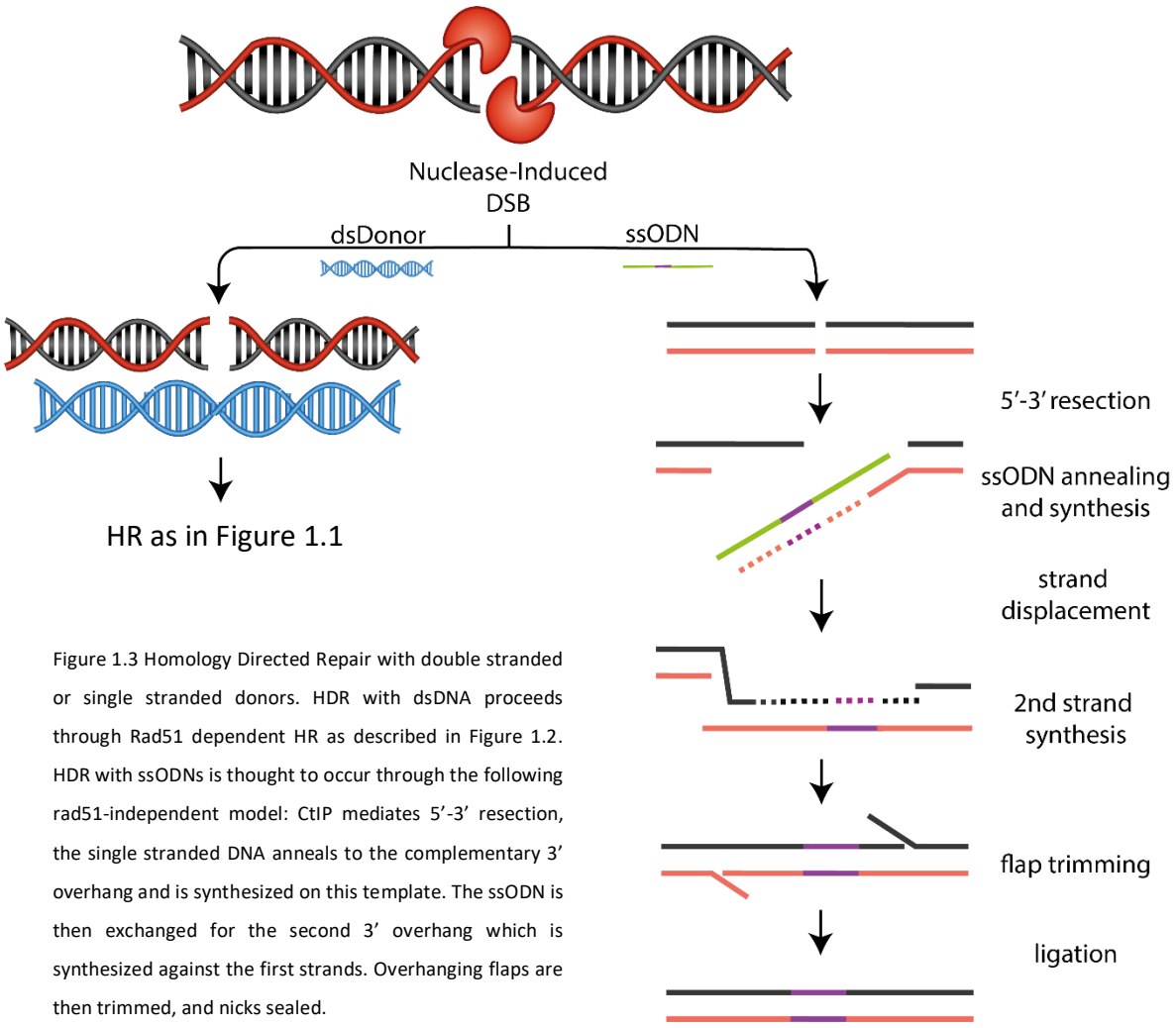


Figure 1.3 Homology Directed Repair with double stranded or single stranded donors. HDR with dsDNA proceeds through Rad51 dependent HR as described in Figure 1.2. HDR with ssODNs is thought to occur through the following rad51-independent model: CtIP mediates 5'-3' resection, the single stranded DNA anneals to the complementary 3' overhang and is synthesized on this template. The ssODN is then exchanged for the second 3' overhang which is synthesized against the first strands. Overhanging flaps are then trimmed, and nicks sealed.

Repair of targeted nuclease induced DSBs using double stranded donor DNA has been shown to occur through the Rad51-dependent non-deleterious HR mechanisms[106,107,108]. However, double stranded exogenous donor repair, either with linear dsDNA or plasmids, has been shown to be relatively inefficient[109]. For SNPs, small insertions and deletions, single stranded template repair (SSTR) using single stranded oligo donors (ssODNs), has been shown to be more efficient[110–112].

However, the repair pathway through which SSTR acts is poorly characterised. Most evidence suggests SSTR acts through a pathway similar to SSA, as repair with exogenous ssODNs is Rad51-independent and Rad52 and CtIP dependent[37]. Additionally, supressing Rad51 has been shown to enhance SSTR, whereas overexpression decreases SSTR[37,106,107]. Storici et al found that only the SSA domain of Rad52 was required for SSTR and proposed a mechanism for the resolution of DSBs from single stranded

templates[37]. The proposed mechanism for SSTR via the SSA pathway involves Rad52-mediated annealing of one resected 3' tail and the ssODN, followed by elongation of the 3' tail against the exogenous template (fig 1.3). Strand exchange then occurs, where the elongated 3' tail dissociates from the ssODN and anneals to the second 3' tail. The second 3' tail is then synthesized against this template, before flap trimming and ligation[37]. However, in human cells, efficient SSA-mediated repair is thought to require over 200 bp of homology, but many researchers have achieved efficient SSTR with only 100 bp of homology[113,114].

This mechanism is further supported by experiments which show the asymmetric nature of editing efficiency[115]. Kan et al found that efficiency of edit incorporation from ssODNs exhibited a gaussian like distribution biased strongly in the 3' direction as the distance of the edit from the break site increased. This supports the mechanism shown in figure 1.3 where the ssODN anneals to the 3' ssDNA tail, and templates 5'-3' synthesis. Therefore, sequence alterations on the donor-targeting 3' end of the cut site are far more likely to be incorporated than those on the 5' side.

Richardson et al have shown that the FA pathway has some role in SSTR, as multiple components of the FA pathway are required for Cas9-induced SSTR but not Cas9-induced NHEJ[116]. They also showed that FANCD2 was enriched at Cas9 DSBs. The specific mechanisms underlying FA involvement are not clear. However, as SSTR is Rad51-independent, resolution of DSBs and incorporation of precise edits must proceed through a different pathway.

Pathways involved in repair of SSBs generated by nickases have also been co-opted for precise genome editing, a process known as 'Nick-HDR'. This process is lower efficiency than HDR, but advantageous as nicks are not a substrate for NHEJ and show generate a much lower rate of indels[117]. Nick-HDR mirrors HDR, in that when a double stranded donor is provided HDR is Rad51-dependent, and when a single stranded donor is provided, repair is Rad51-independent[106,118–120]. Nick-HDR with ssODNs is thought to act through different pathways depending on whether the donor is complementary to the nicked or intact strand[106,118]. If the donor is complementary to the intact strand, it displaces the 5' and

3’ flaps on either side of the nick and anneals to the intact strand. The flaps are then trimmed, and nicks sealed, so that the donor DNA is incorporated into the genome. The genome edits can then be copied to the intact strand via mismatch repair. Alternatively, ssODNs complementary to the nicked strand initiate repair by annealing to the 3’ flap of the SSB. Extension of the 3’ flap proceeds using the ssODN, followed by 5’ flap trimming, nick sealing and mismatch repair. In this manner, changes to the 3’ side of the nicked strand cannot be introduced using a donor complementary to the nicked strand.

Researchers have also found that RNAs can be used to template DSB repair, in both yeast and human cells, albeit at a much lower efficiency than ssDNA templated repair[121,122]. Repair of DSBs can also be achieved with an endogenous RNA, and is thought to be a mechanism which originally operated during transcription: utilising nascent RNA to remedy DSBs in the transcribed DNA[123].

Specific knowledge of the proteins and processes involved in each step of DSB repair in different nuclear contexts, cellular backgrounds and with different donor templates, will be helpful to researchers trying to improve the efficiency of HDR. Currently, it is not clear exactly how SSTR operates.

1.5 Methods for improving HDR efficiency

Delivery of an exogenous donor does not guarantee incorporation of the intended edit as the DSB is repaired, in fact in mammalian cells, repair by NHEJ is the much more efficient pathway[124]. Therefore, many researchers have focused on attempting to improve the efficiency of HDR (Table 1.1). In order to do this, we must first understand what barriers are limiting efficiency. These barriers appear throughout the editing process: from delivery of editing machinery, to sgRNA cutting, to DSB repair.

NHEJ repair and the eventual formation of indels, disrupt Cas9 cutting so the correct edit can no longer be inserted[11]. Therefore, the interplay between the alternative DSB repair pathways is crucial in achieving precise gene editing as opposed to gene knockout and many methods focus on biasing repair away from NHEJ. However, the extent of involvement of different repair pathways in gene editing is poorly understood and varies between cell lines, loci and with different synthetic donor types.

Principle	Method	Gene Editor	Donor	HDR Efficiency increase	Cell line (s)	Author
Donor availability and effective concentration	Modified ss oligos	Cas9, TALEN (plasmids)	ssODN	4-10 fold	U2OS, RPE1	Renaud et al (2016)
	Asymmetric Donors	Cas9 RNP	ssODN	~60% absolute	HEK293	Richardson et al, 2016
	sgRNA aptamer + biotinylated ssODN	Cas9 RNP	ssODN	3-4 fold	HPSCs, HEK293	Carlson-Stevermer et al (2017)
	Covalent linkage of donor and Cas9	Cas9 RNP	ssODN	24 fold	HEK293T	Savic et al (2018)
	Covalent linkage of donor to Cas9	Cas9 RNP	ssODN	30 fold	HEK293T	Aird et al (2018)
G2-specific	Cas9-Geminin	Cas9 (plasmid)	plasmid	< 2 fold	HEK293T	Gutschner et al, 2016
	Block in G2 (Nocodazole)	Cas9 RNP	ssODN	2-5 fold	HEK293T, fibroblasts, H9 ES cells	Lin et al (2014)
	Block in G2/M (ABT)	TALENs, ZFNs, Cas9 (plasmids)	plasmid	3-6 fold	hPSCs	Yang et al (2016)
NHEJ inhibition	Ligase IV inhibtion by SCR7	Cas9 (genome-integrated)	plasmid	19 fold	Mammalian cells / mouse embryos	Maruyama et al (2015)
	siRNA of NHEJ Components, SCR7, Viral inhibitors	Cas9 (plasmid)	plasmid	4-5 fold	HEK293, NHI3T3	Chu et al (2015)
	DNA-Pkc inhibition	Cas9 (plasmid)	ssODN	2 fold	HEK293	Robert et al (2015)

Table 1.1 Summary of common methods to improve the efficiency of HDR.

NHEJ occurs independently of cell-cycle stage, compared to SSA and HR, which are all restricted to the S/G2 phases of the cell cycle[125]. SSA is thought to be the predominant method of DSB repair when using ssODN donors, although evidence for this is inconclusive[126]. Thus, some methods for improving the efficiency of HDR focus on optimising the timing of editing machinery expression using cell-cycle inhibitors[94,127] or cell-cycle specific protein degradation domains[128]. However, cell-cycle inhibitors, like nocodazole, are associated with apoptotic cell death and their usage may be incompatible with some forms of gene editing[129–131].

Furthermore, the addition of extra domains onto Cas9 increases plasmid size, hindering delivery, and reduces flexibility in the type of Cas9 used (i.e. RNP, modified Cas9 alternatives, etc).

While the cell cycle restricts availability of HR repair proteins, availability of synthetic donors can be affected by ineffective delivery, nuclease degradation, and the cytotoxicity of transfected DNA. Therefore, there is a fine balance to be struck between maximising donor availability and optimising cell viability; this equilibrium inevitably varies between different cell types. To increase donor availability, Renaud et al used nuclease-resistant ssODNs with phosphorothioate-modified ends and found that they increased gene editing efficiency by up to four times in two cell lines and three animal models[110]. Phosphorothioate-modified oligonucleotides have been shown to have toxic effects in human cells, however, these effects seem to be cell-type specific[132].

A 2016 paper from the Corn lab found that use of asymmetric donors could improve efficiency of gene editing by up to three times by exploiting the mechanics of the Cas9-DNA interaction[114]. They found that the PAM-distal, non-target strand is released first from the Cas9-DSB complex. This knowledge enabled them to enhance the effective concentration of the template molecule by using donors which annealed more effectively to this strand of the DSB.

The effective concentration of donor DNA in the vicinity of the DSB can be more directly increased through tethering of the ssODN to the guided nuclease[133]. Spatial and temporal colocalization of the donor and DSB increases the probability of repair via HDR instead of NHEJ. For Cas9, this can be

achieved by tethering of the ssODN to either the sgRNA or the Cas9 protein, by using RNA aptamers, SNAP-tags, biotin-avidin fusions, and phosphotyrosine bonds[134–137].

The above barriers all feed into the same basic dichotomy: the choice between HR and NHEJ. In mammalian cells, the balance favours NHEJ, while in yeast HR is preferred[124,138]. In practical terms, the choice relies on whether the NHEJ or HR machinery has preferential access to the DSB to initiate their pathway. Thus, repair pathway choice can be swayed directly towards HDR by inhibiting NHEJ. This has been done by inhibiting or knocking down NHEJ-specific proteins, such as Ku70/Ku80, DNA-PKcs, or DNA Ligase IV[101,102,139,140]. The ligation stage of NHEJ can be inhibited by SCR7, a DNA ligase IV inhibitor. This abrogation of the NHEJ repair process has been shown to improve the efficiency of gene editing at multiple loci in multiple cell types. In contrast, the small molecule RS-1 can be used to enhance the probability of HR by improving the propensity of RAD51 to bind to DNA[141]. Other small molecules, L755507, a β3 adrenergic receptor agonist, and Brefeldin-A, an inhibitor of protein transport from the endoplasmic reticulum to the Golgi apparatus, also enhance HDR[142]. L755507 and Brefeldin-A were identified as part of a large screen and their mechanism of action remains unclear.

Another way to sway the balance towards HR may be to use targeted nucleases to generate DSB substrates which are more suited to repair by HR than repair by NHEJ. This can be achieved using paired nickases to generate staggered double strand breaks with single stranded overhangs[99,143–146]. Paired nickases have mostly been found to be useful because of their dramatic reduction of off-target mutations, compared to individual Cas9 approaches. However, although double nicking has been found to produce reasonable levels of HDR, in general it has not been found to be more efficient than wild type Cas9[147]. This could be due to the reduced on-target DSB-production efficiency of paired nickases compared to wild type Cas9, as paired nickases compound the inefficiencies of both sgRNAs. Surprisingly, it was found that gene editing producing 3' single stranded overhangs, reminiscent of the resected intermediate generated during HR, resulted in less effective HDR than paired nickases which generated 5' overhangs[143]. However, they did find that the ratio of HDR:NHEJ was much higher for

nickases generating 3' single stranded overhangs than 5' single stranded overhangs. Thus, the discrepancy here could be due to an overall reduction in cutting efficiency. Therefore, it seems that the paired nickase approach may be useful to minimise off-targets and indels, but not to explicitly improve the efficiency of HDR. Additionally, there is evidence that location of target DNA within different nuclear domains affects its accessibility to different repair pathway machinery[148,149].

One of the greatest factors for determining efficiency of gene editing, is efficiency of sgRNA cutting. sgRNA cutting efficiency can vary greatly for sgRNAs even when in close proximity, possibly due to inherent sgRNA-target interactions, guide structure, or due to nucleosome inhibition of Cas9 cutting[150]. Therefore, it is important to use both *in silico*-based sgRNA prediction software, and to test multiple sgRNAs for a single locus.

1.6 Selecting for Correctly Edited Cells

In mammalian cells, it is probable that the efficiency of HDR will always be limited due to cell-cycle restrictions and prevalence of NHEJ. In most cases, edited cells are phenotypically indistinguishable from unedited or incorrectly edited cells. However, it is possible to substantially increase the proportion of correctly edited cells in a population by using various enrichment techniques. Enriching for correctly edited cells improves efficiency and reduces labour when using gene editing to produce a new cell line and increases the efficacy of treatment when editing cells for therapeutic purposes. The most common enrichment methods are briefly summarised in Table 1.2.

The first, most basic method to enrich correctly edited cells, is to use transfection positive markers, for example a secondary plasmid expressing fluorescent proteins or antibiotic resistance genes. This method is particularly relevant when attempting to edit primary cells or other cells with limited transfectability. FACS and antibiotic selection can then be used respectively, to enrich for transfected cells more likely to contain the editing machinery[99,151].

A more sophisticated version of this technique is to use a self-cleaving peptide between the targeted nuclease and selective marker to enable expression positive enrichment[152]. A prime example of this

are the px458 and px459 plasmids developed by the Zhang lab to enable synchronised expression of Cas9 and either GFP or puromycin resistance alongside Cas9 expression[153]. Additionally, Cas9 RNPs have been fluorescently labelled, by tagging with either a fluorescent dye or fusion to a fluorescent protein[154,155]. This allows FACS for nuclease positive cells when using RNPs instead of plasmids.

However, presence of the nuclease within the cell does not guarantee presence of the donor template. Double stranded donors can be delivered as a plasmid expressing a selectable marker, but many researchers now prefer to use ssODNs. Lee et al developed modified single stranded donors tagged with fluorescent dye which allowed donor positive cells to be sorted by FACS, increasing the proportion of correctly edited cells by ~80-100%[156].

Nevertheless, although cells receive and/or express editing components, it does not ensure that active targeted nuclease complexes are formed. Researchers have developed extrachromosomal surrogate reporters to sort cells based on nuclease activity[157]. These surrogate reporters typically encode a fluorescent reporter sequence and a nuclease target site identical to the editing target site. They come in two main forms: NHEJ-based and SSA-based. NHEJ based reporters normally encode a transfection positive control, a nuclease target site, and a frameshift fluorescent protein[158]. In nuclease-active cells, indels at the nuclease target site will shift the frameshifted protein into the correct reading frame in one third of cases, resulting in expression. However, sensitivity here is relatively low as only one third of reporters edited will have the correct frameshift. Ramakrishna et al, improved the sensitivity of NHEJ-based reporters by using a dual frameshift cassette, with two GFP sequences in different frames[159].

The second surrogate reporter type is SSA-based and consists of a non-functional reporter interrupted by two long direct repeats and a nuclease target sequence[160,161]. Nuclease-induced DSB repair by SSA results in deletion of one of the direct repeats and reformation of the functional reporter. Results are contested on whether SSA or NHEJ-based reporters are more sensitive in enriching for nuclease active cells, but this may be because different cell lines and constructs were used in each study[162,163].

However, editing of plasmids and reporters, does not accurately simulate editing in the genome, as it disregards the impact of chromatin. It is not certain that editing of a reporter within a cell, will mean that the cells genome is also edited. Some genes are exceptional, in that they are of interest to the genome editing field and editing them provides a detectable phenotype. One prime example is the editing of CCR5 and CXCR4, both of which are surface markers, whose knockout provides resistance to HIV[164]. NHEJ-based knockdown of CCR5 and/or CXCR4 can be selected for by fluorophore or magnetic bead-tagged antibody detection and cell sorting[165,166].

Principle	Method	Enriching for:	Enrichment	Cell Line(s)	Author
Expression positive cells	Self-cleaving peptide	All editing	N/A	HEK293FT, HUES9	Ran et al, 2013
	Fluorophore-labelled sgRNA	All editing	< 2 fold	Jurkat, hiPSCs, $CD34^+$	Nasri et al, 2019
	Fluorescent protein Cas9 fusion	All editing	N/A	Zebrafish embryos	Burger et al, 2016
Donor positive cells	Fluorescently-labelled ssODN	HDR	< 2 fold	HEK293, K562, Mouse Myoblasts	Lee et al, 2017
Nuclease active cells	NHEJ-based surrogate reporter	NHEJ	3-4 fold	HEK293T	Ramakrishna et al, 2014
	SSA-based surrogate reporter	All editing	N/A	HEK293T, MCF-7	Zhou et al, 2016
Endogenous Reporters	Targeting cell-surface antigens	NHEJ	N/A	HEK293T, TZM-bl	Liu et al, 2017
	Co-targeting of loci	NHEJ	N/A	Schwann cell line, $CD34^+$	Moriarity et al, 2014
	Coincident insertion	HDR	8-10 fold	K562	Agudelo et al, 2017
Integration of Selection Cassettes	Puromycin selection and PiggyBac-mediated excision	HDR	N/A	iPSCs	Xie et al, 2014
	Puromycin selection and PiggyBac-mediated excision	HDR	N/A	hPSCs	Yusa et al, 2013
	GFP insertion and FACS and Cre-mediated excision	HDR	N/A	HEK293	Kühn et al, 2015

Table 1.2 A summary of common editing enrichment methods. Where no enrichment value is given, the scope of the experiments was not set up to directly measure this.

Not all gene editing targets present convenient phenotype changes. However, editing of one locus greatly enhances the probability of editing at a second independent locus[167]. Thus, co-targeting of a locus of interest and a second endogenous marker locus, followed by sorting on the edited marker phenotype results in enrichment of cells with an edited locus of interest[167,168].

HDR-based editing can be enriched for in a similar manner[169]. Coincident insertion of two HDR-driven edits was found to be common in mouse ES cells[170]. The difficulty here is that most endogenous marker genes are selective when they are knocked out[171,172]. Therefore, distinguishing HDR competent cells from NHEJ governed cells is complicated. This problem was elegantly solved by Agudelo et al by targeted cutting within an intron of the ATPA1 gene, thus, indels were likely silent and HDR-driven edits made precise changes in the exon to produce a non-functional protein[169]. Selecting cells for ouabain resistance, imparted by ATPA1 knockout, gave a 10-fold enrichment of cells with a correctly edited second independent locus.

Despite this innovative approach, the above techniques are still best suited to enriching for knockout cells. Enrichment of large knock-ins with plasmid or linear dsDNA donors is typically achieved by inserting selectable marker genes as cassettes alongside the edit of interest. This method sensitively and precisely enriches for correctly edited cells[173,174]. However, the presence of selectable marker genes may affect cells in unpredictable ways and make them unsuitable for therapeutic applications.

Therefore, selection cassettes can be excised using methods like the Cre/LoxP system, the Piggybac transposase and even a second round of targeted nuclease editing[175–177]. Excision of the selection cassette can be selected for by negative selection, for example the loss of the conditional lethal marker, thymidine kinase. Excision of Cre/LoxP sites leaves a scar, whereas both the Piggybac and Nuclease editing can be scarless. Thus, these techniques are suitable for dsDNA donor editing in cells robust enough to withstand two selection steps. However, cassettes of these size cannot be introduced with ssODNs, and these techniques still involve integration of large cassettes into the genome, which may have unforeseen effects on cell behaviour.

1.7 Therapeutic Potential of Genome Engineering

Gene editing is a powerful tool both in the lab and in the clinic, but what specifically is its potential and what has already been achieved with these tools for human therapeutics? Gene editing-based treatments have been in multiple clinical trials since 2015 and may, in the future, have a great impact on reducing the burden of human diseases.

When considering gene editing it is important to make the distinction between somatic and germline editing. Somatic editing involves editing adult cells, and any genetic alterations cannot be passed on to offspring. Germline editing, on the other hand, typically involves editing embryos, making changes which affect sex cells and are therefore hereditary. Germline editing is, therefore, a much greater ethical issue due to the permanent changes it makes in the gene pool. In addition, many diseases which could be cured through germline editing can be more effectively solved by pre-implantation embryo screening. Recently, Ma et al attempted to mutate the CCR5 gene in the germline, in an effort to produce HIV-resistant humans[178]. This was achieved by direct injection of RNPs and sperm into human oocytes and culminated in the birth of twin girls, one with mosaicism, who is still vulnerable to HIV infection, and the other who is CCR5null homozygous.

Somatic editing has the potential to cure or reduce the symptoms of patients suffering from a number of diseases, without making permanent changes to the gene pool. Currently, researchers are focused on mendelian diseases as these are the best characterised and simplest to engineer. Research is also focused on serious diseases for which there is little alternative treatment, as the potential harm of off-target mutations caused by targeted nucleases is still unknown. One of the advantages of gene editing over current donor-based methods is the much-reduced probability of autoimmune rejection, as the patient's own cells can be used.

On the frontline of potential targets are blood disorders, as hematopoietic stem cells are relatively easy to isolate from the bone marrow, and, thus, editing can be accomplished *ex vivo*. In 2014, the first clinical trial of ZFN-edited $CD4^+$ T-Cells was undergone in persons infected with HIV[179]. In 2019,

phase 1 and 2 clinical trials began for a CRISPR-based treatment against β-thalassemia and sickle cell disease (SCD), which used Cas9 to upregulate the expression of foetal haemoglobin[180].

However, there is also potential to edit cells *in vivo*. Currently *in vivo* targets consist mainly of the liver and retina, as these are relatively easily targeted by viral vectors[181,182]. A phase I clinical trial in a patient with Hunter's disease, a metabolic liver disease, was completed in 2018[181], and preliminary results are reassuring. Phase I and II clinical trials are also underway for a CRISPR-based treatment for Lens Congenital Amaurosis, a degenerative retinal disease which leads to blindness[183]. Development of precise tissue-targeting viral vectors will advance the scope of *in vivo* editing.

One of the most exciting prospects of gene editing is the use of chimeric antigen receptor (CAR) T cells to fight cancers. Two clinical trials began in 2018 to test the efficacy of CAR T cells on both mesothelin positive solid tumours and multiple myelomas[184,185].

Ultimately, the future of gene editing is exciting for both its applications to human disease and for science in general. The therapeutic potential can be increased by improving the efficiency of gene editing and further developing methods to select for currently edited cells.

Chapter 2: Materials and Methods

Oligonucleotide sequences and maps of plasmids are provided in the appendix of this thesis. Oligonucleotides were purchased from Integrated DNA technologies (lyophilised with standard desalting) and resuspended in Nuclease-Free Water (Ambion©) to 100uM. Antibodies were purchased from Biolegend and enzymes were purchased form New England Biolabs unless otherwise stated.

2.1 Cloning

Enzymes used for cloning were purchased from New England Biolabs, except for BpiI (Thermo Fisher) and were used according to manufacturer's instructions.

sgRNA plasmids were produced using the pSpcas9(BB)-2A-GFP (px458) plasmid (addgene: 48138) from Dr Feng Zhang's lab, following the Zhang lab protocol[186].

Plasmids for CCR5 expression in S2 cells (pAC-360-Syn21-CCR5-P10, pAC-360-Syn21-CCR5-D32-P10) and HEK293T Cells (pll3.7_bi_dir_CCR5_dECFP, pll3.7_bi_dir_CCR5-D32_dECFP, pll3.7_bi_dir_CCR5-insert_dECFP) were kindly provided by Markus Toegel. The CTS1-dEYFP reporter was also cloned by Markus Toegel.

Most circuit component plasmids (CTS1-EYFP reporter + tracrRNA, MCP-P65-HSF1, dCas9-VP64, U6-placeholder-crRNA, hutRNAgln-CTS1-crRNA) were kindly provided by David Knapp. Other circuit component plasmids were obtained from addgene: piRFP670 (addgene: 45457), and px330 (addgene: 42230). Additional spacer length variant plasmids were cloned into the p385 backbone digested using NheI and EcoRI, using annealed oligo pairs with appropriate single stranded overhangs. Antarctic phosphatase was used to dephosphorylate the digested vector before it was gel purified using the QIAquick Gel Extraction Kit (Qiagen). Oligo pair sequences are listed in the appendices and they were annealed and phosphorylated using T4 PNK. Inserts and vector were ligated for 30 min at room temperature using T4 DNA ligase, before transformation into Subcloning Efficiency DH5α Competent

Cells (Thermo Fisher) via heat shock. Plasmid DNA was purified using the QiaPrep Spin Miniprep Kit (Qiagen) according to manufacturer's instructions. Cloned plasmids were verified by Sanger sequencing (Eurofins Genomics).

2.2 Cell culture

Human cells were incubated in standard sterile cell culture conditions at 37°C in 95% air/100% humidity. Cell lines were mycoplasma tested post-thaw, prior to experimental procedures, and every 3 months. HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (thermo fisher) with 10% FBS (Thermo Fisher). HUDEP-2 cells were kindly provided by the Higgs Lab and were cultured according to the expansion protocol from the Orkin and Bauer labs[187]. CD34+ mobilised peripheral blood cells were kindly provided by the Milne Lab, under HTA regulations, and cultured in expansion phase according to the protocol from the Porteus lab, in 300 ng/ml SCF (Peprotech), 100 ng/ml TPO (Peprotech), 300 ng/ml Flt3 ligand (Peprotech) and 60 ng/ml IL3 (Peprotech) albeit with serum-free StemSpan (Stem Cell Tech) in place of GMP SCGM[188,189]. OCI-AML5 cells (both wild-type and GFP) were gifted from Connie Eaves through David Knapp and were expanded in RPMI 1640 (Thermo Fisher) supplemented with 10% FBS. $CD4^+$ T cell clonal lines were produced and expanded, with HTA approval, by Giogio Napolitani, Jili Chen and Hilary Shepherd and irradiated LG2 feeder cells were provided by Jili Chen. T cells were cultured and expanded according to the cited protocol, although 10 IU/ml IL-7 (Peprotech) was added alongside IL-2 (Peprotech) when cells were fed[190]. S2 cells were received from Markus Toegel and were incubated in 100% air at 26°C, in Schneider's *Drosophila* Medium (Thermo Fisher) containing 10% heat-inactivated FBS.

2.3 CRISPR-Cas9 gene editing of HUDEP-2 and $CD34^+$ Cells

2.3.1 MCS-R2 sgRNA and ssODN Design

sgRNA were designed using CRISPRko from the Broad Institute (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design)[191]. Donors were designed to be centred on the Cas9 cut site and homology arm lengths were measured from this point outwards.

The insert barcode sequence was designed to have minimal similarity to the surrounding area to minimise qPCR aberrations and to incorporate a BamHI site to evaluate HDR efficiency by restriction digest.

2.3.2 HUDEP-2 CRISPR-Cas9 genome editing

HUDEP-2 cells were electroporated using the Nucleofector 2b Device (Lonza) with the Human CD34^{+} Cell Nucleofector Kit (Lonza) using the U-08 program. Cells were electroporated with 2.5 µg px458-sgRNA and 200 pmol ssODN.

2.3.3 SCR7 and RS-1

SCR7 pyrazine (HPLC) and RS-1 (HPLC) were purchased from Sigma-Aldrich, resuspended in DMSO (Thermo Fisher) and stored at -20°C. In relevant experiments SCR7 and RS-1 were added to the post-nucleofection media at the specified concentrations.

2.3.4 Sony Cell Sorting of HUDEP-2 Cells

72 hours post nucleofection, HUDEP-2 cells were sorted on GFP expression using the SH800Z Cell Sorter and 100 µm sorting chip (Sony). Cells were gated using SSC-A/FSC-A, single cells were gated using FSC-H/FSC-A, viable cells were gated using DAPI/FSC-A and transfected GFP expressing cells were sorted using GFP/FSC-A (fig 2.1). Untransfected cells were used to set the sorting gates. Cells were sorted into FACS Buffer in 1.5mL NoStick tubes (SSIbio) ready for pelleting and gDNA extraction.

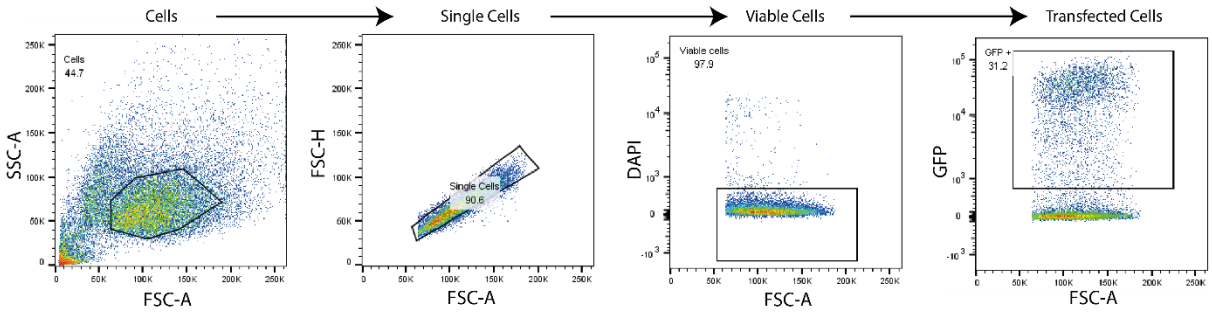


Figure 2.1 Gating Strategy for Sony SH800 cell sorting of GFP positive HUDEP2 cells.

2.3.5 CD34$^+$ Cell CRISPR-Cas9 genome editing

48 hours post-thaw, CD34$^+$ cells were electroporated using the Nucleofector 4D System (Lonza) with the P3 Primary Cell 4D-NucleofectorTM X Kit S (Lonza) using program CA-137. sgRNAs were purchased from Synthego and resuspended in nuclease-free TE to a concentration of 5 µg/µL. CD34$^+$ cells were edited using 5 µg Engen Spy Cas9 NLS (New England Biolabs), 2.5 µg sgRNA and 60 pmol ssODN according to the protocol from the Porteus lab[189]. gDNA was harvested 48 hours after electroporation.

2.4 Genomic DNA extraction

Cells were harvested, pelleted and resuspended in 100 µL Lysis Buffer. Cells were incubated overnight in lysis mix at room temperature. gDNA was precipitated by addition of 10 µL 8 M LiCl with 100 µL isopropanol and incubation for 1 hour at -20°C. Sample were then centrifuged at 20000g for 30min to remove the supernatant and then washed with 70% ethanol. gDNA was resuspended in nuclease free water (Ambion) and quantified using the Qubit high sensitivity dsDNA kit (Thermo Fisher). Concentration of gDNA was adjusted to 20-30 ng/µL.

2.5 sgRNA Efficiency evaluation using T7 Endonuclease 1 and TIDE software

2.5.1 T7E1 Assay

The R2 locus was amplified from the gDNA using the Phusion High Fidelilty Master Mix with GC Buffer (Thermo Fisher) and the primers (R2 fwd 2 and R2 rev 2) specified in the appendix. 20-30 ng of gDNA and 250 nmol of each primer were included in each reaction mix. The following cycling conditions were used: initial denaturation (98°C for 30sec), 30 amplification cycles (98°C for 10sec, 70°C for 20sec, 72°C for 30sec), and final extension (72°C for 5min). Size and purity of the locus was evaluated by agarose gel electrophoresis on a 1% agarose gel with GelRed Nucleic Acid Gel Stain (Biotium) using Bio-Rad Gel Doc XR+ system. The PCR product was then purified using the MinElute PCR purification kit (Qiagen).

The PCR product was then melted by heating to 95°C for 5 minutes and reannealed by cooling to 25°C at a rate of 4°C/min. 60 ng of DNA was then incubated with T7 endonuclease 1 at 37° in NEB Buffer 2 to cleave heteroduplexes. DNA was then run on a 1% agarose gel with GelRed Nucleic Acid Gel Stain (Biotium) to evaluate band size and brightness. Gels were imaged using Bio-Rad Gel Doc XR+ system, ensuring no image saturation, and band brightness was quantified on ImageJ.

2.5.2 TIDE Assay

The PCR product was sequenced by Eurofins Genomics overnight sequencing service. Samples were submitted according to the Eurofins sample submission guide. The .ab1 files received were then evaluated using the TIDE software(TIDE: Easy quantitative assessment of genome editing experiments by sequence trace decomposition, https://www.tide.deskgen.com)[192,193]. The sequenced PCR product from untransfected cells was used as a baseline to measure the sgRNA-transfected cells against.

2.6 Quantification of Editing efficiency

2.6.1 Editing Evaluation by BamHI Restriction Digest

60 ng of the R2 PCR product was digested by BamHI in NEB CutSmart Buffer according to manufacturer's instructions. The product was then run on a 1% agarose gel with GelRed Nucleic Acid Gel Stain (Biotium) and imaged with the Bio-Rad Gel Doc XR+ system ensuring no image saturation by optimising exposure time. Band brightness was then quantified using ImageJ analysis software.

2.6.2 Editing Evaluation by HDR Barcode RT-qPCR

qPCR was performed on a CFX384 real-time thermocycler (BioRad) with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). qPCR Control region primers and the forward test region primer were designed using Primer3 (http://primer3.ut.ee/). The Control Region was selected as it was 6kb from the MCS-R2 editing site, but also in an intron of NPRL3. Thus, it was hoped the genomic context was similar but that it would not be affected by sgRNA-cutting at the editing site.

qPCR primer efficiency was evaluated using an 8 point serial dilution series with a dilution factor of 3. Ct values were plotted against Log Copy Number on GraphPad Prism and the software was used to calculate the slope of the linear regression curve. The primer efficiency (E) was calculated using the following calculation:

$E = -1 + 10^{(-1/slope)}$

Three technical replicates were performed for each biological sample. Editing efficiency was calculated using the following equation: 2^-(Ct test – Ct control).

2.6.3 Editing Evaluation by Next Generation Sequencing

To create amplicon libraries compatible with the illumina sequencing machines, the R2 region was first amplified from gDNA using primers with additional adaptor sequences (R2 fwd 2 seq and R2 rev 2 seq). The initial PCR product was gel-purified using QIAquick Gel Extraction Kit (Qiagen) and diluted 50x before being amplified by a second PCR to append TruSeq index sequences and adaptors to each amplicon. A dual barcoding strategy was used to identify each biological sample using a unique combination of forward and reverse index primers (for details see appendices). These PCRs used Phusion High-Fidelity PCR Master Mix with GC Buffer and the following cycling conditions: initial denaturation (98 °C for 30 s), 13 amplification cycles (98°C for 10 s, 62°C for 10 s, 72°C for 10 s) and final extension (72°C for 5 min). Agencourt AMPure XP beads (0.75X, Beckman Coulter) at a ratio of 0.75 beads:DNA were used to purify the amplicon libraries which were then quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher). The samples were sequences using a 150bp PE v2 sequencing kit on the MiSeq (Illumina).

