

https://dspace.org/

Modification of the human KRAS gene using CRISPR /Cas9 system

Fok, Ezio Tony

2015-04-17
http://hdl.handle.net/10539/17425

Downloaded from a DSpace repository. https://dspace.org/

Modification of the human *KRAS* gene using the CRISPR/Cas9 system

A dissertation by

Ezio Fok

Submitted in fulfilment of the requirements for the degree of Master of Science to the Faculty of Health Sciences at the University of the Witwatersrand, Johannesburg, South Africa.

November 2014

Declaration

I, Ezio Fok, declare that this dissertation is my own work. It is being submitted for the degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.



……

…17th… day of November 2014.

Publications and presentations

Presentations

Fok, E. T., Ruff, P., Weinberg, M.S., Penny, C. Gene editing using CRISPR: Functional characterisation of CRISPR-mediated modification of the human *KRAS* locus. South African Society of Biochemistry and Molecular Biology (SASBMB), 6-9 July 2014, Cape Town, South Africa.

Abstract

The genome is comprised of a simple quaternary code that serves as the "software" which governs all the complex processes of life. The successful manipulation of this code holds enormous potential for applications involving genome engineering. However, the tools needed to navigate the complex landscape of the genome and efficiently and precisely introduce engineered modifications are lacking. The clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system found in prokaryotes functions to recognise and silence foreign pathogenic DNA by double-stranded break (DSB) DNA digestion. The type II CRISPR system of *Streptococcus pyogenes* has recently been reconstituted to function in mammalian cells as a highly programmable and efficient gene-targeting nuclease platform. Through the heterologous expression of the CRISPR associated (Cas) 9 endonuclease and a small single-guide RNA (sgRNA) molecule, specific gene sequences can be targeted for a DNA DSB. The repair of this targeted DNA damage can be exploited to mutate gene sequences or reconstitute break sites according to a homologous repair template for precise gene modifications. In this study, the capabilities of the CRISPR/ Cas9 system for genome editing was tested by using it to precisely mutate the *KRAS* proto-oncogene. A panel of five *KRAS* targeting sgRNAs were designed around a common G>T mutation in codon 12 and characterised. The ability of the CRISPR/ Cas9 system to stimulate DSBs in this region was assessed using the Surveyor Assay, which is able to detect the non-homologous end joining repair of these breaks. A maximum cleavage activity of 6.01 and 2.74% was detected, upstream and downstream of the mutation site, respectively. Simultaneous cleavage by these two sgRNAs was able to successfully introduce a locus-specific micro-deletion. The homology directed repair of these DSBs according to a 90-mer single-stranded oligodeoxynucleotide repair template was shown with the RFLP assay. Analysis of these results implicated guide sequence composition and DSB repair pathway bias as potential factors which may affect the efficiency of desired gene-editing outcomes. These characterised sgRNAs were then applied to generate selectable G>T *KRAS* mutants. A "dual-cut" strategy, which was designed to overcome gene conversion

limitations was employed, and the outcomes were measured with qPCR. The results show that 0.123% of transfected cells were successful recombinants, demonstrating that the use of a “dual-cut” strategy with the CRISPR/Cas9 system was functional and efficient for the generation of knock-in mutants. The CRISPR/Cas9 system has proven to be efficient and robust in modifying the human *KRAS* locus in various ways. With its modularity and simplicity, CRISPR/Cas9 is a powerful tool that will allow for the modification and interrogation of gene function.

Acknowledgements

Thank you to my supervisor Dr Clement Penny for his support throughout my time in his laboratory.

Thank you to my co-supervisor Prof Marco Weinberg who has been instrumental in this work. I cannot thank you enough for your endless guidance and support which has really gone far beyond your call of duty. The opportunity to work with you has been an invaluable learning experience. Your passion for science is inspirational. Thank you for everything!

Thanks to the Musa Mhlanga lab for all their advice and their generosity with their reagents.

I would also like to thank Dr Carol Crowther for providing me with the HEK 293 cell line, Dr Abdullah Ely for the pCI-NeoR-eGFP plasmid and Prof Marco Weinberg, again, for creating and providing the pcDNA.H1.sgRNA expression vector.

To my funders, who have made this financially possible. Thank you to the University of the Witwatersrand, The National Research Foundation and the Canon Collins Trust.

A tremendous thank you to all my colleagues in the laboratory who have become great friends along the way.

To Angie, for all your support, love and care. I really appreciate it all.

And lastly, to my parents who have made and continue to make great sacrifices for me to do what I love. Thank you for everything.

Table of contents

Declaration i

Publications and presentations ii

Abstract iii

Acknowledgements v

Table of contents vi

List of figures and tables ix

Abbreviations xi

1. Introduction 1

1.1 Genome Engineering 1

1.2 Gene Targeting 2

1.3 Designer nucleases 6

1.3.1 RNA-guided endonucleases 8

1.4 Targeting KRAS with CRISPR/Cas9 17

1.5 Aims and objectives 17

2. Materials and Methods 19

2.1 Tissue Culture 19

2.2 Cloning 20

2.2.1 sgRNA expression plasmids 20

2.2.1.1 Preparation of the expression plasmid 21

2.2.1.2 Guide sequence ligation 21

2.2.2 dsDNA donor cassette 24

2.2.2.1 Amplification of the dsDNA donor components 24

2.2.2.2 dsDNA donor assembly 29

2.2.2.3 Blunt-end ligation of the dsDNA donor cassette into pJet 1.2/blunt Cloning Vector 31

2.3 Surveyor Assay 33

2.3.1 Transfection 34

2.3.2 gDNA extraction 34

2.3.3 PCR and heteroduplex formation 35

2.3.4 Heteropduplex digestion and polyacrylamide gel electrophoresis (PAGE) 36

2.3.5 Analysis 37

2.4 Micro-deletion detection 38

2.5 Assessing HDR efficiency with ssODNs 38
2.6 Testing the “dual cut” strategy for the production of G>T KRAS mutants 40
2.6.1 Linear dsDNA donor preparation 40
2.6.2 gDNA template preparation 40
2.6.3 Quantitative real-time PCR (qPCR) 41
2.6.4 Analysis 43
2.7 Statistical Analyses 43
3. Results **44**
3.1 Cloning of sgRNA expression plasmids 44
3.2 Surveyor Assay 47
3.3 Micro-deletion detection 50
3.4 HDR efficiency with ssODNs 52
3.5 Testing the “dual cut” strategy for the production of G>T KRAS mutants 55
3.5.1 dsDNA donor assembly 58
3.5.2 HDR efficiency with the dsDNA donor using the “dual cut” strategy 60
4. Discussion **66**
4.1 sgRNA guide sequence composition affects CRISPR/Cas9 indel forming activity 66
4.2 CRISPR/Cas9-induced DSBs and micro-deletions can be predominantly repaired by different DNA repair pathways 71
4.3 A “dual cut” CRISPR/Cas9 strategy can mediate the production of selectable G>T mutant KRAS cells 73
5. Conclusion **76**
6. References **78**
7. Appendix **83**
7.1 Supplementary data 83
7.2 Supplementary sequences 87
7.2.1 dsDNA donor components 87
7.3 Reagents and recipes 89
7.3.1. Freeze medium 89
7.3.2 Transformation buffer 89
7.3.3 Luria Bertani (LB) broth 89
7.3.4 Ampicillin 1 000X stock (100 mg/ml) 90
7.3.5 Ampicillin agar plates 90
7.3.6 Ampicillin-LB broth 90

7.3.7 TAE buffer (50X) 91
7.3.8 TBE buffer (5X) 91
7.3.9 EDTA solution (0.5M, pH 8.0) 91
7.3.10 PBS (1X) 92
7.3.11 Tris-HCl (1M, pH 8.0) 92
7.3.12 TE buffer (10X) 92
7.3.13 Autoclaving conditions 92
7.4 Protocols 93
7.4.1 Competent E. coli Stbl3 cells 93
7.4.2 QIAprep Spin Miniprep Kit 93
7.4.3 Qiagen Plasmid Midi Kit 94
7.4.4 QIAamp DNA Mini Kit 95
7.4.5 GeneJet Gel Extraction Kit 95
7.4.6 GeneJet PCR Purification Kit 96
7.5 Plasmid maps 97
7.5.1 pcDNA.H1.sgRNA 97
7.5.2 pCI-neo-eGFP 97
7.5.3 pJet 1.2 98
7.5.4 pX330 98
7.6 Ethics approval waiver 99

List of figures and tables

Figure 1: A schematic illustrating the two common repair pathways of DSBs in mammalian cells 6

Figure 2: A schematic showing the function of the CRISPR/Cas9 system in *S. pyogenes* 11

Figure 3: The evolution of the tracrRNA:crRNA complex design for the efficient and simple use of CRISPR/Cas9 in mammalian cells 14

Figure 4: The cloning strategy of the guide sequence into the pcDNA.H1.sgRNA expression plasmid 24

Figure 5: A schematic outlining the production of the eGFP fragment 26

Figure 6: A schematic outlining the production of the SV40-NeoR fragment 27

Figure 7: A schematic outlining the production of the KRS fragment 28

Figure 8: A schematic of the 5′ and 3′ HAs 28

Figure 9: The overlap extension PCR assembly strategy for the dsDNA donor construct 30

Figure 10: The design and screening of the sgRNA constructs 45

Figure 11: Functional assessment of KRAS targeting sgRNAs with the Surveyor Assay 48

Figure 12: PCR detection of CRISPR/Cas9 mediated micro-deletions 50

Figure 13: HDR of the individual DSBs and the micro-deletion with ssODNs 53

Figure 14: The envisioned “dual cut” strategy versus the transitional “single cut” strategy to modify the *KRAS* gene 56