Next generation sequencing data were analysed using R (version 3.4.2). Fastq files were filtered for q scores of greater than 20 and reads were mapped to the human genome using the Burrows Wheeler alignment tool (bwa mem) and unmapped reads were filtered out. Adapter sequences were trimmed using trimLRPatterns, Biostrings package (version 2.44.2). Correctly edited reads were quantified by searching the filtered reads using the matchpattern function with max.mismatch of 2 for the Barcode

insert, Biostrings package (version 2.44.2). The editing efficiency was determined by dividing these correctly edited reads by the total number of locus reads.

2.7 Significance tests

Significance of efficiency of variant editing conditions compared to standard conditions were evaluated using the Dunnett test in GraphPad Prism.

For the ON:OFF ratio of molecular beacons, significance was calculated using t-tests with the Holms-sidak method with $\alpha = 0.05$. The null hypothesis assumed the target would have no effect on beacon fluorescence and thus the ON:OFF ratio would have a value of 1.

2.8 Molecular Beacon Handling and Design

2.8.1 Beacon Synthesis

GFP_7-21-0 was taken from the following paper[194], and manufactured by IDT with standard and phosphorothioate modified DNA bases (Na+ salt exchange, HPLC purification)[194]. Henceforth beacons were synthesized by Afaf El-Sagheer in Tom Brown's group, University of Oxford, or by ADTBio.

2.8.2 Beacon Handling and Storage

Beacons were resuspended in Nuclease-Free Water and aliquoted into 10 µL aliquots for short term storage at -20°C or 100 µL aliquots for long term storage at -80°C. Beacons were kept in foil covered boxes and exposure to light was minimised.

2.8.3 GFP Beacon Design

GFP Beacons (5-19-0, 4-19-0, 5-11-8) were designed manually using RNAfold to search for low-structure areas in the GFP mRNA. 5' stem sequences were added as appropriate. Internal DABCYL quenchers must be attached to a Thymine base, thus, it was ensured that for the wyvern beacon, 5-11-8, there was an appropriately placed T.

2.7.4 Wyvern Beacon Design

CCR5 Wyvern Molecular Beacons were designed using an R script created by David Knapp. The major constraint on beacon sequences was a T for the internal DABCYL quencher. The R script scanned the inverse CCR5 sequence for Ts identified those with a surrounding sequence which would fold appropriately with the of an exogenous 5' stem. Varying Beacon structures outlined in Table 2.1 were inputted into the script. The script scanned either the entire CCR5 (for CCR5 control region) or the region directly bridging the CCR5 D32, searching for complementary sequences with a high probability of folding into the given structure. Beacons were then selected from the output sequences. The control region was selected from the available beacon regions using RNAup program (Vienna RNA webservice) to quantify beacon folding with CCR5 mRNA and selecting the best candidates. The CCR5 WT sequence had a more restricted potential sequence space (the 32 base CCR5 deletion), so was designed by eye in a similar way to the GFP beacons.

Structural Notation	Stem (nt)	Loop (nt)	Tail (nt)
(((((.....))))).....	5	10	10
(((((.....))))).....	5	15	15
(((((.....))))).....	5	20	20
((((.....)))).....	4	10	11
(((((.....))))).....	5	12	15

Table 2.1 Input structures for the Molecular Beacon design script.

2.8.4 *In Silico* Analysis of Beacon Properties

RNA properties, were examined *In Silico* using the ViennaRNA Web Services (http://rna.tbi.univie.ac.at/): RNAcofold to evaluate the free energy of beacon target interaction, beacon ensemble diversity and frequency of structure, and RNAup to evaluate accessibility of the target mRNA.

DNA properties (for GFP_7-21-0) were examined using DNA folding form in mfold (http://unafold.rna.albany.edu/?q=mfold/DNA-Folding-Form).

Prior to use in experiments (both *in vitro* and *in vivo*), beacons were refolded using an annealing protocol. Beacons were diluted to 5 µM in 1x annealing buffer (1x PBS, 10mM $MgCl_2$) and refolded by heating to 85°C for 5 minutes and then by cooling to 25°C at a rate of 4°C/min. Beacons were then kept on ice in the dark until use.

2.8.5 Unique Landing Pad Design

The unique landing pad sequence was designed using an R Script created by David Knapp. Using the RNAfold package RNAInverse the script searched for the 'ideal' beacon structure: (((((.....)))))..... or 5-15-15. The script was constrained with the parameters that it must start with an A and have an U in the correct position for the internal DABCYL quencher. The script outputted 10 billion sequences which were then filtered to remove sequences with a GC content outside the ideal 50-60% and to remove sequences which were present in the human transcriptome (Hsapiens.UCSC.hg19.knownGene) with up to 12 mismatches. The remaining sequences were then refolded with the additional stop sequences, in all three reading frames, and XhoI restriction sequence (CTCGAG) in varying arrangements. Sequences where less than 35 bases were unstructured were removed. The remaining filtered sequences were inspected and the sequence with the highest probability of correct folding was selected.

2.9 In Vitro *Molecular Beacon analysis*

Beacons were diluted to 0.5 µM in 1x annealing buffer and then DNA or RNA targets were added at a concentration of 2.5 µM to a final volume of 25 µL. 20 µL of the beacon:target mix was added to a 384 well black, µClear, medium-binding microplate (Greiner Bio-One) and incubated in the dark for 10 min. Fluorescence intensity was analysed on a ClarioStar Microplate Reader (BGS). The machine was calibrated to 90% of maximum using 0.2 µM Rhodamine Dye (for TAMRA measurements) and 0.5 µM AF647 Goat anti-mouse IgG antibody (biolegend) (for AF647 measurements). For experiments with cellular RNA, cellular RNA was added to a final concentration of 6.5 µg/µL.

ON:OFF ratios were calculated by dividing the beacon fluorescence intensity with the target by the beacon fluorescence intensity without the target. Short DNA targets were ordered from IDT.

2.9.1 Long RNA Target production using *In Vitro* Transcription

Long RNA was produced from a template dsDNA PCR product manufactured using Phusion High Fidelilty Master Mix with GC buffer. The T7 promoter was appended and sequences were amplified from plasmids using the following primers, for GFP and CCR5 variants respectively: T7-EGFP fwd and EGFP-PA rev, T7 CCR5 pAC360 fwd and CCR5 PA rev. The mMessage mMachine T7 transcription kit (Thermo Fisher) was then used to produce full length mRNA, as per the manufacturer's instructions using an overnight incubation step at 37°C. DNA template was digested by addition of 1 µL of Turbo DNase per reaction and incubation for 15 min. RNA was precipitated by addition of 30 µL of Nuclease Free Water and 30 µL of 7.5 M LiCl per reaction, vortexing and incubating at -20°C for 30 min. The RNA was then pelleted by centrifuging at 20000g for 30 min at -4°C. The supernatant was removed and the pellet was washed twice with 70% ethanol, then dried and resuspended in nuclease free TE buffer. Long RNA was quantified using the Nanodrop 2000 Spectrophotometer (Thermo Fisher).

2.9.2 Short RNA Target production using *In Vitro* Transcription

Short RNA was produced with the MEGAscript T7 transcription kit (Thermo Fisher) from annealed template DNA oligos – T7 promoter F and respective T7 reverse oligos (sequences in appendix). 2 µM stock solution of annealed template oligos were produced by addition of 1 µL of each 100 µM oligo to 48 µL nuclease free water. The stock solution was then melted by heating to 95°C for 5 minutes and then annealed by cooling to 25°C at a rate of 4°C/min. 1 µL of this stock solution was then added to the MEGAscript reaction mix following the manufacturer's instructions. The IVT reaction was incubated at 37°C overnight. DNA template was digested by addition of 1 µL of turbo DNase per reaction and incubation for 15 min. The RNA was then extracted using phenol-chloroform precipitation. 1 µL of glycogen (to aid in pelleting of short RNA), 111 µL of nuclease free water and 15 µL of ammonium acetate stop solution were added to each IVT reaction and vortexed. 150 µL of Acid

Phenol Chloroform 5:1 (Ambion) was then added and vortexed. The phases were separated by centrifuging at 1000g for 5 min at 4°C. The supernatant phase was then removed and vortexed with 150 µL isopropanol to precipitate RNA, before incubation at -20°C for 1 hour. The RNA was pelleted by centrifuging at 20000g for 30 min at 4°C. The supernatant was removed, and the pellet washed twice with 70% ethanol and dried, before resuspension in nuclease free TE buffer. Short RNA was quantified using the Qubit™ RNA HS Assay Kit (Thermo Fisher).

2.9.3 Long ssDNA Target production using Lambda Exonuclease

The long ssDNA production protocol was kindly provided by Yavor Bozhilov. dsDNA templates were produced by amplifying DNA with 5'Phosphorylated forward and 5'Phosphorothioated reverse primers using Phusion High Fidelilty Master Mix with GC Buffer. PCR reactions were scaled up 50x 20 µL to produce ~50 µg dsDNA template, which were purified using the Qiagen PCR Purification kit. The 5'phosphorylated strand was then digested using lambda exonuclease (NEB) in lambda nuclease buffer according to manufacturer's instructions at 37°C for 1 hour. The single stranded DNA was then purified using the Qiagen Gel extraction kit and minElute columns (using QG buffer instead of the PCR purification PB buffer). 1% Agarose gel electrophoresis with GelRed Nucleic Acid Gel Stain (Biotium) was then used to visualise the ssDNA to check the size and purity. To denature ssDNA, samples were diluted in RNA loading dye (2x) (NEB) and heated to 95°C for 2 min prior to loading on gel.

2.10 CCR5 and Landing Pad Plasmid Transfection

2.10.1 S2 cell TransIT-2020 transfection

S2 cell transfection was shared between myself and Markus Toegel. 24 hours before transfection $1x10^6$ S2 cells were seeded in 2ml *Drosophila* Schneider's medium with 10% heat-inactivated FBS in a 6 well plate. Cells were all transfected with 1 µg pmCerulean and then 1 µg of either pAC-360-Syn21-CCR5-D32-P10 or pAC-360-Syn21-CCR5-P10. Plasmids were mixed with H_2O to a final volume of 47 µL in sterile conditions. 3 µL room temperature TransIT-2020 (MirusBio) was then added and mixed by

flicking gently. The DNA:Transit2020 mix was incubated at room temperature for 40 minutes before transferring dropwise onto cells incubating in 1 ml *Drosophila* Schneider's medium with 2% heat-inactivated FBS. 4 hours post-transfection media was topped up with 1ml of *Drosophila* Schneider's medium with 10% heat-inactivated FBS. Beacons were electroporated into cells 48 hours post-transfection.

2.10.2 HEK293T PEI transfection

24 hours before transfection $5x10^4$ HEK293T cells were seeded in 1 ml DMEM + 10% FBS in a 12 well plate. Cells were transfected with 0.25 µg of either pll3.7_bi_dir_CCR5_dECFP, pll3.7_bi_dir_CCR5-D32_dECFP, or pll3.7_bi_dir_CCR5-insert_dECFP. The plasmids were mixed in sterile conditions with Gibco Opti-MEM (Thermo Fisher) to a final volume of 48.5 µL and then vortexed with 1.5 µL 1 µg/µL polyethylenimine (PEI). The plasmid:PEI mix was incubated at room temperature for 20-30 min before being transferred dropwise onto cells incubated in DMEM + 2% FBS. 4 hours post-transfection, transfection media was removed and replaced with DMEM + 10% FBS. Beacons were electroporated into cells 48 hours post-transfection.

2.11 Molecular Beacon Electroporation

Prior to electroporation, beacons were refolded according to the above protocol and stored on ice in the dark until use. Beacons were electroporated into cells using the Neon Transfection System and 10 µL Neon Transfection kit (Thermo Fisher) according to the manufacturer's instructions.

For experiments in OCI-AML5 cells with GFP beacons, GFP and WT cells were counted and mixed in equal ratios prior to cell pelleting. For experiments in S2 cells, T cell clones, and HEK293T cells, cells were not mixed prior to pelleting. For each sample $2.2x10^5$ cells were pelleted and washed with PBS, then resuspended in Neon Buffer with 100 nM Beacon and electroporated according to the conditions stated in the Table 2.2. Cells were immediately ejected into warm FACS buffer and incubated at 37°C for 20 min. Viability dye (specified in table below) was then added to the cells and they were stored

on ice in the dark whilst analysed by flow cytometry. Cells were filtered with 70 µM cell strainers prior to flow cytometry analysis.

Cell Line	Buffer	Voltage (mV)	Pulse	Time (ms)	Viability Dye
OCI-AML5	R	1700	1	20	DAPI
S2	R	1700	1	20	Zombie Dye-NIR
T cell clones	T	1600	3	10	DAPI
HEK293T	R	1150	2	20	PI

Table 2.2: Neon Electroporation Conditions and Viability Dye for each cell line.

2.11.1 CCR5 Antibody Staining

For CCR5 beacon experiments, T cell clones and S2 cells were stained for CCR5 expression prior to neon electroporation. Cells were washed once with PBS, then stained in PBS for 15 min on ice with AF647 anti-human CD195 (CCR5) at a dilution of 1:100. After staining, cells were washed twice with FACS buffer and electroporated as normal.

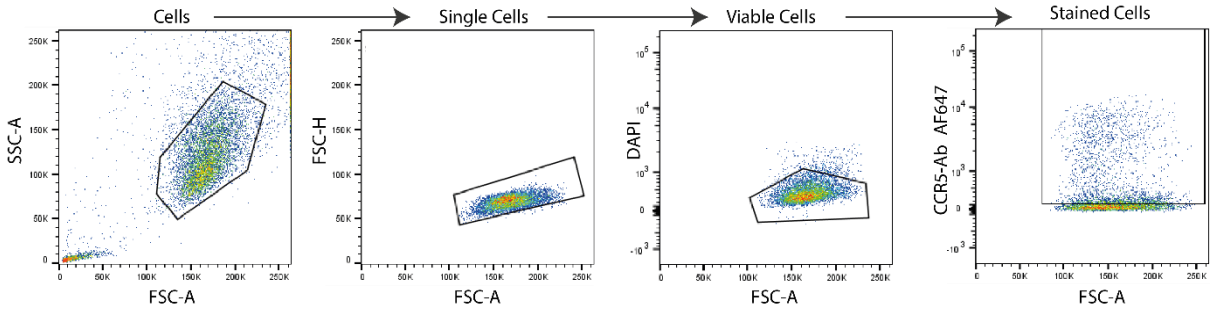


Figure 2.2 Gating strategy for CCR5-Ab AF647 stained S2 cells

2.11.2 *In Vivo* Molecular Beacon Anaylsis by Flow Cytometry

Unless otherwise stated, all flow-cytometry experiments were performed on the BD LSR Fortessa Analyser. For flow cytometry analysis, David Knapp assisted with machine set up and running, and compensation. Untransfected, unelectroporated or unstained cells were used to set analysis gates. Cells were gated using SSC-A/FSC-A, single cells were gated using FSC-H/FSC-A, and viable cells were gated using Viability Dye (specified in Table 2.2)/FSC-A. CCR5 positive cells were gated using AF647/FSC-A (fig 2.2). Flow cytometry data was analysed using FlowJo v10.6.1.

2.12 CCR5 expression analysis in T cell clones

T cell clone RNA was extracted within 24 hours of the respective beacon electroporation and CCR5 expression was analysed.

2.12.1 RNA extraction and Reverse Transcription

RNA was extracted from 2.5x10^5 cells using the ChargeSwitch Total RNA Cell Kit (Thermo Fisher) according to the manufacturer's instructions, albeit using 0.25x volumes (25 µL magnetic beads and 50 µL binding buffer). cDNA was produced using the Quantitect Reverse Transcription Kit (Qiagen), following the manufacturer's instructions and using 1 µg RNA. cDNA was then diluted 1:5 in nuclease free water and used for RT-qPCR.

2.12.2 CCR5 RT-qPCR

RT-qPCR was performed on a CFX384 real-time thermocycler (BioRad) with SsoAdvanced Universal SYBR Green Supermix. Housekeeping GUSB qPCR primers and CCR5 primers were designed using Primer3 (http://primer3.ut.ee) (see appendix for qPCR primer sequences: GUSB Fwd 1 and GUSB rev 1, CCR5-D32-F1 and CCR5-D32-R2). qPCR primer efficiency was evaluated as discussed above in Section 2.6.2.

Three technical replicates were performed for each biological sample.

2.13 Qiazol RNA extraction

Concentrated whole cell RNA for molecular beacon background RNA *in vitro* experiments was extracted from 3x10^7 HEK293T cells using QIAzol Lysis Reagents (Qiagen). Cells were trypsinized and resuspended in DMEM + 10% FBS, then pelleted and weighed. Cells were resuspended in 1 ml of QIAzol lysis Reagent was added per 100mg tissue and incubated at room temperature for 5 min. 0.2 ml of Chloroform was added per 1 ml of QIAzol lysis reagent and tubes were vortexed vigorously, before further incubation at room temperature for 3 min. The upper aqueous phase was taken and vortexed with 0.5 ml isopropanol per 1ml QIAzol lysis reagent. Tubes were then incubated at room

temperature for 10 min, before centrifugation at 12000 g for 10min at 4°C to pellet the precipitated RNA. RNA pellets were washed twice with 70% ethanol before drying and resuspending in 50 µL nuclease free water. RNA was quantified using the Nanodrop 2000 Spectrophotometer (Thermo Fisher).

2.14 Circuit optimisation experiments

2.14.1 Plasmids used in circuit experiments

Experiment	Mix	Plasmids in Mix	Sample Specific Plasmids
Spacer Length variants	Cas9	px330 (no guide) MCP-P65-HSF1 piRFP670 CTS1-EYFP_tracrRNA	U6_placeholder-crRNA 20 bp CTS1 crRNA 15 bp CTS1 crRNA 14 bp CTS1 crRNA 13 bp CTS1 crRNA 12 bp CTS1 crRNA 11 bp CTS1 crRNA 10 bp CTS1 crRNA
	dCas9	dCas9 MCP-P65-HSF1 piRFP670 CTS1-EYFP_tracrRNA	U6_placeholder-crRNA 20 bp CTS1 crRNA 15 bp CTS1 crRNA 14 bp CTS1 crRNA 13 bp CTS1 crRNA 12 bp CTS1 crRNA 11 bp CTS1 crRNA 10 bp CTS1 crRNA
Cas9 Activator Variants	Cas9	px330 (no guide) MCP-P65-HSF1 piRFP670 CTS1-EYFP_tracrRNA	U6_placeholder-crRNA 20 bp CTS1 crRNA 11 bp CTS1 crRNA
	dCas9	dCas9 MCP-P65-HSF1 piRFP670 CTS1-EYFP_tracrRNA	U6_placeholder-crRNA 20 bp CTS1 crRNA 11 bp CTS1 crRNA
	SAM	dCas9-VP64 MCP-P65-HSF1 piRFP670 CTS1-EYFP_tracrRNA	U6_placeholder-crRNA 20 bp CTS1 crRNA 11 bp CTS1 crRNA
	VPR	dCas9-VP64-p65-RTA MCP-P65-HSF1 piRFP670 CTS1-EYFP_tracrRNA	U6_placeholder-crRNA 20 bp CTS1 crRNA 11 bp CTS1 crRNA

sgRNA exchange	Reporter	CTS1-EYFP_tracrRNA piRFP670 MCP-P65-HSF1 CTS2-ECFP reporter	CTS1 sgRNA-Cas9 (RNP – not plasmid) 11 bp CTS2 crRNA
Time Course	Negative Control	piRFP670 MCP-P65-HSF1 dCas9-VP64 CTS1-EYFP_tracrRNA	
	Circuit	piRFP670 MCP-P65-HSF1 dCas9-VP64 11 bp CTS1 crRNA CTS1-EYFP_tracrRNA	
Editing Detection	Positive Control	piRFP670 MCP-P65-HSF1 dCas9-VP64 CTS1-EYFP_tracrRNA 11 bp CTS1 crRNA	
	Circuit	piRFP670 MCP-P65-HSF1 dCas9-VP64 CTS1-EYFP_tracrRNA	Cas9 (protein) Cas9 + BRCA1 sgRNA (RNP) BRCA1-CTS1 donor

Table 2.3 Plasmid Mastermixes and sample-specific plasmids and RNPs for each experiment.

2.14.2 Transfection of circuit components into HEK293T cells

Cells were seeded and circuit components were transfected with PEI as described in section 2.10.2. Equimolar amounts of each plasmid (specifics in Table 2.3) were mixed prior to optimum and PEI addition so each plasmid had a final concentration of 0.1 nM in the 250 µL transfection media. Cells were harvested using trypsin, washed with PBS and resuspended in FACS buffer with DAPI for flow cytometry analysis.

Untransfected cells were used to set analysis gates. Cells were gated using SSC-A/FSC-A, single cells were gated using FSC-H/FSC-A, viable cells were gated using DAPI/FSC-A, transfected and reporter-active cells were gated using iRFP670/EYFP. Flow cytometry data was analysed using FlowJo v10.6.1.

2.14.3 sgRNA exchange experiments

2.14.3.1 CTS1-sgRNA production by IVT

The 11 bp CTS1 sgRNA was amplified and T7 promoter added using the Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB) and T7-CTS1 fwd and sgRNA rev primers (sequences in appendices), according to the PCR protocol in section 2.6.3. The PCR product was purified using the MinElute PCR purification kit (Qiagen) and quantified using the Nanodrop 2000 Spectrophotometer (Thermo Fisher). 1 µg of PCR product was added to the T7 MEGAscript reaction mix following the manufacturer's instructions. The IVT reaction was incubated at 37°C overnight. DNA template was digested by addition of 1 µL of turbo DNase per reaction and incubation for 15 min. RNA was purified according to the phenol-chloroform precipitation described in section 2.9.2.

2.14.3.2 sgRNA experimental scheme

On day 1 plasmid-encoded components (Reporter mix - Table 2.3) were transfected as described in section 2.10.2 and 2.14.1. On day 2, CTS1 sgRNA-Cas9 RNPS were electroporated into cells using the Neon Transfection System and 10 µL Neon Transfection kit (Thermo Fisher). 5 µg Engen Spy Cas9 NLS (New England Biolabs), 2.5 µg 11 bp CTS1 sgRNA were mixed and incubated at room temperature for 20 min. Transfected cells from day 1 were harvested with trypsin and counted. $2x10^5$ cells for each sample were aliquoted, pelleted, washed with PBS, re-pelleted and resuspended in 12 µL Neon Buffer R. Cells were then mixed with the RNP and electroporated using the program 1150mV, 2 zaps, 20ms, according to the manufacturer's instructions. On day 4, cells were harvested as described in section 2.14.1 and analysed by flowcytometry.

Untransfected cells were used to set analysis gates. Cells were gated using SSC-A/FSC-A, single cells were gated using FSC-H/FSC-A, viable cells were gated using PI/FSC-A, CTS1-EYFP and CTS2-ECFP reporter-active cells were gated using EYFP/ECFP. Flow cytometry data was analysed using FlowJo v10.6.1.

2.14.4 Time Course Experiments

Circuit plasmids (see Table 2.3) were transfected into HEK293T cells as described in section 2.10.2 and 2.14.1. Plasmid, optimum and PEI transfection mixes were scaled up to 13x master mixes for

transfecting 12 wells for 6 time points each for dividing cells and non-dividing cells. 4 hours post transfection, media in dividing wells was exchanged for 500 µL DMEM + 10% FBS. Media in non-dividing wells was topped up with 250 µL DMEM + 2% FBS. Wells with dividing cells were trypsinized and diluted 1 in 4 on days 3 and 6. Media in non-dividing wells was replaced on days 3 and 6. Cells were harvested and analysed by flow cytometry as in section 2.14.1 within a time margin of 1-2 hours of the following time points post transfection: 1 day, 2 days, 3 days, 4 days, 6 days, 8 days.

Time course graphs were plotted as % total viable cells that are EYFP positive against time using GraphPad Prism. Lines of best fit were calculated using non-linear regression, where x was linear and y was logarithmic, and confidence intervals of 95% are shown on the graphs.

2.15 Detection of Editing using Circuit system

2.15.1 Sequential electroporation of editing and circuit components

2.15.1.1 Delivery of editing components

HEK293T cells were split on Day 0 and electroporated with editing components on Day 1, to ensure cells would be dividing rapidly when editing components were delivered. Plasmids were electroporated into cells using the Neon Transfection System and 10 µL Neon Transfection kit (Thermo Fisher). 500 ng px459, with or without BRCA-1 sgRNA (addgene #62988), or Cpf1 plasmid (addgene #69988), with or without Cpf1 crRNA plasmid (modified addgene #78957 with BRCA-1 crRNA) (spacer sequences in appendices) were mixed together with 50 pmol BRCA1-CTS1crRNA oligo donor, prior to electroporation. Cells were harvested with trypsin and counted. $1x10^5$ cells for each sample were aliquoted, pelleted, washed with PBS, re-pelleted and resuspended in 12 µL Neon Buffer R. Cells were then mixed with the plasmid:donor mixes and electroporated using the program 1150mV, 2 zaps, 20ms, according to the manufacturer's instructions. Post electroporation cells were incubated in 1 ml warm DMEM + 10% FBS in a 12 wells plate.

2.15.1.2 Delivery of Circuit components

3 days post editing component electroporation, cells were harvested, counted and $2x10^5$ cells were electroporated again with circuit detection components (see Table 2.3). Equimolar concentrations of plasmids with a combined mass of 750 ng were used and cells were electroporated as above in section 2.15.1.1. 3 days post electroporation cells were harvested and FACS was performed. An aliquot of cells was taken prior to sorting and placed on ice until gDNA extraction as described in section 2.15.3.1.

2.15.2 Bulk Sorting of Edited Cells using Circuit System

FACS was performed on the BD FACSAria™ Fusion (BD Biosciences) by Kevin Clark. Untransfected cells were used to set analysis gates for cells, single cells and viable cells (fig 2.3). Positive control cells were used to set the gate for iRFP670. Negative control cells were used to set the gates for EYFP. Cells were gated using SSC-A/FSC-A, single cells were gated using FSC-H/FSC-A, viable cells were gated using DAPI/FSC-A, transfected and reporter-active cells were gated using iRFP670/EYFP. Cells were sorted on the 'high purity' setting into PBS + 2% FBS. 200000 iRFP670+ EYFP- cells were sorted for each sample. All iRFP670+ EYFP+ cells were sorted for each sample (500-2000 cells).

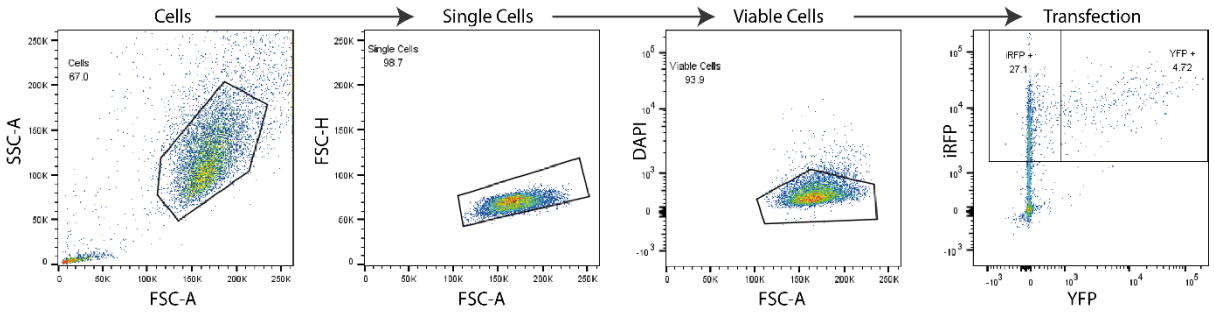


Figure 2.3 Gating strategy on BD FACSAria Fusion for bulk sorting of iRFP670+ EYFP- and iRFP670+ EYFP+ cells.

2.15.3 Editing Analysis of Sorted Cells

2.15.3.1 gDNA extraction and PCR of BRCA*-1 locus*

Unsorted, EYFP-, and EYFP+ cells were pelleted by centrifuging at 300g for 5 min. Cell pellets were resuspended in gDNA extraction buffer (50 µM Tris-HCl pH 8, 1 µM EDTA pH 8, 0.5% Tween-20, and 16 U/mL freshly added Proteinase K (NEB)). Unsorted and EYFP- cell pellets were resuspended in 100 µL gDNA extraction buffer and EYFP+ pellets were resuspended in 20 µL gDNA extraction buffer, prior

to being incubated for 60 min at 37°C. Samples were then incubated at 95°C for 10 min. 1 µL of gDNA and 0.5 µM BRCA1 forward and reverse primers (sequences in appendices) were added to a final volume of 50 µL Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB) and the following cycling conditions: initial denaturation (98 °C for 30 s), 33 amplification cycles (98 °C for 10 s, 64 °C for 12 s, 72 °C for 12 s) and final extension (72 °C for 5 min). PCR products were run on a 1% agarose gel with GelRed Nucleic Acid Gel Stain (Biotium) beside a 100 bp DNA ladder (NEB) to verify correct product amplification. PCR products were then purified using the MinElute PCR purification kit (Qiagen) and quantified using the Nanodrop 2000 Spectrophotometer (Thermo Fisher).

2.15.3.2 TIDE analysis of sorted cells

The PCR product was sequenced by Eurofins Genomics overnight sequencing service. Samples were submitted according to the Eurofins sample submission guide. The .ab1 files received were then evaluated using the TIDE software(TIDE: Easy quantitative assessment of genome editing experiments by sequence trace decomposition, https://www.tide.deskgen.com)[192,193]. The sequenced PCR product from untransfected cells was used as a baseline to measure the sorted cells against.

2.15.3.3 Agarose Gel analysis of sorted cells

50ng of PCR product was visualised on a 1% agarose gel with GelRed Nucleic Acid Gel Stain (Biotium) beside a 100 bp DNA ladder (NEB). ~330 bp (edited) and 230 bp (unedited) bands were quantified using imageJ analysis software and HDR percentages were calculated by dividing volume of ~330 bp bands, by total volume of both bands for each sample.

2.15.4 Simultaneous Delivery of Editing and Circuit Components

Components were delivered exactly as in the second stage of 2.15.1.2, except prior to electroporation the Cas9 (IDT) and BRCA-1 sgRNA (synthego) RNP was added to the circuit component mix. 0.5 µL Cas9 (62 µM) and 0.5 µL BRCA-1 sgRNA (125 µM) were premixed at RT for 10 minutes before mixing with plasmids and donor. Flow cytometry proceeded as in 2.15.2.

Chapter 3: Evaluating published methods of improving editing efficiency in HUDEP-2 and CD34⁺ cells

3.1 Introduction

In this chapter I describe my efforts to evaluate current methods for improving efficiency of HDR. To do this I focus on a single therapeutically relevant locus, the MCS-R2 enhancer, in two different cell types: HUDEP-2 and CD34+ cells.

3.1.1 Methods for improving the efficiency of gene editing

As discussed in the introductory chapter, the extent of involvement of different repair pathways in gene editing is poorly understood and varies between cell lines, loci and with different synthetic donor types. Therefore, for simplicity, in this chapter I focus on editing in only the MCS-R2 locus with short oligo (ssODN) donors (100-200 nt).

I attempted to evaluate the effect on editing on four of the most commonly used methods discussed in section 1.4: SCR7, the DNA Ligase IV inhibitor[101,102,195]; RS-1, the Rad51 stimulator[107,196]; phosphorothioate modified ssODNs[110]; and asymmetric ssODNs[114]. There is a high degree of variance

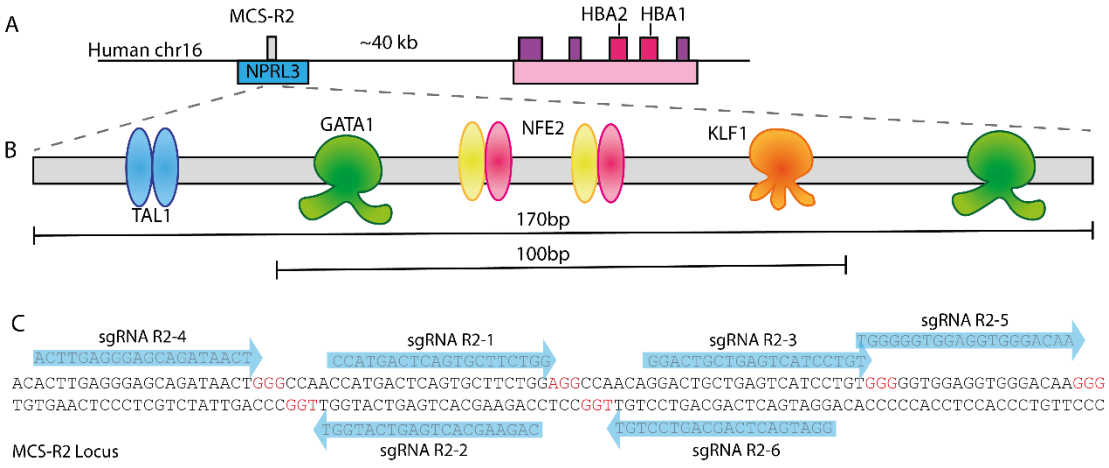


Figure 3.1. MCS-R2 sgRNA design. (A) Schematic depicting MCS-R2 (grey) within intron 5 of NPRL3 (blue) relative to the alpha globin genes (HBA1 and HBA2 in pink). (B) Enlarged diagram of MCS-R2. Transcription factor binding sites for TAL1 (blue), GATA1 (green), NFE2 (pink and yellow), and KLF1 (orange) are shown. (C) 100 bp central sequence of MCS-R2. sgRNA sequences are shown as blue arrows pointing 5'-3'. PAM sequences are coloured red.

in the success of these different methods across different papers, different cell types, and different loci. I attempted to systematically compare their efficacy to improve the efficiency of gene editing.

3.1.2 HUDEP-2 cells as a model for CD34+ cells

To test these different methods, I selected the MCS-R2 locus: a potential therapeutic target for the treatment of β-thalassemia. The proposed method for treatment of thalassemia patients involves harvesting the patient's bone marrow, sorting for the $CD34^+$ cells, editing these cells to delete regions of the MCS-R2 enhancer, and transplant these edited cells back into the patient[197]. $CD34^+$ cells are notoriously difficult and expensive to culture. Therefore, I completed my initial optimisation experiments in an immortalised $CD34^+$ cell line, HUDEP-2[198].

HUDEP-2 cells were produced by lentiviral insertion of the HPV16-E6 and -E7 proteins, which drive immortality[199]. They were originally created as an *in vitro* source of red blood cells, as they can differentiate into reticulocytes. However, since then they have been used extensively to study blood diseases, e.g. sideroblastic anaemia, β-thalassemia, and sickle-cell disease, and have proven to be a good stand-in for $CD34^+$ cells in these instances[200,201].

In this chapter, I describe our efforts to improve the efficiency of gene editing at the MCS-R2 locus, evaluating my results through a combination of NGS, RT-qPCR and restriction digest.

3.2 Results

3.2.1 Testing sgRNAs

I first designed six sgRNAs which spanned the central 100 bp of MCS-R2, covering the central four transcription factor binding sites (fig 3.1.A)[202]. To do this I used the CRISPRko tool from the Broad Institute (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design)[191]. As efficiency of Cas9 cutting varies greatly, even for sgRNAs in close proximity[150], I endeavoured to both use *in silico*-based sgRNA prediction software, and to functionally test multiple sgRNAs for a single locus. I selected

sgRNAs from the resulting list ensuring I selected sgRNAs targeting both the forward and reverse strand with relatively different sequences (fig 3.1.B). I hypothesized that this would give me the best chance of finding an effective sgRNA.

To test the cutting efficiency of the MCS-R2-targeting sgRNAs, I first delivered a plasmid expressing the sgRNA and Cas9 into HUDEP-2 cells via electroporation (fig. 3.2.A). I used the Zhang lab plasmid

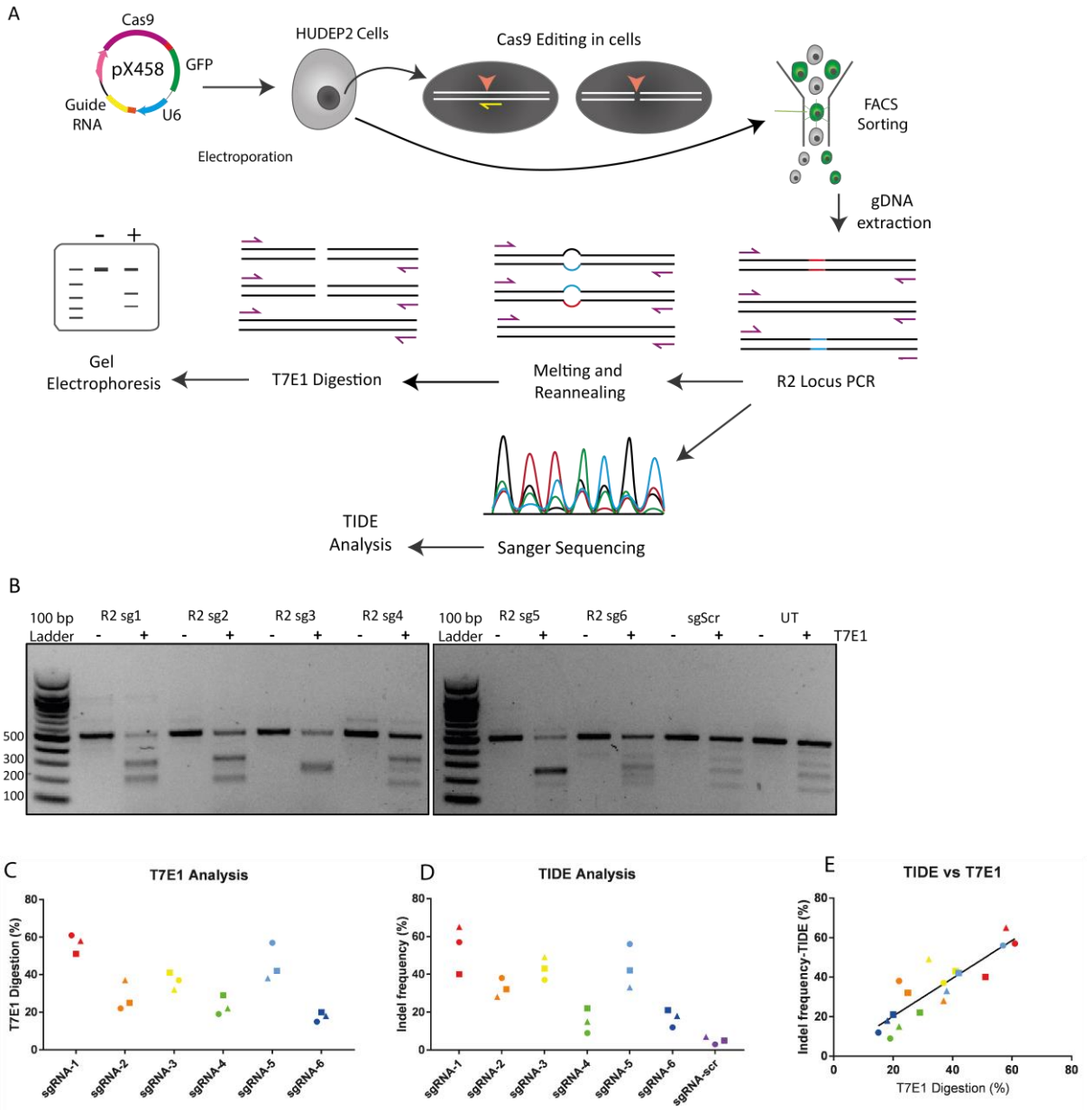


Figure 3.2 Evaluating sgRNA efficiency. (A) Schematic depicting the protocol used for editing, processing and analysing HUDEP-2 cells. (B) Typical agarose gel of the MCS-R2 amplicon, ± T7E1, for cells transfected with each sgRNA. (C) Quantification of T7E1 digestion for three biologicsl replicates. (D) Plot of Indel Frequency calculated using TIDE software for three biological replicates of each sgRNA. (E) Graph showing the Indel frequency calculated by TIDE vs the indel frequency calculated by T7E1 assay for each biological replicate.

pX458 which expresses Cas9 and GFP simultaneously using a T2A self-cleaving peptide. This enabled me to sort Cas9-expressing cells on GFP expression. 72 hours after electroporation, I sorted the cells and amplified the R2 locus from the extracted gDNA. I evaluated the sgRNA cutting efficiency using both the T7 endonuclease 1 (T7E1) assay, which cleaves DNA heteroduplexes, and TIDE software, which quantifies editing using sanger sequencing trace decomposition (fig. 3.2.B). TIDE and T7E1 measurements were highly correlated with an R^2 value of 0.76 (fig 3.2.E). sgRNA cutting efficiency varied greatly from ~15% for R2sg6 and ~60% for R2sg1 (fig 3.2.C, 3.2.D). Thus, for editing experiments, I decided to use the sgRNA R2sg1.

3.2.2 Effect of Different Donors on Editing Efficiency

I wanted to evaluate the effect of different donor templates on the efficiency of editing. I designed five different oligo donors, the first of which, R2 D 45-45 T, was the standard DNA donor, with two 45 bp arms, homologous to the target strand and, therefore, complementary to the non-target strand (fig 3.3.A). The subsequent four donors I designed to mimic the format of published donor templates, described above, which had been shown to improve the efficiency of editing[114,126]. I then quantified and compared the editing efficiency of these donors, to see if these published findings were relevant in HUDEP-2 cells at the MCS-R2 locus.

Following the findings of Richardson et al, I designed two asymmetric donors with a longer homology arm on the PAM-distal side, which were complementary to the non-target strand[114]. The most successful donors found in this paper had arms of 91 nt and 36 nt, and 26 nt and 71 nt, respectively. I attempted to mirror these findings by testing one donor with 90 nt and 30 nt arms, (R2 D 90-30), and one donor with 15 nt and 75 nt arms, (R2 D 75-15).

I also wanted to test the hypothesis that target strand-homologous ssODNs were more effective than nontarget strand-homologous ssODNs[114]. Thus, I included the nontarget strand homologous R2 D 45-45 NT ssODN: a reverse complement of R2 D 45-45 T.

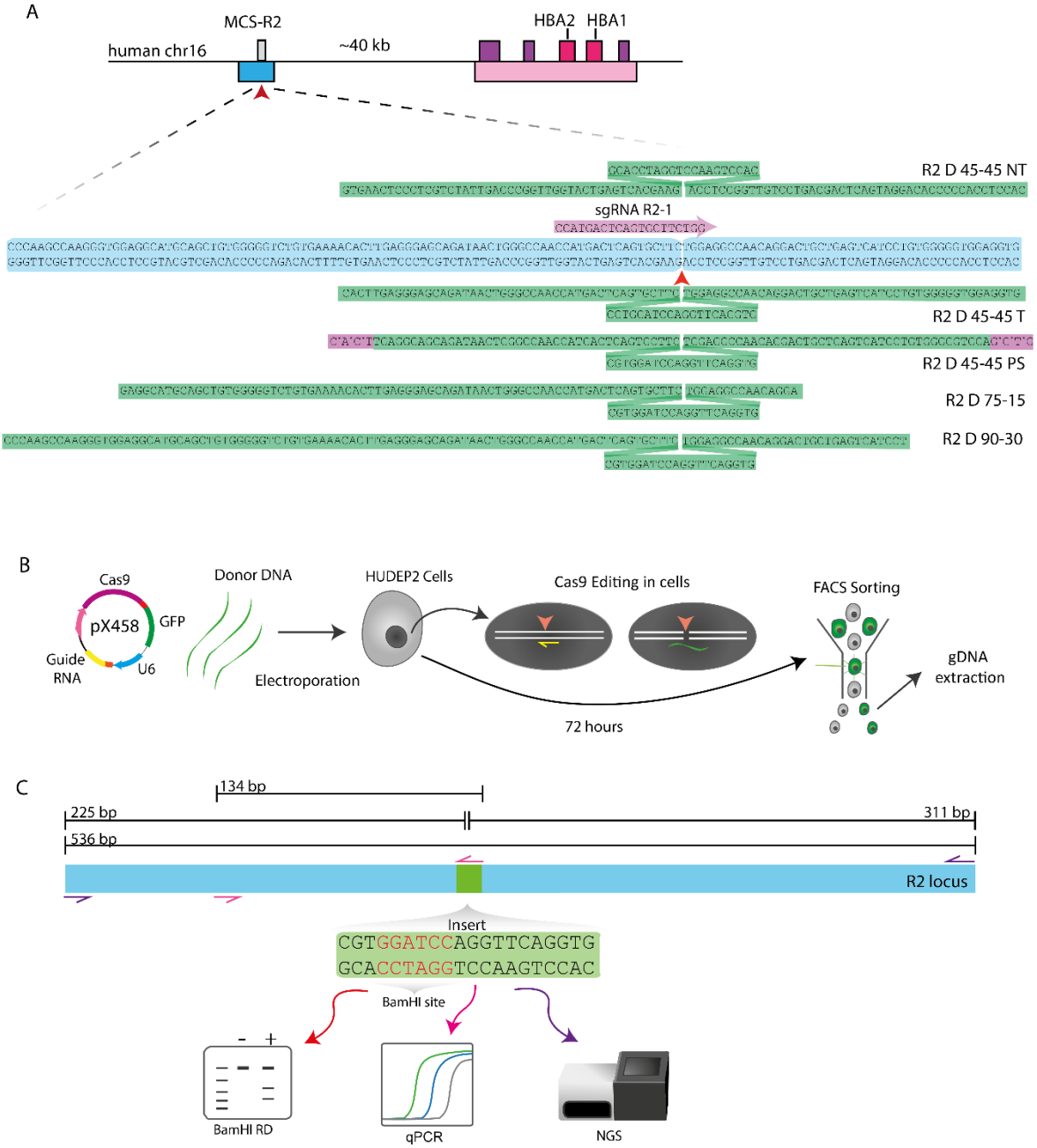


Figure 3.3. Donor Design and Analysis Protocol (A) Schematic depicting the MCS-R2 locus, R2sg1 (pink) and R2 donors (green) in the context of human chromosome 16. The red arrow shows the cut site. Phosphorothioated bases are coloured red. (B) Diagram showing the sequence of events between electroporation of cells and gDNA extraction. (C) The Editing Evaluation Methods. The R2 locus is shown in blue and the insert in green. The BamHI restriction site is coloured red. R2 amplification primers are purple and qPCR primers are pink.

My final donor, R2 D 45-45 PS, was identical to R2 D 45-45 T, except at both the 3' and 5' ends the final 3 nucleotides were phosphorothioate modified.

In order to rapidly and efficiently quantify editing efficiency, I included a 20 bp barcode insert sequence in the donor templates (fig. 3.3.C). The insert sequence has a dual purpose: it acts as a specific binding site for a reverse qPCR primer; and it includes a BamHI restriction site to allow analysis by restriction digest. Therefore, I was able to quantify editing efficiency rapidly through two parallel methods. I also initially compared editing efficiency quantified via restriction digest and qPCR to editing efficiency quantified using next generation sequencing (NGS), to determine if these two methods produced accurate results (fig 3.5.D). I designed the insert sequence to have low homology to the surrounding MCS-R2 sequence, in order to reduce non-specific qPCR primer binding to the surrounding area. However, the low homology and relatively large size of the insert (20 bp) could potentially reduce editing efficiency.

I delivered the oligo donor and plasmid-encoded sgRNA and Cas9 into HUDEP-2 cells via electroporation (fig 3.3.B). 72 hours later, I sorted cells on GFP expression and extracted gDNA for analysis (for detailed methods and sorting gates see section 2.3.4). I amplified the MCS-R2 locus from the gDNA, digested with BamHI and analysed agarose gel band absorption. BamHI digestion of correctly edited bands gave two bands of ~200 bp and ~300 bp, respectively, whereas, unedited bands were 500 bp. I calculated editing efficiency by dividing the combined emission of both digested bands by the total emission of all three bands.

Alternatively, gDNA was used directly in qPCR. The reverse primer was specific to the insert sequence, while the forward primer bound outside of the edited region. I used control primers, positioned in the MCS-R3 enhancer in the 7th intron of NPRL3[203], to normalise for initial DNA input (fig 3.5.A and 3.5.B). I hypothesized that this control region was likely in a very similar genomic context to MCS-R2 but was

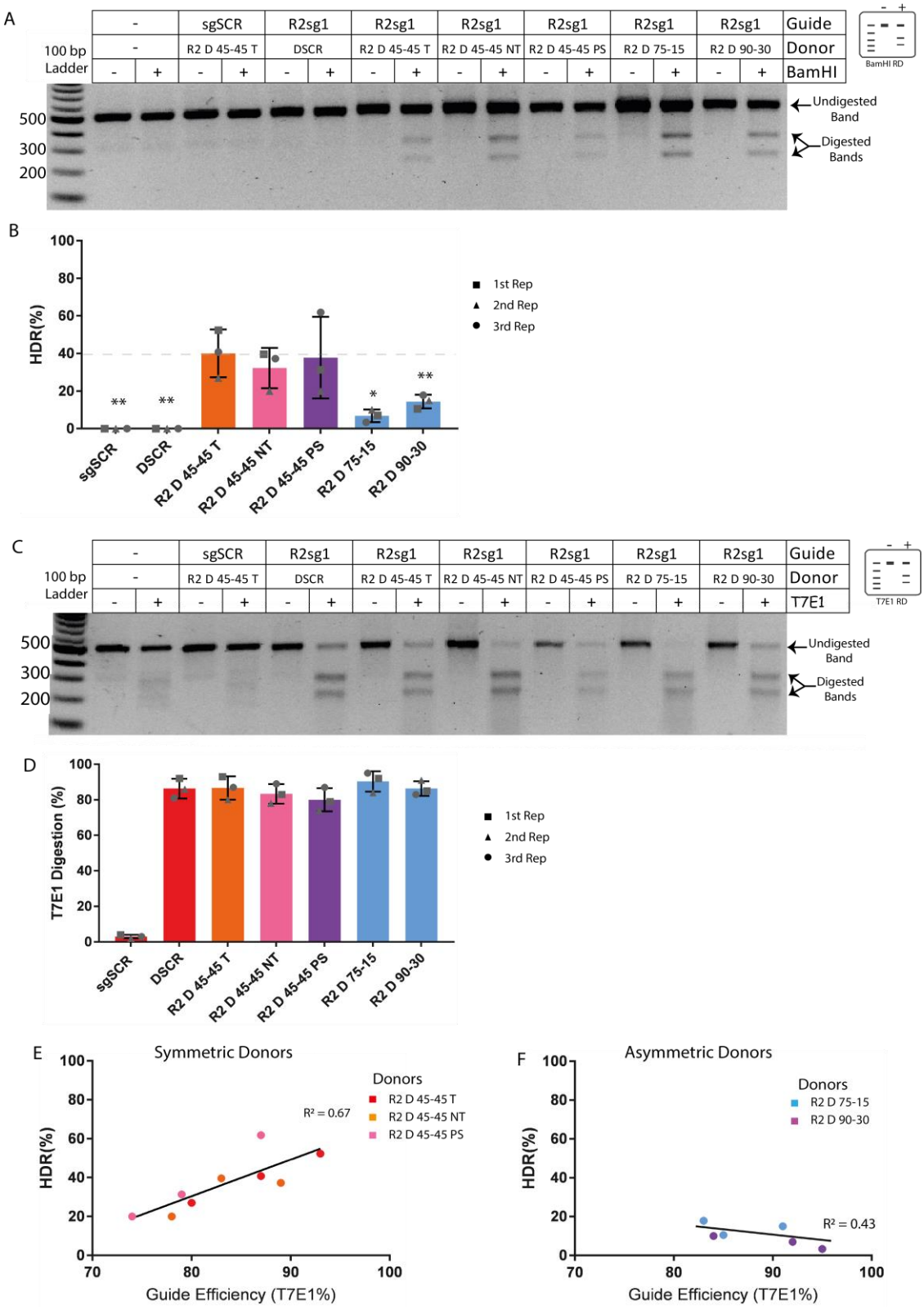


Figure 3.4 Evaluating editing efficiency of different donor templates using restriction digest. (A and C) Example of a typical agarose gel showing the products of the MCS-R2 amplicon ± BamHI (A) and T7E1 (C). Undigested and digested bands are labelled with arrows. sgRNA and Donor used for editing are shown in the table above the gel. (B and D) Band absorptior quantification for A and C. Each point represents an individual biological replicate. Error bars show SD. Bars show mean. Grey dashed line represents the mean efficiency of editing with R2 D 45-45 T. * represent significance as compared to editing with R2 D 45-45 T. * P ≤ 0.05. ** P ≤ 0.01. (E and F) Graphs showing the editing efficiency from B plotted against the sgRNA cutting efficiency from D for each individual replicate. Plots are separated into symmetric donors (E) anc asymmetric donors (F).

far enough away that it would not be influenced or altered by editing at the MCS-R2 locus.

I found that editing with modified, asymmetric, and non-template donors did not significantly improve editing efficiency, as measured by all three methods. In fact, when quantified using the restriction digest (RD) method, both asymmetric donors (R2 D 75-15 and R2 D 90-30) significantly reduced the editing efficiency by over 50% (fig 3.4.A and 3.4.B). However, when I quantified editing efficiency by qPCR and NGS, I found that although a drop in efficiency was visible for R2 D 75-15, neither modified, asymmetric, or non-template donors produced a significant change from R2 D 45-45 T (fig 3.5.C and 3.5.E).

Use of different donors had no observable effect on cutting efficiency of the sgRNA measured using the T7E1 assay (fig 3.4.C and 3.4.D). When I plotted sgRNA cutting efficiency against RD-calculated editing efficiency for the 3 symmetric donors (R2 D 45-45 T, R2 D 45-45 NT, R2 D 45-45 PS), I found they were positively correlated with an R^2 value of 0.67 and a P value of < 0.01 (fig 3.4.E). However, when I plotted the same variables for the two asymmetric donors, I found they were not positively correlated (fig 3.4.F).

In order to determine whether the RD and qPCR methods for quantifying editing efficiency were sufficient to generate accurate estimates of editing efficiency, I produced pairwise comparison plots for each quantification method (fig 3.5.F, 3.5.G, 3.5.H). For each biological method, the editing efficiency calculated by one method was plotted on the y axis and for an alternate method was plotted on the x axis. Editing efficiency calculated by NGS was, on average, slightly lower than the value calculated by RD. Each method was well correlated with each other method with R^2 values of 0.70 for RD vs NGS, 0.81 for RD vs qPCR, and 0.69 for qPCR vs NGS. I felt confident that the qPCR and RD methods were reasonable methods for estimating editing efficiencies and in the future used these methods instead of the more expensive and laborious NGS.

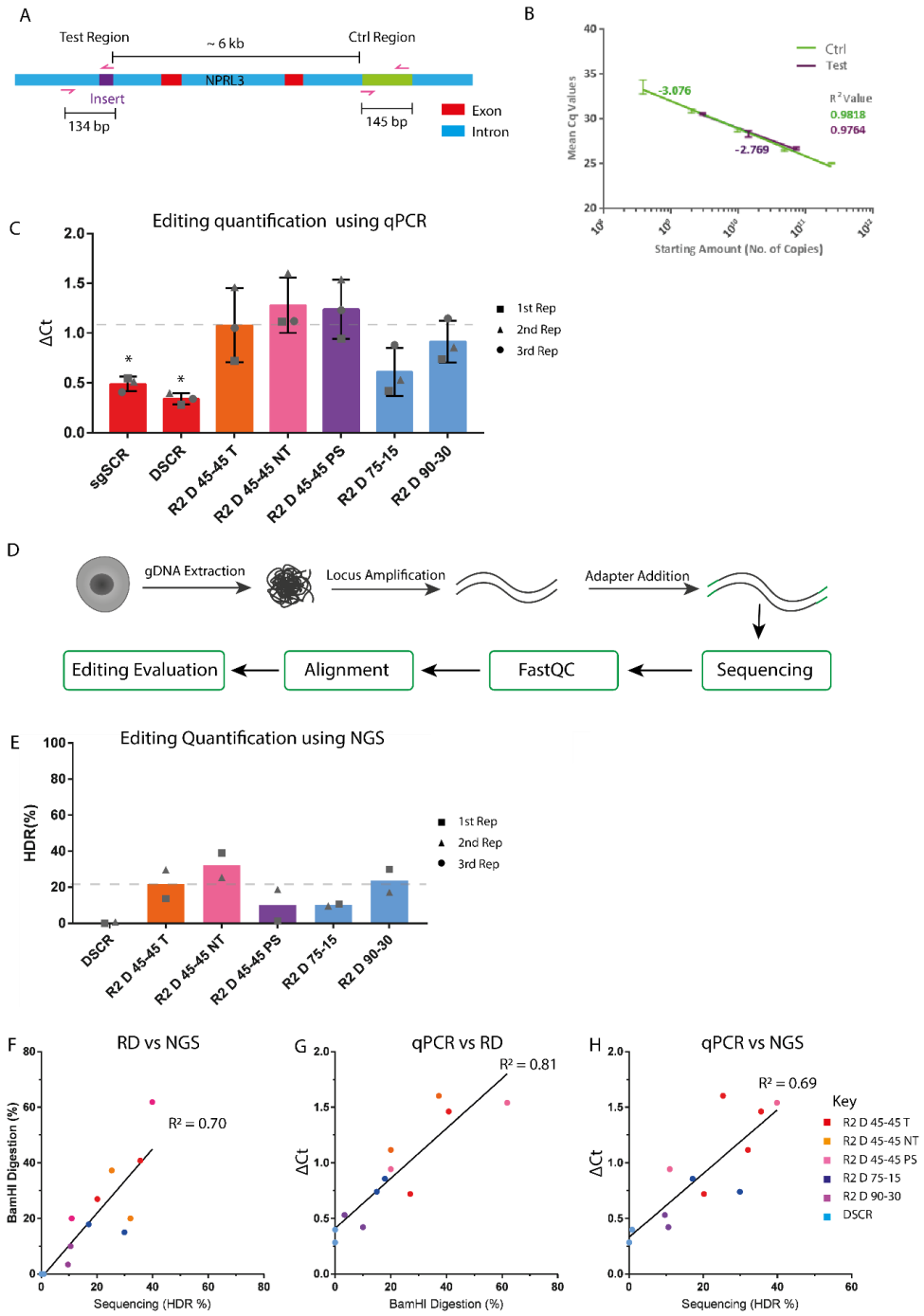


Figure 3.5 Evaluating Editing Efficiency with Different Donors using qPCR and NGS. (A) Diagram showing the positioning of MCS-R2 and Control qPCR primers (pink) in NPRL3. Introns are shown in blue, exons in red. The insert is shown in purple and the MCS-R3 region is shown in green. (B) Graph showing results of qPCR dilution series with Mean Ct value plotted against initial DNA copy no. (log scale) for MCS-R2 (purple) and control (green) qPCR primers. R^2 values are shown for both. (C) Quantification of editing efficiency using qPCR for each donor. Calculated by dividing Mean (technical) Ct for control qPCR by Mean (technical) Ct for MCS-R2 qPCR. Each point shows a separate biological replicate. Error bars show SD. Bars show mean for biological replicates. Grey dashed line represents the mean editing efficiency with R2 D 45-45 T. Significance calculated compared to editing efficiency with R2 D 45-45 T. * - $P \leq 0.05$. (D) NGS preparation and analysis pipeline. (E) Quantification of editing efficiency using NGS for each donor. Only two biological replicates shown. (F, G, H) Plots comparing quantification of editing efficiency by each method for the first two biological replicates. Different colours of points indicate the donor used.

3.2.3 Effect of RS-1 and SCR7 on editing efficiency

In this section I evaluated the effect of the small molecules, SCR7 and RS-1 (fig 3.6.A and 3.7.A), on editing efficiency. For each molecule I used a range of concentrations, distributed above and below optimum concentrations found in published data[102,107]. I carried out delivery of editing components as described previously, electroporating DNA oligo and plasmid encoded Cas9 and sgRNA into HUDEP-2 cells (fig 3.3.B). I diluted RS-1 and SCR7 to the appropriate concentration in media, into which the cells were transferred 15 min after electroporation.

I found that at the MCS-R2 locus in HUDEP-2 cells SCR7 did not cause a significant increase in editing efficiency at any concentration. In fact, at the highest concentration, 10 μM, SCR7 significantly decreased the editing efficiency as measured by using the BamHI RD method. I also found that using higher concentrations of SCR7 produced more variable results, possibly due to lower cell viability and thus smaller cell populations. When I calculated editing efficiency using the qPCR method, I found that SCR7 had no effect. SCR7 has previously been shown to reduce the occurrence of Cas9-induced indels[102]. My results confirmed this finding, although T7E1 cutting was only significantly reduced at 10 μM SCR7.

Similarly, I found that RS-1 did not significantly increase the editing efficiency when measured by both RD and qPCR (fig 3.7.C and 3.7.E). However, 20 μM RS-1 did significantly reduce the efficiency of sgRNA cutting by approximately 30%.

Given these findings, I decided not to attempt to test either SCR7 or RS-1 in $CD34^+$ cells. This decision was cemented by reports of both SCR7 and RS-1 causing toxicity in some cell lines, albeit at higher concentrations of RS-1 than those tested here[101,204,205]. Furthermore, I observed a reduction in viability at high concentrations of both SCR7 and RS-1 and this toxicity is reflected in the greater variability of editing efficiencies at these concentrations.

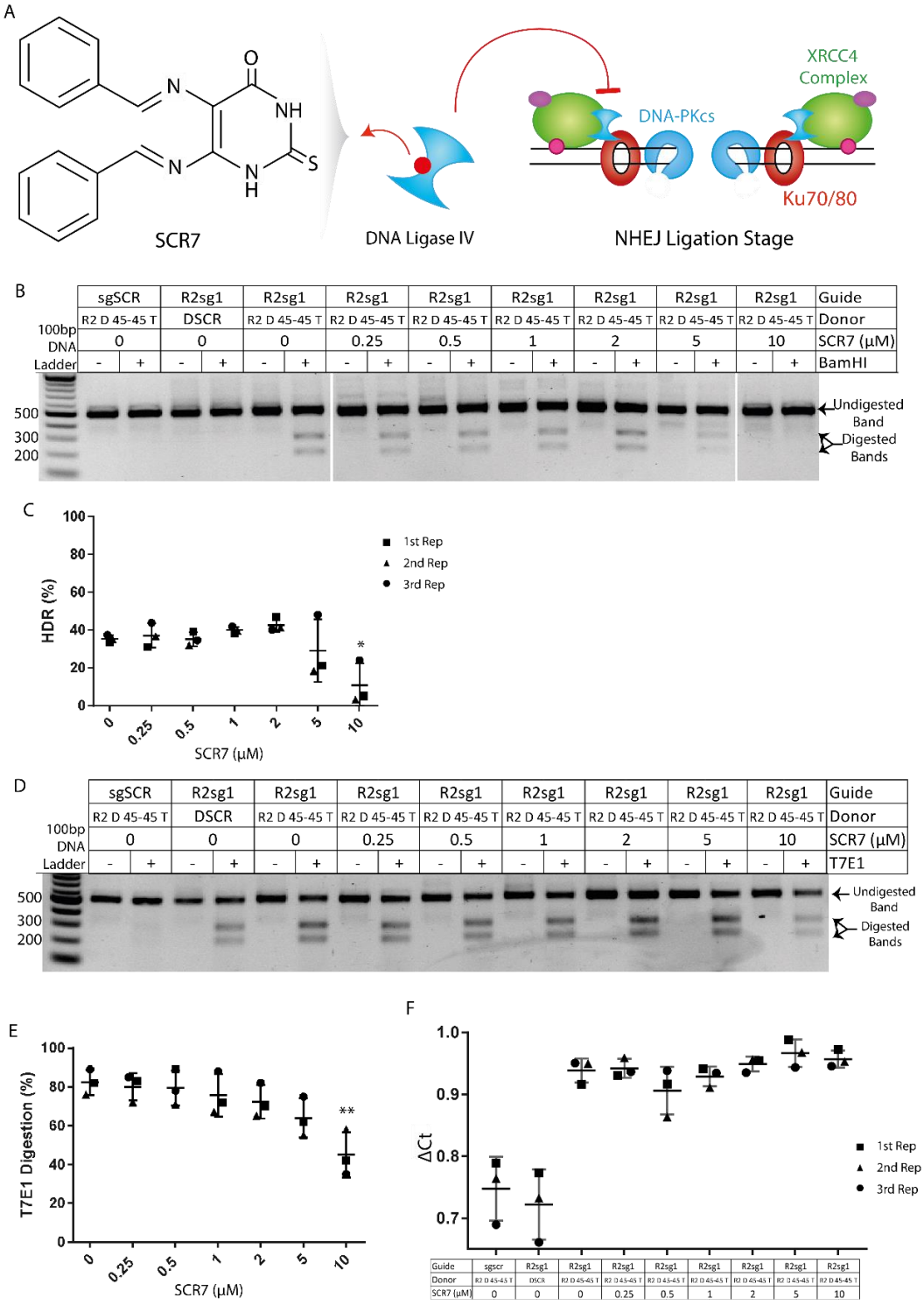


Figure 3.6 Evaluating Editing Efficiency with SCR7 using RD and qPCR. (A) Molecular structure of SCR7 pyrazine and inhibition of ligase IV and thus the ligation stage of NHEJ. SCR7 is shown as a red circle. Damaged DNA is black. Other components are labelled. (B and D) Example of a typical agarose gel of MCS-R2 amplicon, ± BamHI (B) and T7E1 (D), for different concentrations of SCR7. (C and E) Band absorption quantification for B and D. Each point represents a biological replicate. Error bars show SD. Significance calculated compared to editing efficiency with R2 D 45-45 T. * - P ≤ 0.05. ** - P ≤ 0.01. (F) Quantification of editing efficiency using qPCR for each concentration of SCR7. Calculated by dividing Mean (technical) Ct for control qPCR by Mean (technical) Ct for MCS-R2 qPCR. Each point shows a separate biological replicate. Error bars show SD. Bars show mean for biological replicates.

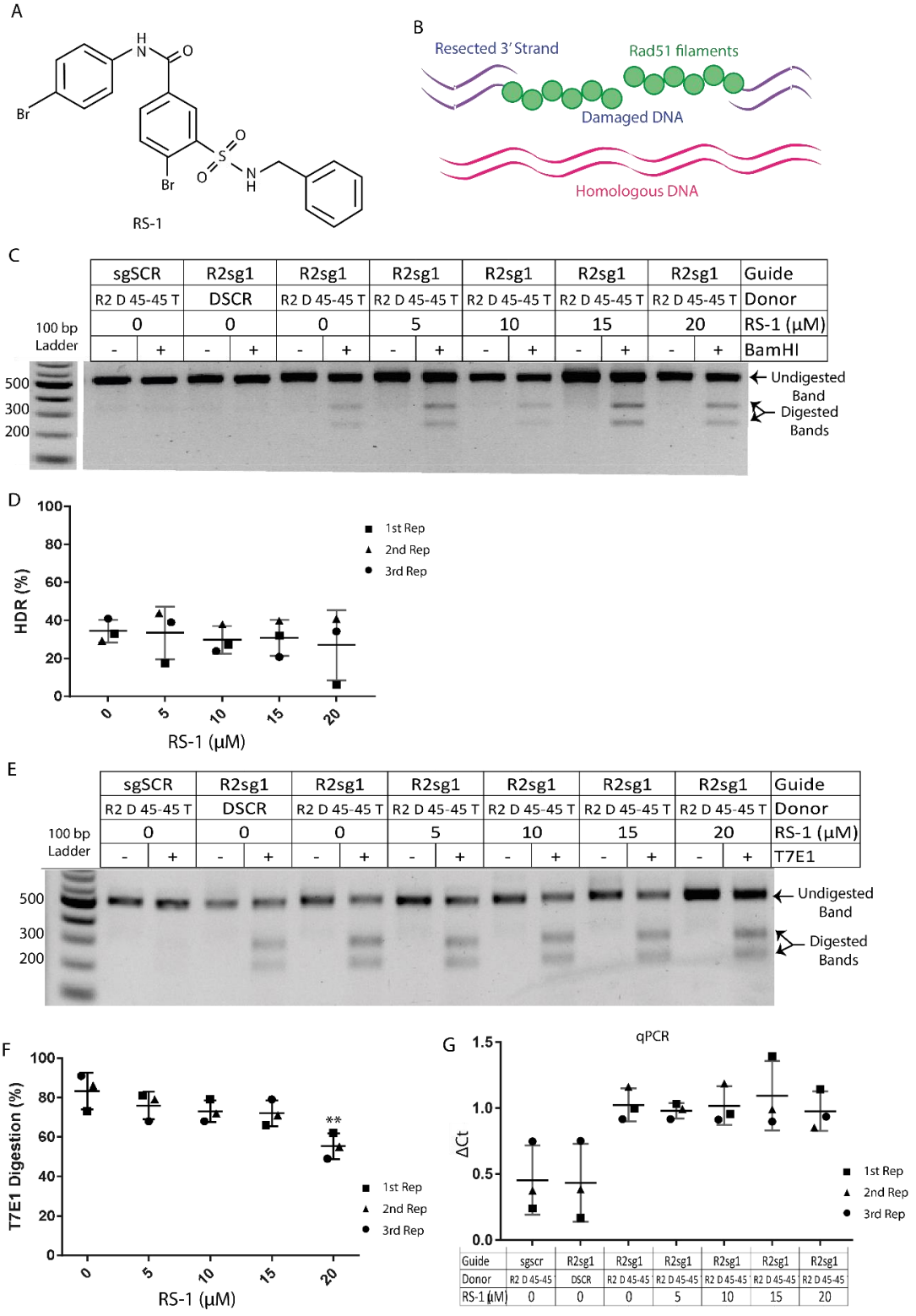


Figure 3.7 Evaluating Editing Efficiency with RS-1 using RD and qPCR. (A) Molecular structure of RS-1. (B) Diagram of Rad51 filaments on damaged, resected DNA in the process of repair by HR. (C and E) Example of a typical agarose gel of MCS-R2 amplicon, ± BamHI (C) and T7E1 (E), for different concentrations of RS-1. (D and F) Band absorption quantification for C and E. Each point represents a biological replicate. Error bars show SD. Significance calculated compared to editing efficiency with R2 D 45-45 T. ** - P ≤ 0.01. (G) Quantification of editing efficiency using qPCR for each concentration of RS-1. Calculated by dividing Mean (technical) Ct for control qPCR by Mean (technical) Ct for MCS-R2 qPCR. Each point shows a separate biological replicate. Error bars show SD. Bars show mean for biological replicates.

3.2.4 Effect of Different Donors on Editing Efficiency in CD34⁺ Cells

I wanted to investigate whether the results I observed in HUDEP-2 cells with donor templates were replicated in mobilised peripheral $CD34^+$ cells. In this section I compare the efficiency of phosphorothioate-modified, asymmetric and reverse oriented (NT) donor templates in $CD34^+$ cells.

I nucleofected $CD34^+$ cells 48 hours post-thaw with Cas9-sgRNA ribonucleoprotein (RNP) and oligo donor(fig 3.8.A)[189]. With this method I did not include a transfection or expression control, so it was not possible for me to sort cells, as I had previously with the HUDEP-2 cells. 48 hours post-nucleofection, I harvested cellular gDNA and analysed editing efficiency and sgRNA cutting efficiency with BamHI RD, qPCR and the T7E1 assay.

I found when analysing using the RD assay, that most of the replicates fell below the threshold for detection on agarose gels. I determined the threshold for accurate detection to be ~2% editing and only 3 of a total 15 replicates were above this threshold. Two of these measurable replicates were from the phosphorothioate-modified donor and averaged ~2.3% editing efficiency. The final R2 D 90-30 replicate (shown below fig 3.9.B and C) I quantified as ~12%. I determined this was not a PCR error using technical replicates. qPCR measurements confirm that the final R2 D 90-30 replicate had the highest editing efficiency in the third experimental replicate, however, the difference was relatively small and was not greater than some measurements from the second experimental replicate (fig 3.8.D). This finding may be due to technical variability in qPCR set-up between replicates. Overall the qPCR data showed that none of the donors result in a significant change in editing efficiency compared to R2 D 45-45 T.

The T7E1 assay revealed that sgRNA efficiency was reasonably consistent albeit much lower than in HUDEP-2 cells at 20-30% total cutting (fig 3.8.E and 3.8.F). I found that single nucleotide polymorphisms in the wild-type MCS-R2 sequence could be seen in the T7E1 assays, as three fainter bands at ~350, ~250 and ~150 bp, respectively. I confirmed the existence and location of these two SNPs by Sanger sequencing, and they did not overlap with the sgRNA sequence. These extra bands

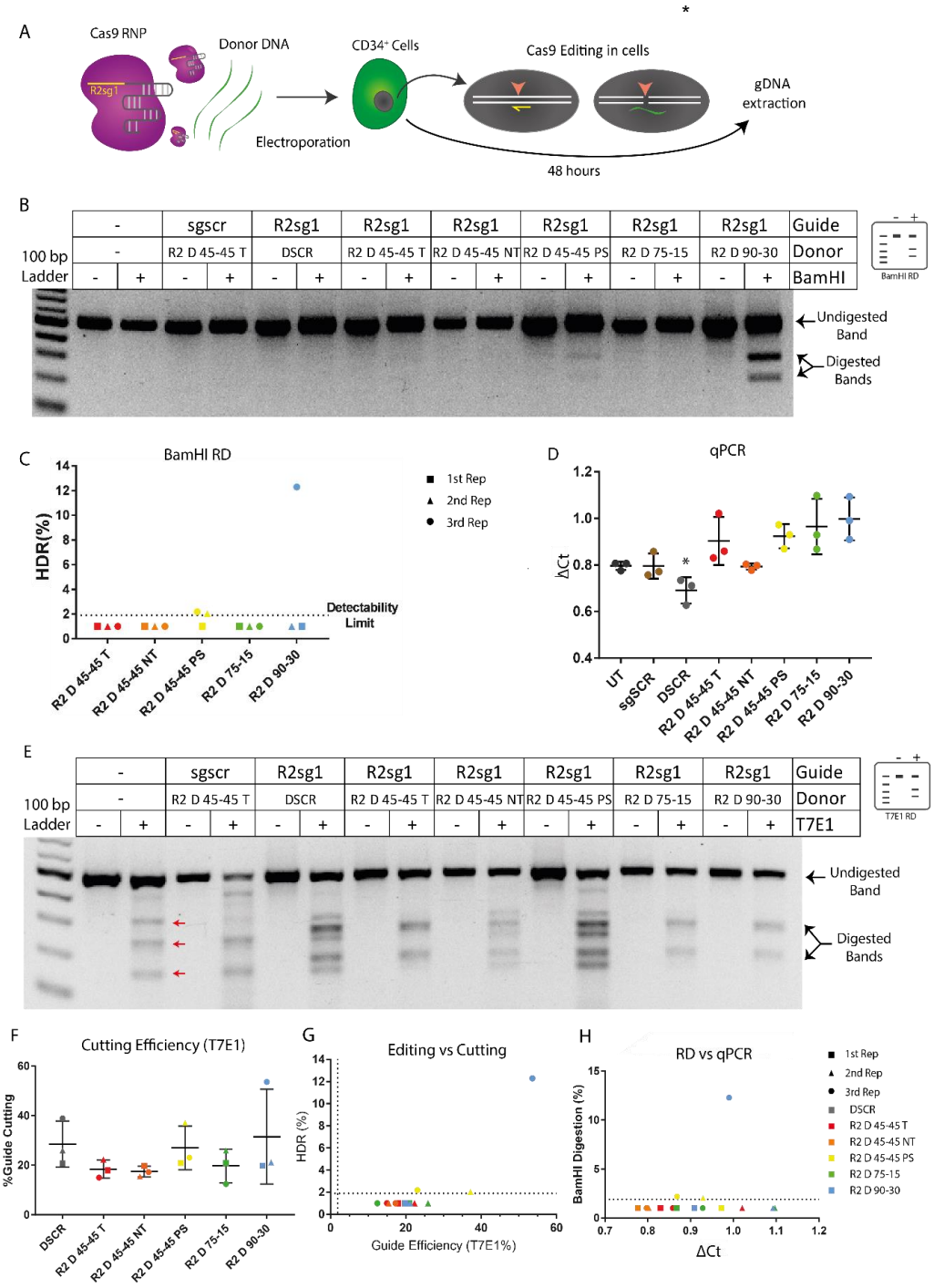


Figure 3.8 Evaluating editing efficiency with different donors in CD34⁺ Cells. (A) Schematic showing the experimental procedure for editing CD34⁺ cells. (B and E) Example of a representative agarose gel of MCS-R2 amplicons, ± BamHI (B) and T7E1 (E), for each donor. Red arrows indicate bands created by T7E1 digesting single bp bulges created by SNPs. (C) Band absorption quantification for B. Dotted line shows limit of detection for analysis of agarose gels. Each point represents a biological replicate. (D) Quantification of editing efficiency using qPCR for each concentration of SCR7. Calculated by dividing Mean (technical) Ct for control qPCR by Mean (technical) Ct for MCS-R2 qPCR. Each point shows a separate biological replicate. Error bars show SD. (F) Band absorption quantification for E. Each point represents a biological replicate. Error bars show SD. (G) Editing efficiency calculated by RD plotted against Cutting efficiency calculated by T7E1 digest. Dotted lines show limit of detection for analysis of agarose gels. (H) RD-calculated editing efficiency plotted against qPCR-calculated editing efficiency.

interfered with T7E1 analysis as their emission varied between samples and they are positioned relatively close to the sgRNA-created bands.