Figure 15: Assembly and verification of the dsDNA donor construct 59

Figure 16: Integration efficiency of the dsDNA donor 63

Figure 17: Analysis of the guide sequence composition of sgRNA function 67

Supplementary figure 1: Optimisation of ssODN dosage for maximum HDR 81

Supplementary figure 2: Optimisation of the primer sets for qPCR 82

Supplementary figure 3: Melt curve analysis to validate specific primer amplification 83

Supplementary figure 4: The effect of different dsDNA donor concentrations and formats on HDR-mediated construct integration 84

Table 1: List of primer sequences used for sgRNA cloning 23

Table 2: List of primer sequences used for the production of the dsDNA donor 32

Table 3: List of primer sequences used for the testing of the sgRNAs 36

Table 4: List of primer sequences used for qPCR 41

Abbreviations

HR	Homologous recombination
HSV-*tk*	Herpes Simplex Virus thymidine kinase
AAV	Adeno-associated virus
rAAV	Recombinant adeno-associated virus
DSB	Double-stranded break
NHEJ	Non-homologous end joining
Indels	Insertions and deletions
HDR	Homology-directed repair
ZFN	Zinc-finger nuclease
TALEN	Transcription activator-like nuclease
RVD	Repeat variable diresidue
CRISPR	Clustered regularly interspaced palindromic repeats
Cas	CRISPR associated
pre-crRNA	Precursor-CRISPR-RNA
crRNA	CRISPR-RNA
tracrRNA	Trans-activating CRISPR RNA
ssDNA	Single-stranded DNA
dsDNA	Double-stranded DNA
ITR	Inverted terminal repeat
sgRNA	Single guide RNA
PAM	Protospacer adjacent motif
ERK	Extracellular signal-regulated kinase
EGFR	Epidermal growth factor receptor
HEK	Human embryonic kidney
ssODN	Single-stranded oligodeoxynucleotide
Bp	Base pair
Pen/Strep	Penicillin/Streptomycin
dCas9	Deactivated Cas9
DMEM	Dulbecco's Modified Eagle Medium

FBS	Foetal bovine serum
PBS	Phosphate buffered saline
ddH_20	Deionised water
TAE	Tris-acetate EDTA
PCR	Polymerase chain reaction
KRS	*KRAS* reconstitution sequence
HA	Homology arm
gDNA	Genomic DNA
PEI	Polyethylenimine
PAGE	Polyacrylamide gel electrophoresis
APS	Ammonium persulphate
TEMED	Tetramethylethylenediamine
qPCR	Quantitative real-time PCR
IPC	Inter-plate calibrator
TBE	Tris-borate EDTA

1. Introduction

1.1 Genome Engineering

The ability to understand and exploit the genetic code that governs all the complex functions of life, relies heavily on our ability to manipulate the molecules that constitute it. Genome engineering involves applying targeted genetic modifications into the genome of a living system for this purpose. The manipulation of DNA began in the 1970s when our basic understanding of this code was born. This led to the emergence of recombinant DNA technologies, allowing biologists to manipulate DNA and study the function of genes and gene systems, and also employ them for human benefit. These technologies have been particularly relevant with the increasing availability and capacity to rapidly read entire genomes. Genome engineering using recombinant DNA technologies have allowed us to functionally decipher large volumes of gene sequence data generated by high-throughput sequencing. Through the process of systematic reverse engineering, the function of many causal genetic elements have been elucidated and validated. In this way, interesting and useful biological molecules and pathways have been uncovered and studied. Furthermore, the understanding of these genomic elements and the ability to manipulate their code, has allowed for the routine reprogramming and engineering of life.

Despite the exponential growth in the capacity to synthesise, manipulate and analyse DNA (a trend reminiscent to Moore's Law), the successful engineering of DNA in the context and scale of the genome has been rather limited (Carr and Church, 2009). Tools that are capable of navigating the complex landscape of the genome, at the resolution of the single nucleotide level, and facilitating the efficient introduction of precise genomic perturbations have been the rate-limiting step. Without these tools, hypotheses based on observational data derived from the genome cannot be rapidly or easily validated, preventing these biological models from being applied to advance basic research, biotechnology and medicine.

1.2 Gene Targeting

The technology to make precise targeted alterations to the mammalian genome first came about in the 1980s, when Mario Capecchi and Oliver Smithies showed that newly introduced DNA constructs could be stably inserted into specific genomic loci by homologous recombination (HR). Before this, the introduction of novel DNA into the genome was mediated by retroviral vectors, which sometimes integrated their DNA payload into random sites, causing unwanted insertional mutagenesis (Wirt and Porteus, 2012). Gene targeting, however, involved the integration of a gene targeting DNA construct into the target sites of interest within the genome, that were highly homologous. Successful recombination events resulted in the linear transfer of genetic material from the exogenous donor DNA to the endogenous target site, in a process known as gene conversion (Smithies et al., 1985). This breakthrough technique has been the cornerstone of genome engineering, enabling molecular biologists to generate knock-out and knock-in cell and animal models for decades (Mussolino and Cathomen, 2012).

Despite this, the efficiency of gene targeting is extremely poor, as spontaneous and targeted HR with exogenous gene targeting DNA occurs at a rate of 10^{-6}-10^{-8} in cultured mammalian cells (Wirt and Porteus, 2012). Because of the rarity of this event, the use of gene targeting to produce and isolate isogenic cell lines that recapitulate genetic perturbations of interest is very challenging. Current strategies involve the enrichment of cell populations that have successfully undergone the targeted gene modification, by concurrently knocking-in a gene that would imbue the cell with an advantageous phenotype under selective conditions. From this, clonal cell colonies can be isolated and expanded into recombinant cell lines. This was first demonstrated in cultured human cells by work done by Oliver Smithies, who showed that the *neo* gene could be stably integrated into the β-globin locus by HR, and expressed to allow for the selection of recombinants using the aminoglycoside antibiotic G418 (Smithies et al., 1985). However, the high frequency of non-homologous recombination in mammalian cells often results in the selection of clones that have recombined at illegitimate genomic sites. This produces a large number of false positive clones that add to the screening process. In order to overcome this, a positive-negative selection system was developed to decrease the

yield of illegitimate recombinants (Capecchi, 1989). This system is made up of two independent expression cassettes. The *neo* expression cassette is designed to be flanked by two homology arms that will direct its integration into the target site by HR. The Herpes Simplex Virus thymidine kinase (HSV-*tk*) gene is placed adjacent to the region of homology, and its integration is excluded when on-target HR occurs. Only off-target integration of the gene targeting DNA, by non-homologous recombination, will retain the HSV-*tk* cassette. In the first round of positive selection, G418 is used to select for cells that have experienced recombination and harbour the *neo* cassette. The HSV-*tk* expression cassette is then used to negatively select for cells that have specifically undergone targeted HR. In the presence of gancylcovir, illegitimate recombinants that express the HSV-*tk*, as a result of random integration, are killed off. The use of this dual selection strategy has improved the rate at which positive clones are selected. However, this process requires the production of large gene targeting vectors to accommodate the two selection cassettes and the two homology arms, which are often between 2-4 kb in length each, to achieve sufficient HR (Capecchi, 1989). Furthermore, the requirement for two subsequent selection steps makes this a very time consuming process.

An alternative to preventing the enrichment of random recombinants is the use of a promoter trap strategy (Sedivy and Dutriaux, 1999). This relies on the integration of the selection cassette, in frame, into an exon of the targeted gene so that its expression can be driven by the endogenous upstream promoter. Translation of the selection cassette can then subsequently occur from the native start codon to produce a fusion protein. This strategy has been shown to decrease the selection of random recombinants as the chances of the targeting vector integrating in frame into the exon of a random active off-target gene is quite low (Sedivy and Dutriaux, 1999). However, this strategy is only limited to genes that are being actively expressed in the cells used. Also, the integration of the selection cassette will result in the loss of function of the targeted gene, and provisions must be made to replace the disrupted exon or remove the selection cassette, after the selection process, to maintain gene functionality. Both these strategies involve the production of large and cumbersome gene targeting vectors, which simply serve to enhance the recovery and enrichment of the rare cells that harbour the desired genetic modifications. Methods that enhance the rate of HR at the target site, to

increase the obtainable yield of successful recombinants, may be more favourable for gene editing applications.

The delivery of gene targeting DNA constructs into cells using recombinant adeno-associated virus (rAAV) vectors has been shown to increase the rate of targeted HR by a factor of 25-fold compared to naked plasmid-based approaches (Topaloglu et al., 2005). This is because the AAV genome has been found to be intrinsically highly recombinogenic, aiding the HR of gene targeting DNA with the genomic target sequence. AAVs are composed of a 4.7 kb single-stranded DNA genome which is flanked by two short stretches of double-stranded DNA (dsDNA) known as the inverted terminal repeats (ITRs) (Vasileva et al., 2006). It is thought that the single-stranded nature, and its transition from being single-stranded to double-stranded at the ITRs, of AAV genomes, recruit DNA repair proteins or complexes which facilitate and enhance viral DNA integration via HR (Russell and Hirata, 1998, Vasileva et al., 2006). Furthermore, rAAVs are able to infect a large number of different cell types and efficiently deliver intact gene targeting DNA into the cell, which further improves the efficiency at which HR can occur. rAAVs have been shown to be able to target genes at an efficiency of 1% under optimal conditions (Khan et al., 2011). But even with this high efficiency of gene targeting, successfully modified cells still need to be selected for from the entire cell population (Porteus, 2007). This necessitates the inclusion of a selection cassette along with the two homology arms and the intended genetic modification in the viral gene targeting DNA construct. However, this can be challenging as the size of the construct is limited by the small packaging capacity of AAVs. The tradeoff often made to include all these elements within a virion particle is the use of shorter homology arms, which sacrifices HR efficiency. Furthermore, it has been shown that rAAVs are relatively inefficient at targeting silent loci, which limits its versatility for gene editing applications. The production of highly concentrated and pure rAAV stocks, a requisite for efficient transduction, is also technical and time consuming (Khan et al., 2011). Even though rAAVs are able to significantly stimulate HR, several factors limit the ease and versatility at which rAAVs can be used for gene targeting.