3.3 Discussion

In this chapter, I established the following:

1) SCR7 and RS-1 treatment does not have a significant effect on editing efficiency in HUDEP-2 cells at the MCS-R2 locus.

2) Use of phosphorothioate-modified and asymmetric donor oligos does not significantly improve editing efficiency in HUDEP-2 or CD34$^+$ cells at the MCS-R2 locus.

3) qPCR and BamHI RD are reasonable but imperfect methods for measuring editing efficiency compared to NGS.

3.3.1 HUDEP-2 cells as a model for CD34$^+$ cells

Although I could not directly compare editing in HUDEP-2 and CD34$^+$, as I did not account for transfection efficiency, I found differences in the efficiency trends between cell types when using different donors. I found that while in HUDEP-2 cells, levels of editing decreased with the asymmetric donors. I did not see this in CD34$^+$ cells; results were not consistent. This suggests that while HUDEP-2 cells may be a good disease model for CD34$^+$ cells, they are a poor model for DNA repair processes.

A closer look at the immortalisation process that transformed CD34$^+$ cells into the HUDEP-2 line, provides an explanation for this[198]. The HUDEP-2 cells have a tet-inducible copy of the HPV16 proteins E6 and E7 integrated into their genome. This singular integration helps to drive cell cycle progression by interacting with the following targets: HPV16-E6 is involved in targeted degradation of p53; while E7 binds to the retinoblastoma protein (Rb)[199]. p53 is a major coordinator of the DNA damage response and Rb, to a lesser extent, has been shown to play a central role in NHEJ repair[206,207,208]. As

mentioned above, the cell cycle is heavily intertwined with DNA repair. These factors combined mean that HUDEP-2 cells are likely a poor proxy for $CD34^+$ cells in DNA repair.

Additionally, karyotyping of HUDEP-2 cells has shown that they have a modal chromosome count of 51, although this duplication does not include chromosome 16[209]. Most cultured cell lines contain similar duplications: HELA and K562 have a count of 66 and 67, respectively, however, these abnormalities probably also contribute to the differences observed.

Given both the (relatively) high editing efficiency and the correlation between sgRNA cutting and editing efficiency in HUDEP-2 cells (fig 3.4.E), the limiting factor for DNA repair in these cells seems to be sgRNA cutting efficiency. Maybe this explains why neither the different donors nor small molecules result in a significant improvement in editing efficiency.

Additionally, Cas9 and sgRNA were delivered to HUDEP-2 cells via electroporated plasmid, whereas for $CD34^+$ cells they were delivered as an RNP. Thus, the time from delivery of the donor until the first Cas9-induced DSBs is longer in HUDEP-2 than in $CD34^+$cells. Therefore, I would expect that in HUDEP-2 cells compared to $CD34^+$ cells, I would see a more significant rise in efficiency using the phosphorothioate-modified donor, but this was not the case. Perhaps, therefore, toxicity levels or degradation time of the ssODNs differ between these cell types.

3.3.2 SCR7 and RS-1 in HUDEP-2 Cells

Following results in HUDEP-2 cells and a review of recent literature, I decided not to test SCR7 and RS-1 in $CD34^+$ cells. The efficacy of SCR7 and RS-1 varies greatly between different studies and different cell lines[210,211]. They found that while higher concentrations of RS-1 and SCR7 decreased the efficiency of NHEJ, both small molecules had no significant effect on HDR efficiency.

In addition studies have shown that SCR7 usage is toxic, although the severity of this effect is cell type dependent[101,212]. High levels of RS-1, albeit higher levels than I tested here, have also been found to have toxic effects[213].

3.3.3 RT-qPCR and Restriction Digest as editing-assessment tools

According to the correlation plots, the qPCR and RD methods of measuring editing efficiency were good but not perfect methods compared to the NGS gold standard. These differences are likely due to multiple different factors.

Editing efficiency as measured using the RD method consistently overestimated editing efficiency when compared to NGS. I suspect this may be due to incomplete incorporation of the donor sequence into the genome. For detectable digestion, only the 6-nucleotide restriction site needed to be incorporated, whereas, when scanning the sequenced data for the HDR barcode, I only allowed for 2 mismatches. I could test this theory by performing a second sequencing analysis where I searched only for the restriction site.

This reasoning also provides an explanation for why NGS and qPCR measurements may not completely agree. The reverse qPCR primer for the HDR barcode insert will still have reasonable binding efficiency if the insert is not completely correctly incorporated. I consistently observed that the qPCR values for gDNA harvested from cells transfected with the scrambled sgRNA and standard donor were greater than those transfected with the editing sgRNA (R2sg1) and scrambled donor, although these differences were minimal and not found to be significant. This observation may be an indication that the ssODN may be able to partially prime the qPCR amplification. If this is the case, then this would also bias the quantification of editing efficiency by qPCR.

In addition, I found that using ImageJ to quantify DNA abundance from gels was at times subjective and sometimes complicated by shadowing from extra bands (i.e. SNP bands in the T7E1 assay). This

method also had the significant limitation of a relatively high detectability limit imposed by the gel electrophoresis imaging method. One possible way around this would be to use a different imaging method, for example the TapeStation automated electrophoresis system (Agilent), with a higher resolution.

Despite, these limitations, I found that the RD and qPCR methods were a trustworthy alternative to NGS (fig 3.5). None of the methods tested in $CD34^+$ cells were able to improve editing efficiency to a point where it could be consistently detected. Thus, on consultation with my supervisor, I decided that further optimisation of these methods would be unlikely to produce a significant outcome. We decided instead to formulate a method that would select for cells which were correctly edited. Many methods to optimise editing efficiency have diverse effects in different cell lines and different loci, and some cell lines are extremely reticent to editing. We theorised that instead of making incremental improvements to editing efficiencies, a more effective, universal technique would be to find a way to enrich for correctly edited cells. It is probable that a successful method able to select for correctly edited cells will be more universally effective than a method which improves the efficiency of editing. Many techniques rely on shifting the balance between the DNA repair pathways. However, this balance varies widely between cell types, meaning many efficiency-improving techniques are highly variable. Enriching for edited cells is a post-editing process and likely to produce more consistent results.

Chapter 4: A Molecular Beacon-Based Method to Enrich for Correctly Edited Cells

3.1 Introduction

In this chapter, I describe my efforts to select for correctly edited cells using molecular beacons. My goal was to identify correct mRNA produced from the edited locus, and sort cells into an enriched correctly edited population based on this classification. In addition, any changes apart from the intended edit would ideally be translationally silent. Therefore, this method would be minimally invasive and more palatable for the clinic.

3.1.1 Molecular Beacons

Molecular beacons are nucleic acid-based hybridisation probes, consisting of a hairpin sequence folded to bring fluorophore and quencher molecules, positioned at opposing ends, into close proximity (fig 4.1.A)[214]. The beacon sequence is designed to be homologous to a target molecule (typically RNA or DNA), unfolding and annealing to this sequence when the beacon and target come into contact. Unfolding of the beacon separates the fluorophore and quencher, allowing transmission of beacon fluorescence. The thermodynamic balance of intra-molecular stem hybridisation and inter-molecular beacon-target hybridisation, results in beacons being finely tuned to mutations in the target sequence[215].

Therefore, one of the key advantages of molecular beacons is that in previous work they have been shown to differentiate SNPs well *in vitro*[216]*.* There is also evidence, although less so, that they may be able do so *in vivo*[217]. *In vitro*, molecular beacons are mostly used as an alternative to taqman-probes in qPCR to give a real-time fluorescent readout of allele-specific mRNA levels. These assays have been used for multiple different purposes, for example: analysis of drug resistance in *Mycobacterium tuberculosis*[218,219]*;* quantification of miRNA[220]; and copy number estimation of mitochondrial (mt)DNA

in biopsy samples[216]. The main advantages of using this molecular beacon-based method are that it is rapid and specific.

Molecular beacons have also been used *in vivo* in live cell imaging, allowing mRNAs to be visualised in a temporally and spatially specific manner, and in cell sorting, allowing different cell populations to be isolated. Chen et al succeeded in visualising a single modified mRNA, with a cassette of 8 target sites, within single cells[221]. Fluorescence microscopy has also been used to image β-actin[222]. Flow cytometry has been used alongside molecular beacons to simultaneously detect oct4 mRNA in stem cells[223], and to sort successfully differentiated cardiomyocytes from human IPSCs[224].

There are some restrictions which must be considered when designing molecular beacons. Firstly, one must ensure that the fluorophore-quencher pair used are compatible. Secondly, many fluorophores (e.g. TAMRA) are quenched by guanine nucleotides, and thus, Gs should be avoided in the fluorophore-adjacent sequence[225]. Internal Dabcyl quenchers must be attached to a thymine nucleotide when used in a non-terminal position[226]. Finally, beacons should be synthesized with quencher, not fluorophore, at the 3' end. Therefore, any incomplete beacons will not have unquenched fluorophores and will not contribute to background signal.

3.1.2 Motivation

We hoped to use molecular beacons to detect specific edits in mRNA made using the Cas9-sgRNA system. We planned to electroporate beacons into cells after the editing process, and then sort cells on beacon fluorescence using FACS. One caveat of this approach is that it works poorly on weakly expressed mRNA. We hypothesized that if this technique was successful, we would be able to apply it to non-coding and moderately/non-expressed mRNA by amplifying our target sequence using a minimal RNA Pol III promoter and terminator.

4.2 Results

4.2.1 Detection of GFP mRNA with Molecular Beacons

First, I wanted to test if I could detect mRNA in cells by flow cytometry using molecular beacons. I began by using a GFP model system, where GFP intensity is an accurate readout for mRNA expression, and, is therefore a good control for the efficacy of molecular beacons for detecting different levels of mRNA[227].

4.2.1.1 Design of GFP Molecular Beacons

I ordered a replica of a phosphorothioate-modified GFP beacon, hereafter referred to as GFP_7-21-0 (notation indicates "target_stem-loop-tail"), described in a paper from Tang et al[228]. In their paper they showed that, when injected into COS7 cells, the beacon was activated ~4x more in GFP-transfected cells than untransfected cells. I wanted to determine if this beacon would work as well when electroporated into GFP-expressing AML5 cells.

However, when we initially examined GFP_7-21-0 *in silico,* we found that while the correctly folded form (fig 4.1.A) was most prevalent, the second most prevalent form (fig 4.1.B) folded so that the quencher and fluorophore (AF647) were separated by about ~15 nucleotides. Therefore, *in silico* models suggest approximately 20% of GFP_7-21-0 would be non-specifically activated at any one time. This may have contributed to the background fluorescence observed by Tang et al in their paper[229]. We also realised that phosphorothioate-modified DNA-RNA hybrids have a lower T_m than unmodified DNA-RNA hybrids and, despite reduced sensitivity, are still substrates for RNase H-mediated degradation[230,231].

Therefore, we wanted to see if we could design optimised GFP beacons with a lower probability of mis-folded secondary forms. We first used RNAfold to find the most open and accessible regions of the GFP mRNA and designed our beacons to be complementary to the most accessible region. In order to optimise the ΔG of folded beacons, we created beacons with different stem lengths of 4 and 5 bp, GFP_4-19-0 and GFP_5-19-0 respectively (fig 4.1.C and 4.1.D). We decided to use 2'O

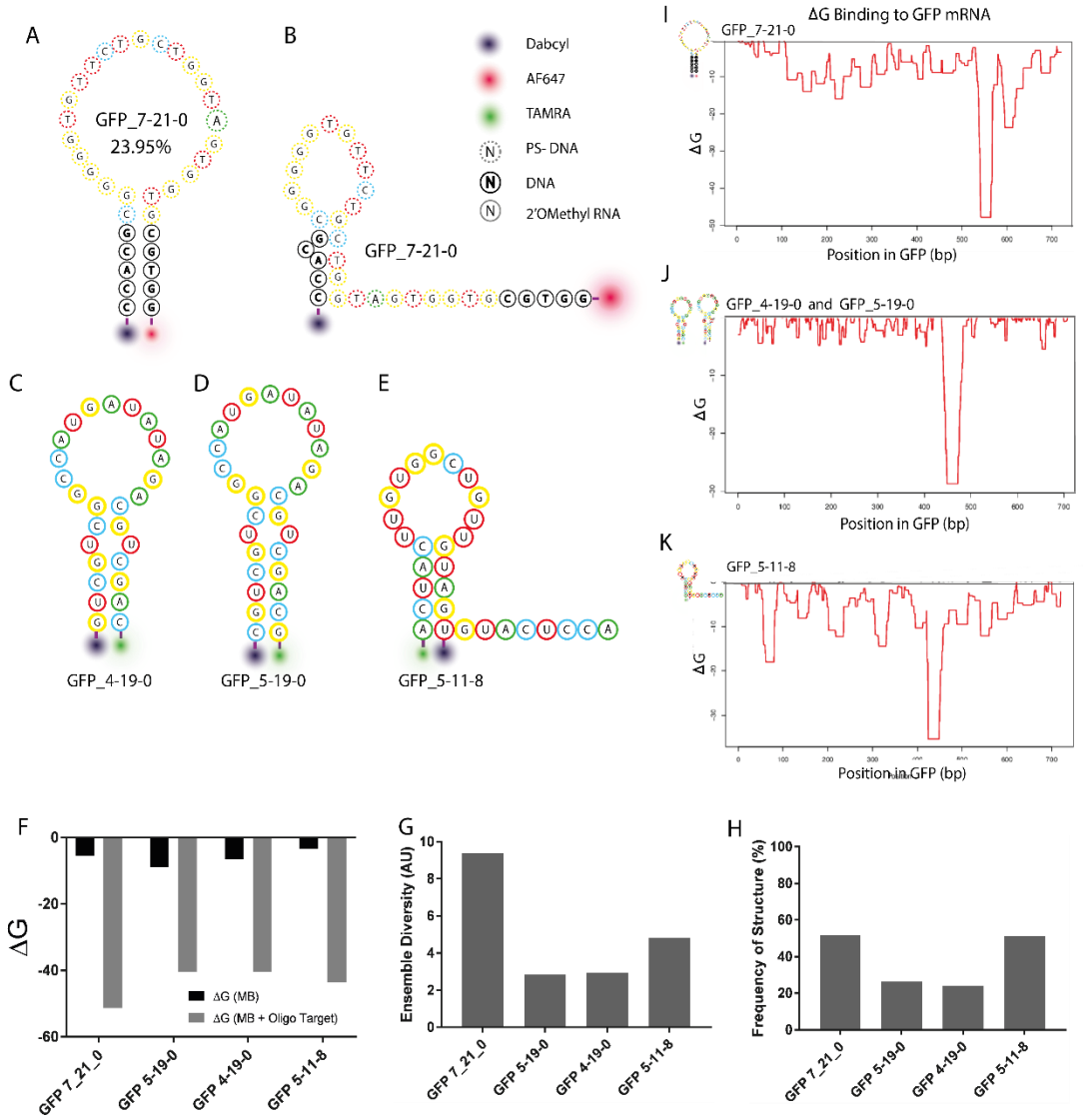


Figure 4.1 Design and *in silico* analysis of GFP Beacons. Structural diagrams of folded beacons: GFP_7-21-0, correctly folded (A), and incorrectly folded (B), GFP_4-19-0 (C), GFP_5-19-0 (D), GFP_5-11-8 (E). DNA bases are outlined in black. Phosphorothioate-modified DNA bases (dashed outlines) and 2'OMethyl RNA bases (solid outlines) are outlined in red (uracil and thymine), yellow (guanine), green (adenine), and blue (cytosine). Quenchers glow black. Fluorophores: AF647 and TAMRA, glow red and green respectively. (F) Graph plotting ΔG for each GFP beacon alone (black) and complexed with oligo target (grey). Ensemble Diversity (G) and Frequency of Structure occurrence (H) for each GFP beacon. ΔG of beacon binding to each position in GFP mRNA for: GFP_7-21-0 (I), GFP_5-19-0 and GFP_4-19-0 (J), and GFP_5-11-8 (K).

Methyl RNA bases to synthesize these beacons as RNA-RNA hybrids are not substrates of RNAse H and 2'O Methyl RNA bases have been found to improve stability of RNA duplexes[232].

I also tested a new molecular beacon design, courtesy of David Knapp: the wyvern beacon or GFP_5-11-8 (fig 4.1.E). This design stems from the hypothesis that two closed-loop, folded sequences may be

less likely to interact as there is no sequence readily available to initiate binding. Dr Knapp theorised that a toehold, or 'tail', sequence would stimulate initial interaction and unfolding in the presence of the target sequence[233].

4.2.1.2 In Silico Analysis

To facilitate and justify our design choices, we collected *in silico* data in parallel using both RNA fold and mfold. We calculated the ΔG of the folded beacon sequence and used this as a rough estimate for the energy needed to unfold the stem of each beacon (fig 4.1.F). We hypothesized that there would be a 'Goldilocks' stem strength where the ΔG was neither too small, which could result in non-specific binding, or too great, which could mean the beacon may not be able to open correctly in the presence of the target. Thus, we ensured we had a reasonable ΔG range within our four beacons to aid in optimisation of stem strength. We also compared these values to the ΔG of the unfolded beacon annealed to the target sequence (fig 4.1.F). Increased ΔG of target-beacon interaction improves the probability of a strong true positive signal. However, increasing the length of the beacon to maximise the ΔG increases the possibility of off-target interactions which will produce background fluorescence and confound results.

We found that GFP_5-11-8, had the highest ratio between ΔG of folded beacon and ΔG of annealed beacon and target (fig 4.1.F). This likely means that this beacon has a higher probability of unfolding in the presence of the correct target but could also mean that this beacon has a higher probability of non-specific binding. Overall GFP_7-21-0, as the longest sequence, had the highest ΔG of annealed beacon and target. GFP-5-19-0 had the highest ΔG of folded beacon and thus may be recalcitrant to target interaction.

We also determined the frequency of structure (%) (incidence of the correctly folded structure within the pool of structures) and the ensemble diversity (average distance between base pairs within the pool of structures) using RNAfold (fig 4.1.G and 4.1.H). We found that GFP_7-21-0 had the highest ensemble diversity, which fits with the secondary misfolded structure in fig 4.1.B. GFP_5-11-8 had a

higher ensemble diversity than GFP_5-19-0 and GFP_4-19-0, possibly due to its flexible tail. However, the frequency of the correctly folded structure within the structure pool for GFP_7-21-0 and GFP_5-11-8 (~50%) was higher than for GFP_5-19-0 and GFP_4-19-0 (~25%). But, as ensemble diversity for GFP_5-19-0 and GFP_4-19-0 is relatively low, it is likely that the next most common structures are variations of the correctly folded form.

These *in silico* results provided a very rough estimate of what we might expect *in vitro* and allowed us to filter our design choices. However, there are multiple caveats to this approach. These *in silico* models are limited and cannot take into account the base modifications, like phosphorothioate and 2'O Methyl bases, and the effect these may have on binding[230,232].These models are heavily simplified and it is very difficult to predict how beacons will react in cells, where numerous other factors, like off-target binding and protein interactions, may affect beacon and target interaction. In addition, the GFP_7-21-0 was composed of DNA bases, and whilst most of the *in-silico* data for this beacon was collected using the DNA folding form from unafold, ensemble diversity could only be determined using RNAfold. Therefore, this data point should be considered sceptically.

4.2.1.3 In Vitro Analysis

Prior to testing beacons in cells, I used *in vitro* analysis to attain a more in-depth understanding of the specifics and dynamics of beacon interaction with different types of targets. To achieve this, I combined beacon with target molecules in excess, in an annealing buffer meant to mimic the ionic and pH conditions within cells and measured fluorescence output using a Clariostar Microplate reader. I used four different target types: short DNA, long RNA, short RNA and long DNA; here, short means the same length as the beacon and long means the same length as the full-length mRNA. I wanted to see the relative contribution of length, and thus complexity of folding, and type of duplex to beacon activation. I normalised each replicate internally by calculating the ON:OFF ratio (fluorescence with target/fluorescence without target). This also allowed me to compare GFP_7-21-0 with

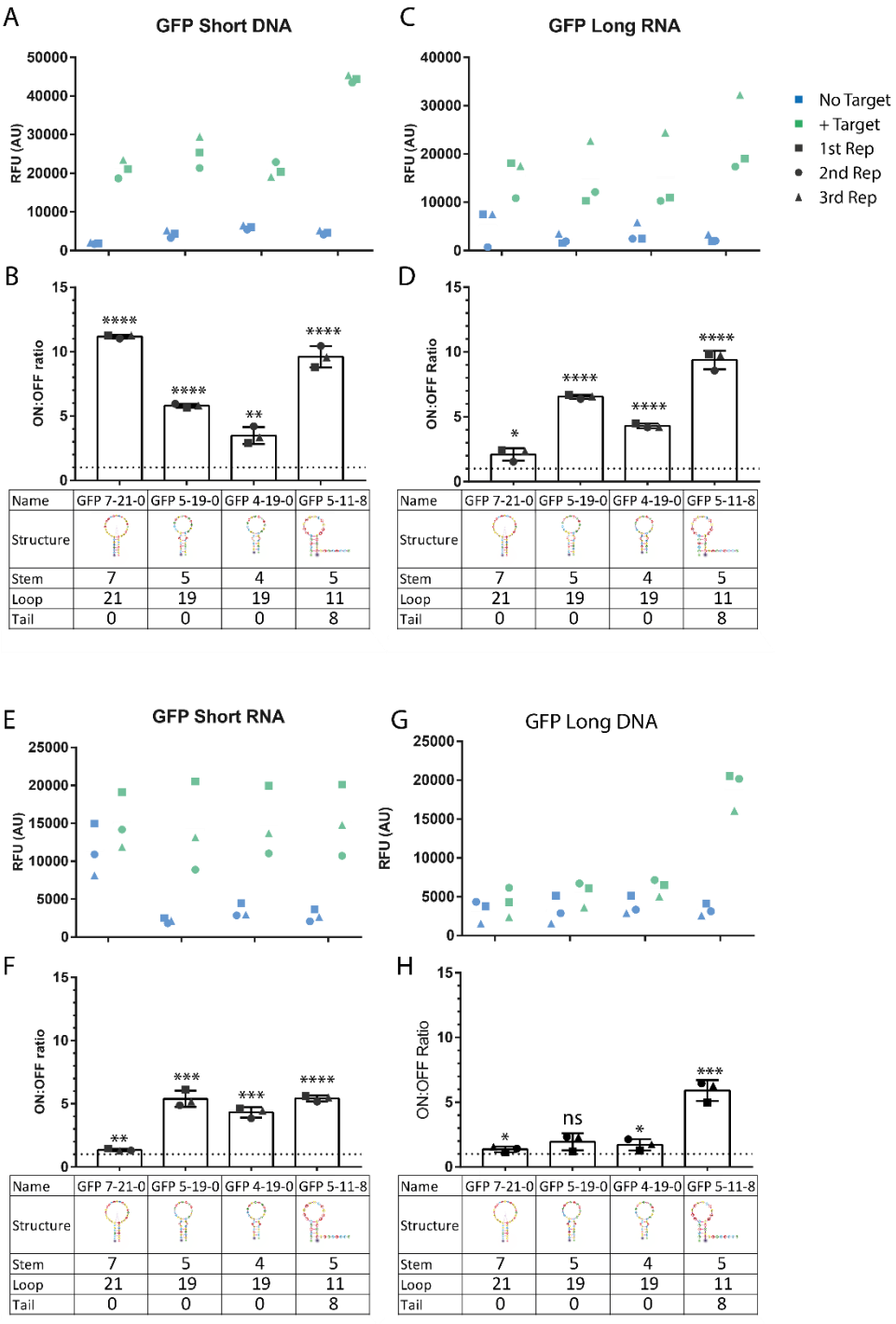


Figure 4.2 *In vitro* Analysis of GFP Beacons. (A, C, E, G) Raw fluorescence measurements for each beacon with no target (blue) and with specified target (green). (B, D, F, H) ON:OFF ratio of Target/No Target measurements for each beacon and each specified target. Error bars show SD. * shows significance compared to null hypothesis. ns = not significant. * = P ≤ 0.05. ** = P ≤ 0.01. *** = P ≤ 0.001. **** = P ≤ 0.0001. Dashed lines show an ON:OFF ratio of 1 (null value).

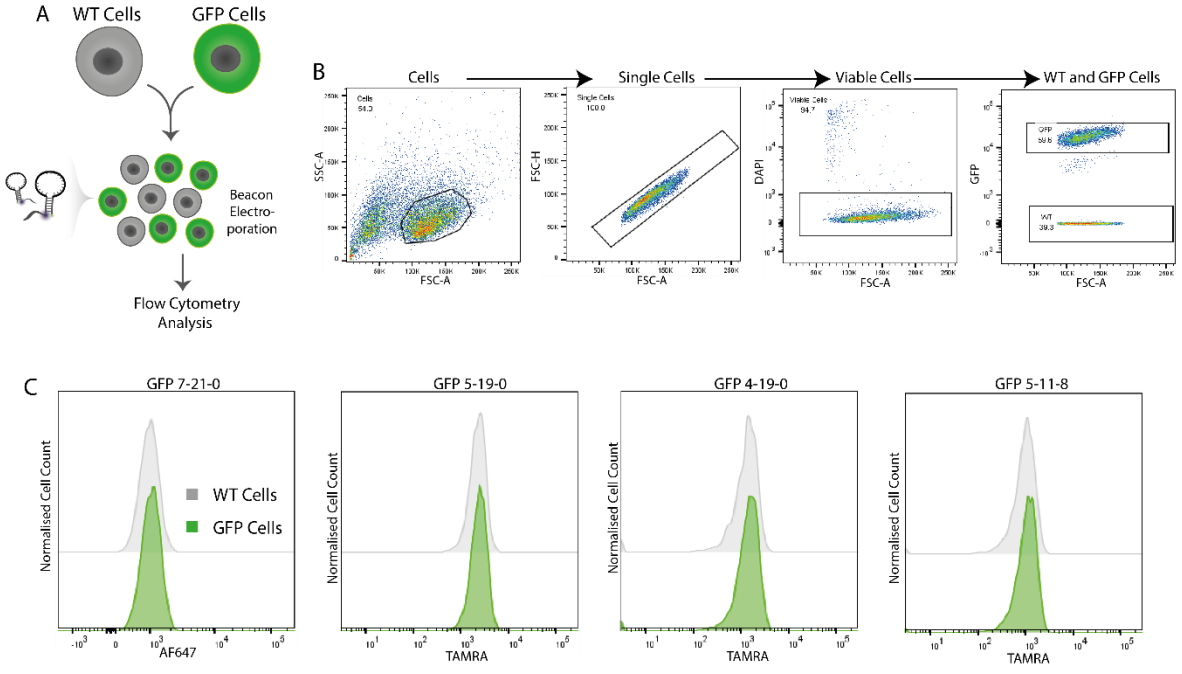
the other GFP beacons as they used different fluorophores, AF647 and TAMRA, respectively, and thus could not be measured in the same assay.

I found that although GFP_7-21-0 had the highest ON:OFF ratio with the short DNA target, it performed poorly with all other targets (fig 4.2.B, 4.2.D, 4.2.F, 4.2.H). Unfortunately, as the other beacons used a different quencher-fluorophore pair, I could not compare the raw background values of GFP_7-12-0 with the other beacons in the off-state (fig. 4.2.A, 4.2.C, 4.2.E, 4.2.G). Therefore, it was difficult to determine from the *in vitro* data if our efforts to reduce background fluorescence from misfolded beacons had much effect.

GFP_5-19-0, GFP_4-19-0 and GFP_5-11-8 showed similar ON:OFF ratios in long RNA and short DNA, compared to the GFP_7-21-0. This suggests that our efforts to target the beacons to more accessible areas of the GFP RNA were successful. GFP_4-19-0 had a noticeably lower ON:OFF ratio than GFP_5-19-0 with most targets due to a higher off rate. I believe this background fluorescence is due to non-specific opening of the weaker 4 bp stem.

The wyvern beacon, GFP_5-11-8, outperformed GFP_4-19-0 and GFP_5-19-0 on both the long DNA and long RNA targets: with an ON:OFF ratio of ~2x greater with the long RNA target and 3.5x greater with the long DNA target (fig 4.2.D, 4.2.H). This result suggested that in this simplified scenario, the wyvern beacon may be better at accessing longer, more intricately folded targets.

I hypothesized the long RNA results were the most relevant to the ultimate success of the beacons, as these would be the best indicator of how well the beacon will bind to the mRNA in cells (fig 4.2.C, 4.2.D). Generally lower ON:OFF ratios with short RNA and long DNA targets, compared to short DNA and long RNA targets, may be due to the lower quality of these homemade targets produced with less well-established techniques.



Name	GFP 7-21-0	GFP 5-19-0	GFP 4-19-0	GFP 5-11-8
Structure				
Stem (nt)	7	5	4	5
Loop (nt)	21	19	19	11
Tail (nt)	0	0	0	8

Figure 4.3 *In vivo* analysis of GFP beacons in AML5 cells. (A) Schematic: electroporation of beacons into mixed WT and GFP AML5 cells. (B) Gating strategy for flow cytometry analysis. (C) Flow cytometry histograms showing cell counts against beacon fluorescence for each GFP beacon. (D) MFI for beacon fluorophore in GFP population divided by the MFI in WT population. Individual points show 3 biological replicates. Lines show mean. Error bars show SD. * shows significance compared to null hypothesis. ns = not significant. * = $P \leq 0.05$. Dashed lines show an ON:OFF ratio of 1 (null value).

4.2.1.4 In Vivo Analysis

To test the efficacy of the beacons for detecting GFP mRNA *in vivo,* I used a flow-cytometry readout system. I delivered beacons into a mixed population containing both WT and GFP-expressing AML5 cells. GFP protein fluorescence could be used as a control to identify cells containing GFP mRNA and cells with no GFP mRNA (WT). Within these two cell populations, I observed beacon fluorescence to see if there was a difference in beacon activation in the presence of GFP mRNA.

For these experiments, I first mixed the two cell populations in a 1:1 ratio and then delivered the beacons to the cells using electroporation (fig 4.3.A). I then analysed the cells by flow cytometry, looking specifically at the fluorophore intensity in the two different cell populations (fig 4.3.B, 4.3.C). I then quantified the beacon *in vivo* ON:OFF ratio by dividing the fluorophore MFI in the GFP population (ON) by the fluorophore MFI in the WT population (OFF). I determined the significance of these results by comparing the data to the null hypothesis where the ON:OFF ratio is 1.

I found that for all GFP beacons there was a marginal increase in beacon fluorescence in GFP cells (fig 4.3.D). This increase was only significant ($P < 0.05\%$) for GFP_4-19-0 and GFP_5-11-8 and was more highly variable in GFP_7-21-0. Technically GFP_5-11-8 performed best but it was clear that in order to be able to sort cells effectively, the beacons needed to be further optimised.

4.2.2 Detection of CCR5 mRNA with Molecular Beacons

We decided to further optimise beacon structure using an endogenous mRNA and relevant gene editing target. Thus, we settled on exploring the use of the wyvern beacons to detect the presence of the D32 mutation in CCR5 mRNA. As the D32 mutation results in loss of cell surface expression of CCR5 protein[234,235], and this loss is detectable by flow cytometry using CCR5 antibodies, we again have a control for detecting mutation of WT to D32 CCR5 mRNA.

We designed beacons to detect three different regions within the D32 and WT CCR5 mRNA: a D32-spanning region, a WT region (found only in WT CCR5), and a control region (found in both mutant and WT mRNA) (fig 4.4.A). The D32-spanning region inevitably included the WT sequence, which binding simulations predicted would not be an issue. Thus, we hoped that by measuring brightness of D32 or WT-specific beacons against control beacons, we would be able to normalise for variable mRNA expression and sort cells into D32 and WT populations. Ultimately, for this method to work, we would need to use compatible fluorophores with the control beacons and the D32 or WT beacons, however, for simplicity in the initial optimisation phase we used the same fluorophore for all beacons.

4.2.2.1 Design of CCR5 Molecular Beacons

We wanted to generate a more efficient and accurate system for finding the optimal beacon for a designated mRNA or region of an mRNA. Thus, David Knapp created an R script to automate beacon

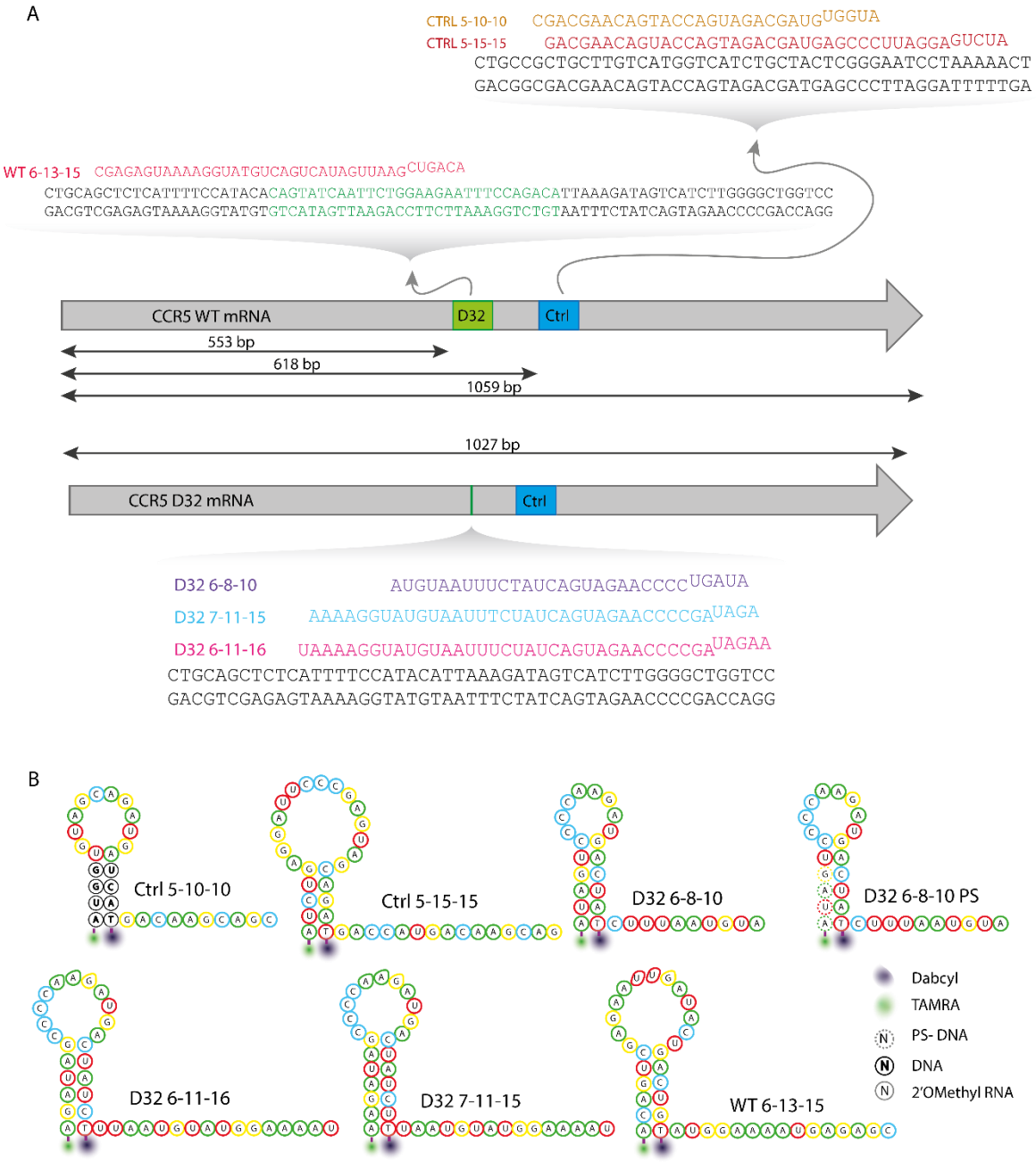


Figure 4.4 Design of CCR5 Molecular Beacons. (A) Schematic: CCR5 WT and D32 Sequence and beacon placement. (B) Structures of CCR5 beacons. DNA bases are outlined in black. Phosphorothioate-modified DNA bases (dashed outlines) and 2'OMethyl RNA bases (solid outlines) are outlined in red (uracil and thymine), yellow (guanine), green (adenine), and blue (cytosine).

design: where input factors included mRNA sequence, optimal beacon structure, and other structural restrictions. We selected output sequences on their propensity to fold correctly.

In addition, for the control region we selected the best region by observing competitive folding of the candidate beacons with the CCR5 mRNA. We selected a control region outside of the area potentially affected by sgRNA-mediated deletions.

We wanted to further test the effect of loop length, tail length and base modifications on beacon ON:OFF ratios. For the control region, where we had more flexibility of sequence and thus structure, we designed a longer structure (Ctrl_5-15-15) and a shorter structure (Ctrl_5-10-10) (fig.4.4B). For the WT and D32 regions our sequence, and thus potential structures, were much more limited. In the D32 region, we designed one shorter structure (D32_6-10-8) and two longer structures (D32_6-11-16 and D32_7-11-15). The sequences of these longer structures differed by only one base pair. For the WT region we tested only one structure (WT_6-13-15), which was designed manually as this region was too short to contain multiple viable targets. We did not compare the beacon structures between regions as we suspected that the mRNA region targeted would have a significant effect on beacon efficacy.