Targeted DNA double-stranded breaks (DSBs) have been shown to increase the frequency of HR at the DSB site by a factor of 1 000-fold or more in mammalian cells (Chen et al., 2007). This is because chromosomal breaks stimulate the high fidelity homology-directed repair (HDR) response, which relies on a homologous template to seamlessly reconstitute the damaged site (Yang et al., 2013). Therefore, the efficiency of gene targeting can be dramatically enhanced by introducing a targeted DSB along with a homologous repair template to stimulate HDR and subsequent gene conversion (Chen et al., 2007). In this way, homologous repair templates can contain 200-800 bp long homology sequences, as the stimulation of HDR by the targeted DSB, is able to sufficiently compensate for the use of shorter homology arms. This makes the construction of donor repair templates significantly simpler than traditional gene targeting vectors (Chen et al., 2011). Furthermore, these repair templates are not limited by any design constraints, such as the size limits associated with the use of rAAVs, allowing for multiple selection cassettes and gene modifications to be included, and various selection strategies to be employed. The repair of DNA DSBs by the alternative error prone non-homologous end joining (NHEJ) pathway further expands the achievable repertoire of targeted gene modifications in mammalian cells (Kim and Kim, 2014). NHEJ occurs by the re-ligation of the broken DNA ends at the DSB site, in the absence of a repair template. In this process, small insertions and deletions (indels) are introduced into the repair site, which results in the targeted mutagenesis of the gene (Kim and Kim, 2014) (Figure 1). The resolution of DNA DSBs by these common mammalian repair pathways each yield distinct genetic modifications that are useful for genome engineering. By introducing a DSB at a specific target site, a variety of genetic modifications can be made, ranging from gene mutagenesis and gene deletions, to the efficient introduction of very precise gene alterations using donor repair templates.

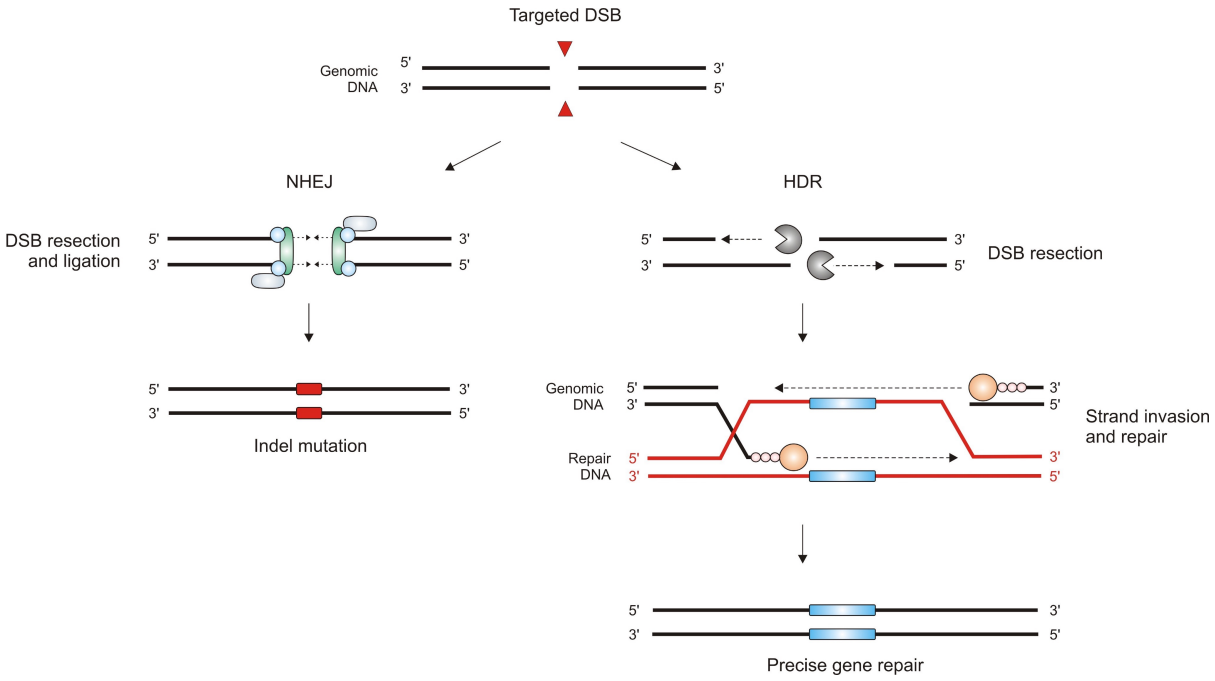


Figure 1: A schematic illustrating the two common DNA repair pathways of DSBs in mammalian cells. In the absence of a homologous repair template, NHEJ is stimulated. This involves the re-ligation of the broken DNA ends together and in the process small, mutagenic, insertions and deletions are included in the repaired gene sequence. HDR involves the re-sectioning of the broken DNA ends, which invade a homologous repair template. The damaged DNA is re-synthesised according to the repair template, allowing for the introduction of precise gene modifications.

1.3 Designer nucleases

Targeted chromosomal DSBs, for the stimulation of HDR in mammalian cells, can be achieved through the use of endonucleases to cleave specific genomic sites. This was first demonstrated with the yeast-derived homing endonuclease I-*Sce*I (Wirt and Porteus, 2012). This homing endonuclease recognises and cleaves an 18 bp non-palindromic sequence, which is predicted to occur only every 7×10^{10} bp, assuming that the base pairs of the genome being targeted are randomly distributed (Wirt and Porteus, 2012). This small target space severely limits the use of this homing endonuclease for gene targeting purposes. Because of this, the use of I-*Sce*I has mostly served as a proof-of-concept to demonstrate that targeted DSBs can stimulate HDR for gene editing. In these studies, the I-*Sce*I recognition sequence was integrated into dysfunctional *neo* or *eGFP* expression cassettes, which were then targeted for HDR correction. These repaired constructs were then detected through the selection of cells that displayed the relevant phenotype (Wirt and Porteus, 2012). In order to introduce

chromosomal breaks at specific target sites of interest within the genome, endonucleases that have programmable sequence recognition characteristics are needed. Designer nucleases, such as chemical nucleases, meganucleases, zinc-finger nucleases (ZFNs) and transcription activator-like nucleases (TALENs), can be programmed to recognise and target specific genomic sequences. By employing these nucleases in genome engineering strategies, the endogenous DNA repair pathways found in mammalian cells can be exploited for the versatile and efficient modification of genomic gene sequences.

ZFNs are one of the platforms that have successfully gone into the mainstream for genome engineering purposes. ZFNs are chimeric proteins that have been shown to be able to introduce targeted DSBs within the genome, and stimulate HDR (Porteus and Carroll, 2005). These nucleases are based on the ubiquitous mammalian Cys2-His2 zinc-finger DNA recognition domain, which is fused to the catalytic domain of the *Fok*I endonuclease for DNA DSB induction. ZFNs are designed to function as obligate pairs, which dimerise at the target site to form an active *Fok*I homodimer (Porteus and Carroll, 2005). The target specificity of each ZFN pair is dictated by a tandem array of 3-6 monomers, which each bind to 3-4 nucleotides (Chhabra, 2014). This particular DNA recognition resolution has limited the target space of ZFNs to only 1 site in every 100 bp on average, in the human genome. Furthermore, the construction of ZFNs is a labour intensive process, that is not feasibly scalable to large genomic studies (Carroll, 2011). Despite extensive engineering to improve these nucleases, their adoption for genome engineering applications has been relatively poor.

TALENs have proven to be far more robust than ZFNs, and currently serve as the gold standard in the gene editing toolbox. TALENs have been adapted from plant bacterial pathogens which use these DNA binding molecules to modulate host gene expression during infection (Mussolino and Cathomen, 2012). The DNA binding domain of TALENs consist of tandem arrays of 33-35 amino acid repeats. Each repeat module is able to recognise a single nucleotide through a corresponding variation of two amino acids at the twelfth and thirteenth position of each repeat. These are known as the repeat variable diresidues (RVDs) (Kim and Kim, 2014). Following a simple amino acid to DNA cipher, TALENs have been constructed with 18-20 repeat modules and can be potentially configured to target any user-defined

genomic space (Boch et al., 2009). Similar to ZFNs, the DNA recognition module of TALENs is fused to the *Fok*I cleavage domain and function as obligate pairs for activation. This robust architecture has been widely adopted for various applications and even scaled up for high-throughput applications. However, the construction of these molecules still remains very time consuming and technical (Li et al., 2014).

The use of these designer nucleases to stimulate DNA repair pathways at specific genomic targets sites has been the driving force of genome engineering. This strategy is currently the most efficient way to precisely edit the human genome (Yang et al., 2013). However, despite the modular design and programmability of both ZFNs and TALENs, the need to completely re-engineer the protein-based recognition module to target different genes is inefficient. Also, recognition module position bias inherent to these two platforms, has been demonstrated to adversely affect their robustness. It has been shown that the activity of certain module combinations are unpredictable and inefficient (Carroll, 2011, Carroll and Beumer, 2014). Furthermore, the technical complexities and intense labour requirements to generate a single pair of these nucleases render these platforms undesirable and unfeasible for the systematic interrogation of entire genomes, or the rapid construction of model systems that recapitulate genetic elements of interest. Currently, there is the need for the development of new genome engineering tools that can robustly and efficiently mediate various gene perturbations in a highly specific manner. Moreover, in the current genomics era, these tools need to be highly programmable and easily scalable, through a construction and implementation process that is simple, rapid and cost-effective (Perez-Pinera et al., 2012). The development of such powerful, yet simple tools, will allow the exploration of the genetic code to expand to the scale of the genome.

1.3.1 RNA-guided endonucleases

The clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated (Cas) proteins constitute an adaptive immune system found in bacteria and archaea, which functions to silence foreign DNA by RNA-guided nucleolytic digestion (Barrangou et al., 2007, Brouns et al., 2008, Marraffini and Sontheimer, 2008). There are

three types (I-III) of the CRISPR/Cas system that have been identified in various prokaryotic species, which each have the same general mode of function to achieve acquired immunity (Barrangou and Marraffini, 2014). In each type, the organism is immunised through the integration of short fragments of the invading nucleic acids into the CRISPR locus. Expression and maturation of these sequences (also known as spacers) yields individual CRISPR RNAs (crRNAs), which each contain the exogenously-derived spacers for target recognition by Watson and Crick base-pairing. DNA interference occurs through the crRNAs binding with the effector Cas protein component and guiding it to the cognate DNA sequence (also known as the protospacer) for digestion (Barrangou and Marraffini, 2014). Successful binding of the Cas ribonucleoprotein complex to the target DNA sequence requires the recognition of the protospacer adjacent motif (PAM). This is a short recognition sequence directly adjacent to the target sequence, which serves in the identification of potential spacer sequences in the immunisation process, as well as foreign target DNA recognition to avoid auto-immunity (Mojica et al., 2009, Paez-Espino et al., 2013). The sequences of these motifs have been identified in various species that make use of the type I and type II systems, however, they still need to be characterised for the type III system (Barrangou and Marraffini, 2014).

The CRISPR/Cas system types are defined by the various Cas proteins involved in achieving this kind of immune response. The type I and type III system rely on the assembly of a large multisubunit complex of Cas proteins for crRNA maturation and target cleavage activity (Brouns et al., 2008, Hatoum-Aslan et al., 2013). The type II system is able to execute these functions through the use of a single Cas polypeptide, Cas9, making it the most minimalistic CRISPR/Cas system identified and the most extensively studied (Gasiunas et al., 2012). The type II CRISPR locus encodes the Cas9 endonuclease and two non-coding RNAs: the precursor-CRISPR-RNA (pre-crRNA) array, which consists of unique exogenously-derived guide spacer sequences that are interspaced by short direct repeats, and the trans-activating CRISPR RNA (tracrRNA) (Fonfara et al., 2014). The process in which these components function to achieve adaptive immunity with the type II system was first demonstrated in *Streptococcus pyogenes* (*S. pyogenes*) SF 370 and involves the following 3 phases (Figure 2) (Deltcheva et al., 2011):

1. *Adaptation phase* - Under a viral or plasmid challenge, *S. pyogenes* is able to respond by integrating fragments of the foreign DNA into the CRISPR array to acquire new spacer/ guide sequences that serve to recognise the target. Accompanying the integration of these new spacers into the 5′ end of the array, the repeat sequence is duplicated so that the architecture of the CRISPR array is maintained (Jinek et al., 2012).

2. *Expression phase* - The entire CRISPR array is expressed to generate a pre-crRNA transcript that is made up of multiple spacer-repeat units in tandem. Expressed tracrRNA molecules bind to the repeat regions of the pre-crRNA array by 25 conserved and complementary nucleotides. This initiates the maturation process, which involves the tracrRNA-pre-crRNA array complex being cleaved by RNase III, in the presence of Cas9, to yield mature individual tracrRNA:crRNA units. Mature crRNAs are 39-42 nucleotides long. On the 5′ end of these, is a unique 20 nucleotide guide sequence derived from the spacer regions of the array, for target recognition. The 3′ end consists of a 14 nucleotide long CRISPR repeat sequence that is bound to the mature tracrRNA, forming the repeat:anti-repeat duplex. The tracrRNA is 89 nucleotides long, with the remaining unbound nucleotides forming a tracrRNA tail that folds itself into a secondary structure that consists of a tetra-loop followed by 3 stem loops. These structural features have been found to be important for interactions with Cas9 (Jinek et al., 2012, Nishimasu et al., 2014).

3. *Interference phase* - The mature crRNA is then loaded into the Cas9 endonuclease, which is comprised of two catalytic domains. The HNH domain and the RuvC-like domain cleave the complementary and the non-complementary strand of double-stranded DNA (dsDNA), respectively (Jinek et al., 2012). The loading of the crRNA activates the Cas9 endonuclease, through conformational rearrangements of its catalytic cores (Jinek et al., 2014). This forms an RNA-protein complex that is able to recognise target DNA sequences that are complementary to the 20 nucleotide guide sequence of the loaded crRNA. The successful binding of the crRNA-Cas9 complex to the target, or protospacer sequence, is mediated by the recognition of an obligatory PAM. This is located directly 3′

of the target sequence. With this, the crRNA-Cas9 complex destabilises the target DNA duplex and initiates strand invasion, whereby the appropriate strands of the target are loaded into the Cas9 catalytic cores for cleavage. A DSB is generated exactly 3 bps upstream of the PAM site, to produce blunt-ended products (Sternberg et al., 2014). For the type II system in *S. pyogenes*, the PAM recognition motif is NGG, but cleavage activity with a NAG motif has also been observed, albeit at a much lower efficiency (Hsu et al., 2013).

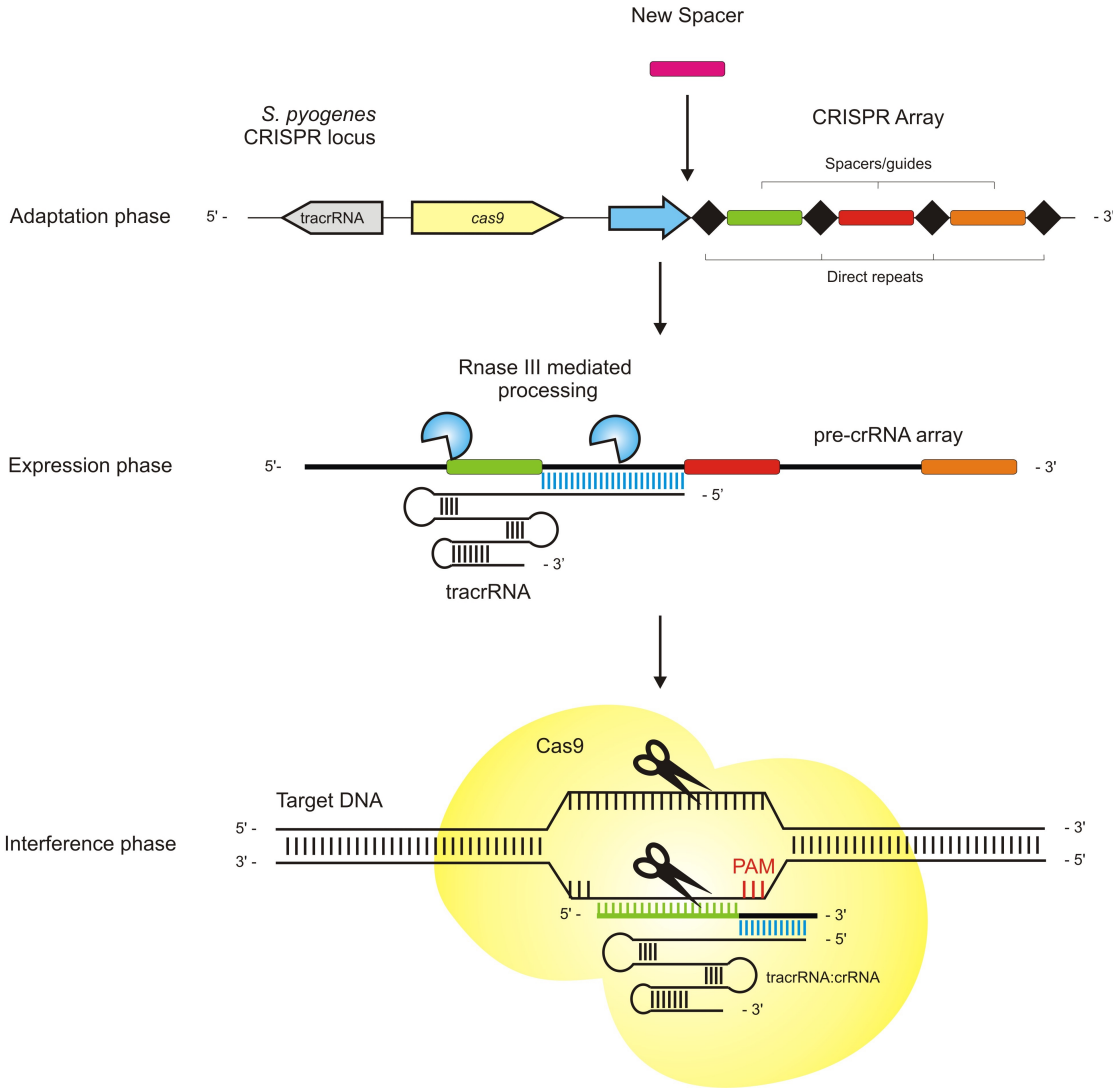


Figure 2: A schematic showing the function of the CRISPR/Cas9 system in *S. pyogenes*. The type II CRISPR/Cas9 system found in *S. pyogenes* is comprised of two non-coding RNA components (the crRNA and the tracrRNA) and the catalytic Cas9 endonuclease. crRNA is produced from the CRISPR array, which is expressed as a single transcript that gets processed into individual, active crRNA:tracrRNA complexes. These tracrRNA:crRNA complexes are able to activate the Cas9 endonuclease and direct it to cleave DNA sequences that are complementary to the guide or spacer sequence (green) and are directly upstream of a PAM recognition motif (red).