Furthermore, we also tested the effect of modifying the 5' half of the beacon stem with phosphorothioate bases, by directly comparing D32_6-8-10 and D32_6-8-10_PS. For the stem of Ctrl_5-10-10 we used normal DNA bases instead of RNA 2'O methyl bases, to see if this would affect the ON:OFF ratio. Although, this addition provided us with more information using fewer beacons, it also made it more difficult to directly compare the two control beacons based on structure alone.

4.2.2.2 In vitro testing of CCR5 beacons

As with the GFP beacons, I first observed the dynamics of the CCR5 beacons *in vitro* using the ClarioStar plate reader (fig 4.5). Results again showed that beacon activation with short RNA and long DNA was

generally lower than with long RNA and short DNA, possibly for the reasons discussed above. These wyvern beacons also showed similar ON:OFF ratios in long RNA and short DNA (fig 4.5.B and fig 4.5.D). Furthermore, I observed similar patterns in beacon activation between different targets: Ctrl_5-10-

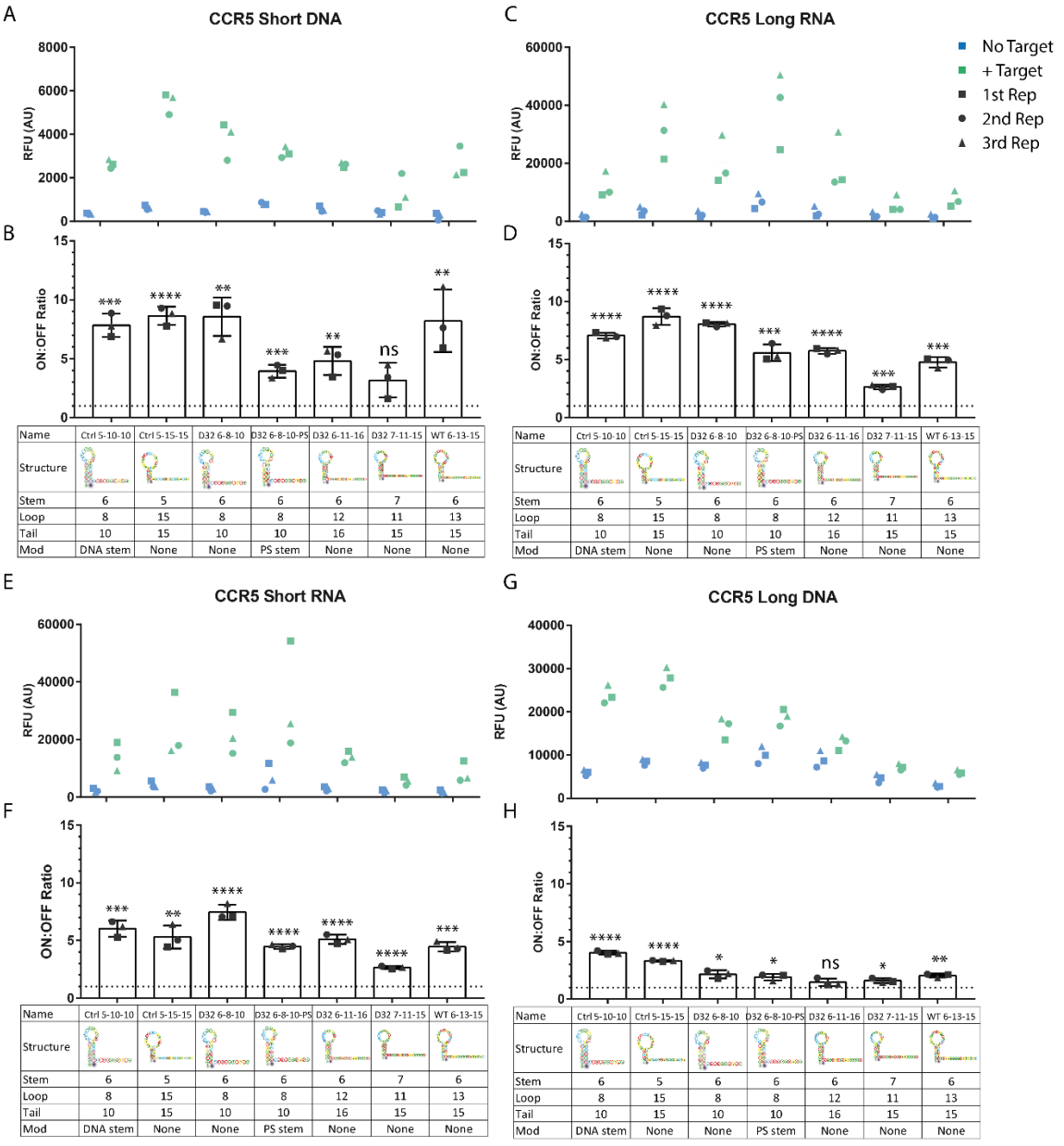


Figure 4.5 *In vitro* Analysis of CCR5 beacons. (A, C, E, G) Raw fluorescence measurements for each beacon with no target (blue) and with specified target (green). (B, D, F, H) ON:OFF ratio of Target/No Target measurements for each beacon and each specified target. Error bars show SD. * shows significance compared to null hypothesis. ns = not significant. * = P ≤ 0.05. ** = P ≤ 0.01. *** = P ≤ 0.001. **** = P ≤ 0.0001. Dashed lines show an ON:OFF ratio of 1 (null value).

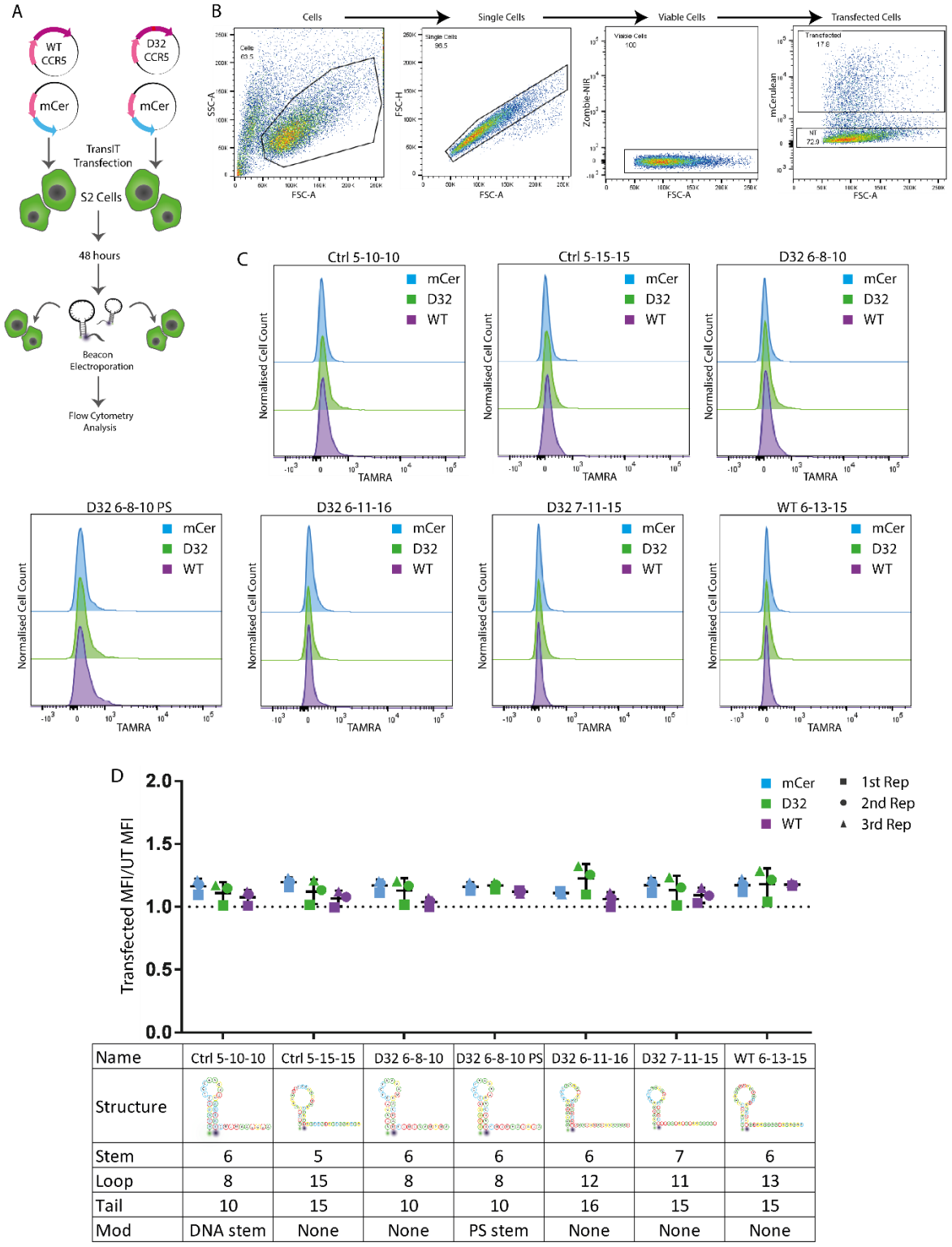
10, Ctrl_5-15-15 and D32_6-8-10 tended to perform better, whilst D32_6-8-10-PS and D32_6-11-16 tended to perform worse.

The control region beacons performed consistently better than the D32 and WT region beacons, perhaps because the D32 and WT regions had a much more restricted sequence space. Therefore, D32 and WT beacons are less likely to be optimally structured. Ctrl_5-15-15 performed 10-20% better than Ctrl_5-10-10 on short DNA and long RNA (fig 4.5.B and fig 4.5.D). This result could be due either to the longer tail and loop of Ctrl_5-15-15, or to the DNA stem of Ctrl_5-10-10. However, OFF values for both control region beacons are similar, suggesting their stems have similar melting temperature and the difference in ON:OFF ratio is more likely due to the more optimal structure of Ctrl_5-15-15 (fig 4.5.A and fig 4.5.C).

However, in the D32 region, the longer beacons, D32_6-11-16 and D32_7-1-15, performed consistently worse than the shorter D32_6-8-10. This suggests that longer beacons are not always the better option and for every region, several beacons may need to be trialled and optimised. In addition, despite being an elongation of only one nucleotide in the stem, D32_7-11-15 performed 1.5-2.5x worse than D32_6-11-16 with all targets except long DNA. Thus, small changes in sequence can lead to significant changes in beacon efficacy.

D32_6-8-10-PS had a consistently lower ON:OFF ratio than D32_6-8-10 due to higher off values, suggesting that the phosphorothioate modifications contributed to a lower stem Tm and non-specific beacon activation.

Despite limited *in silico* optimisation of the WT beacon, WT_6-15-15, it performed reasonably well with all targets, although consistently underperformed by 20-50% compared to Ctrl_5-15-15 and D32_6-8-10.



Name	Ctrl 5-10-10	Ctrl 5-15-15	D32 6-8-10	D32 6-8-10 PS	D32 6-11-16	D32 7-11-15	WT 6-13-15
Structure							
Stem	6	5	6	6	6	7	6
Loop	8	15	8	8	12	11	13
Tail	10	15	10	10	16	15	15
Mod	DNA stem	None	None	PS stem	None	None	None

Figure 4.6 *In vivo* analysis of CCR5 beacons in S2 cells. (A) Schematic: transfection of S2 cells with CCR5 and mCer plasmids, followed by electroporation of beacons. (B) Gating strategy for flow cytometry analysis. (C) Flow cytometry histograms showing cell counts against beacon fluorescence for each CCR5 beacon. mCer only transfected cells are shown in blue, WT CCR5 + mCer in purple and D32 CCR5 + mCer in green. (D) MFI for beacon fluorophore in transfected population divided by the MFI in untransfected population. Individual points show 3 biological replicates. Lines show mean. Error bars show SD. Dashed line indicates the null value, 1.

4.2.2.3 In vivo testing of CCR5 Molecular Beacons in S2 cells

As CCR5 is an endogenous protein which is expressed at varying levels in many human cell lines (http://www.proteinatlas.org), I first tested the beacons in a CCR5-free system using *Drosophila* S2 cells. The advantage of S2 cells were two-fold: firstly, they provided a clean background to transfect either WT or D32 CCR5 copies; secondly, transfected WT CCR5 was translated, expressed and could be detected on the S2 cell surface (Fig 2.2).

We transfected S2 cells with either WT CCR5 or D32 CCR5, together with an mCerulean transfection control (fig 4.6.A). 2 days post-transfection, we electroporated each cell population with each beacon and used flow cytometry to analyse the results. I gated viable cells into transfected and untransfected populations and recorded levels of TAMRA fluorescence in each population (fig 4.6.B and fig 4.6.C). I quantified beacon ON:OFF ratios by dividing TAMRA MFI in the transfected population by TAMRA MFI in the untransfected population.

I found that although the ON:OFF ratios for all beacons were visibly above the null value of 1, these results were not significant (fig 4.6.D). In most cases the mean ON:OFF ratio for beacons was higher in the mCerulean only negative control, than with WT CCR5 or D32 CCR5 cells. This result could be explained as a flow cytometry artefact, where transfected cells have a slightly higher level of autofluorescence. Alternatively, it is possible that there may be an off-target binding site for the beacons within the mCerulean or one of the plasmids. Staining of cells showed that WT CCR5 cells exhibited strong CCR5 expression (fig 2.2), thus low ON:OFF ratios were likely not due to low expression levels of CCR5.

Despite discouraging results in S2 cells, we decided it would be valuable to test the most promising beacons in a human cell line.

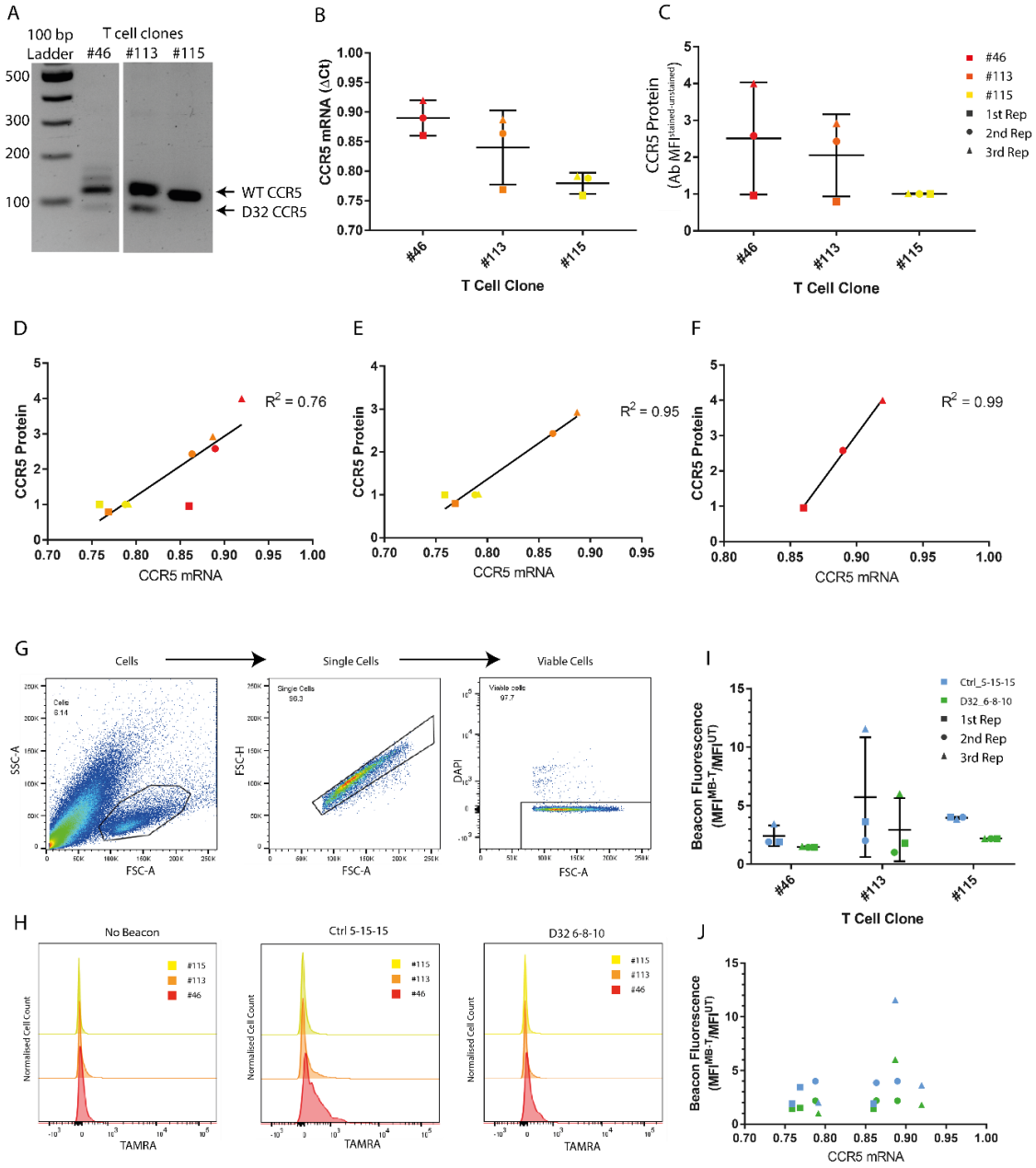


Figure 4.7 Analysis of CCR5 Beacons in T cell clones. (A) Genotyping gel of the T cell clones. Arrows indicate bands from the PCR products of WT and D32 CCR5, respectively. (B) qPCR data for CCR5 mRNA expression levels for the T cell clones. (C) CCR5 protein expression levels calculated from flow cytometry data by dividing MFI for stained cells by MFI from unstained cells. Each point for B and C represents a different biological replicate. For each replicate mRNA for qPCR was harvested within 24 hours of flow cytometry analysis. (D, E, F) Plot showing protein expression against RNA expression for each replicate for: all T cell clones (D), #113 and #115 only (E), and #46 only (F). Lines of best fit calculated by linear regression. R^2 values are indicated. (G) gating strategy for flow cytometry analysis. (H) Histograms of TAMRA fluorescence in each T cell clone, for: untransfected cells, Ctrl_5-15-15 transfected and D32_6-8-10 transfected cells. (I) Quantification of data from (H), calculated by dividing TAMRA MFI from beacon-electroporated cells by TAMRA MFI from un-electroporated cells. (J) Plot showing beacon activation (H) against RNA expression (C) for each biological replicate. For each replicate mRNA for qPCR was harvested within 24 hours of flow cytometry analysis.

4.2.2.4 CCR5 beacons in T cells

Editing of the D32 region in CCR5 in $CD4^+$ T cells, to introduce D32-like indels, has been shown to provide resistance to HIV-1 infection[236]. We attempted to use the beacon to distinguish between $CD4^+$ cells with and without the D32 mutation. We obtained 3 T cell clonal lines (#46, #113, #115), a kind gift from the Cerundolo lab (University of Oxford), which we used to carry out these experiments.

We obtained two clonal T-cell lines heterozygous for D32 (#46 and #113) and one homozygous wild–type line (#115). I confirmed this using a genotyping PCR (fig 4.7.A). I also determined CCR5 mRNA expression levels using RT-qPCR and CCR5 protein expression levels using antibody staining and flow cytometry (fig 4.7.B and 4.7.C). To account for differing levels of mRNA expression at different points in the T cell life cycle, cells used for each corresponding qPCR and flow cytometry replicate were taken at the same time point. I found that at corresponding time points, mRNA and protein expression correlated well (fig 4.7.D) and when separated into replicates from #46 and those from #113 and #115, they correlated very well (fig 4.7.E and 4.7.F). This suggests that in the different T cell clones, efficiency of CCR5 mRNA translation into protein differs. Across the three different time points, mRNA and protein levels in #115 remained consistently low, whereas in both #113 and #46 CCR5 expression varied over time. At the first time point, 7 days post-thaw, both mRNA and protein expression level were low for all clonal lines (fig 4.7.B and 4.7.C). mRNA and protein expression for both #113 and #115 was highest at 11 days post thaw. Generally, CCR5 mRNA levels were highest in #46 and lowest in #115. Although #115 was a homozygous WT line, no CCR5 was detectable on the cell surface, whereas in the two heterozygous lines, this was not the case. Thus, these three cell lines provided a varied foundation for exploring beacon properties.

Due to limited cell numbers, I decided to test only the most promising beacons from the *in vitro* analyses: Ctrl_5-15-15 and D32_6-8-10. Beacon activation was determined by flow cytometry at the same time points as qPCR, so differences in mRNA expression could be accounted for (fig 4.7.G and fig 4.7.H). I calculated beacon activation by dividing TAMRA MFI in beacon-electroporated cells by the

baseline MFI in non-electroporated cells. There was no clean background alternative in these cell lines, thus, I did not account for non-specific beacon opening. I found that despite high CCR5 mRNA expression levels, beacon activation in #46 cells was low (fig 4.7.I). In #113 cells, I found that beacon activation was highly variable, whereas in #115 cells, beacon activation was consistent. Ctrl_5-15-15 had a higher mean activation than D32_6-8-10 in all clonal lines. This could be explained by the

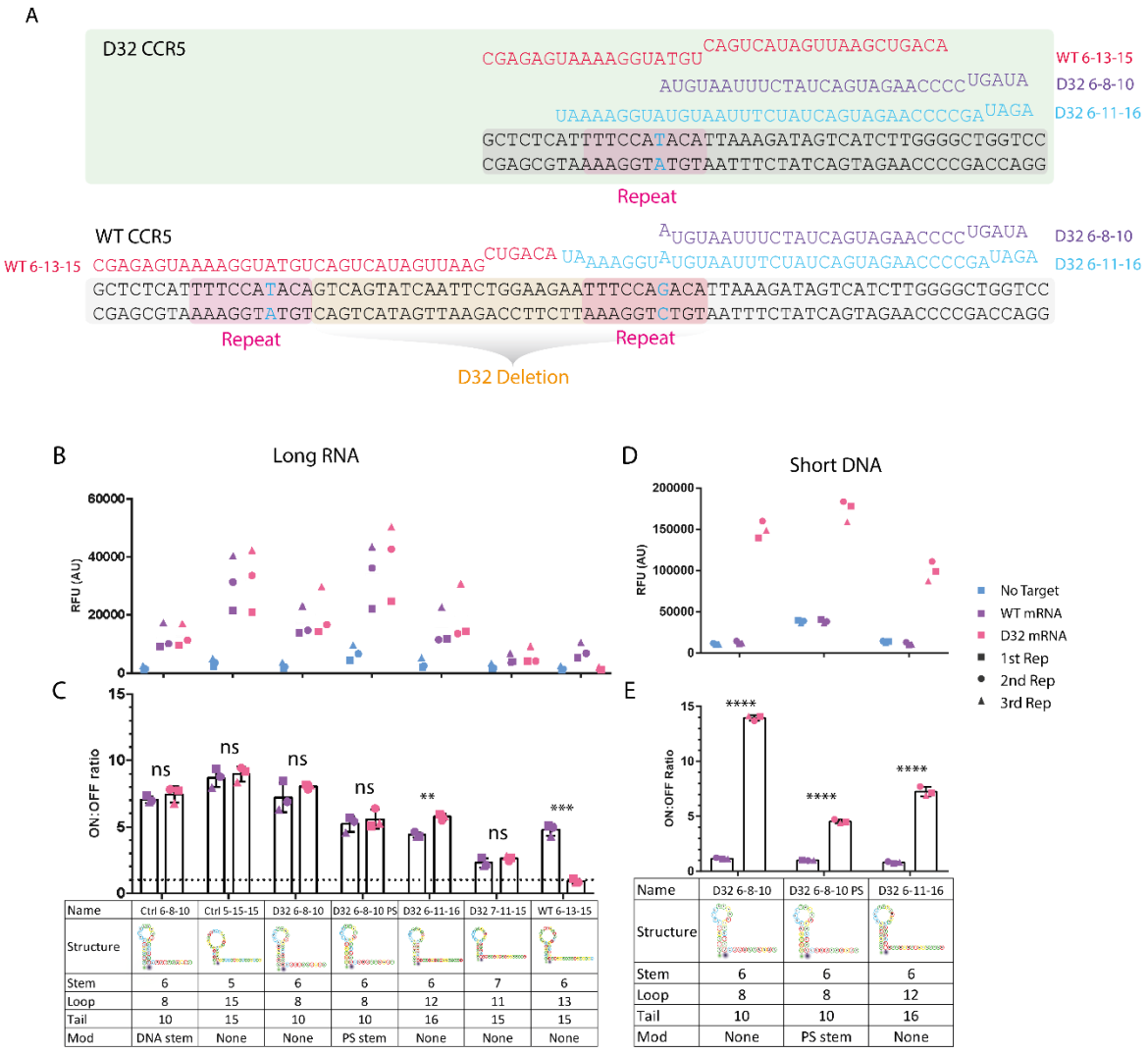


Figure 4.8 The D32 repeat. (A) Schematic depicting the repeat region in both WT and D32 sequences and showing WT and D32 beacon homology to this repeat. (B, D) Raw fluorescence measurements for each beacon with no target (blue), with WT Target (purple) and D32 target (pink) for long RNA (B) and short DNA (D). (C, E) ON:OFF ratio of Target/No Target measurements for each beacon and WT target (purple) and D32 Target (pink) for long RNA (C) and short DNA (E). Error bars show SD. * shows significance compared to null hypothesis (WT and D32 targets activate beacons to the same degree). ns = not significant. * = $P \leq 0.05$. ** = $P \leq 0.01$. *** = $P \leq 0.001$. **** = $P \leq 0.0001$. Dashed lines show an ON:OFF ratio of 1 (null value).

presence of more WT than D32 CCR5 mRNA or by the presence of more off target sequences for Ctrl_5-15-15. D32_6-8-10 was significantly activated in WT-homozygous #115 cells, suggesting that this beacon was non-specifically activated by either the WT CCR5 mRNA or by other off-targets. Disappointingly, beacon activation correlated poorly with CCR5 qPCR data from the same timepoints (fig 4.7.J). This suggests that both beacons tested, D32_6-8-10 and Ctrl_5-15-15, were activated non-specifically.

4.2.2.5 Activation of CCR5 beacons by WT and D32 mRNA

We decided to further investigate the possibility that the WT CCR5 mRNA may be capable of non-specifically activating the D32 beacons. Thus, we interrogated the sequence of WT CCR5 mRNA, looking for regions of homology to the beacon sequence. We found a 10 bp repeat with one T-G mismatch on either side of the D32 deletion. Therefore, when the deletion occurs the 30-40 bp beacon-binding region is reformed with only 1-2 mismatches (fig 4.8.A). Our *in silico* tests prior to beacon synthesis did not identify this potential problem because the mismatches for the beacons coincidentally occur at the 3' base. This issue does not affect the WT beacon. This microhomology repeat may also explain the origin of the D32 mutation, although I have not found any discussion of this in the literature.

I tested the effect of this finding *in vitro* by analysing the ON:OFF ratio of the D32 and WT beacons with both WT and D32 mRNA (fig 4.8.B and 4.8.C). I found there was no visible difference in beacon activation with WT and D32 mRNA for D32_6-8-10, D32_6-8-10-PS, and D32_7-11-15. The ON:OFF ratio of D32_6-11-16 increased by 20% with D32 mRNA compared with WT mRNA. I found this to be quite surprising, as the RNA binding sequence of D32_7-11-15 and D32_6-11-16 is the same (fig 4.4.A). The crucial difference between these two molecular beacons was the extra base pair in the stem of D32_7-11-15, showing that stem melting temperature has a significant effect on ON:OFF ratio.

I also compared the ON:OFF ratios of D32 beacons with WT and D32 short DNA targets (fig 4.8.D and 4.8.E) and found that there was no visible difference between beacon fluorescence with WT DNA and

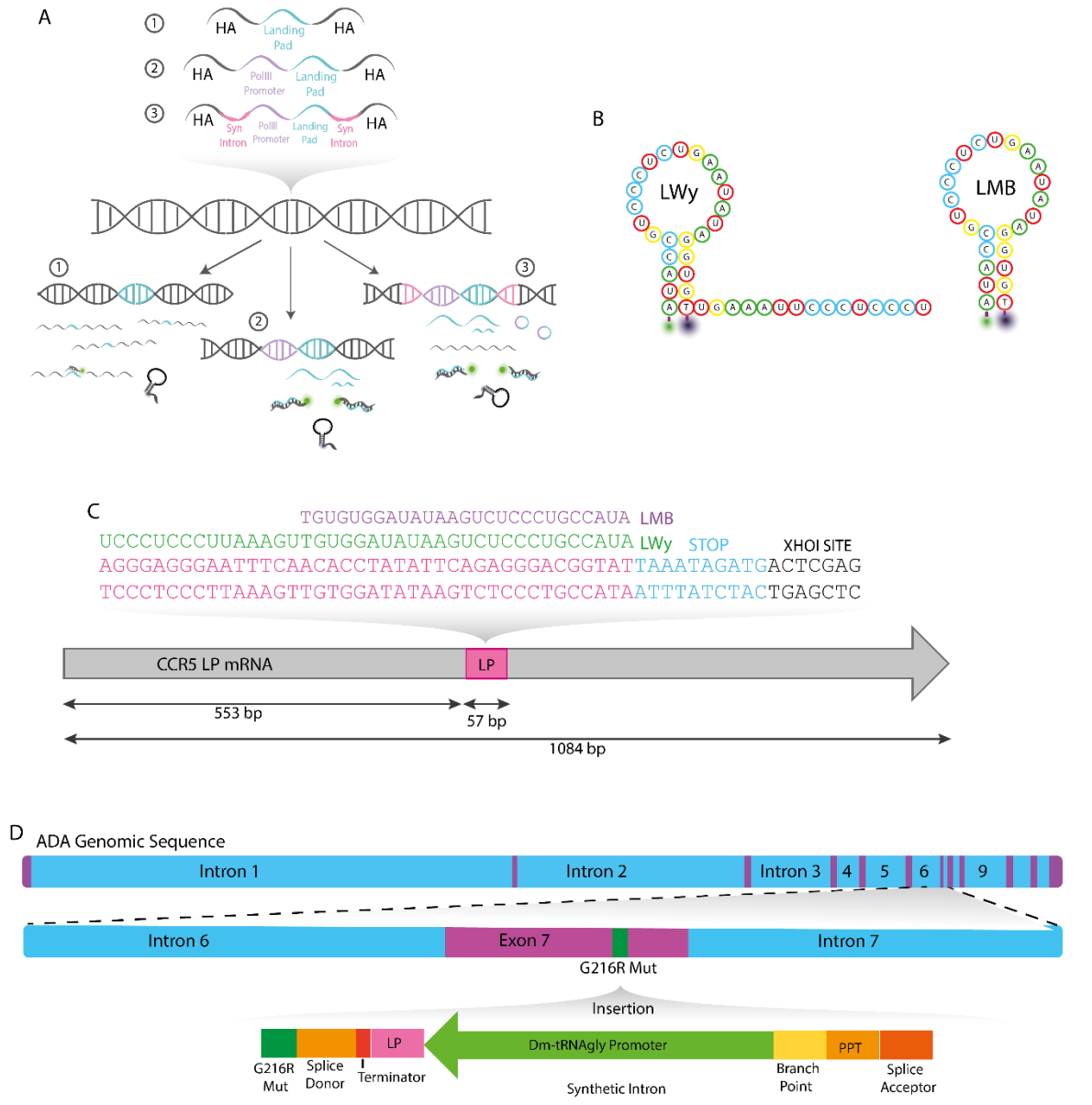


Figure 4.9 Unique landing pad implementation. (A) Schematic showing the different forms of landing pad implementation for different editing scenarios. 1) Landing pad expressed as part of a highly expressed mRNA for knockout purposes (e.g. CCR5). 2) Landing Pad amplified from minimal Pol III promoter for moderately expressed mRNAs or non-coding regions (e.g. MCS-R2). 3) Landing Pad with Pol III promoter and synthetic intron for modifications in coding regions (e.g. ADA). (B) Sequence and structure of Landing Pad Wyvern Beacon (LWy) and Landing Pad Molecular Beacon (LMB). (C) Schematic showing the landing pad sequence in the CCR5 mRNA along with the stop codons and restriction site. (D) Schematic showing the landing pad sequence (pink) within the ADA genomic sequence along with the Pol III promoter (green) and terminator (red), and the synthetic intron components.

beacon fluorescence with no target. Therefore, it is likely that activation of D32 beacons with WT mRNA is due to this homologous repeat and not due to non-specific binding to other parts of the mRNA.

I found the WT beacon, WT_6-13-15, was highly specific to the WT mRNA despite the 18 bp of its sequence which bind to both WT and D32 mRNA. One possible explanation for this is that the homology to D32 mRNA does not extend into the loop region of WT_6-13-15, and so may not be sufficient to activate the beacon.

In conclusion, the CCR5 target region presented unforeseen problems. Our simplified *in vitro* tests show that some degree of homology is acceptable (i.e. for WT_6-13-15). However, our *in vivo* tests suggest that off-target binding events are much more common than we originally thought, and that any new direction would need to take this into account.

4.2.3 A unique Landing Pad sequence for Molecular Beacons

In response to these findings, David Knapp conceived an alternative approach which would hopefully allow us to reduce non-specific beacon activation and amplify signal from specific beacon activation. Our approach consisted of incorporating a unique 'landing pad' (LP) sequence alongside the intended edit (fig 4.9). This landing pad sequence was designed to overlap as little as possible with the known human transcriptome and be the ideal structure for binding molecular beacons. We designed variations of the incorporated landing pad cassette to make it applicable to most editing scenarios (fig 4.9.A):

1) Knock outs: For simple knock out scenarios, like for the CCR5 mutation, we simply included the landing pad sequence alongside an 11 bp sequence containing a stop codon in each ORF.

2) Non-coding regions: For edits in non-coding regions (e.g. MCS-R2 enhancer) or knock outs in mRNAs with low expression levels, we included a tRNA Pol III promoter[237]. Thus, with the addition of ~70bp sequence we could achieve high levels of short RNA expression levels.

3) For SNPs/Small changes in coding regions: For edits where we did not want to significantly alter the coding sequence (e.g. ADA), we determined that we could place the landing pad and Pol III promoter within a synthetic intron (fig 4.9.D). Therefore, although we would be able to detect the landing pad within cells, the changes would be translationally silent (except, of course, for the intended SNP).

We hoped this new method could be a universal system for all editing situations, where scientists would have no need for designing and optimising new molecular beacons for their areas of interest.

4.2.3.1 Design of unique landing pad

The unique landing pad sequence was designed using an R script created by David Knapp. The script identified sequences which had minimal homology to sequences in the human transcriptome and which folded optimally for the beacon 5-15-15 and 5-15-0 structures. The script also identified sequences which optimised folding with the addition of a stop codon cassette and restriction site, to aid editing efficiency quantifications (fig 4.9.C).

We hypothesized that with the new landing pad sequence, which was designed to be unfolded, the toehold tail on the wyvern beacon may not be necessary. However, folding with the surrounding mRNA sequence was not taken into account, which may also have an impact. Therefore, we wanted to compare the performance of a wyvern beacon and a standard molecular beacon in the landing pad system (fig 4.9.B).

4.2.3.2 In vitro testing of landing pad beacons

For *in vitro* testing of the landing pad (LP) beacons, I used the CCR5 mRNA, with the landing pad stop cassette in place of the D32 deletion (LP CCR5), as the long RNA target. I found that the beacons performed very well with both short DNA and long RNA targets and that the LP molecular beacon

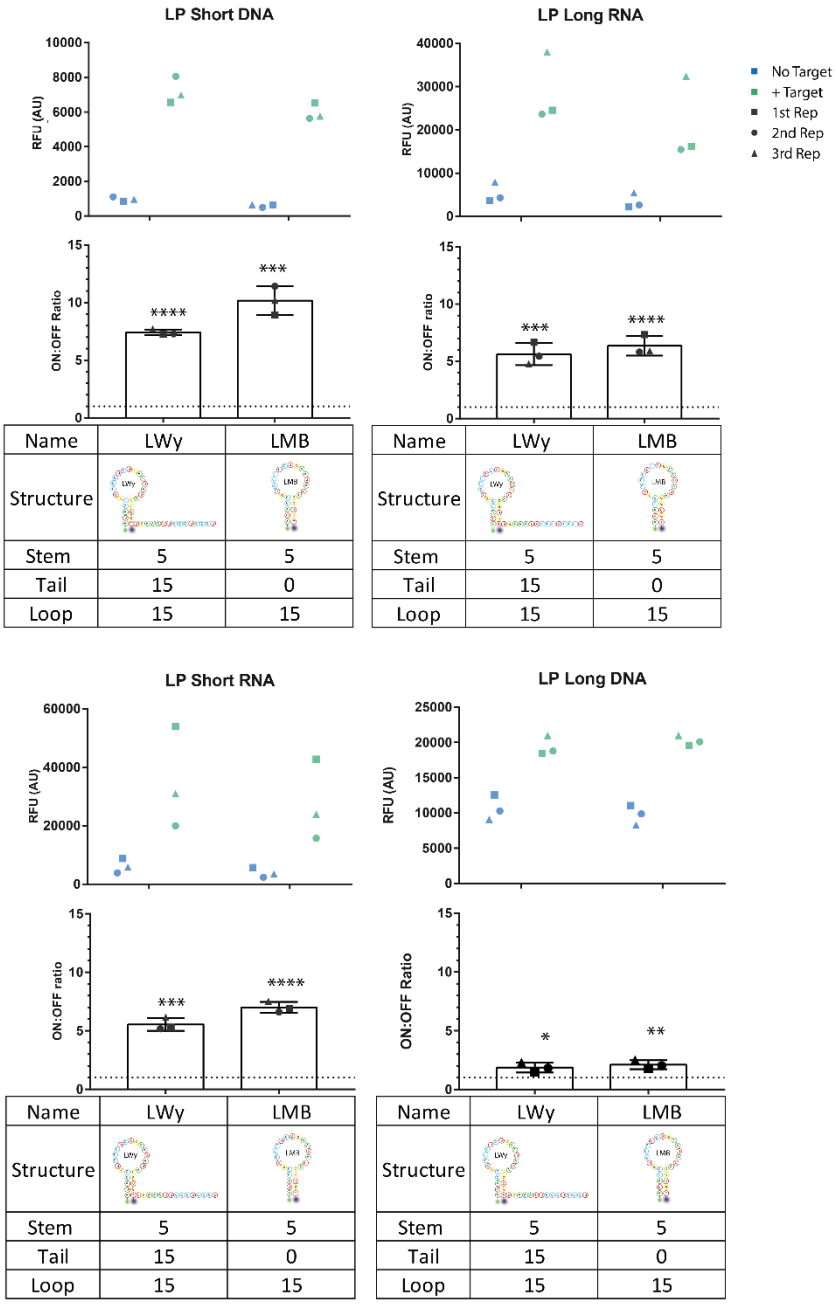


Figure 4.10 *In vitro* Analysis of landing pad beacons. (A, C, E, G) Raw fluorescence measurements for each beacon with no target (blue) and with specified target (green). (B, D, F, H) ON:OFF ratio of Target/No Target measurements for each beacon and each specified target. Error bars show SD. * shows significance compared to null hypothesis. ns = not significant. * = P ≤ 0.05. ** = P ≤ 0.01. *** = P ≤ 0.001. **** = P ≤ 0.0001. Dashed lines show an ON:OFF ratio of 1 (null value).

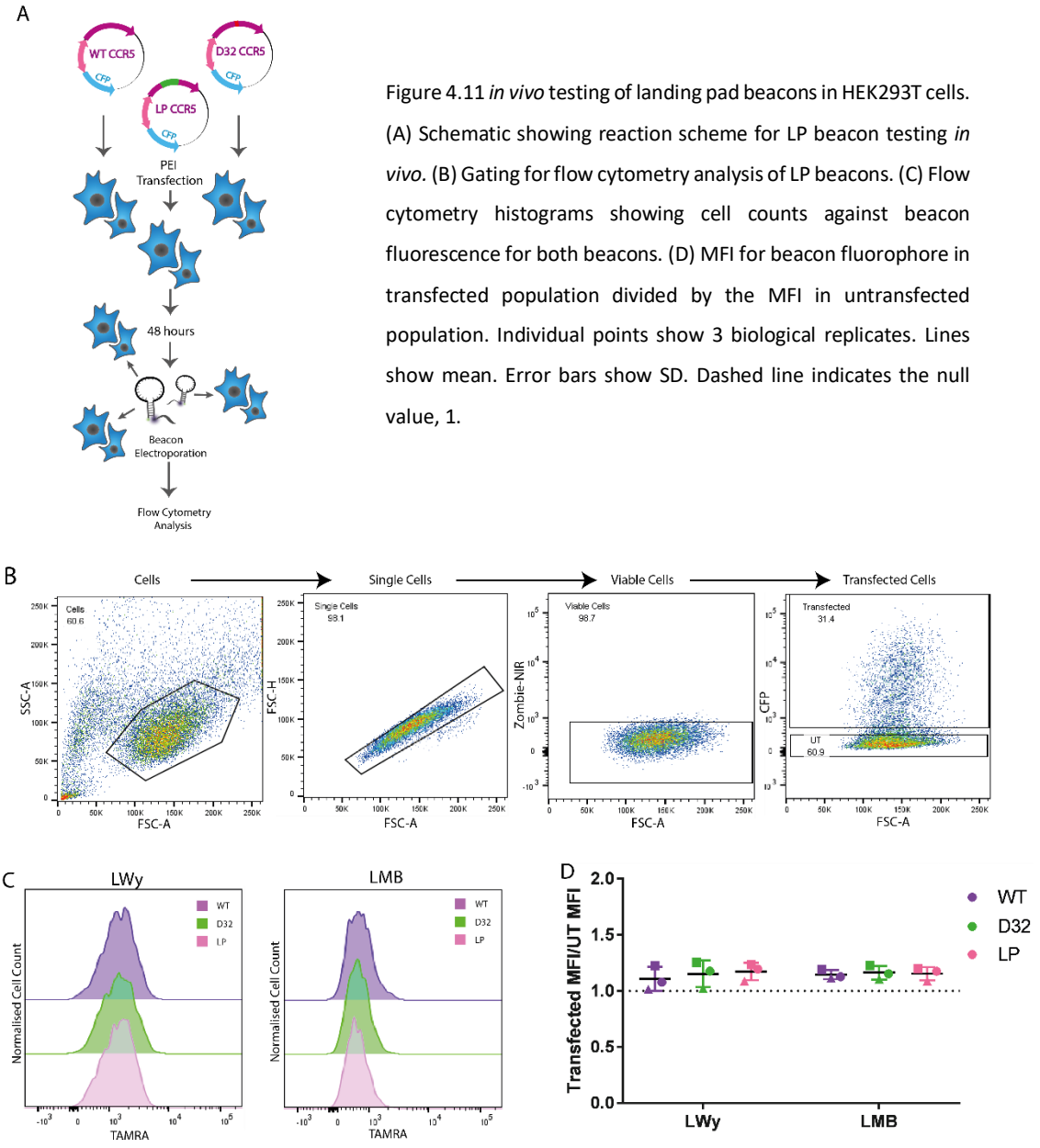


Figure 4.11 *in vivo* testing of landing pad beacons in HEK293T cells. (A) Schematic showing reaction scheme for LP beacon testing *in vivo.* (B) Gating for flow cytometry analysis of LP beacons. (C) Flow cytometry histograms showing cell counts against beacon fluorescence for both beacons. (D) MFI for beacon fluorophore in transfected population divided by the MFI in untransfected population. Individual points show 3 biological replicates. Lines show mean. Error bars show SD. Dashed line indicates the null value, 1.

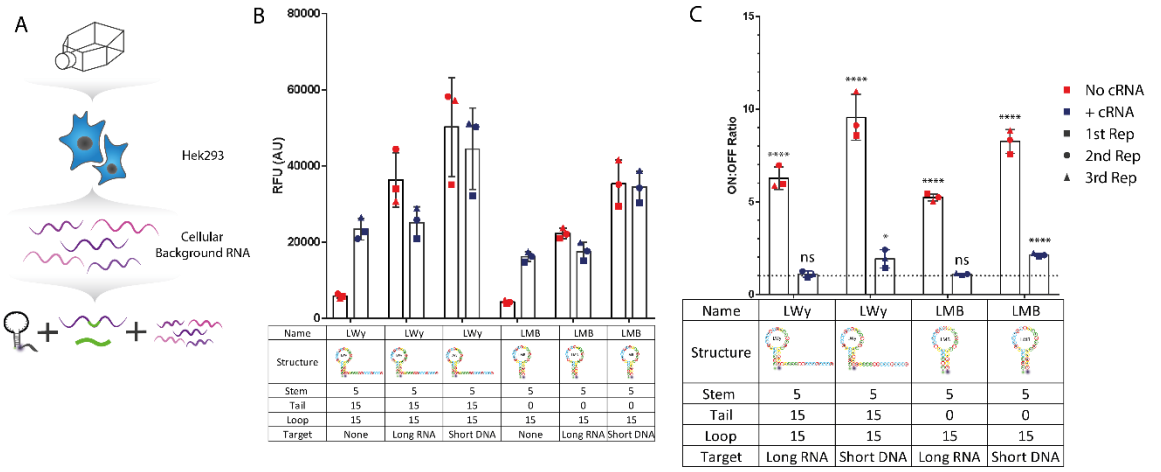
(LMB) slightly outperformed the LP Wyvern Beacon (LWy) with ON:OFF ratios 10-30% greater (fig 4.10). This difference, is possibly due to higher off values of the LWy, compared with the LMB. Both beacons again performed worse with long DNA targets, possibly for the reasons discussed previously.

4.2.3.3 In vivo testing of Landing Pad Beacons

I tested LP beacons *in vivo* using HEK293T cells transfected with plasmids expressing either: WT, D32 or LP CCR5; and ECFP, as a transfection control, from a bidirectional promoter. 48 hours post

transfection, I electroporated cells with the beacons and then observed the beacon activation using flow cytometry.

I found that beacon activation, calculated by dividing beacon fluorescence in transfected cells by beacon fluorescence in untransfected cells, was not significantly higher in LP than in WT or D32 CCR5-transfected cells. The beacon ON:OFF ratio in all transfected cell populations was observably above the null hypothesis value of 1, but this again may be due to a flow cytometry artefact. The lack of a significant change in beacon activation between transfected and untransfected cells was vastly different from the ON:OFF ratio found in *in vitro* experiments. This suggests that despite our attempts to reduce this effect, the beacons are opening non-specifically in the presence of cellular RNA.



Name	LWy	LWy	LWy	LMB	LMB	LMB
Structure						
Stem	5	5	5	5	5	5
Tail	15	15	15	0	0	0
Loop	15	15	15	15	15	15
Target	None	Long RNA	Short DNA	None	Long RNA	Short DNA

Name	LWy	LWy	LMB	LMB
Structure				
Stem	5	5	5	5
Tail	15	15	0	0
Loop	15	15	15	15
Target	Long RNA	Short DNA	Long RNA	Short DNA

Figure 4.12 *In vitro* testing of landing pad beacons with cellular RNA. (A) Schematic showing cRNA extraction and use in *in vitro* assays. (B) Raw fluorescence measurements for each beacon with specified target. Red points show raw values with no cRNA and blue points show raw values with cRNA background (C) ON:OFF ratio of Target/No Target measurements for each beacon and each specified target. Red points show ON:OFF ratio with no cRNA, blue points show ON:OFF ratio in cRNA background. Error bars show SD. * shows significance compared to null hypothesis. ns = not significant. * = P ≤ 0.05. ** = P ≤ 0.01. *** = P ≤ 0.001. **** = P ≤ 0.0001. Dashed lines show an ON:OFF ratio of 1 (null value).

4.2.3.4 In vitro testing of Landing Pad Beacons with cellular RNA

In order to more thoroughly investigate non-specific opening of beacons with cellular RNA, I performed more *in vitro* experiments where I attempted to more faithfully replicate cellular conditions (fig 4.12). For these experiments I performed the earlier ± target experiments, but this time in the presence or absence of physiological levels of cellular HEK293T RNA (for effective cellular RNA concentration calculations, see appendix).

I found that in the absence of cellular RNA ON:OFF ratios were similar to those found previously with both long RNA and short DNA. With a cellular RNA background, I found that there was no significant difference between fluorescence values with no target and with an mRNA target. I found that with a short DNA target there was some activation of the beacons in the presence of cellular RNA, but this activation was 4-5x less than the activation in the absence of cellular RNA. These results suggest that the landing pad beacons are almost fully activated by non-specific cellular RNA, despite our attempts to limit off-targets within the transcriptome.

4.3 Discussion

In this chapter I established the following:

- All beacons tested are activated non-specifically *in vivo*
- Wyvern Beacons may be a preferable option to consider when optimising molecular beacons mRNA

4.3.1 Wyvern Beacons

Unfortunately, I have not been able to produce much useful data regarding the wyvern beacons *in vivo*, and therefore, my analysis of their efficacy was limited to *in vitro* findings. When directly compared (LMB vs LWy), I found that the LWy typically had higher OFF values than LMB, suggesting that the wyvern tail may contribute to non-specific beacon opening. However, for the GFP beacons, I

found that this was not the case and that the wyvern beacon gave much higher ON values. Whereas the LP sequence was specifically designed to be accessible, the GFP sequence represents a more typical mRNA. This suggests that any benefits from the wyvern beacon may be mRNA specific, and likely depends on the accessibility of the target region. Thus, it may be a worthwhile consideration when the mRNA of interest is highly structured and target region choice is limited.

4.3.2 Beacon non-specificity

It was disappointing that I was not able to show specific activation of the beacons *in vivo.* There are several possible explanations for this and consequently several possible adaptations which may lead to improved results.

The GFP_7-21-0 was shown in the Tang et al paper to give 4x greater fluorescence in target cells compared to non-target cells[228]. However, in this case, the beacon was microinjected into cells and fluorescence levels were recorded by fluorescence microscopy. Fluorescence microscopy allows fluorescence levels to be separated into cytoplasmic and nuclear fluorescence and for bright-spots caused by aggregated beacons to be removed. Thus, artefacts caused by aggregation and nuclear accumulation of beacons can be removed. Delivery of beacons by both microinjection and electroporation is common in the literature, however, analysis of beacon fluorescence by flow cytometry is less common, although still present[223,238,224]. This suggests that it may be harder to produce beacons which show specificity by flow cytometry. For our purposes, fluorescence microscopy is not possible, and microinjection is too low throughput and damaging, thus these approaches would not work.

In their 2008 paper, Chen et al experienced similar levels of non-specificity and determined it to be due to nuclear sequestration of molecular beacons, a phenomena also seen in other papers[239,238,222,240,241]. These papers determine that molecular beacons are sequestered in the nucleus within 2 minutes of delivery, regardless of delivery method, and that most non-specific fluorescence

comes from the nucleus. This may explain why the landing pad beacons, which were designed to have minimal off targets in the transcriptome, still opened non-specifically. The Qiazol RNA extraction method used in Figure 4.12 to gather cellular RNA, was not specific to cytoplasmic RNA, and thus non-specificity of the beacon *in vitro* was shown with total cellular RNA, not cytoplasmic RNA.

To counteract nuclear sequestration, scientists have altered molecular beacons in the following ways to facilitate nuclear export or block nuclear import: addition of tRNAs[240]; linkage with cytosolic proteins[222]; supplementing with siRNA-like domains[239]; and linking to quantum dots[241]. These methods, are all potential ways for us to combat these false-positive signals, should we continue working with molecular beacons.

It's also possible that our beacons may be insufficiently optimised and that usage of modified nucleic acids, such as LNAs, which have been shown previously to aid specificity of molecular beacons, may improve this situation[217,242].

I consistently observed that transfected cells, including controls, had higher beacon fluorescence values than non-transfected cells. There are several possible explanations for this. It may be that transfected cells are more susceptible to beacon electroporation, due to more penetrable membranes. It may also be that transfected cells have a higher intracellular RNA concentration and therefore more opportunity for non-specific binding. Differing electroporation efficiencies could be controlled for by including an unquenched fluorescent probe with the beacon mix to normalise beacon fluorescence against. However, for this method to work I would need to ensure the control fluorophore used does not have FRET interference with the beacon fluorophore or quencher.

4.3.3 CCR5 D32 Mutation

One of the less relevant but more interesting findings in this chapter was the microhomology repeat within the CCR5 D32 mutation. I initially hypothesized that this microhomology may explain the prevalence of this deletion in the population (~10-16% in people of European origin)[243]. However, on review of the literature, I found that the high occurrence of the D32 mutation is probably due to

genetic drift in the European populace, rather than multiple occurrences of the same mutation[244]. Although the first occurrence was probably due to this microhomology.

4.3.4 Landing Pad

Despite initial difficulties with non-specific activation of landing pad beacons, I believe that the idea of the landing pad has many benefits. Currently, using molecular beacons for *in vivo* purposes requires trialling multiple beacons and intensive optimisation. A universal molecular beacon and target sequence is applicable to many situations, for example, single-molecule imaging of modified mRNAs.

It is possible that if we selected and trialled more sequences from the landing pad script, combined these with the nuclear-export additives, or refined the landing pad script to minimise off-targets within the nucleus, we would find a landing pad beacon with specificity *in vivo.* For detection of editing, the landing pad system has many advantages: it is flexible, applicable to most editing cases, amplifiable and translationally silent.

However, we have found that specificity for small targeting RNA sequences often comes by combination with a protein partner, i.e. for the miRNA, siRNA and CRISPR systems. Therefore, we decided on a new approach, utilising the landing pad system in combination with a synthetic Cas9 circuit.

Chapter 5: A Cas9-circuit based system for enriching for correctly edited cells

5.1 Introduction

The landing pad method discussed above is a translationally silent method to detect correctly edited cells, using signal amplifying small RNAs. These small RNA-generating constructs could be embedded within most Cas9-based edits undertaken with ssODNs. In Chapter 4 we found that molecular beacons may not be the most effective method for detecting these small RNAs. Therefore, David Knapp devised a new technique to identify correctly edited cells by adapting the small RNA into a crRNA. As discussed in section 1.4, crRNAs are small (typically 36 bp) RNA sequences that provides the specific part of the targeting capability of the crRNA:tracrRNA:Cas9 complex[68,245,246]. We could then use a transiently transfected Cas9-reporter circuit, specifically activated by the crRNA, to amplify the signal and efficiently detect crRNA-expressing correctly edited cells (fig 5.1).

5.1.1 Cas9 Activators

The Cas9 nuclease had been revolutionary not just because of its impact on the gene editing field but also due to its potential in synthetic biology. The significant appeal of Cas9 here, is not its cutting capacity, but its ability to specifically localise to sequences in DNA, in an easily programmable manner. Functional versatility of Cas9 was developed by the addition of extra modules to the RNP, with fusion of protein domains to the Cas9 and RNA aptamers to the sgRNA[143,247–250]. Therefore, the Cas9 system can be used to specifically colocalise all three of the major cellular biopolymers: DNA, RNA and protein. This is useful both for the building of sophisticated synthetic circuits and the interrogation of cellular mechanisms.

The first Cas9 fusions created involved fusing of activator or repressor domains, such as VP64 and KRAB, to the dCas9 protein[143,247,250]. The fusions were particularly useful as part of CRISPRa (activation) and CRISPRi (interference) screens, as they used a single regulator that could be guided to multiple specific locations. Other researchers appended RNA aptamers such as the MS2 loop to compatible regions of the sgRNA, which could then recruit MS2 fusion proteins to the Cas9 RNP[143,248,249,251]. Fusion

to the sgRNA allowed locus-differential regulation using a single Cas9 protein with multiple different modified guides. Evolutions of these systems produced synergistic activators, such as the SAM system which recruited three different activator proteins, VP64, p65 and HSF1, to the same locus[249]. This was achieved using a dCas9-VP64 fusion and MS2-p65-HSF1 fusions recruited to MS2 loops in the stem loop 2 and tetraloop of the sgRNA. These activating proteins were posited to each recruit different components of the transcription machinery and so were expected to have a synergistic activating effect on repressed loci. The VPR system uses a similar principle but recruits the VP64-P65-RTA proteins only by fusion to dCas9[252]. The efficacy of these different systems has not been directly compared, but all have been shown to work effectively in mammalian cells.

5.1.2 Synthetic Cas9 Circuits

dCas9-regulator fusions have become an important tool in the creation of synthetic cellular circuits. Many circuits involve reporter-based readout of cellular or environmental inputs, effected by some form of transcriptional activation or repression.

Prior to the CRISPR revolution, ZFNs and TALENs were first used as effectors in synthetic circuits. However, for more complex circuits Cas9 is advantageous as it can target multiple different loci simultaneously. Two early examples of Cas9 for synthetic circuits showed how use of one sgRNA to drive expression of a second sgRNA could create regulation cascades with multiple outputs[78,79].

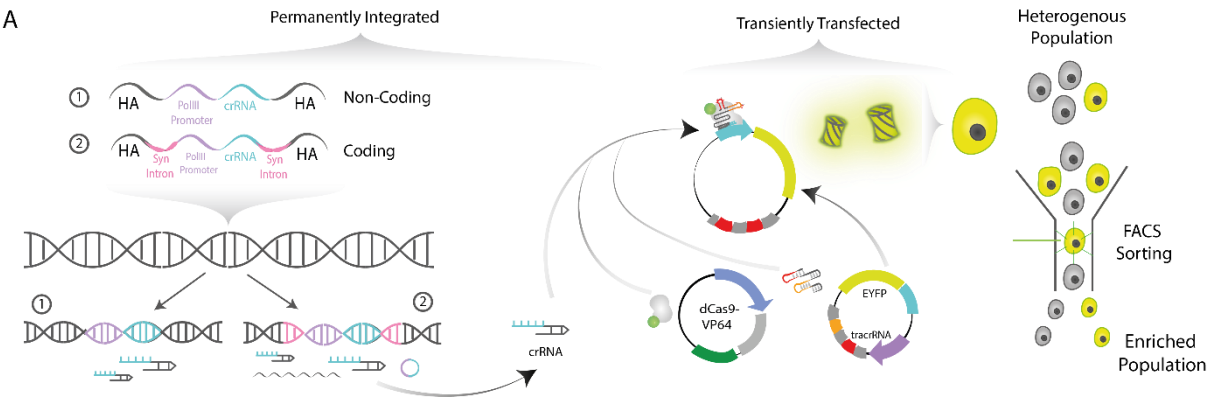


Figure 5.1 A synthetic circuit for FACS-based selection of correctly edited cells from a heterogenous population. The permanently integrated, crRNA combines with the transiently transfected tracrRNA and dCas9-activator to form a targeted activating complex which then stimulates expression of the EYFP reporter. EYFP-expressing cells can then be sorted by FACS.

5.1.3 Circuit Design

Our circuit-based editing detection system relies on the permanent integration of a crRNA alongside the intended edit, which becomes the input for a transiently expressed Cas9 based reporter circuit (fig 5.1). Combination of the crRNA with transient tracrRNA and dCas9-activator results in recruitment to target sites on the reporter sequence and expression of the reporter. The circuit output is an enhanced yellow fluorescent protein (EYFP) reporter, which allows reporting of correctly edited cells by FACS. This system has two amplification steps: amplification of the crRNA by the tRNA Pol III promoter; and subsequent activation of the EYFP reporter by the dCas9-activator. Therefore, we hoped it would produce a robust signal. In theory, both editing and circuit components can be co-delivered, although in practice this may be harder to achieve. As discussed for the landing pad in section 4.2.3, the crRNA can be made translationally silent by use of synthetic intron sequences and is relevant for most editing situations.

5.2 Results

5.2.1 Optimising crRNA spacer length

Before attempting to use the circuit to detect edited cells, we first wanted to determine the inherent efficiency of a synthetic Cas9 circuit at detecting and reporting the presence of our chosen crRNA and assess potential problems with the system. We used a crRNA sequence previously developed in the lab which has been proven to have no known targets within the human genome, CRISPR targeting sequence 1 (CTS1)[253]. For these preliminary optimisation experiments, we expressed the CTS1 crRNA sequence on a plasmid using the human glutamine tRNA, which has been shown to be an efficient Pol III promoter[237]. . The crRNA scaffold we used was a homemade minimal version to further reduce the size of the insert. We based our tracrRNA on the modified sgRNA produced by Zalatan et al in their 2015 paper[248], which contained 3' MS2 and F6 loops, both of which recruit the MS2 protein. These MS2 loops recruited the MS2-p65-HSF1 fusion proteins first described in the SAM system[249].

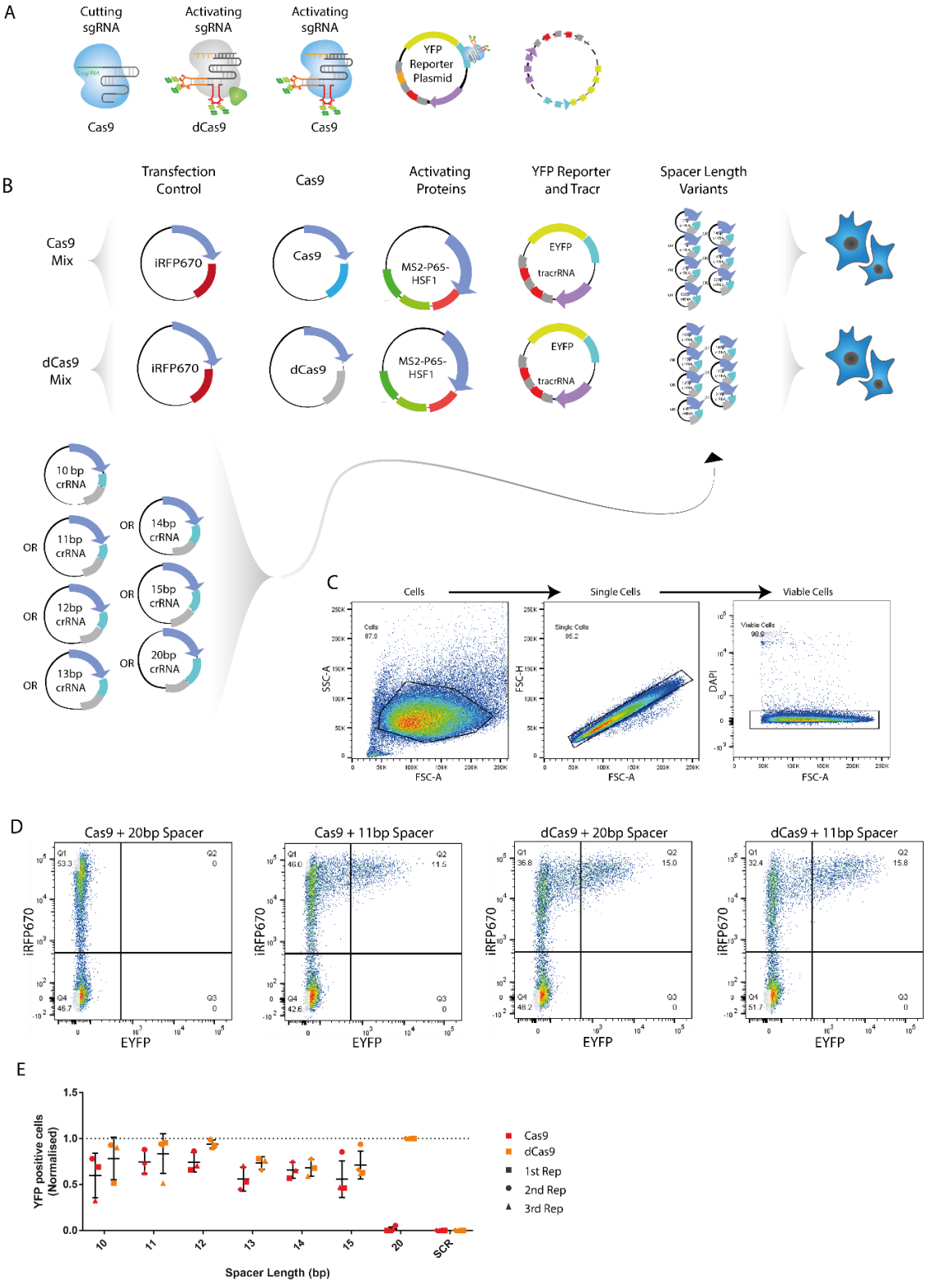


Figure 5.2 Effect of spacer length on sgRNA efficiency. (A) Diagram demonstrating the problem of guide exchange between editing and activating guides. Targeting of the Cas9 to the YFP reporter leads to reporter destruction instead of activation (B) Schematic showing transfection scheme for experiments. Active or Dead Cas9 are transfected with a YFP reporter, tracrRNA 2.0, and activating proteins. For each of these mixes 7 spacer length variants were tested. (C) Flow cytometry gating strategy. (D) Typical flow cytometry plots showing percentage activation of reporter in transfected cells for Cas9 mix and spacer length indicated. (E) Quantification of plots from C. Percentage of YFP positive cells in transfected iRFP670 positive population were calculated. For each replicate data were normalised by dividing by YFP+ % for 20 bp spacer with dCas9. The dotted line shows y = 1.

We transfected the crRNA plasmid alongside an iRFP670 transfection control and the circuit component plasmids, which encoded the Cas9-activator, tracrRNA, and the EYFP reporter. To reduce the required plasmid number, the EYFP reporter and tracrRNA were encoded on the same plasmid. The EYFP reporter consists of 8 repeats of the CTS1 crRNA target site, separated by 22 bp spacer sequences, followed by a minimal promoter sequence from the adenovirus major late promoter (MLP), before the Kozak and EYFP sequence. Cas9 variants and modified tracrRNA-targeted activator proteins were delivered on separate plasmids. Using this controlled system, we established the inherent efficiency or activation of the circuit by determining the percentage of EYFP-positive cells within the iRFP670-positive transfected population.

One potential issue we envisaged with combining a Cas9-based editing approach and a dCas9 circuit-based selection approach, was the potential for exchange of sgRNAs between the different Cas9 complexes, a phenomenon which has been reported before[254]. Guiding of the editing Cas9 by our reporter-activating sgRNA would lead to reporter destruction and subsequent circuit silencing (fig 5.2.A). Truncation of spacer sequences to 14-15 bp has been shown to maintain targeting capacity of the sgRNA-Cas9 complex while abolishing the nuclease activity of the Cas9[255]. Truncating the crRNA sequence therefore has the advantage that where the CTS1 crRNA forms a complex with the editing Cas9, the RNP will be unable to cut and destroy the reporter but will still activate EYFP expression. Shortening the crRNA spacer length has the additional advantage of reducing the size of the insert sequence required and thus most likely increasing the efficiency of editing[256]. Thus, the first component of the circuit we chose to evaluate and optimise was the spacer length of the crRNA.

To investigate the effect of spacer length on circuit activation, we transfected plasmids containing crRNA spacer length variants of 10-15 bp and 20 bp into HEK293T cells alongside one of two different circuit mixes (fig 5.2.B). The cutting mix was composed of plasmids expressing: an iRFP670 transfection control, Cas9, the tracrRNA and EYFP reporter, and MS2-P65-HSF1 fusion proteins. The dCas9 mix was nearly identical to this, albeit with dCas9 in place of Cas9. Three days after transfection, I evaluated

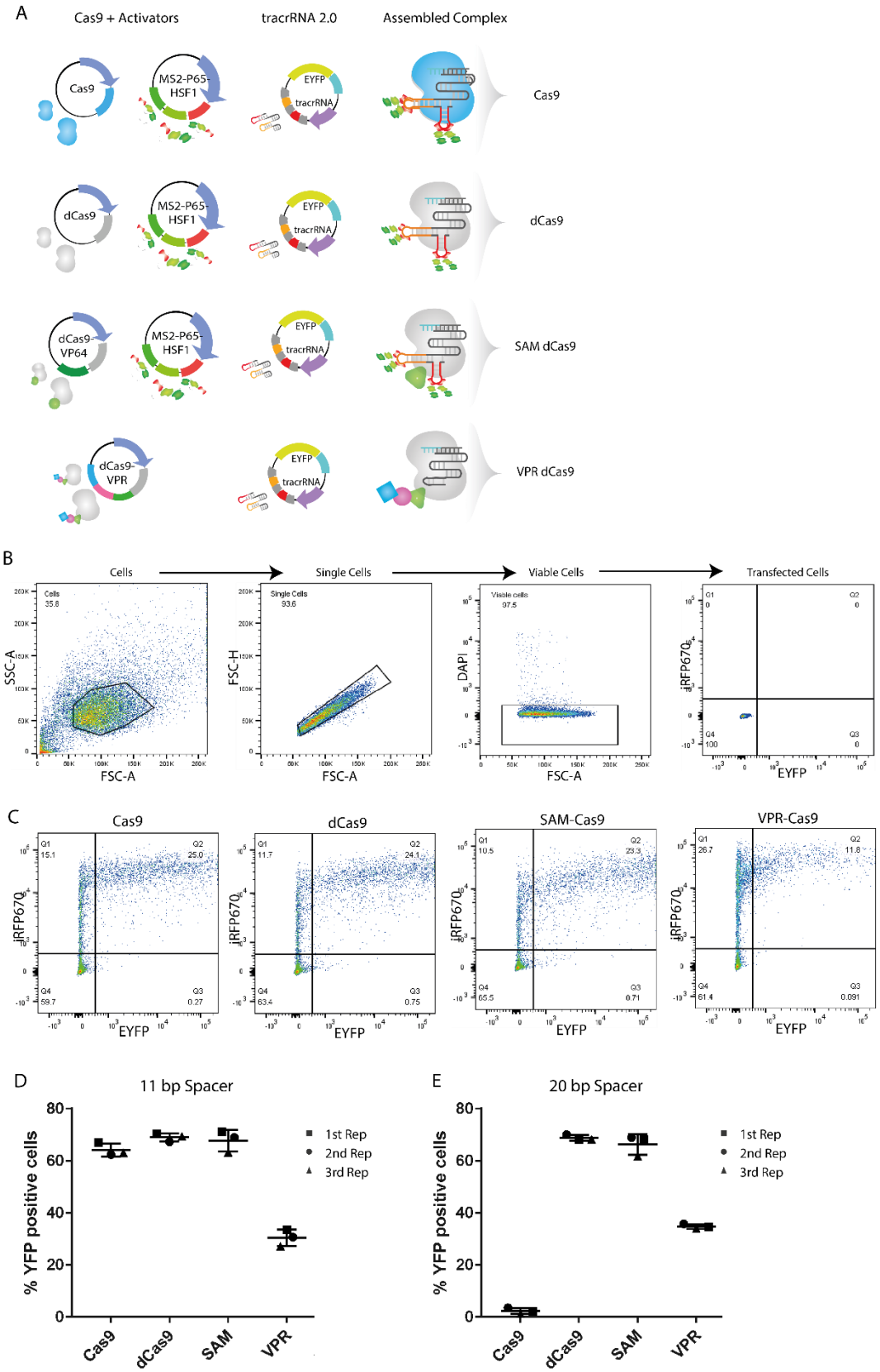


Figure 5.3 Optimisation of Cas9-Activator Complex. (A) Schematic showing the plasmids and components involved for each Cas9-Activator Variant. (B) Flow Cytometry Gating Strategy. (C) Representative flow cytometry plots showing iRFP670 fluorescence against EYFP expression, for each Cas9-activator variant. (D and E) Quantification of plots from C for 11 bp spacer (D) and 20 bp spacer (E). Y axis shows percentage of EYFP positive cells in iRFP670 positive population. Three biological replicates are shown. Error bars show SD.

iRFP670 and EYFP expression with flow cytometry (fig 5.2.C and D). I then compared the percentage EYFP activation for each spacer length with that from the optimal combination, a 20 bp spacer and dCas9 mix (fig 5.2.E).

As expected, I found that with Cas9, the 20 bp spacer had negligible EYFP activation, most likely due to cutting and degradation of the reporter plasmid. Surprisingly I found that EYFP activation was lower for spacers of length 13-15 bp than for those of length 11-12 bp with both Cas9 and dCas9. For spacer lengths of 11 and 12 bp I found EYFP activation with Cas9 was nearly on par to EYFP activation with dCas9 and was close to the maximal activation achieved by the 20bp spacer with dCas9. Thus, I settled on using the 11 bp spacer for future experiments.

5.2.2 Optimising choice of Cas9-Activator Variant

I attempted to further enhance activation of the circuit using different Cas9 activators variants. As described in section 5.1.2, interest in this area has led to the development of a diverse array of Cas9-activators[88,247,249,252]. I compared circuit activation with four different Cas9-activator variants: Cas9 + tracrRNA 2.0 + MS2-P65-HSF1 (Cas9); dCas9 + tracrRNA 2.0 + MS2-P65-HSF1 (dCas9); dCas9-VP64 + tracrRNA 2.0 + MS2-P65-HSF1 (SAM); dCas9-VP64-P65-RTA (Cas9-VPR) (fig 5.3.A).

As with previous experiments, I transfected HEK293T cells with plasmids expressing the Cas9-activator variants, an iRFP670 transfection control, the tracrRNA 2.0 and EYFP reporter, and the CTS1 crRNA with either an 11 bp or 20 bp spacer. Three days later I analysed the cells by flow cytometry (fig 5.3.B and C) and determined the percentage of EYFP positive cells in the transfected population (fig 5.3.D and E). I found that using either Cas9, dCas9 or SAM-Cas9 had little effect on circuit activation, with 65-70% of transfected cells expressing EYFP for both the 11 and 20 bp spacer. However, I found that the Cas9-VPR activator consistently underperformed compared to the other activators, with only ~30% of transfected cells expressing EYFP for both 11 and 20 bp spacers. Therefore, for future experiments, I settled on using SAM-Cas9.

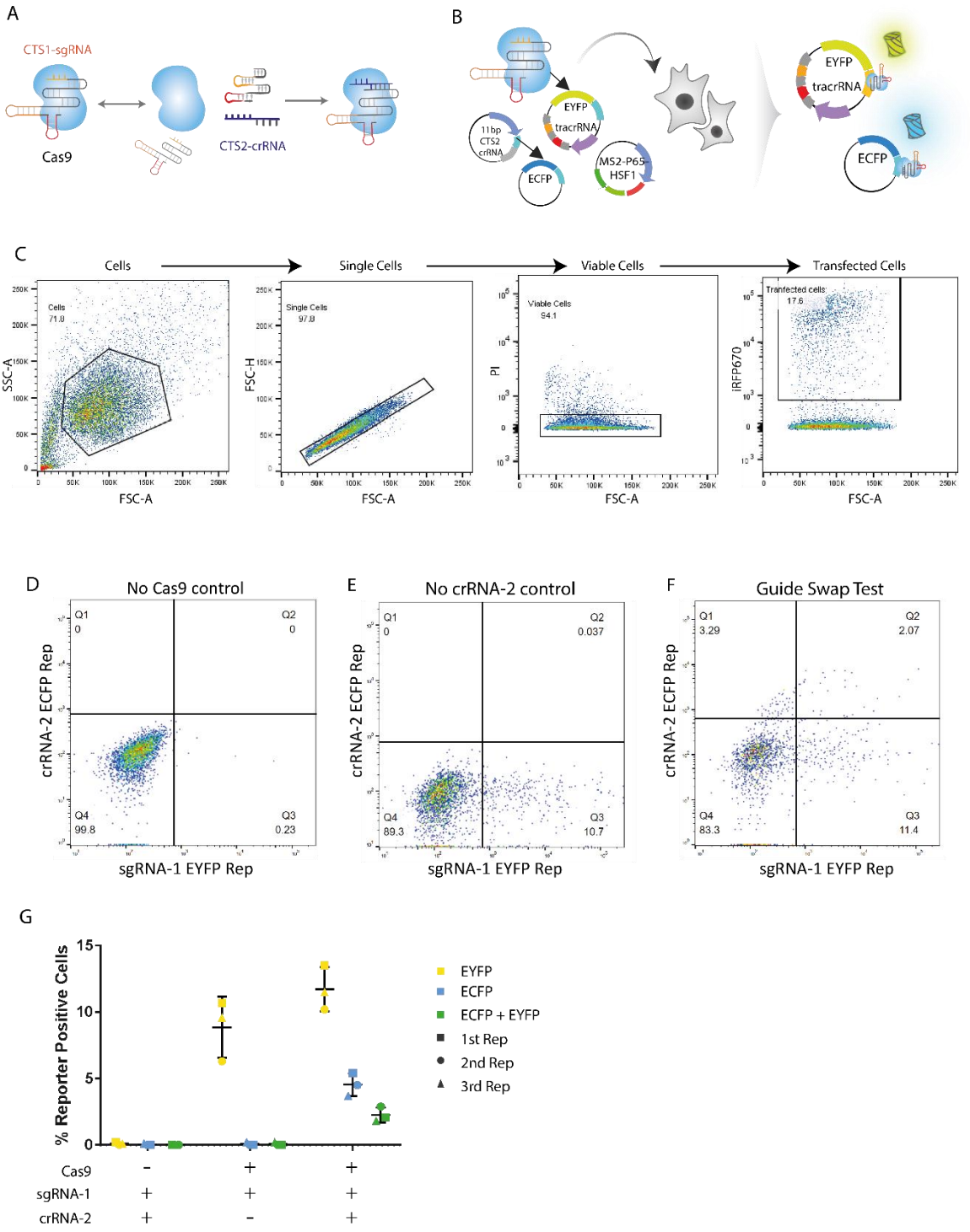


Figure 5.4 sgRNA and crRNA-tracrRNA exchange in Cas9. (A) Schematic potential mechanism for exchange of sgRNA for crRNA-tracrRNA on Cas9 within cells (B) Schematic for experimental plan for sgRNA exchange test. ECFP and EYFP reporters, CTS1 sgRNA-Cas9 RNP, activating proteins and CTS2 crRNA are delivered to HEK293T cells. The yellow and blue barrels represent EYFP and ECFP, respectively. (C) Flow cytometry gating strategy. (D, E, F) flow cytometry plots showing EYFP and ECFP reporter fluorescence for: Cas9-negative control (D), crRNA-2 negative control (E), and Cas9+ and crRNA-2+ population. (G) Quantification of EYFP, ECFP and EYFP+ECFP, reporter positive cells from D, E and F.