Recently, various type II CRISPR/Cas9 systems (from *S. pyogenes*, *Streptococcus thermophilus* and *Neisseria meningitidis*) have been successfully reconstituted to function in mammalian cells through the heterologous expression of Cas9 and the tracrRNA and pre-crRNA components (Fonfara et al., 2014). In a system developed by Zhang *et al.*, a human codon-optimised version of *S. pyogenes* Cas9 and the pre-crRNA and tracrRNA components were expressed from independent promoters from a single plasmid. The Cas9 component was flanked by two nuclear localisation signals and was expressed from the elongation factor 1α promoter. The two RNA components were expressed from their own U6 Pol-III promoters (Figure 3A) (Cong et al., 2013). This system has been shown to successfully mimic the native CRISPR/Cas9 system of *S. pyogenes* in mammalian cells, whereby endogenous RNase III is able to mediate the maturation of the pre-crRNA and tracrRNA to form an active complex with the Cas9 endonuclease and guide it to the specified target for DSB induction (Cong et al., 2013).

In an attempt to develop a more efficient and facile system, two independent groups used a tetra-loop to fuse the 5′ end of the tracrRNA and 3′ end of the crRNA component into a single chimeric molecule, which was expressed from one promoter. This single-guide RNA (sgRNA) v1.0 was composed of a full length, mature crRNA (42 nucleotides) and a truncated tracrRNA (RNA scaffold), which retained the minimum number of nucleotides (position 23-48) to maintain functionality (Figure 3B) (Jinek et al., 2012, Cong et al., 2013). The use of sgRNA v1.0 was expected to improve the cleavage efficiency of the CRISPR/Cas9 system, as its function bypasses any bottlenecks associated with the expression and complexing of the crRNA and tracrRNA as individual components, and the process of tracrRNA:crRNA maturation. However, it was shown that this chimeric sgRNA v1.0 design consistently performed poorly compared to the tracrRNA:crRNA duplex configuration when targeting the *EMX1* gene in mammalian cells (Cong et al., 2013). Subsequent optimisation of this sgRNA architecture produced the sgRNA v1.1 design, which is a chimeric molecule that more closely resembles the wild-type (WT) tracrRNA:crRNA duplex. This was achieved by lengthening the RNA scaffold tail to the native 89 nucleotides, which significantly improved the cleavage efficiency of the system to levels that were comparable or exceeded that seen with the tracrRNA:crRNA duplex configuration (Hsu et al., 2013). This can be attributed to the RNA

scaffold tail assuming the secondary structural loops which have been shown to be important for sgRNA loading and activation of the Cas9 protein (Nishimasu et al., 2014). In other work done by Huang *et al.*, the CRISPR/Cas9 system was repurposed to label and visualise specific genomic loci in living cells. By tagging eGFP to a deactivated form of Cas9 (dCas9), the improved sgRNA design was still not able to efficiently form distinct observable puncta, which represent target-bound sgRNA:dCas9 complexes. From this it was found that the sgRNA architecture was still not being efficiently expressed or loaded into the dCas9 to form an active complex. A putative Pol-III terminator sequence (4 consecutive U's) in the crRNA sequence of the sgRNA was identified. In order to prevent this from prematurely stopping the expression of sgRNA v1.1, the terminator sequence was abolished by an A-U base swap in the repeat:anti-repeat duplex region. To further improve the loading of sgRNA v1.1, the repeat:anti-repeat duplex region was also extended so that it more closely mimicked the length used in the tracrRNA:crRNA duplex design (Figure 3C). These improvements have been shown to significantly improve the production of active sgRNA:dCas9 complexes (Chen et al., 2013). It is predicted that the benefits of these changes to the sgRNA architecture will carry through to its use with the active Cas9 endonuclease for gene targeting.

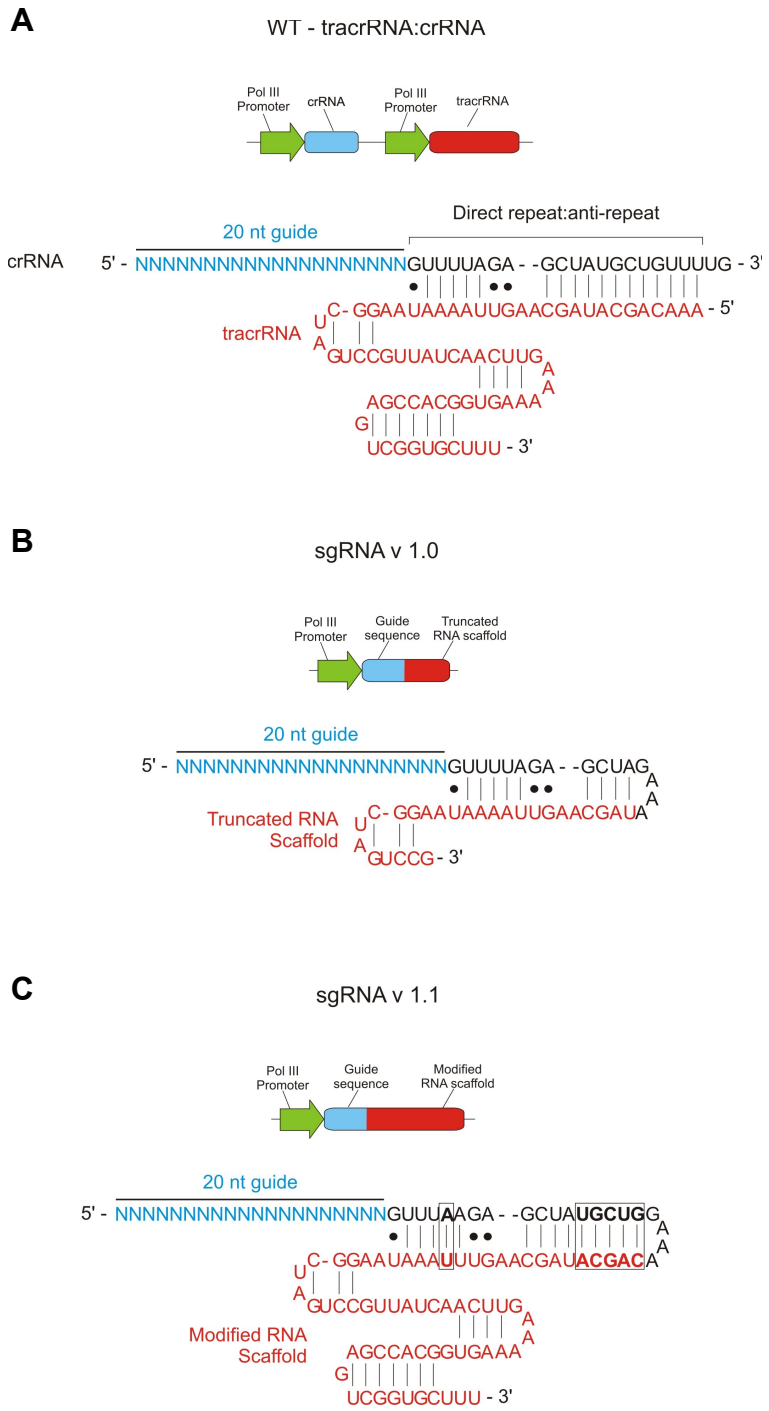


Figure 3: The evolution of the tracrRNA:crRNA complex design for the efficient and simple use of CRISPR/Cas9 in mammalian cells. (**A**) The original tracrRNA:crRNA design was a copy of the WT architecture found in *S. pyogenes*. This involved the expression of the crRNA and tracrRNA as separate components from two different promoters. The function of this design relied on the formation of the tracrRNA:crRNA complex through base-pair interactions between the direct repeat:anti-repeat region of each RNA component and its subsequent maturation. (**B**) The sgRNA fused the crRNA and tracrRNA components together via a tetra-loop, to produce a very minimalist chimeric molecule, which could be expressed from a single promoter and mimicked the natural crRNA:tracrRNA architecture. (**C**) Improvements to the sgRNA design involved removing a premature Pol-III terminator sequence and extending the repeat:anti-repeat stem, as well as the RNA scaffold.

The evolution of the CRISPR/Cas9 system for its use in mammalian cells has produced a very minimalistic and efficient system for genome engineering. By simply expressing Cas9 and a sgRNA containing the guide sequence of interest, almost any genomic sequence can be targeted for modification. This presents a completely new tool that is easy to engineer and implement for the introduction of targeted DSBs in the mammalian genome. Comparing this system to existing platforms, CRISPR/Cas9 has been estimated to have 1 target site in every 8 bp on average in the human genome, if the NGG PAM constraint of *S. pyogenes* is adhered to (Cong et al., 2013). This statistic is increased to every 1 in 4 bp if the degenerate NAG PAM is also considered (Hsu et al., 2013). The use of Cas9 derived from different bacterial species, which have different PAM requirements, may further improve the size of this target space. A recent study provides a collection of divergent PAM sequences recognised by several species of bacteria, which could be used to improve the genome editing capabilities of the CRISPR/Cas9 system (Fonfara et al., 2014). For example, the Cas9 of *Neisseria meningitidis* requires a PAM sequence of NNNNGATT, which is significantly different to the PAM of *S. pyogenes*, expanding the targetable sequence space in the mammalian genome (Hou et al., 2013). Alternatively, the PAM sequence for *Francisella novicida* is NG, which is extremely short and non-specific, placing minimal constraints on target sequence selection (Fonfara et al., 2014). Therefore, through the development and use of Cas9 orthologs, which have evolved to use different PAM sequences, it is likely that the entire mammalian genome can be targeted with the CRISPR/Cas9 system. Furthermore, as the understanding of the interaction between the sgRNA:Cas9 complex and the target DNA PAM site improves, the Cas9 protein could potentially be engineered to have different PAM specificities by rational design (Anders et al., 2014). The activity of this system has also been shown to be comparable, and in some cases, exceed that of TALENs, making it a suitable candidate to replace this gold standard (Hendel et al., 2014).