5.2.3 Evidence of sgRNA exchange

As discussed in section 5.2.1, it is possible that exchange of sgRNA between Cas9 and dCas9 may occur(fig 5.2.A)[254]. To minimise damage to the circuit system, I determined in section 5.2.1 that use of an 11 bp spacer maximises reporter targeting but minimises reporter cutting with Cas9. However, it is still important to know whether there is any leakage between the editing system and reporter system, as this may affect the desirability of this enrichment method in some instances.

Therefore, we designed an experiment to determine whether exchange can occur between a crRNA-tracrRNA and a sgRNA-Cas9 complex, resulting in a crRNA-tracrRNA-Cas9 complex (fig 5.4.A and 5.4.B). I transfected HEK293T cells with two plasmids encoding two different fluorescent reporter constructs, respectively: EYFP with CTS1 target sites and tracrRNA, and ECFP with CTS2 target sites. I also delivered a CTS1 sgRNA-Cas9 RNP and a plasmid expressing the CTS2 crRNA. I then examined expression of both reporters by flow cytometry (fig 5.4.C, D, E and F). I controlled for non-specific reporter activation by including a Cas9-negative and CTS2 crRNA-negative control.

I found that in samples with both CTS2 crRNA and Cas9, ~5% of transfected cells activated the CTS2-ECFP reporter and ~2% of cells activated both the CTS1-EYFP and CTS2-ECFP reporters (fig 5.4.G). In comparison, for both samples containing Cas9 CTS1-EYFP reporter expression was ~10% of transfected cells. Thus, I concluded that the likelihood of sgRNA exchange was relatively high and researchers looking to use this system should be aware of this.

5.2.4 Clearance of transient circuit components from system

One of the advantages we predicted from our circuit enrichment system is that, with the exception of the permanently integrated crRNA, all the components are transient. We wanted to verify that plasmids would indeed disappear from cells. I investigated this by observing the loss of the reporter plasmid from the system. To more accurately visualise plasmid, as opposed to protein, disappearance, I used an EYFP reporter tagged with a destabilising domain (dEYFP), which targeted the protein for degradation.

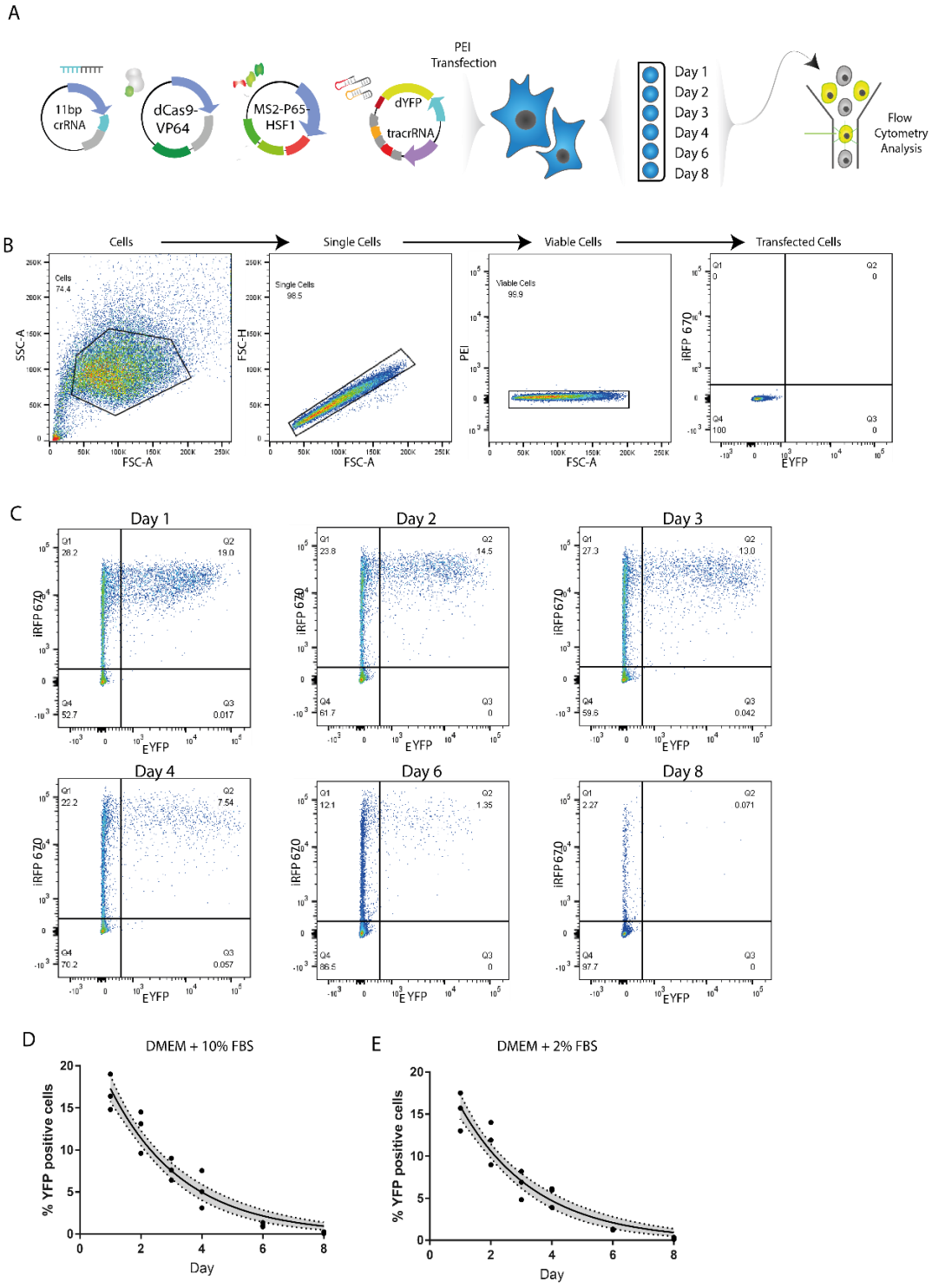


Figure 5.5 Investigating Circuit Longevity. (A) Graphic showing the experimental.scheme for this time course. (B) Flow Cytometry Gating Strategy. (C) Representative flow cytometry plots showing iRFP670 fluorescence against YFP expression, for each time point in cells in 10% FBS. (D and E) Quantification of plots from C for cells: in 10% FBS (D), and 2% FBS (E). Y axis shows percentage of YFP positive cells in the total population. Points show three biological replicates. Best fit curve calculated using non-linear regression. Shaded area shows 95% confidence levels.

One assumption I made here is that all plasmids would be degraded at a similar rate, which may not be the case according to some studies[257,258]. Intracellular degradation depends on the presence of labile sites which the plasmids used may have to a greater or lesser extent. In addition, rate of plasmid degradation likely varies between different cells lines and different delivery methods. Nevertheless, reporter loss is a reasonable proxy for determining that the circuit system is indeed transient, if not for determining exact kinetics of all components in individual systems.

To observe reporter loss, I first transfected cells with the dEYFP reporter plasmid, the plasmids encoding SAM-Cas9 components, and the 11 bp crRNA plasmid (fig 5.5.A). I harvested and analysed iRFP670 and EYFP expression at subsequent time points ranging from one to eight days post transfection (fig 5.5.B and C). Additionally, I conducted experiments in media predicted to stimulate different growth rates, using different proportions of FBS. 10% FBS was predicted to drive more rapidly dividing cells, and 2% FBS was predicted to give slower growth rates[259,260].

Surprisingly, I found there was little difference in the plasmid clearance kinetics between cells in 10% and 2% FBS (fig 5.5. D and E). I do not think this was due to maintenance of the plasmid within the HEK293T cells as none of the plasmids used contained the SV40 replication site. I calculated plasmid clearance by observing the EYFP positive cells in the total population instead of in the transfected population, as the iRFP670 protein was also lost from cells overtime, albeit with differing kinetics. I observed that at the 8-day timepoint, a negligible number of cells were EYFP-positive.

I also observed differing kinetics with the dEYFP reporter to the EYFP reporter, due to more rapid protein degradation. With the dEYFP reporter, I found that the proportion of EYFP positive cells was greatest at the 1-day timepoint and declined exponentially over time (fig 5.5.D). However, with the EYFP reporter I found that the greatest proportion of EYFP positive cells was found at the 2-day timepoint (data not shown). This is likely due to the greater accumulation of EYFP protein with the EYFP reporter as opposed to the dEYFP reporter.

5.2.5 Enriching for correctly edited cells with the circuit system

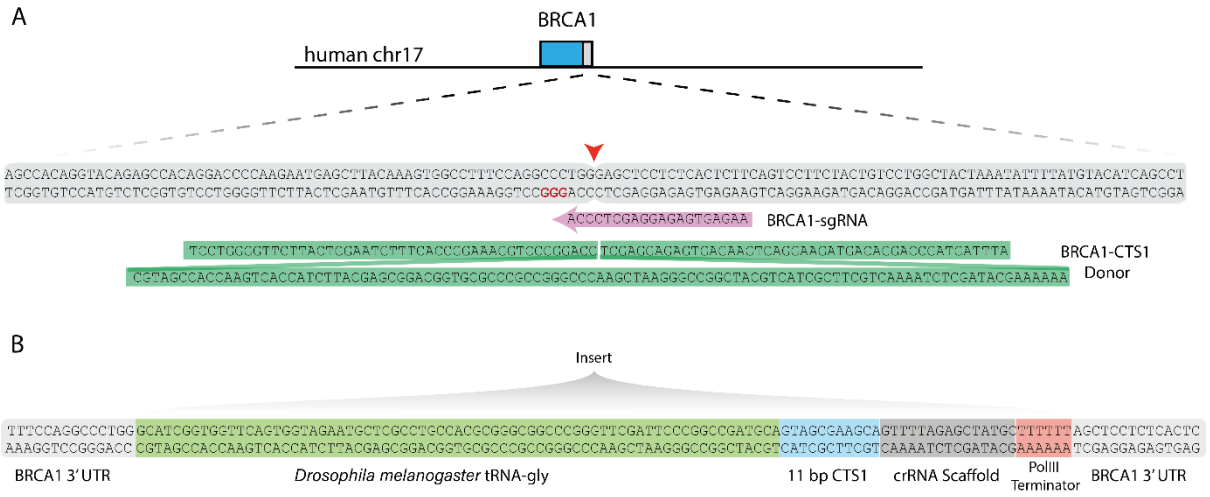


Figure 5.6 Editing the BRCA1 Locus. (A) Schematic of BRCA1 locus. The gene body is shown in blue and 3'UTR in grey. The sgRNA is pink, PAM coloured red, and donor shown in green. The red arrow indicates the Cas9 cut site. (B) Diagram detailing the correctly edited locus with tRNA promoter, CTS1 crRNA and Pol III terminator.

Following optimisation of the circuit components, we sought to test the efficacy of our system at detecting correctly edited cells. We edited the BRCA-1 3'UTR in HEK293T cells, a locus our lab had experience in editing[261]. We wanted to use HEK293T cells at this stage as they are resilient and efficiently electroporated and transfected.

We designed a single stranded oligo donor which would introduce the 11 bp CTS1 spacer and crRNA scaffold under the control of a minimal Pol III promoter. For the minimal Pol III promoter we settled on the *Drosophila melanogaster* tRNA-gly, which has been shown to promote expression of small RNAs at a similar level to that of the U6 promoter[237]. We chose to use a *Drosophila* tRNA instead of a human tRNA to reduce the likelihood of tRNA-mediated HDR within the host cell. Downstream of the crRNA scaffold we inserted a Pol III terminator. To reduce the likelihood of false positives, we designed the single stranded donor to be antisense, as the CTS1-crRNA sequence would then only be present within the nucleus after incorporation into the genome. We used homology arms of 45 bp, for an

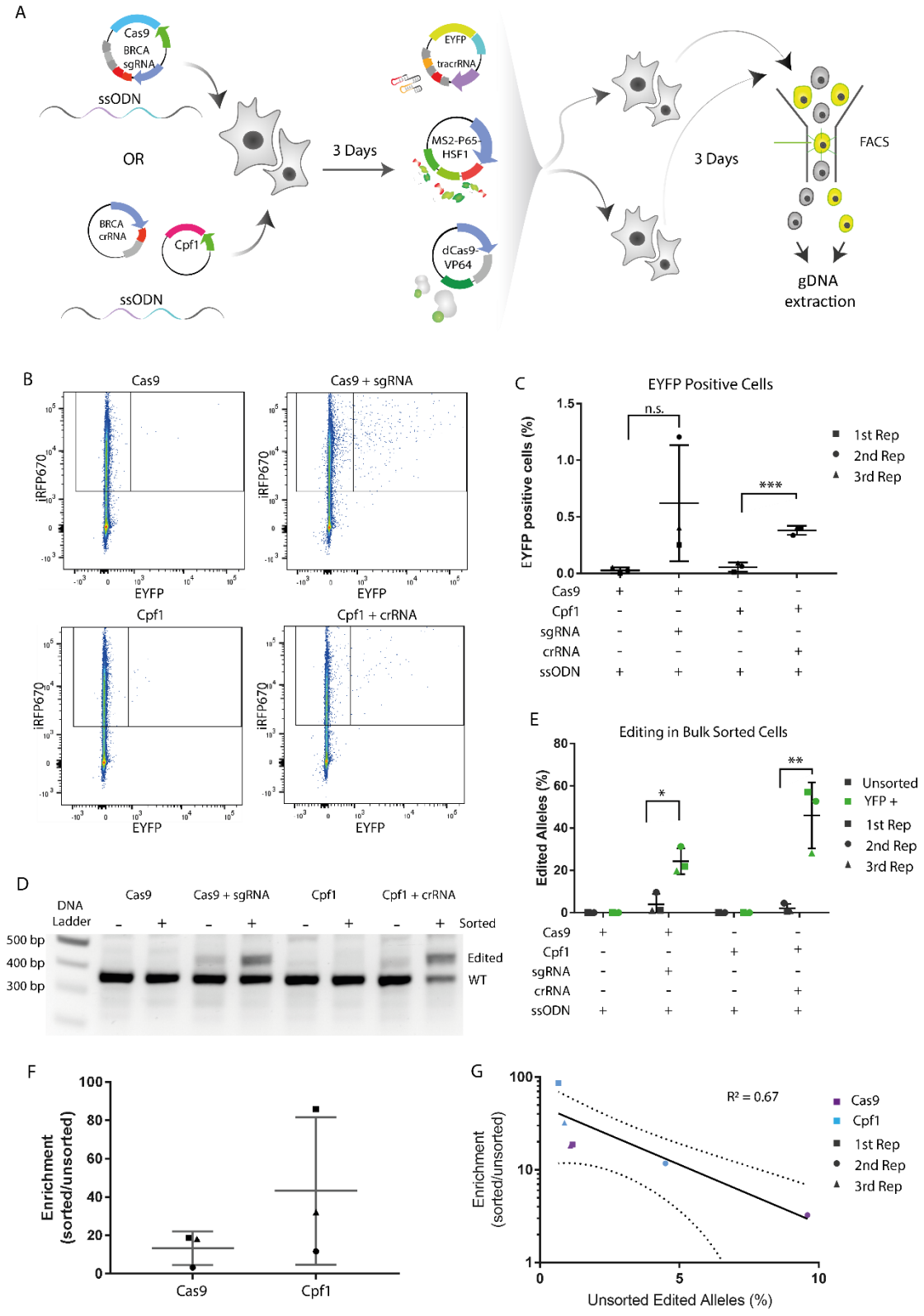


Figure 5.7 Enrichment of correctly edited cells using a Cas9 synthetic circuit. (A) Experimental scheme for editing and sorting of correctly edited cells. (B) Example flow cytometry plots for Cas9 and Cpf1-based editing, ± guide RNAs. (C) Percentage of YFP+ cells within the iRFP670+ population, for Cas9 and Cpf1-based editing, ± guide RNAs. Results are from 3 biological replicates (D) Example agarose gel of BRCA1 3'UTR for YFP+ cells and unsorted cells from Cas9 and Cpf1-based editing, ± guide RNAs. Both the edited (~450 bp) and wild type (~350 bp) bands are labelled. (E) HDR% calculated from (D) for three biological replicates. (F) Fold enrichment (HDR% in sorted/HDR% in unsorted) for Cas9 and Cpf1 calculated from (E). (G) Plot showing relationship between fold enrichment and unsorted HDR% for each sample. Line of best fit was calculated using least squares fit. Dashed lines show confidence interval 95%.. ns = not significant. * = $P \leq 0.05$. ** = $P \leq 0.01$. *** = $P \leq 0.001$

insert of 103 bp. We expected incorporation of an insert of this size to be relatively inefficient with an ssODN, and thus, a good test for our enrichment system.

We hypothesized that any crossover between the editing and circuit components would be reduced by using the Type V effector nuclease, Cpf1. Although targeting of the Cas9 with the truncated reporter guide is unlikely to result in cutting, it will likely reduce reporter activation. Therefore, we decided to test the editing potential of Cpf1 alongside Cas9. Transfection of the editing and circuit components was done sequentially (fig 5.7A). First the editing components were electroporated in as plasmid-encoded Cpf1 or Cas9 alongside the ssODN. As a control, components were delivered without the guide RNA, so false positive rates could be observed. Three days post-transfection, the circuit components, as optimised above, were electroporated into cells. After another three days, cells were analysed by flow cytometry (fig 5.7B).

I determined the percentage of EYFP positive cells within the transfected iRFP670 population. I found that levels of EYFP+ cells in the guide RNA-negative controls was consistently much lower (0.01-0.1% of transfected cells) than in the sgRNA+ populations (fig 5.7C). However, EYFP-bright false positives were still visible on the flow cytometry plots. I found that levels of EYFP+ cells in the Cpf1 edited population were very consistent (~0.45% of transfected cells), whereas levels of EYFP+ cells in the Cas9 edited population were much more variable (0.3-1.2% of transfected cells). Nevertheless, I was reassured by the observation that the edited cell populations consistently had a much higher percentage of positive cells than the negative controls.

I then sorted the cells to isolate the iRFP670+ EYFP+ population and extracted the gDNA from both this population and the unsorted cells. I amplified the BRCA1 3'UTR and observed the edited and wild-type bands on an agarose gel (fig 5.7D). I quantified HDR percentages by dividing edited band volume by total band volume (fig 5.7E). I found that unsorted cells had HDR percentages in the range of 1-5% for both Cas9- and Cpf1-edited populations. FACS-sorted Cas9-edited cells contained 20-30% HDR

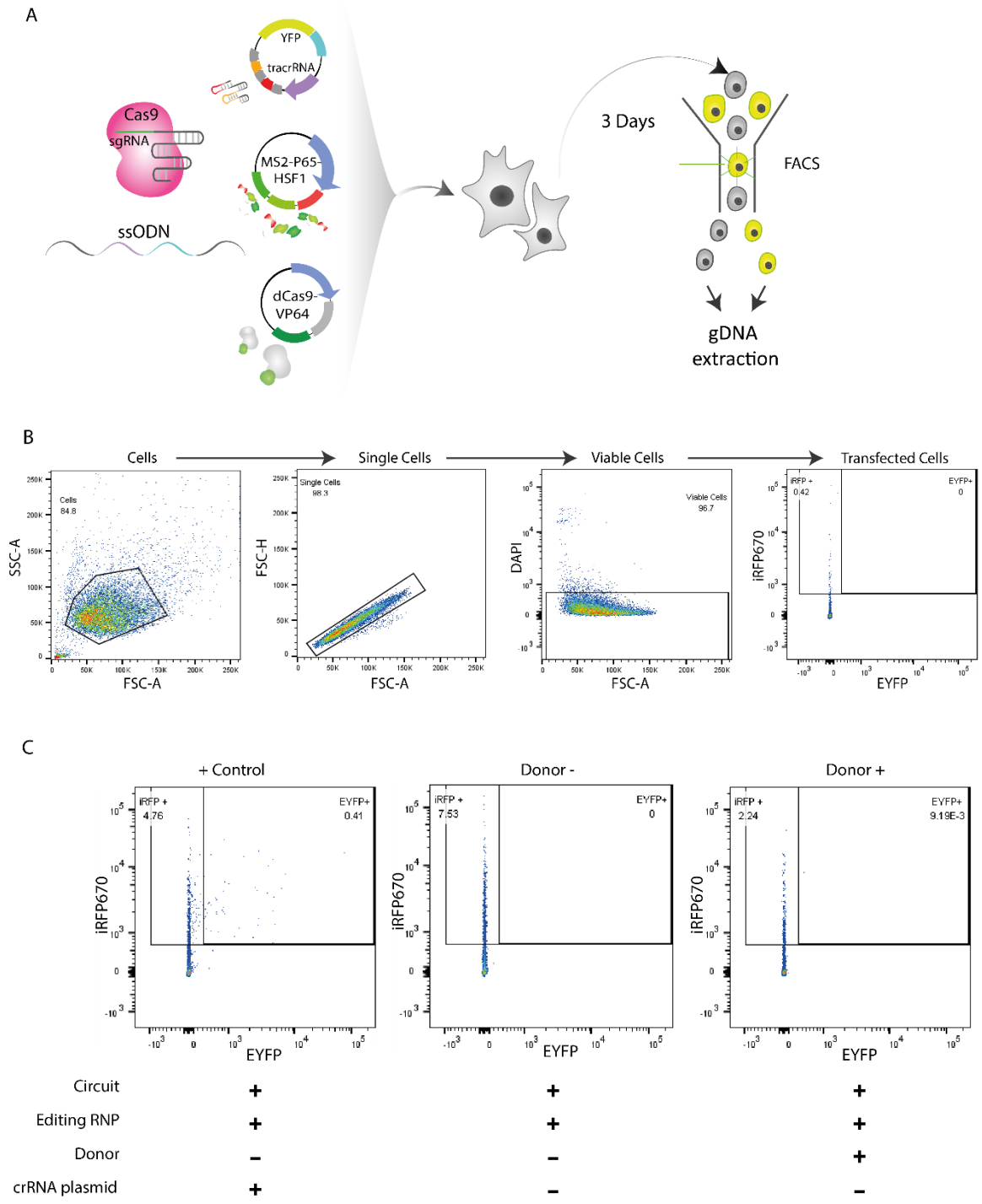


Figure 3.8 Simultaneous delivery of editing and circuit components. (A) Diagram showing experimental scheme for delivery of editing and circuit components. (B) Flow cytometry gating strategy. (C) Representative flow plots for iRFP670(transfection control) and EYFP (reporter) with electroporated components shown below.

positive alleles, whereas FACS-sorted Cpf1-edited cells contained 30-60% HDR positive alleles, suggesting up to 100% of cells could be heterozygous.

I calculated fold enrichment by dividing % HDR in sorted cells by % HDR in unsorted cells (fig 5.7F). I found that enrichment of Cas9 edited cells was reasonably consistent at 13.5-fold (± 8-fold). However, I found that enrichment of Cpf1 edited cells was much more variable, although typically much greater, at 43-fold (± 38-fold). I also plotted log enrichment against initial HDR efficiency in unsorted cells and found there is a negative correlation between these variables, likely due to inflation of fold estimations when denominators approach zero (fig 5.7.G).

Finally, I attempted to deliver both editing and circuit components simultaneously (fig 5.8). I used an editing sgRNA and Cas9 RNP and ssODN, and electroporated them alongside the circuit plasmid components (fig 5.8.A). Three days later I observed the percentage of EYFP positive cells by flow cytometry (fig 5.8.B and 5.8.C). However, I found simultaneous delivery of components drastically reduced the percentage of EYFP+ cells in both positive control populations, and test populations (fig 5.8.C). The percentage of false positive YFP+ cells in the ssODN-, negative population was also reduced. Additionally, the total percentage of iRFP+ transfected cells was reduced, suggesting that simultaneous delivery of editing and circuit components massively reduced transfection efficiency.

5.3 Discussion

In this chapter I have achieved the following:

1. Synthetic Cas9 circuits can enrich for correctly edited cells by 20-40 fold
2. Optimised circuit components including crRNA length and Cas9-activator
3. Shown that circuit plasmids are eliminated from the system within 8 days

5.3.1 False positives and false negatives

Key features for a successful enrichment system are high degrees of sensitivity and accuracy, or a low proportion of false negatives and false positives. I initially predicted from the low percentage of EYFP+ cells on the guide RNA- flow cytometry plots, that false positive cells would be <10% of the final

population. This base-line level of expression could be due to spontaneous incorporation of the ssODN into the genome, resulting in reporter crRNA expression, or non-specific EYFP activation.

However, I observed the edited bands on the gel were only 20-60% of the total band volume. Although this effect could be partially due to heterozygosity: a completely heterozygotic population would mean the maximum proportion of edited alleles would be 50%. However, a solely heterozygotic population is unlikely, due to the co-incident insertion effect[262]. From a purely heterozygous population with <10% false positive rate, one would expect edited bands to be 45-50% of total band volume. Therefore, this possibility does not completely account for the extra false positive cells predicted from the gel-based assay. Another possibility is that the editing sgRNA may increase the potential of the ssODN to integrate in off-target sites, by generating off target DSBs with some homology to the donor.

It is also important to understand the sensitivity of our circuit enrichment system. The scope of these experiments is only deep enough to speculate on the proportion of false negatives. Enrichment of edited alleles after sorting is roughly 20-40 fold. Therefore, if all EYFP+ cells were true positive and there were no false negatives, we would expect the percentage of EYFP+ cells within the total cell population to be 2.5-5%. However, the percentage of EYFP+ cells in the transfected population is 0.5-1% and in the total population is only 0.25-0.5%. Based on these numbers I predict we did not detect ~80% of true positive cells. In optimisation experiments, ~60% of transfected cells and ~25% of total cells were EYFP+ when the reporter crRNA was expressed from a plasmid. This corresponds roughly with the numbers observed in editing and sorting experiments, meaning that this loss of sensitivity could be explained as a delivery issue. I had hoped that there may be some 'transfection competent' cells which were more likely to receive both editing and circuit components and so reduce the effect of inefficient delivery. However, without more information it is difficult to speculate on whether this phenomenon exists and whether it would have an impact. The major difference between the optimisation experiments and the editing and sorting tests, is that the reporter crRNA is expressed

from the genome instead of from a plasmid. Copy number is lower and thus crRNA expression is reduced. Furthermore, it may be that some of the edited alleles detected in the gel from the unsorted population (and to a lesser extent the sorted population) may be incorrectly copied, artificially reducing the fold change and thus causing overestimation of false negatives.

However, it is difficult to predict the proportions of false positives and negatives and their root causes without more information. There are multiple experiments that will help to address these issues. Deep sequencing of the BRCA1 3'UTR in the unsorted, EYFP+, and EYFP- fractions will help to determine the proportion of false negative and false positive cells. I could identify off-target donor integration using a fragment PCR and NGS method, reminiscent of GUIDE-seq[92]. Additionally, I hope to use index sorting and single cell clonal analysis to map the EYFP intensity of both heterozygous and homozygous cells. I could also use these results to build a probability purity curve to help advise the drawing of flow cytometry gates.

Additionally, a better test of the efficacy of my system would be to test the final editing efficiency post-enrichment against the editing efficiency of a donor with a much smaller insert. This would more accurately represent any advantage given by the system, as the larger insert size required to express the crRNA will likely reduce editing efficiency[263].

5.3.2 Activator and Editor Choice

I observed that choice of activator had a relatively small effect on EYFP output, except in the case of dCas9-VPR which performed much worse. I found this interesting because dCas9-VP64 has fewer activators. Is it possible that in this context p65 and RTA have an inhibitory effect? dCas9-VPR was primarily designed to synergistically activate transcription at reticent loci[252]. It's possible that for simple plasmid-based transcription dCas9-VP64 is sufficient and the extra transcription factors hinder the targeting capacity of dCas9.

My final experiments suggest that, although I did not test it in any of the optimisation experiments, Cpf1 may be the superior choice of editor. Some papers have found that Cpf1 exhibits improved HDR

efficiency with ssODNs compared to SpCas9[84]. However, I did not see a difference between the initial HDR efficiency of Cpf1 and Cas9 in unsorted cells. Instead I observed that the enrichment of edited alleles with Cpf1 tended to be higher with Cpf1 than Cas9. This could be due to the improved orthogonality with Cpf1.

5.3.3 Timing and Delivery of Components

One adjustment which would enhance the scope of this technique is simultaneous delivery of both the editing and circuit components. Sequential delivery of components is not ideal when dealing with cells that are difficult to culture or transfect, i.e. primary cells.

However, early experiments show achieving enrichment with simultaneous delivery requires optimisation of component delivery (fig 5.8). Originally, we hypothesized this could be due to non-productive interactions between the systems: dCas9 obstructing Cas9 cutting and vice versa. However, when we trialled simultaneous delivery with Cpf1 as an editor we found similar issues (data not shown). It may be that delivery load on the cells is too great and causes toxicity or incomplete delivery of all components. This hypothesis, is supported by the reduced percentage of iRFP+ cells in Figure 5.8.C. It is also possible that timings may not be fully optimised, and expression of the crRNA from the integrated site occurs when most of the reporter plasmid is already degraded.

In summary, initial results using the circuit-based system to enrich for precisely edited cells are promising. However, further exploration may help us to better characterise the efficacy and increase the applicability of this system.

Chapter 6: Final conclusions and future directions

6.1 Summary of Findings

The ultimate goal of the work described in this thesis was to develop a method to produce cell populations with a high proportion of correct CRISPR-mediated edits. Such a method would be useful in several contexts: from increasing the therapeutic potential of gene editing, to reducing the labour required to produce cell lines for basic research.

This goal can be achieved by improving efficiency of editing or by enriching for correctly edited cells. These two strategies occur at different points in the editing timeline. Methods which improve the efficiency of editing typically occur before or alongside editing, whereas, methods to enrich for edited cells typically occur after cells are edited.

I began my work by investigating common methods to improve the efficiency of editing and found these strategies had little impact. I then shifted my focus to developing a novel, minimally invasive method to select for edited cells using molecular beacons. This was an attempt to directly recognise the desired changes in the genome, without requiring the insertion of a large reporter cassette. I found that the beacons were highly susceptible to off-target binding. Therefore, I developed a small RNA expression system to be integrated into the genome alongside the desired edit and express the beacon target sequence at a high level. However, I found the beacons still exhibited non-specific binding which obscured the identification of true-positive cells. Finally, I developed a Cas9 circuit-based enrichment method using an adapted form of the small RNA expression system to produce a reporter-activating crRNA. I established this system is capable of producing a 20-40 fold enrichment for correctly edited HEK293T cells with edited alleles at a frequency of 20-60% in the final population.

6.2 Potential of circuit-based enrichment

6.2.1 Previous work

The most effective methods for enriching for HDR-mediated edits detect insertion directly by integrating a selection cassette into the genome[175–177,262]. The reporter can be inserted directly into the target locus[175–177]. Alternatively, the reporter can be integrated into a separate safe harbour locus and enrich correct editing at the target locus using the principle of coincident insertion[262]. Selection cassettes can then be excised using the PiggyBac transposon system, the Cre/LoxP recombinase system, CRISPR-mediated HDR, or Cas9-assissted MMEJ[175,176,264–266]. The PiggyBac system is scarless, provided that the cassette is inserted into a TTAA site; the Cre/Lox recombinase leaves 34 bp LoxP scars; and CRISPR requires a PAM sequence[267,268].

A prime example of this method was the seamless correction of β-thalassemia mutations in patient iPSCs[176]. The authors inserted a puromycin and thymidine kinase selection cassette between PiggyBac inverted terminal repeats (ITRs). Xie et al positively selected correctly edited cells using puromycin and produced clonal lines. 50% of clonal lines contained the correct insertion. Cells were then transiently transfected with a transposase and underwent negative selection with FIAU, which thymidine kinase converts into a toxic substrate. Clonal lines were produced again, 12.5% of which had the correct excision. The method took a total of 3 months to complete.

A second example of selection cassette-based selection used a first round of EGFP-based selection to enrich for cells with a snap tag inserted at the TERT locus[175]. This was followed by the Cre/Lox method of excision to remove the EGFP selection cassette, and finally negative selection on the absence of EGFP. This method left LoxP site 'scars' after excision, but as the edit was located in the 3' UTR and introduced an exogenous tag, this was not an issue. After the initial EGFP+ sort, 66% of clonal lines developed were true positives. After the final EGFP- sort, 100% of clonal lines had the selection cassette successfully removed.

Finally, scars at the target locus can be avoided using the coincident insertion method. Shy et al inserted a neomycin selection cassette flanked by PiggyBac ITRs into a control site and used this reporter to select for EGFP insertion at a separate target site[262]. They found that after neomycin-

selection, edited GFP-expressing cells were 40 ± 5% of the total population, an enrichment of ~30-fold. They also observed an inflation of fold estimations with low initial editing efficiencies, when using lower concentrations of linear dsDNA donors. The editing was performed on mouse ES cells, which are a more relevant disease and genome editing model and more difficult to transfect than HEK293T cells[269]. Following editing, enrichment and clonal cell line analysis, PiggyBac mediated excision was stimulated and enriched for with negative thymidine kinase selection. However, they found that only 1-3% of clonal lines exposed to PiggyBac became resistant to thymidine kinase selection, showing that PiggyBac-mediated excision is not very efficient.

In summary, these methods produce a high frequency of correctly edited cells post-enrichment, but are laborious, time-consuming processes, requiring 1-3 months, and the production of multiple clonal lines. These methods would be inappropriate cell types which are difficult to transfect or culture, i.e. primary cells[270]. The multiple screening and enrichment steps may also be inappropriate for some cell types. The literature often does not directly report on the frequencies of false positives and/or false negative cells. However, the above methods are advantageous as they can be completely scarless, in the case of PiggyBac-mediated excision[264]. In addition, if the target locus does not have an appropriately positioned TTAA site for scarless editing, coincident insertion can be used by co-targeting a separate safe-harbour site[262].

6.2.2 Advantages and limitations of circuit-based enrichment

In comparison, my circuit-based enrichment system is a much faster process: requiring less than one week from editing to enriched population, or three weeks if clonal lines are required. This may be advantageous for cells which are difficult to culture, e.g. $CD34^+$ cells[270]. In addition, circuit-based enrichment requires only one round of enrichment, compared to the two required to remove the reporter cassette using selection-excision enrichment.

Circuit-based editing produces enrichment folds similar to the coincident insertion enrichment values. In addition, loss of false negatives in the secondary selection step, may mean that circuit-based enrichment gives an improved absolute number of true positive cells in the final population.

Excision efficiencies with cassette-selection methods appear to be highly variable: In Xie et al's 2014 study, PiggyBac-mediated excision was only successful in 12.5% of clones; whereas in Yusa et al's 2013 study Cre/Lox mediated excision had a 100% success rate. Therefore, with cassette-based selection and excision, the compromise may be between inefficient excision with PiggyBac and scarring excision with LoxP. However, I anticipate circuit-based enrichment to be translationally scarless without requiring excision. This assumption rests on the efficiency of splicing of the artificial intron. Evidence suggests that artificial introns are spliced efficiently in mammalian cells, however, we will need to test efficiency in our system[271]. Synthetic introns have been shown to affect transcription efficiencies of genes, and this alteration in gene expression may also have an unforeseen effect on cells[272]. Additionally, splice sites impose a restriction on the target sequence: an exonic 3'AG and 5'GT[273]. This limitation is similar to the TTAA of PiggyBac mediated excision. However, SpCas9 editing already requires an NGG PAM sequence, which fulfils half of this limitation.

Circuit-based enrichment introduces permanent transcriptional changes within the cell, and it is unclear what the effects of continued expression of a crRNA may be. In this thesis I have assumed that expression of a small RNA is preferable to expression of proteins, but this is not definitive. Although protein antigens may be the primary substrates of the innate immune system, RNA sequences also commonly activate immune responses[274].

Circuit-based enrichment is principally limited to ssODN-based editing, as the tRNA Pol III promoter is active on dsDNA, regardless of genomic integration[237]. This limitation can be potentially avoided by elongating the time between the first and second transfections, so that when circuit components are delivered, editing components have degraded. However, this solution elongates the time that cells are in culture, counteracting one of the advantages of this enrichment system.

Both cassette-based selection and circuit-based enrichment cannot distinguish between on target and off target insertion. EYFP-expressing positive cells may be activated by crRNA generated from the targeted insertion or from a random off-targeted insertion.