Besides having robust activity, the simplicity of reprogramming the Cas9 target specificity, through the expression of different sgRNAs, is highly advantageous for genome engineering applications. This has been demonstrated in several recent studies. For example, the high-throughput capability of this system was demonstrated in a genome-wide knock-out screen. In this work, a library of sgRNAs designed to target over 18 000 genes was used for a

negative selection screen to identify important genes for cell viability and proliferation. It was found that sgRNA species that targeted genes essential for ribosomal function were depleted, as these cells failed to survive or proliferate robustly within the cultured cell population. This library was further used in a positive selection screen, in which genes that play a role in the resistance to vemurafenib, a therapeutic for treating melanoma, were identified. In this screen, sgRNA species that were enriched were used to identify genes important for drug resistance, as their loss of function provided the cells with a survival advantage under treatment conditions (Shalem et al., 2014). From this, it was demonstrated that the CRISPR/Cas9 system is able to be feasibly employed for genome-wide knock-out screens. Moreover, by interfering with the gene sequence directly, complete knock-out was achieved, which cannot be done using the conventional RNA interference approach. The CRISPR/Cas9 system has also been used to generate animal models with multiple gene mutations in a single step. This was done by co-injecting the Cas9 mRNA and different sgRNAs into mouse zygotes to generate mice with biallelic mutations in multiple genes (Wang et al., 2013). This method circumvents the tedious process of cross-breeding mice with single mutations to achieve the same result. CRISPR/Cas9 is also not limited to only generating targeted DSBs. By inactivating the catalytic domains of the Cas9 endonuclease to produce a dCas9, (which still maintains its reliance on sgRNAs for specific sequence targeting), the CRISPR/Cas9 system has been used to modulate and explore the genome in other ways. These include fusing transcription factors or repressors to dCas9 and guiding this complex to particular genomic loci to modulate gene expression (Qi et al., 2013). Other applications involve using the CRISPR/Cas9 platform for visualising the spatio-temporal architecture of the genome in living cells by tagging dCas9 to fluorescent molecules (Chen et al., 2013). This demonstrates the versatility of the CRISPR/Cas9 system, which can be adapted to be used as a generic gene sequence targeting platform, which presents users with a new tool to easily, efficiently and creatively explore the genome.

1.4 Targeting *KRAS* with CRISPR/Cas9

CRISPR/Cas9 system is a simple, but extremely powerful and versatile tool for modifying and interrogating the genome. To test this technology and explore different strategies to optimise its function for genome engineering, the targeted mutation of the *KRAS* proto-oncogene was used as a model system to investigate this. Kras is part of the extracellular signal-regulated kinase (ERK) signalling pathway (Gollob et al., 2006). This pathway functions to transmit extracellular stimuli, through a signal transduction cascade, into the nucleus of the cell. Kras serves as an intermediary component in the cytoplasm, in which its activation by the epidermal growth factor receptor (EGFR) catalyses the conversion of a bound GDP to GTP. This activated Kras then subsequently activates the downstream kinase activity of Braf, to relay the input signal. The activity of Kras is self-regulated by an intrinsic competing GTPase, which abrogates its transduction activity by removing the GTP molecule. Several activating point mutations have been found to commonly occur at codons 12, 13 and 61 of *KRAS*, which eliminates the regulatory GTPase activity (Gollob et al., 2006, Loupakis et al., 2009). This results in the production of a constitutively active form of Kras, which continuously stimulates the ERK pathway. This has been commonly observed in various cancer types, providing the transformed cells with enhanced growth and proliferative characteristics (Heinemann et al., 2009).

1.5 Aims and objectives

The aim of this study was to assess the functionality of the CRISPR/Cas9 system for genome editing. The system was tested by programming it to target the *KRAS* gene in human cells. The ability of CRISPR/Cas9 to precisely target and modify this gene locus was assessed based on the efficiency of cleavage and the subsequent repair outcomes of the DSBs. Subsequent to this, a "dual cut" strategy using the CRISPR/Cas9 system was implemented to recapitulate an oncogenic *KRAS* mutant variant in a cell-based model system. In order to achieve this, the following general objectives were set out:

1. Testing the CRISPR/Cas9 system for the modification of the *KRAS* gene

Five sgRNAs (using the sgRNA v1.1 architecture) were designed to mediate DNA DSBs around a common oncogenic G>T mutation hotspot in codon 12, exon 2 of the human *KRAS* gene (Loupakis et al., 2009). The cleavage activity of each sgRNA was assessed by measuring the NHEJ-mediated repair of DSBs. The ability of the CRISPR/Cas9 system to stimulate HDR, for the precise modification of the *KRAS* gene, was also tested by repairing individual DSBs, as well as a micro-deletion, using short single-stranded oligodeoxynucleotides (ssODNs) as repair templates.

2. Using the CRISPR/Cas9 system to generate selectable G>T *KRAS* mutants

The CRISPR/Cas9 system was used to modify the *KRAS* gene to recapitulate the oncogenic G>T mutation in recombinant cells. A “dual cut” strategy, which was designed to facilitate the co-integration of a selection cassette and the point mutation, was simulated. The outcomes of this gene editing strategy was assessed to determine the efficiency of producing selectable *KRAS* mutant recombinants, which can be developed into an isogenic cell line model for further study.

2. Materials and Methods

2.1 Tissue Culture

Human embryonic kidney (HEK) 293 cells were cultured in a laminar flow cabinet under sterile conditions. A 2 ml tube of frozen cells (stored in freeze medium (Appendix 7.3.1)) was quickly thawed and added to 10 ml of complete DMEM, which was composed of Dulbecco's Modified Eagle Medium (DMEM): F 12 mixture, containing glucose, HEPES buffer and L-glutamine (Lonza, USA), and was further supplemented with 10% heat-inactivated foetal bovine serum (FBS) (Hyclone, USA) and 0.2% penicillin-streptomycin (pen/strep) antibiotic mix (Lonza, USA). The cells were pelleted by centrifuging the cell suspension at 200×g for 3 minutes (Minor MSE), after which the supernatant was completely discarded and the pellet resuspended in 10 ml of complete DMEM and added to a T-75 (75 cm^3) flask. HEK 293 cells were incubated in a 37°C humidified incubator (Thermo Forma, USA), with 5% CO_2.

Cell proliferation was periodically monitored with an inverted light microscope (Zeiss Axiovert 25, Germany). Cultures that reached 70-80% confluency were sub-cultured. This involved discarding the cell culture medium and washing the cells with 5 ml of 1X phosphate buffered saline (PBS) (Appendix 7.3.10). The adherent cells were detached by adding 2.5 ml of Trypsin-EDTA (Hyclone, USA) and incubating them for 2-3 minutes at 37°C. After this, 2.5 ml of complete DMEM was added to neutralise the trypsin. The cells were then pelleted by centrifuging the mixture at 200×g for 3 minutes (Minor MSE, UK). The supernatant was discarded and the pellet resuspended in 10 ml of complete DMEM and split between two flasks. A further 5 ml of complete DMEM was added to each flask to obtain a final volume of 10 ml for each culture.

For experiments, cell counts were determined by mixing a small volume of the harvested cells with Trypan Blue (Hyclone, USA) at a 1:1 ratio, and loading 10 µl of this into a counting chamber of a disposable haemocytometer slide (Immune Systems, UK). Because Trypan Blue permeates the cell membranes of dead cells, they stain blue. The "white", living

cells, within 5 large squares were counted and the concentration of cells calculated according to the following formula:

$$cells\ per\ ml = \frac{Total\ number\ of\ cells}{Number\ of\ squares\ counted} \times 10^4 \times Dilution\ factor$$

The required number of cells for a particular experiment would be obtained by taking the appropriate volume of this cell suspension for further use.

Cultured HEK 293 cells were periodically frozen to maintain a stock of low passage cells. This was done by detaching and harvesting the cells, as described above. For a T-75 flask, the cells were resuspended the in 16 ml of freeze medium (Appendix 7.3.1), and made into 2 ml aliquots which were stored in a -80°C freezer.

2.2 Cloning

2.2.1 sgRNA expression plasmids

Five different sgRNA guide sequences targeting the exon 2 region of the human *KRAS* locus found on chromosome 12p12.1 (ENSG00000133703) were designed using the CRISPR design tool provided by the Feng Zhang group (http://crispr.mit.edu/). The tool identified all the potential sgRNA target sites within a 250 bp region of the G>T mutation site (125 bp in each direction) and generated a score that ranked the potential guide sequences according to various target specificity parameters (Hsu et al., 2013). Of the five sgRNAs chosen, three of these were designed to direct the Cas9 endonuclease to a region that is upstream of the G>T transversion in exon 2, to induce 5′ cleavages (sgRNA 224-226), and the remaining two to induce 3′ cleavages (sgRNA 227-228) that are downstream to the G>T modification. Each guide sequence was cloned into the pcDNA.H1.sgRNA vector (Appendix 7.5.1) for sgRNA expression.

2.2.1.1 Preparation of the expression plasmid

The pcDNA.H1.sgRNA expression plasmid was shipped to the laboratory from the Scripps Research Institute on pieces of dried filter paper. To reconstitute the plasmid in solution, the filter paper was soaked in 20 µl of elution buffer from the QIAprep Spin Miniprep kit (Qiagen, Germany) containing 1.25 M NaCl, 50 mM Tris-Cl (pH 8.5) and 15% isopropanol (v/v), for 1 hour in a 1.5 ml microcentrifuge tube.

For the bulk preparation of this plasmid, 2 µl of this DNA suspension was used to transform 100 µl of competent *E. coli* Stbl3 cells (Appendix 7.4.1) that were thawed on ice. This mixture was allowed to rest on ice for 30 minutes before it was heat shocked at 42°C for 90 seconds, exactly. The cells were allowed to recover on ice for 5 minutes. An additional outgrowth step was added to aid the survival of transformed bacteria under ampicillin selection conditions. 1 ml of LB broth (Appendix 7.3.3) was added to the transformation mixture and incubated for an hour at 37°C. The cells were then collected by centrifugation at 720×g for 3 minutes (Eppendorf Centrifuge 5402, Germany). 800 µl of the supernatant was discarded, and the remaining volume was used to resuspend the pellet of bacterial cells. The cells were plated onto an ampicillin-LB agar plate (100 µg/ml final ampicillin concentration) (Appendix 7.3.5) and incubated for 14-16 hours in a 37°C incubator. Successfully transformed cells formed colonies that appeared on the antibiotic supplemented plate. A single colony was picked and inoculated into 100 ml of ampicillin-LB broth (100 µg/ml final ampicillin concentration) (Appendix 7.3.6). This bacterial culture was placed in a shaking incubator, set at 180 rpm, for 14-16 hours at 37°C, before the plasmid was extracted using the QIAprep Plasmid Midi kit (Qiagen, Germany), as directed by the manufacturer's protocol (Appendix 7.4.3). The extracted plasmid was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA) and diluted to 1 µg/µl stocks.

2.2.1.2 Guide sequence ligation

1 µg of the pcDNA.H1.sgRNA plasmid was digested with 10 U of *BsmB*I (10 U/µl stock concentration) (Thermo Scientific, USA) in a 20 µl reaction containing 10X Tango Buffer

(which contains 33 mM Tris-acetate (pH7.9 at 37°C), 10 mM Mg-acetate, 66 mM K-acetate and 0.1 mg/ml BSA) and deionised water (ddH_2O). This reaction was incubated at 37°C for 1 hour. After this, the entire reaction was mixed with 4 µl of 6X DNA Loading Dye (Thermo Scientific, USA) and loaded into a 50 ml 1% (w/v) agarose gel stained with 2 µl of ethidium bromide (10mg/ml stock concentration). 5 µl of Kapa Universal Ladder (Kapa Biosystems, USA) was used as a size marker. Electrophoresis was performed in 1X Tris-acetate EDTA (TAE) buffer (Appendix 7.3.7), at 100 V for 1 hour using a Bio-Rad Power Pac Basic (Bio-Rad, USA). The gel was visualised over a UV trans-illuminator, and the linearised plasmid was excised and extracted from the gel using the GeneJet Gel Extraction Kit (Thermo Scientific, Lithuania) according to the protocol provided by the manufacturer (Appendix 7.4.5). The eluted product was quantified using the Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA).

For the each of the five CRISPR/Cas9 target sequences, a pair of complementary oligonucleotides were synthesised by Integrated DNA Technologies (IDT, USA). Each pair of oligonucleotides were composed of a 20 nucleotide guide sequence for their target and appropriate 5′ overhangs that allowed for sticky-end ligation into the *BsmB*I digested pcDNA.H1.sgRNA expression plasmid (Figure 4A). See Table 1 for the oligonucleotide sequences used.

Each oligonucleotide pair was annealed and phosphorylated in individual reactions containing 1 µl dATP (10 mM stock concentration), 1 µl of each oligonucleotide (100 µM stock concentration), 1 µl 10X T4 Ligation Buffer (New England Biolabs, Germany) containing 500 mM Tris-HCl, 100 mM $MgCl_2$, 10 mM ATP and 100 mM DTT (pH7.5 at 25°C), 0.5 µl T4 polynucleotide kinase (New England Biolabs, Germany) and 6.5 µl of ddH_2O. This was incubated in a Bio-Rad MJ Personal Thermal Cycler (Bio-Rad, USA) that was programmed to incubate the reaction at 37°C for 30 minutes, and then denature the oligonucleotides at 95°C for 5 minutes. After this, the thermal cycler was switched off and allowed to cool to room temperature for the complementary oligonucleotides to anneal. Each guide sequence duplex was then ligated into the linearised pcDNA.H1.sgRNA plasmid by mixing 1 µl of a 1:250 dilution of the annealed oligonucleotides with 50 ng of the plasmid

backbone in a 10 µl reaction containing 2.5 U T4 Ligase (Thermo Scientific, USA), ddH_2O and 10X Ligation Buffer (Thermo Scientific, USA) which contains 400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT and 5 mM ATP (pH 7.8 at 25°C). A control reaction, where the annealed oligonucleotide was substituted with ddH_2O, was also included. The reactions were incubated at 22.5°C for 1 hour in the thermal cycler (Bio-Rad, USA). 2 µl of each ligation reaction was used to transform 100 µl of competent *E. coli* Stbl 3 cells as described in section 2.2.1.1. Three colonies from each ampicillin-LB agar plate were picked, and each inoculated into 10 ml of ampicillin-LB broth. This was incubated for 14-16 hours at 37°C in a shaking incubator set at 180 rpm. Following this, the plasmid clones were extracted using the QIAprep Spin Miniprep kit according to the provided manufacturer's protocol (Appendix 7.4.2). The isolated DNA was quantified using the Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA) and each clone was sent for Sanger sequencing at Inqaba Biotech. For this purpose, primer U-90a, which annealed immediately upstream of the H1 promoter was used to read the guide sequences (Table 1) (Figure 4B). Clones that showed the successful ligation of the correct guide sequence were prepared in bulk as described in section 2.2.1.1.

Table 1: List of primer sequences used for sgRNA cloning

Primer Name	Usage	Sequence (5' - 3')
sgRNA F1 224 F	sgRNA guide cloning	GATCCGAGTTTGTATTAAAAGGTAC
sgRNA F1 224 R	sgRNA guide cloning	AAACGTACCTTTTAATACAAACTCG
sgRNA F2 225 F	sgRNA guide cloning	GATCCATGTGTGACATGTTCTAATA
sgRNA F2 225 R	sgRNA guide cloning	AAACTATTAGAACATGTCACACATG
sgRNA F3 226 F	sgRNA guide cloning	GATCCGTTTGTATTAAAAGGTACTGG
sgRNA F3 226 R	sgRNA guide cloning	AAACCCAGTACCTTTTAATACAAACG
sgRNA R1 227 F	sgRNA guide cloning	GATCCGTGAATTAGCTGTATCGTCA
sgRNA R1 227 R	sgRNA guide cloning	AAACTGACGATACAGCTAATTCACG
sgRNA R2 228 F	sgRNA guide cloning	GATCCACAAGATTTACCTCTATTGT
sgRNA R2 228 R	sgRNA guide cloning	AAACACAATAGAGGTAAATCTTGTG
U-90a	sgRNA sequencing	TCCCCGAAAAGTGCCACCTGACG

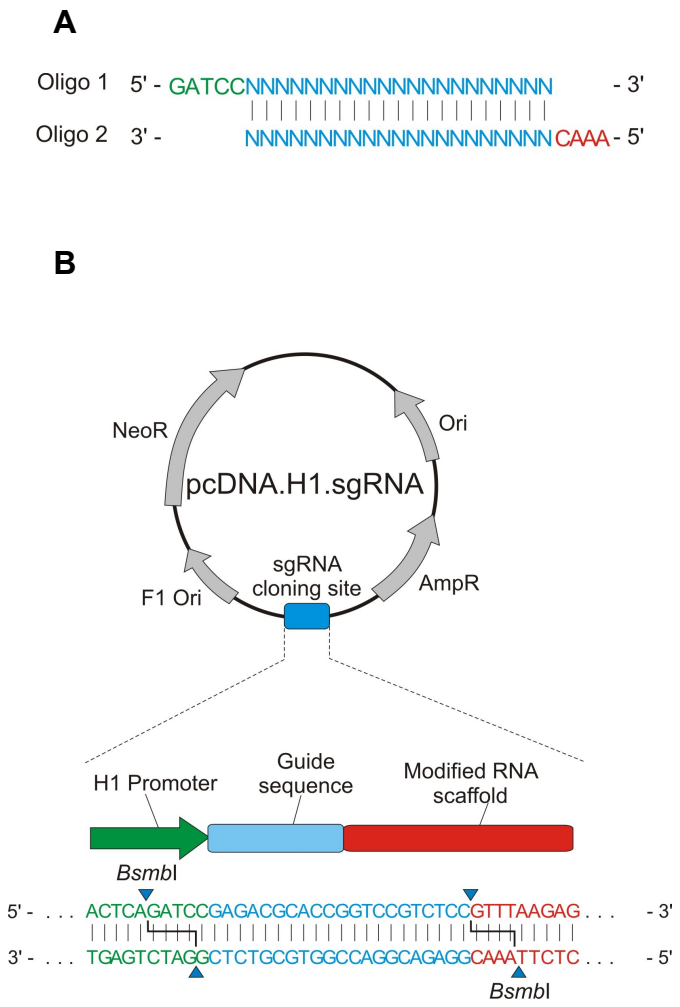


Figure 4: The cloning strategy of the guide sequence into the pcDNA.H1.sgRNA expression plasmid. (**A**) The design of a pair of complementary oligonucleotides that have a 20 nucleotide guide sequence (blue) and the respective 5′ overhangs (green and red). (**B**) A schematic of the pcDNA.H1.sgRNA expression plasmid. The expression plasmid is digested by the type II restriction enzyme *Bsmb*I, to produce overhangs that facilitate the sticky-end ligation of the annealed oligonucleotide pairs.

2.2.2 dsDNA donor cassette

2.2.2.1 Amplification of the dsDNA donor components

The dsDNA donor construct is composed of various DNA parts that were each isolated and amplified using polymerase chain reactions (PCRs). The construction of the dsDNA donor was done by assembling these DNA parts in two sequential overlap-extension PCRs. In this way, the large dsDNA donor construct was assembled very efficiently, without the need for

multiple cloning steps, and produced a seamless product. The primer sequences used are shown in Table 2 and the completed DNA fragment sequences are shown in Appendix 7.2.1.