Perhaps, one of the biggest flaws of circuit-based enrichment is that large inserts have been shown to decrease editing efficiency[263]. Insertion efficiency of linear dsDNA dropped from ~50% to ~30% when insert sized was increased from 57 bp to 336 bp. Thus, use of circuit-based enrichment to enrich for cells with a correct SNP edit may be counter-productive, as it requires incorporation of a 100 bp insert. Alternatively, the observed negative correlation between initial editing efficiency and enrichment fold (fig 5.7.G), may compensate for this drop in efficiency. Where initial editing efficiencies varied greatly, post-enrichment editing efficiencies were much more similar (fig 5.7.E).

Furthermore, prime-editing, the innovative new DSB-independent technique from the Liu lab, is thus far only able to incorporate edits of up to 30 nt[105]. However, this technique is still in its infancy and may be able to incorporate larger edits with further development. Therefore, although currently prime-editing is not compatible with circuit-based enrichment, this may change in the future.

6.2.3 Development of circuit-based enrichment

Further to the characterisation experiments discussed in section 5.3, I would like to suggest the following developments to improve the scope of this method.

One important experiment to increase the relevance of this work is to show that this system still works when the crRNA is placed within a synthetic intron. We plan to test this in the adenosine deaminase (ADA) locus, a therapeutically relevant target for severe combined immune deficiency (SCID)[275]. Numerous point mutations in the coding region of ADA are associated with varying severities of SCID[276]. Addition of synthetic intron sequences expand the insert size by ~30 bp, resulting in a final ssODN size of ~220 bp. This increase will likely result in a drop in editing efficiency which may impact the efficacy of the enrichment system. In addition, I would like to confirm the correct splicing of the synthetic intron from the ADA gene with gel and RT-qPCR-based assays. Confirming the utility of this

method for translationally-scarless editing within the coding region will massively enhance the relevance of circuit-based enrichment.

I wish to test the efficacy of circuit-based enrichment in primary cells. Collaborators of ours in Rod Dunbar's lab, University of Auckland, have already begun to test this system in primary keratinocytes. They hope to use this enrichment method to improve the study and treatment of the skin disease, epidermolysis bullosa. However, as of yet, efficacy seems to be limited by inefficient transfection.

Furthermore, another important question which much be answered is: can circuit-based enrichment be achieved using a single transfection step? This adjustment will cement circuit-based enrichment as an advantageous method. However, it is still unclear why simultaneous delivery of components has failed. I plan to vary the concentrations of components and trial alternative methods of delivery to determine if this can improve results. I will also attempt to use an inducible promoter for the dCas9-activator, to see if timed expression has an impact. Finally, a minimal single plasmid expressing all components of the circuit system may have a positive effect.

One of the limitations I foresee for this ssODN-dependent system, is the prohibitive size of the required insert. I propose that further optimisation of the tRNA Pol III promoter may help to reduce the size of the insert. The Pol III promoter function of tRNAs is under-researched[237]. It may be that the length of the promoter can be optimised to improve the efficiency of editing.

To conclude, the scope of circuit-based enrichment is exciting, but more work must be done to fully realise this potential. Current enrichment methods are time consuming and not well suited to selecting for primary cells. Circuit-based enrichment has the potential to succeed in these areas where cassette-based enrichment falls short. Alternatively, circuit-based enrichment is an innovative approach to a difficult problem, and an interesting example of yet another area where the unique CRISPR mechanism is valuable.

Bibliography

1. Thomas, K. R., Folger, K. R. & Capecchi, M. R. High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419–428 (1986).

2. Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature* **317**, 230–234 (1985).

3. Scherer, S. & Davis, R. W. Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4951–4955 (1979).

4. Rothstein, R. J. One-Step Gene Disruption in Yeast. *Methods Enzymol.* **101**, 202–211 (1983).

5. Capecchi, M. R. Altering the genome by homologous recombination. *Science (80-.).* **244**, 1288–1292 (1989).

6. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).

7. Plessis, A., Perrin, A., Haber, J. E. & Dujon, B. Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* **130**, 451–460 (1992).

8. Choulika, A., Perrin, A., Dujon, B. & Nicolas, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **15**, 1968–1973 (1995).

9. Santiago, Y. *et al.* Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5809–5814 (2008).

10. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes:

When nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**, 198–199 (1998).

11. Bétermier, M., Bertrand, P. & Lopez, B. S. Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLoS Genetics* vol. 10 (2014).

12. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5172–5177 (1998).

13. Thompson, L. H. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: The molecular choreography. *Mutation Research - Reviews in Mutation Research* (2012) doi:10.1016/j.mrrev.2012.06.002.

14. Tsai, A. G. & Lieber, M. R. Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics* **11**, (2010).

15. Chiruvella, K. K., Liang, Z. & Wilson, T. E. Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).

16. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination. *DNA Repair* vol. 3 817–826 (2004).

17. Williams, R. S. *et al.* Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair. *Cell* **135**, 97–109 (2008).

18. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in Cell Biology* vol. 26 52–64 (2016).

19. Parvathaneni, S., Stortchevoi, A., Sommers, J. A., Brosh, R. M. & Sharma, S. Human RECQ1 Interacts with Ku70/80 and Modulates DNA End-Joining of Double-Strand Breaks. *PLoS One* **8**, (2013).

20. Fell, V. L. & Schild-Poulter, C. Ku Regulates Signaling to DNA Damage Response Pathways through the Ku70 von Willebrand A Domain. *Mol. Cell. Biol.* **32**, 76–87 (2012).

21. Thompson, L. H. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: The molecular choreography. *Mutat. Res. Mutat. Res.* **751**, 158–246 (2012).

22. Moynahan, M. E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature Reviews Molecular Cell Biology* vol. 11 196–207 (2010).

23. Yun, M. H. & Hiom, K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nature* **459**, 460–463 (2009).

24. Polato, F. *et al.* CtIP-mediated resection is essential for viability and can operate independently of BRCA1. *J. Exp. Med.* **211**, 1027–1036 (2014).

25. Krasner, D. S., Daley, J. M., Sung, P. & Niu, H. Interplay between Ku and Replication Protein A in the Restriction of Exo1-mediated DNA Break End Resection. *J. Biol. Chem.* **290**, 18806–18816 (2015).

26. Wolner, B., van Komen, S., Sung, P. & Peterson, C. L. Recruitment of the recombinational repair machinery to a DNA double-strand break in yeast. *Mol. Cell* **12**, 221–32 (2003).

27. Ma, C. J., Kwon, Y., Sung, P. & Greene, E. C. Human RAD52 interactions with replication protein a and the RAD51 presynaptic complex. *J. Biol. Chem.* **292**, 11702–11713 (2017).

28. Sak, A., Stueben, G., Groneberg, M., Böcker, W. & Stuschke, M. Targeting of Rad51-dependent homologous recombination: Implications for the radiation sensitivity of human lung cancer cell lines. *Br. J. Cancer* **92**, 1089–1097 (2005).

29. Nimonkar, A. V., Sica, R. A. & Kowalczykowski, S. C. Rad52 promotes second-end DNA capture in double-stranded break repair to form complement-stabilized joint molecules. *Proc. Natl.*

Acad. Sci. U. S. A. **106**, 3077–3082 (2009).

30. Ira, G., Satory, D. & Haber, J. E. Conservative Inheritance of Newly Synthesized DNA in Double-Strand Break-Induced Gene Conversion. *Mol. Cell. Biol.* **26**, 9424–9429 (2006).

31. Costantino, L. *et al.* Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science (80-.).* **343**, 88–91 (2014).

32. Wang, L. C. & Gautier, J. The Fanconi anemia pathway and ICL repair: Implications for cancer therapy. *Critical Reviews in Biochemistry and Molecular Biology* vol. 45 424–439 (2010).

33. Niraj, J., Färkkilä, A. & D'Andrea, A. D. The Fanconi Anemia Pathway in Cancer. *Annu. Rev. Cancer Biol.* **3**, 457–478 (2019).

34. Huertas, P. & Jackason, S. P. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J. Biol. Chem.* **284**, 9558–9565 (2009).

35. Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences* vol. 40 701–714 (2015).

36. Truong, L. N. *et al.* Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Acad. Sci.* **110**, 7720–7725 (2013).

37. Storici, F., Snipe, J. R., Chan, G. K., Gordenin, D. A. & Resnick, M. A. Conservative Repair of a Chromosomal Double-Strand Break by Single-Strand DNA through Two Steps of Annealing. *Mol. Cell. Biol.* **26**, 7645–7657 (2006).

38. Vaze, M. B. *et al.* Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. *Mol. Cell* **10**, 373–385 (2002).

39. Bradley, M. O. & Kohn, K. W. X-ray induced DNA double strand break production and repair in

mammalian cells as measured by neutral filter elution. *Nucleic Acids Res.* **7**, 793–804 (1979).

40. Caldecott, K. W. Single-strand break repair and genetic disease. *Nature Reviews Genetics* vol. 9 619–631 (2008).

41. Maizels, N. & Davis, L. Initiation of homologous recombination at DNA nicks. *Nucleic Acids Res.* **46**, 6962–6973 (2018).

42. Kong, Q. & Maizels, N. DNA breaks in hypermutating immunoglobulin genes: evidence for a break-and-repair pathway of somatic hypermutation. *Genetics* **158**, 369–78 (2001).

43. Holliday, R. A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282–304 (1964).

44. Meselson, M. S. & Radding, C. M. A general model for genetic recombination. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 358–61 (1975).

45. Radding, C. M. Homologous Pairing and Strand Exchange in Genetic Recombination. *Annu. Rev. Genet.* **16**, 405–437 (1982).

46. Muñoz, I. G. *et al.* Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res.* **39**, 729–743 (2011).

47. Krishna, S. S., Majumdar, I. & Grishin, N. V. Structural classification of zinc fingers. *Nucleic Acids Research* vol. 31 532–550 (2003).

48. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science (80-.).* **252**, 809–817 (1991).

49. Bogdanove, A. J., Bohm, A., Miller, J. C., Morgan, R. D. & Stoddard, B. L. Engineering altered protein-DNA recognition specificity. *Nucleic Acids Research* vol. 46 4845–4871 (2018).

50. Deng, D., Yan, C., Wu, J., Pan, X. & Yan, N. Revisiting the TALE repeat. *Protein Cell* **5**, 297–306 (2014).

51. Boch, J. & Bonas, U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function . *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).

52. Holkers, M. *et al.* Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.* **41**, (2013).

53. Esensten, J. H., Bluestone, J. A. & Lim, W. A. Engineering Therapeutic T Cells: From Synthetic Biology to Clinical Trials. *Annu. Rev. Pathol. Mech. Dis.* **12**, 305–330 (2017).

54. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakatura, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).

55. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–821 (2012).

56. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science (80-.).* **339**, 823–826 (2013).

57. Mojica, F. J. M., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* vol. 36 244–246 (2000).

58. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).

59. Brouns, S. J. J. *et al.* Small Crispr Rnas Guide Antiviral Defense in Prokaryotes. *Cancer Epidemiol. Biomarkers Prev.* **2**, 531–535 (1993).

60. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *Cris. J.* **1**, 325–336 (2018).

61. Takeuchi, N., Wolf, Y. I., Makarova, K. S. & Koonin, E. V. Nature and intensity of selection pressure on crispr-associated genes. *J. Bacteriol.* **194**, 1216–1225 (2012).

62. Makarova, K. S., Zhang, F. & Koonin, E. V. SnapShot: Class 2 CRISPR-Cas Systems. *Cell* vol. 168 328-328.e1 (2017).

63. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* (2019) doi:10.1038/s41579-019-0299-x.

64. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* vol. 1 (2006).

65. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* **60**, 385–397 (2015).

66. Gao, P., Yang, H., Rajashankar, K. R., Huang, Z. & Patel, D. J. Type v CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* **26**, 901–913 (2016).

67. Zhang, B. *et al.* Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. *Nat. Commun.* **10**, (2019).

68. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

69. Lee, H., Dhingra, Y. & Sashital, D. G. The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *Elife* **8**, (2019).

70. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740

(2009).

71. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).

72. Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).

73. Hirano, H. *et al.* Structure and Engineering of Francisella novicida Cas9. *Cell* **164**, 950–961 (2016).

74. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–23 (2013).

75. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–9 (2013).

76. Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Research* vol. 23 720–723 (2013).

77. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).

78. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol. Cell* **54**, 698–710 (2014).

79. Kiani, S. *et al.* CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11**, 723–726 (2014).

80. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (80-.).* **343**, 84–87 (2014).

81. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science (80-.).* **343**, 80–84 (2014).

82. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).

83. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

84. Wang, Y. *et al.* Systematic evaluation of CRISPR-Cas systems reveals design principles for genome editing in human cells. *Genome Biol.* **19**, 62 (2018).

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

86. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).

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

88. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).

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

90. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).

91. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human

cells. *Nat. Methods* **12**, 237–243 (2015).

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

93. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).

94. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, (2014).

95. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318 (2015).

96. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature Biotechnology* vol. 33 139–142 (2015).

97. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–8 (2016).

98. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

99. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–308 (2013).

100. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–582 (2014).

101. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–42 (2015).

102. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CrIsPr-Cas9-induced

precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, (2015).

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

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

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

106. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

107. Song, J. *et al.* RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* **7**, 10548 (2016).

108. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends in Genetics* vol. 12 224–228 (1996).

109. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8**, 753–757 (2011).

110. Renaud, J.-B. *et al.* Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.* **14**, 2263–2272 (2016).

111. Yang, L. *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* **41**, 9049–61 (2013).

112. Boel, A. *et al.* CRISPR/Cas9-mediated homology-directed repair by ssODNs in zebrafish induces complex mutational patterns resulting from genomic integration of repair-template fragments. *DMM Dis. Model. Mech.* **11**, (2018).

113. Homology Requirement for Efficient Gene Conversion between Duplicated Chromosomal Sequences in Mammalian Cells. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1203052/.

114. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

115. Kan, Y., Ruis, B., Takasugi, T. & Hendrickson, E. A. Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res.* **27**, 1099–1111 (2017).

116. Richardson, C. D. *et al.* CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* **50**, 1132–1139 (2018).

117. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **339**, 819–823 (2013).

118. Davis, L. & Maizels, N. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Rep.* **17**, 1872–1881 (2016).

119. Metzger, M. J., Stoddard, B. L. & Monnat, R. J. PARP-mediated repair, homologous recombination, and back-up non-homologous end joining-like repair of single-strand nicks. *DNA Repair (Amst).* **12**, 529–534 (2013).

120. Vriend, L. E. M. & Krawczyk, P. M. Nick-initiated homologous recombination: Protecting the genome, one strand at a time. *DNA Repair* vol. 50 1–13 (2017).

121. Storici, F., Bebenek, K., Kunkel, T. A., Gordenin, D. A. & Resnick, M. A. RNA-templated DNA repair. *Nature* **447**, 338–341 (2007).

122. Shen, Y. *et al.* RNA-driven genetic changes in bacteria and in human cells. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **717**, 91–98 (2011).

123. Keskin, H. *et al.* Transcript-RNA-templated DNA recombination and repair. *Nature* **515**, 436–439 (2014).

124. Brandsma, I. & Gent, D. C. Pathway choice in DNA double strand break repair: Observations of a balancing act. *Genome Integrity* vol. 3 (2012).

125. Huertas, P., Cortés-Ledesma, F., Sartori, A. A., Aguilera, A. & Jackson, S. P. CDK targets Sae2 to control DNA-end resection and homologous recombination. *Nature* **455**, 689–92 (2008).

126. Renaud, J.-B. *et al.* Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.* **14**, 2263–2272 (2016).

127. Yang, D. *et al.* Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci. Rep.* **6**, 21264 (2016).

128. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F. & Chin, L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep.* **14**, 1555–1566 (2016).

129. Beswick, R. W., Ambrose, H. E. & Wagner, S. D. Nocodazole, a microtubule depolymerising agent, induces apoptosis of chronic lymphocytic leukaemia cells associated with changes in Bcl-2 phosphorylation and expression. *Leuk. Res.* **30**, 427–436 (2006).

130. Endo, K., Mizuguchi, M., Harata, A., Itoh, G. & Tanaka, K. Nocodazole induces mitotic cell death with apoptotic-like features in *Saccharomyces cerevisiae*. *FEBS Lett.* **584**, 2387–2392 (2010).

131. Orth, J. D., Loewer, A., Lahav, G. & Mitchison, T. J. Prolonged mitotic arrest triggers partial activation of apoptosis, resulting in DNA damage and p53 induction. *Mol. Biol. Cell* **23**, 567–76 (2012).

132. Rios, X. *et al.* Stable gene targeting in human cells using single-strand oligonucleotides with modified bases. *PLoS One* **7**, e36697 (2012).

133. Ruff, P., Koh, K. D., Keskin, H., Pai, R. B. & Storici, F. Aptamer-guided gene targeting in yeast and human cells. *Nucleic Acids Res.* **42**, (2014).

134. Carlson-Stevermer, J. *et al.* Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* **8**, 1711 (2017).

135. Ma, M. *et al.* Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Research* vol. 27 578–581 (2017).

136. Savic, N. *et al.* Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife* **7**, (2018).

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

138. Manolis, K. G. *et al.* Novel functional requirements for non-homologous DNA end joining in Schizosaccharomyces pombe. *EMBO J.* **20**, 210–221 (2001).

139. Robert, F., Barbeau, M., Éthier, S., Dostie, J. & Pelletier, J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* **7**, 93 (2015).

140. Li, G. *et al.* Suppressing Ku70/Ku80 expression elevates homology-directed repair efficiency in primary fibroblasts. *Int. J. Biochem. Cell Biol.* **99**, 154–160 (2018).

141. Jayathilaka, K. *et al.* A chemical compound that stimulates the human homologous recombination protein RAD51. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15848–15853 (2008).

142. Yu, C. *et al.* Small molecules enhance crispr genome editing in pluripotent stem cells. *Cell Stem Cell* **16**, 142–147 (2015).

143. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases

for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–8 (2013).

144. Chiang, T.-W. W. *et al.* CRISPR-Cas9D10A nickase-based genotypic and phenotypic screening to enhance genome editing. *Sci. Rep.* **6**, 24356 (2016).

145. Gopalappa, R., Suresh, B., Ramakrishna, S. & Kim, H. H. Paired D10A Cas9 nickases are sometimes more efficient than individual nucleases for gene disruption. *Nucleic Acids Res.* **46**, (2018).

146. Wang, Y. *et al.* Paired CRISPR/Cas9 Nickases Mediate Efficient Site-Specific Integration of F9 into rDNA Locus of Mouse ESCs. *Int. J. Mol. Sci.* **19**, 3035 (2018).

147. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–9 (2013).

148. Tsouroula, K. *et al.* Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol. Cell* **63**, 293–305 (2016).

149. Lemaître, C. *et al.* Nuclear position dictates DNA repair pathway choice. *Genes Dev.* **28**, 2450–2463 (2014).

150. Yarrington, R. M., Verma, S., Schwartz, S., Trautman, J. K. & Carroll, D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 9351–9358 (2018).

151. Chen, R. *et al.* Enrichment of transiently transfected mesangial cells by cell sorting after cotransfection with GFP. *Am. J. Physiol. - Ren. Physiol.* **276**, (1999).

152. Szymczak-Workman, A. L., Vignali, K. M. & Vignali, D. A. A. Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harb. Protoc.* **7**, 199–204 (2012).

153. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

154. Burger, A. *et al.* Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Dev.* **143**, 2025–2037 (2016).

155. Nasri, M. *et al.* Fluorescent labeling of CRISPR/Cas9 RNP for gene knockout in HSPCs and iPSCs reveals an essential role for GADD45b in stress response. *Blood Adv.* **3**, 63–71 (2019).

156. Lee, K. *et al.* Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. (2017) doi:10.7554/eLife.25312.001.

157. Certo, M. T. *et al.* Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods* **8**, 671–6 (2011).

158. Kim, H. *et al.* Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods* **8**, 941–943 (2011).

159. Ramakrishna, S. *et al.* Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. *Nat. Commun.* **5**, 3378 (2014).

160. Segal, D. J. & Carroll, D. Endonuclease-induced, targeted homologous extrachromosomal recombination in Xenopus oocytes. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 806–810 (1995).

161. Zhou, Y. *et al.* Enhanced genome editing in mammalian cells with a modified dual-fluorescent surrogate system. *Cell. Mol. Life Sci.* **73**, 2543–2563 (2016).

162. Kuhar, R. *et al.* Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks. *Nucleic Acids Res.* **42**, e4 (2016).

163. He, Z. *et al.* Comparison of surrogate reporter systems for enrichment of cells with mutations induced by genome editors. *J. Biotechnol.* **221**, 49–54 (2016).

164. Dean, M. *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter

AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**, 1856–62 (1996).

165. Wang, W. *et al.* CCR5 gene disruption via lentiviral vectors expressing Cas9 and single guided RNA renders cells resistant to HIV-1 infection. *PLoS One* **9**, (2014).

166. Liu, Z. *et al.* Genome editing of the HIV co-receptors CCR5 and CXCR4 by CRISPR-Cas9 protects CD4+ T cells from HIV-1 infection. *Cell Biosci.* **7**, 47 (2017).

167. Kim, H. *et al.* A Co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. *Genetics* **197**, 1069–1080 (2014).

168. Moriarity, B. S. *et al.* Simple and Efficient Methods for Enrichment and Isolation of Endonuclease Modified Cells. *PLoS One* **9**, e96114 (2014).

169. Agudelo, D. *et al.* Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat. Methods* **14**, 615–620 (2017).

170. Shy, B. R., MacDougall, M. S., Clarke, R. & Merrill, B. J. Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells. *Nucleic Acids Res.* **44**, 7997–8010 (2016).

171. Urlaub, G. & Chasin, L. A. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 4216–4220 (1980).

172. Laursen, M., Gregersen, J. L., Yatime, L., Nissen, P. & Fedosova, N. U. Structures and characterization of digoxin- And bufalin-bound Na+,K+-ATPase compared with the ouabain-bound complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1755–1760 (2015).

173. Arias-Fuenzalida, J. *et al.* FACS-Assisted CRISPR-Cas9 Genome Editing Facilitates Parkinson's Disease Modeling. *Stem Cell Reports* **9**, 1423–1431 (2017).

174. Zhang, L. *et al.* Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One* **10**, (2015).

175. Yusa, K. Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyBac transposon. *Nat. Protoc.* **8**, 2061–2078 (2013).

176. Xie, F. *et al.* Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res.* **24**, 1526–33 (2014).

177. Xi, L., Schmidt, J. C., Zaug, A. J., Ascarrunz, D. R. & Cech, T. R. A novel two-step genome editing strategy with CRISPR-Cas9 provides new insights into telomerase action and TERT gene expression. *Genome Biol.* **16**, (2015).

178. Ma, H. *et al.* Correction of a pathogenic gene mutation in human embryos. *Nature* **548**, 413–419 (2017).

179. Tebas, P. *et al.* Gene Editing of *CCR5* in Autologous CD4 T Cells of Persons Infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014).

180. A Safety and Efficacy Study Evaluating CTX001 in Subjects With Transfusion-Dependent β-Thalassemia - Full Text View - ClinicalTrials.gov. https://www.clinicaltrials.gov/ct2/show/NCT03655678?term=CRISPR.

181. Ledford, H. First test of in-body gene editing shows promise. *Nature* (2018) doi:10.1038/d41586-018-06195-6.

182. Yu, W. & Wu, Z. In Vivo applications of CRISPR-based genome editing in the retina. *Frontiers in Cell and Developmental Biology* vol. 6 (2018).

183. Single Ascending Dose Study in Participants With LCA10 - Full Text View - ClinicalTrials.gov. https://www.clinicaltrials.gov/ct2/show/NCT03872479?term=NCT%252303872479&rank=1.

184. Study of CRISPR-Cas9 Mediated PD-1 and TCR Gene-knocked Out Mesothelin-directed CAR-T Cells in Patients With Mesothelin Positive Multiple Solid Tumors. - Full Text View - ClinicalTrials.gov. https://www.clinicaltrials.gov/ct2/show/NCT03545815?term=CRISPR&cond=cancer&cntry=CN&rank=1.

185. NY-ESO-1-redirected CRISPR (TCRendo and PD1) Edited T Cells (NYCE T Cells) - Full Text View - ClinicalTrials.gov. https://www.clinicaltrials.gov/ct2/show/NCT03399448?term=CRISPR&cntry=US&state=US%253APA&rank=1#contacts.

186. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **339**, (2013).

187. Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197 (2015).

188. Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc. Natl. Acad. Sci.* **112**, 201521077 (2015).

189. Park, S. H. *et al.* Highly efficient editing of the β-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gkz475.

190. Matthis, J. & Reijonen, H. Production of primary human $CD4^+$ T cell lines and clones. *Methods Mol. Biol.* **960**, 545–555 (2013).

191. 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).

192. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in

human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–9 (2014).

193. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168–e168 (2014).

194. Tang, H., Yang, X., Wang, K., Tan, W. & Li, W. mRNA detection in living cell using phosphorothioate-modified molecular beacon. *Sci. Bull.* **54**, 1507–1514 (2009).

195. Hu, Z. *et al.* Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR–Cas9 and ssODN in human cancer cells. *Cell Biosci.* **8**, 12 (2018).

196. 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).

197. Mettananda, S. *et al.* Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-Thalassemia. *Nat. Commun.* **8**, (2017).

198. Kaneko, K. *et al.* Establishment of a cell model of X-linked sideroblastic anemia using genome editing. *Exp. Hematol.* **65**, 57-68.e2 (2018).

199. Wong, S. *et al.* Establishment of an erythroid cell line from primary CD36+ erythroid progenitor cells. *Exp. Hematol.* **38**, 994-1005.e1–2 (2010).

200. Traxler, E. A. *et al.* A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat. Med.* **22**, 987–990 (2016).

201. Chung, J. E. *et al.* CRISPR-Cas9 interrogation of a putative fetal globin repressor in human erythroid cells. *PLoS One* **14**, e0208237 (2019).

202. Philipsen, S. & Hardison, R. C. Evolution of hemoglobin loci and their regulatory elements. *Blood Cells. Mol. Dis.* **70**, 2 (2018).

203. Vernimmen, D. Uncovering Enhancer Functions Using the α-Globin Locus. *PLoS Genet.* **10**, e1004668 (2014).

204. Greco, G. E. *et al.* SCR7 is neither a selective nor a potent inhibitor of human DNA ligase IV. *DNA Repair (Amst).* **43**, 18–23 (2016).

205. Mason, J. M. *et al.* The RAD51-Stimulatory Compound RS-1 Can Exploit the RAD51 Overexpression That Exists in Cancer Cells and Tumors. *Cancer Res.* **74**, 3546–3555 (2014).

206. Lakin, N. D. & Jackson, S. P. Regulation of p53 in response to DNA damage. *Oncogene* **18**, 7644–7655 (1999).

207. Huang, P. H., Cook, R. & Mittnacht, S. RB in DNA repair. *Oncotarget* **6**, 20746–7 (2015).

208. Thangavel, C. *et al.* The retinoblastoma tumor suppressor modulates DNA repair and radioresponsiveness. *Clin. Cancer Res.* **20**, 5468–5482 (2014).

209. Moir-Meyer, G. *et al.* Robust CRISPR/Cas9 Genome Editing of the HUDEP-2 Erythroid Precursor Line Using Plasmids and Single-Stranded Oligonucleotide Donors. *Methods Protoc.* **1**, 28 (2018).

210. Zhang, J.-P. *et al.* Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol.* **18**, 35 (2017).

211. Riesenberg, S. & Maricic, T. Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. *Nat. Commun.* **9**, 2164 (2018).

212. Greco, G. E. *et al.* SCR7 is neither a selective nor a potent inhibitor of human DNA ligase IV. *DNA Repair (Amst).* **43**, 18–23 (2016).

213. Mason, J. M. *et al.* The RAD51-stimulatory compound RS-1 can exploit the RAD51 overexpression that exists in cancer cells and tumors. *Cancer Res.* **74**, 3546–55 (2014).

214. Tyagi, S. & Kramer, F. R. Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nat.*

Biotechnol. **14**, 303–308 (1996).

215. Bonnet, G., Tyagi, S., Libchaber, A. & Kramer, F. R. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6171–6 (1999).

216. Szuhai, K. *et al.* Simultaneous A8344G heteroplasmy and mitochondrial DNA copy number quantification in Myoclonus Epilepsy and Ragged-Red Fibers (MERRF) syndrome by a multiplex Molecular Beacon based real-time fluorescence PCR. *Nucleic Acids Res.* **29**, 13e – 13 (2001).

217. Dahan, L., Huang, L., Kedmi, R., Behlke, M. A. & Peer, D. SNP Detection in mRNA in Living Cells Using Allele Specific FRET Probes. *PLoS One* **8**, e72389 (2013).

218. El-Hajj, H. H., Marras, S. A., Tyagi, S., Kramer, F. R. & Alland, D. Detection of rifampin resistance in Mycobacterium tuberculosis in a single tube with molecular beacons. *J. Clin. Microbiol.* **39**, 4131–7 (2001).

219. Lin, S.-Y. G., Probert, W., Lo, M. & Desmond, E. Rapid detection of isoniazid and rifampin resistance mutations in Mycobacterium tuberculosis complex from cultures or smear-positive sputa by use of molecular beacons. *J. Clin. Microbiol.* **42**, 4204–8 (2004).

220. Baker, M. B., Bao, G. & Searles, C. D. In vitro quantification of specific microRNA using molecular beacons. *Nucleic Acids Res.* **40**, e13–e13 (2012).

221. Chen, M. *et al.* A molecular beacon-based approach for live-cell imaging of RNA transcripts with minimal target engineering at the single-molecule level. *Sci. Rep.* **7**, 1550 (2017).

222. Tyagi, S. & Alsmadi, O. Imaging native beta-actin mRNA in motile fibroblasts. *Biophys. J.* **87**, 4153–62 (2004).

223. Rhee, W. J. & Bao, G. Simultaneous detection of mRNA and protein stem cell markers in live cells. *BMC Biotechnol.* **9**, (2009).

224. Wile, B. M., Ban, K., Yoon, Y.-S. & Bao, G. Molecular beacon-enabled purification of living cells by targeting cell type-specific mRNAs. *Nat. Protoc.* **9**, 2411–24 (2014).

225. Seidel, C. A. M., Schulz, A. & Sauer, M. H. M. Nucleobase-Specific Quenching of Fluorescent Dyes. 1. Nucleobase One-Electron Redox Potentials and Their Correlation with Static and Dynamic Quenching Efficiencies. *J. Phys. Chem.* **100**, 5541–5553 (1996).

226. Dabcyl dT Oligo Modifications from Gene Link. http://www.genelink.com/newsite/products/mod_detail.asp?modid=82.

227. Soboleski, M. R., Oaks, J. & Halford, W. P. Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. *FASEB J.* **19**, 440–2 (2005).

228. Tang, H., Yang, X., Wang, K., Tan, W. & Li, W. mRNA detection in living cell using phosphorothioate-modified molecular beacon. *Sci. Bull.* **54**, 1507–1514 (2009).

229. Tang, H., Yang, X., Wang, K., Tan, W. & Li, W. mRNA detection in living cell using phosphorothioate-modified molecular beacon. *Sci. Bull.* **54**, 1507–1514 (2009).

230. Young, S. & Wagner, R. W. Hybridization and dissociation rates of phosphodiester or modified oligodeoxynucleotides with RNA at near-physiological conditions. *Nucleic Acids Res.* **19**, 2463–70 (1991).

231. Gao, W. Y., Han, F. S., Storm, C., Egan, W. & Cheng, Y. C. Phosphorothioate oligonucleotides are inhibitors of human DNA polymerases and RNase H: implications for antisense technology. *Mol. Pharmacol.* **41**, 223–9 (1992).

232. Yildirim, I., Kierzek, E., Kierzek, R. & Schatz, G. C. Interplay of LNA and 2′- *O* -Methyl RNA in the Structure and Thermodynamics of RNA Hybrid Systems: A Molecular Dynamics Study Using the Revised AMBER Force Field and Comparison with Experimental Results. *J. Phys. Chem. B* **118**, 14177–14187 (2014).

233. Zhang, D. Y., Chen, S. X. & Yin, P. Optimizing the specificity of nucleic acid hybridization. *Nat. Chem.* **4**, 208–214 (2012).

234. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–77 (1996).

235. O'Brien, S. J. & Moore, J. P. The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol. Rev.* **177**, 99–111 (2000).

236. Schumann, K. *et al.* Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc. Natl. Acad. Sci.* **112**, 10437–10442 (2015).

237. Knapp, D. J. H. F. *et al.* Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. *Nat. Commun.* **10**, 1490 (2019).

238. Chen, A. K., Behlke, M. A. & Tsourkas, A. Efficient cytosolic delivery of molecular beacon conjugates and flow cytometric analysis of target RNA. *Nucleic Acids Res.* **36**, e69–e69 (2008).

239. Chen, A. K., Davydenko, O., Behlke, M. A. & Tsourkas, A. Ratiometric bimolecular beacons for the sensitive detection of RNA in single living cells. *Nucleic Acids Res.* **38**, e148–e148 (2010).

240. Mhlanga, M. M., Vargas, D. Y., Fung, C. W., Kramer, F. R. & Tyagi, S. tRNA-linked molecular beacons for imaging mRNAs in the cytoplasm of living cells. *Nucleic Acids Res.* **33**, 1902–1912 (2005).

241. Chen, A. K., Behlke, M. A. & Tsourkas, A. Avoiding false-positive signals with nuclease-vulnerable molecular beacons in single living cells. *Nucleic Acids Res.* **35**, e105 (2007).

242. Yanrong Wu, Chaoyong James Yang, Leonid L. Moroz, and & Tan*, W. Nucleic Acid Beacons for Long-Term Real-Time Intracellular Monitoring. (2008) doi:10.1021/AC702637W.

243. Solloch, U. V. *et al.* Frequencies of gene variant CCR5-Δ32 in 87 countries based on next-

generation sequencing of 1.3 million individuals sampled from 3 national DKMS donor centers. *Hum. Immunol.* **78**, 710–717 (2017).

244. Boldt, A. B. W. *et al.* Analysis of the CCR5 gene coding region diversity in five South American populations reveals two new non-synonymous alleles in Amerindians and high CCR5*D32 frequency in Euro-Brazilians. *Genet. Mol. Biol.* **32**, 12–19 (2009).

245. Le Rhun, A. & Charpentier, E. Small RNAs in streptococci. *RNA Biology* vol. 9 414–426 (2012).

246. Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39**, 9275–9282 (2011).

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

248. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).

249. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).

250. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).

251. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–46 (2014).

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

253. Ferry, Q. R. V., Lyutova, R., Fulga, T. A., Wieland, M. & Fussenegger, M. Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. *Nat. Commun.*

8, 14633 (2017).

254. Ting, P. Y. *et al.* Guide Swap enables genome-scale pooled CRISPR–Cas9 screening in human primary cells. *Nat. Methods* **15**, 941–946 (2018).

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

256. Li, K., Wang, G., Andersen, T., Zhou, P. & Pu, W. T. Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS One* **9**, (2014).

257. Rattan, R., Bielinska, A. U. & Banaszak Holl, M. M. Quantification of cytosolic plasmid DNA degradation using high-throughput sequencing: Implications for gene delivery. *J. Gene Med.* **16**, 75–83 (2014).

258. Neves, C., Escriou, V., Byk, G., Scherman, D. & Wils, P. Intracellular fate and nuclear targeting of plasmid DNA. *Cell Biol. Toxicol.* **15**, 193–202 (1999).

259. Effect of Different Fetal Bovine Serum Concentrations on the Replicative Life Span of Cultured Chick Cells on JSTOR. https://www.jstor.org/stable/4292298?seq=1#metadata_info_tab_contents.

260. Mengual Gómez, D. L., Belaich, M. N., Rodríguez, V. A. & Ghiringhelli, P. D. Effects of Fetal Bovine Serum deprivation in cell cultures on the production of Anticarsia gemmatalis Multinucleopolyhedrovirus. *BMC Biotechnol.* **10**, 68 (2010).

261. Michaels, Y. S. *et al.* Precise tuning of gene expression levels in mammalian cells. *Nat. Commun.* **10**, (2019).

262. Shy, B. R., Macdougall, M. S., Clarke, R. & Merrill, B. J. Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells. *Nucleic Acids Res.* **44**, 7997–8010 (2016).

263. Paix, A. *et al.* Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10745–E10754 (2017).

264. 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).

265. Kühn, R. & Chu, V. T. Pop in, pop out: a novel gene-targeting strategy for use with CRISPR-Cas9. *Genome Biol.* **16**, 244 (2015).

266. Kim, S. Il *et al.* Microhomology-assisted scarless genome editing in human iPSCs. *Nat. Commun.* **9**, 1–14 (2018).

267. Scahill, M. D., Pastar, I. & Cross, G. A. M. CRE recombinase-based positive-negative selection systems for genetic manipulation in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **157**, 73–82 (2008).

268. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–7 (2014).

269. Tamm, C., Kadekar, S., Pijuan-Galitó, S. & Annerén, C. Fast and Efficient Transfection of Mouse Embryonic Stem Cells Using Non-Viral Reagents. *Stem Cell Rev. Reports* **12**, 584–591 (2016).

270. *Trouble in the hood: culturing difficult cell types*. http://www.nature.com/naturemethods (2005).

271. Tikhonov, M. V, Maksimenko, O. G., Georgiev, P. G. & Korobko, I. V. Optimal Artificial Mini-Introns for Transgenic Expression in the Cells of Mice and Hamsters. *Orig. Russ. Text ©* **51**, 671–676 (2017).

272. Crane, M. M. *et al.* In vivo measurements reveal a single 5'-intron is sufficient to increase protein expression level in Caenorhabditis elegans. *Sci. Rep.* **9**, (2019).

273. Anna, A. & Monika, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics* vol. 59 253–268 (2018).

274. Chen, N. *et al.* RNA sensors of the innate immune system and their detection of pathogens. *IUBMB Life* vol. 69 297–304 (2017).

275. Arredondo-Vega, F. X., Santisteban, I., Daniels, S., Toutain, S. & Hershfield, M. S. Adenosine Deaminase Deficiency: Genotype-Phenotype Correlations Based on Expressed Activity of 29 Mutant Alleles. *Am. J. Hum. Genet.* **63**, 1049–1059 (1998).

276. Markert, M. L., Norby-Slycord, C. & Ward, F. E. A high proportion of ADA point mutations associated with a specific alanine-to-valine substitution. *Am. J. Hum. Genet.* **45**, 354–61 (1989).