Enhanced Green Fluorescent Protein (eGFP) fragment (875 bp)

The eGFP gene sequence was amplified from the pCI-NeoR-eGFP plasmid (Appendx 7.5.2). Three primers were designed to amplify the eGFP sequence and add overlapping sequences that would aid the subsequent assembly of the dsDNA donor. On the 5′ end of the eGFP fragment was an extension that encoded for half of the 2A self-cleaving peptide sequence. The 3′ end consisted of a synthetic PolyA tail and a complementary linker sequence to the 5′ *KRAS* homology arm. These features were added in two consecutive rounds of PCR (Figure 5). The first round of PCR contained 10 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 1 µl of each eGFP F and eGFP R primer (10 µM stock concentration), 1 µl of the pCI-NeoR-eGFP plasmid (1 ng/µl stock concentration) and 7 µl ddH_2O. Thermal cycling conditions were set so that the initial denaturation of 98°C lasted 10 seconds, followed by 25 cycles of denaturation at 98°C for 1 second, annealing at 64.7°C for 5 seconds, extension at 72°C for 15 seconds, and a final extension at 72°C for 1 minute. 2 µl of this reaction was mixed with 8 µl ddH_2O and 2 µl 6X DNA Loading Dye (Thermo Scientific, USA). The DNA products were resolved on an ethidium bromide stained, 1% (w/v) agarose gel. Electrophoresis was carried out in 1X TAE, at 75 V for 1 hour using a Bio-Rad Power Pac Basic (Bio-Rad, USA). 5 µl of Kapa Universal Ladder (Kapa Biosystems, USA) was used as a size marker. The PCR product was then visualised using a Bio-Rad Gel Doc EZ imaging system (Bio-Rad, USA). The second round PCR was prepared in the same way as the first, except the primer set combination was changed to 1 µl of 5′HA F and 1 µl of eGFP R (10 µM stock concentration). The template used in this reaction was 1 µl of a 1:500 dilution of the round 1 PCR. The cycling conditions remained the same, except the annealing temperature was increased to 70.4°C. Visualisation of this product was carried out in the same way as described for the round 1 reaction. The completed round 2 product was column purified with the GeneJet PCR Purification Kit (Thermo Scientific, Lithuania) according to the provided protocol (Appendix 7.4.6). The eluted DNA was quantified using the Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA).

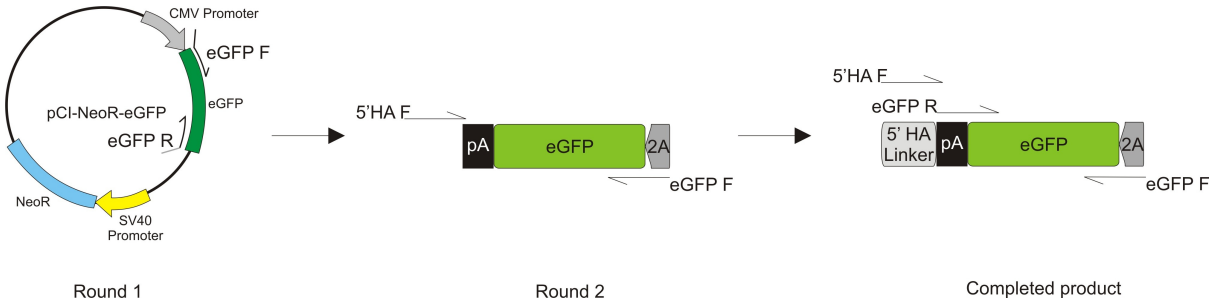


Figure 5: A schematic outlining the production of the eGFP fragment. eGFP was amplified from pCI-NeoR-eGFP. A poly A tail and overlapping linker sequences were added in two subsequent rounds of PCR. The linkers served to fuse this fragment with the upstream 5′HA and the downstream SV40-NeoR fragment.

SV40-NeoR fragment (1 332 bp)

Similarly, the SV40-NeoR gene sequence was amplified from the pCI-NeoR-eGFP plasmid (Figure 6). The forward primer was designed to prime to the 5′ end of the SV40 promoter upstream of the NeoR gene. This primer also provided an overlapping sequence into the *KRAS* reconstitution sequence (KRS) fragment (see section 2.2.2.1 - *KRAS reconstitution sequence (KRS) fragment*). The reverse primer bound to the 3′ end of the NeoR gene sequence and contained the second half of the 2A self-cleaving peptide sequence. The overlapping sequences of this 2A peptide was designed to serve as a linker between the SV40-NeoR and eGFP gene fragments, and allow for the expression of eGFP and NeoR from the single SV40 promoter. The amplification reaction was set up as described in section 2.2.2.1 - *eGFP fragment*, but contained the SV40-NeoR F and SV40-NeoR R primers (10 µM stock concentration) instead. PCR cycling conditions were altered slightly and set so that the initial denaturation at 98°C lasted 10 seconds, followed by 25 cycles of denaturation at 98°C for 1 second, annealing and extension at 72°C for 20 seconds, and a final extension at 72°C for 1 minute. The DNA was then visualised on an ethidium bromide stained agarose gel, column purified and quantified as described in section 2.2.2.1 - *eGFP fragment*.

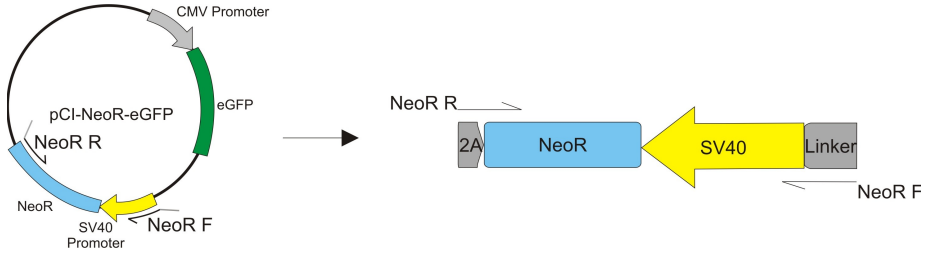


Figure 6: A schematic outlining the production of the SV40-NeoR fragment. The SV40 promoter and the NeoR expression cassette was amplified from pCI-NeoR-eGFP. Terminal linker sequences were added by PCR for its fusion with the eGFP and KRS fragments.

KRAS reconstitution sequence (KRS) fragment (131 bp)

Two consecutive rounds of PCR were used to generate the KRS fragment (Figure 7). The first reaction involved the annealing and extension of two complementary oligonucleotides which contained the KRS. This was done by setting up a PCR with 10 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 1 µl of each KRAS F and KRAS R1 primer (10 µM stock concentration) and 8 µl ddH_2O. Thermal cycling conditions were set so that the initial denaturation of 98°C lasted 10 seconds, followed by 25 cycles of denaturation at 98°C for 1 second, annealing at 54.7°C for 5 seconds, extension at 72°C for 3 seconds, and a final extension at 72°C for 1 minute. The product was subject to electrophoresis and visualisation as previously described in section 2.2.2.1 - *eGFP fragment*. The second round PCR was carried out in a similar manner as round 1. However, primer KRAS R2 (10 µM stock concentration) was used in place of KRAS R1, and 1 µl of a 1:500 dilution of the round 1 reaction was used as the template. Cycling conditions were slightly altered, with the annealing temperature increased to 56°C. The completed fragment consisted of the KRS and was flanked by overlapping linker sequences that were complementary to the SV40-NeoR fragment and the 3′ *KRAS* homology arm. The product was visualised, column purified and quantified as previously described (see section 2.2.2.1 - *eGFP fragment*).

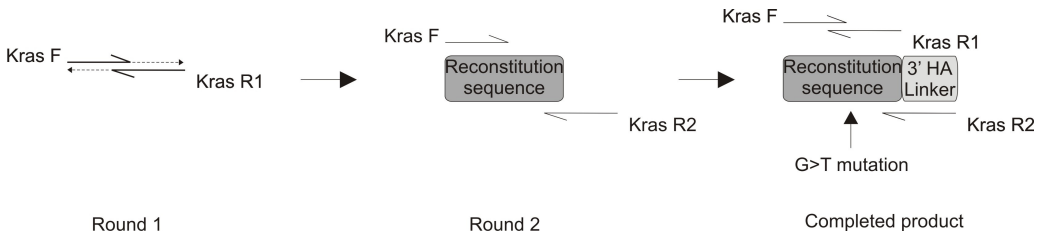


Figure 7: A schematic outlining the production of the KRS fragment. The KRS fragment was generated in two subsequent rounds of PCR. The first involved the primer annealing and extension of primer KRAS F and R1. A 3′ linker sequence was added in the second reaction for its fusion with the downstream 3′ HA.

KRAS homology arms (5′ HA - 437 bp, 3′ HA 432 bp)

Homology arms (HAs) that flanked the 5′ and 3′ DSBs were obtained by PCR from the genomic DNA (gDNA) of HEK 293 cells (Figure 8). HEK 293 cells were harvested and 1×10^6 cells (as determined by cell counting described in section 2.1) were collected for gDNA extraction with the QIAamp DNA Mini Kit (Qiagen, Germany) by following the recommended protocol (Appendix 7.4.4). The gDNA was quantified using the Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA) and diluted to a concentration of 20 ng/µl. The 5′ HA was amplified with 10 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 1 µl of each 5′HA F2 and 3′HA R2 primer (10 µM stock concentration), 2.5 µl of gDNA (20 ng/µl stock concentration) and 5.5 µl ddH_2O. Cycling conditions were set so that the initial denaturation of 98°C lasted 10 seconds, followed by 35 cycles of denaturation at 98°C for 1 second, annealing at 62.7°C for 5 seconds, extension at 72°C for 7 seconds, and a final extension at 72°C for 1 minute. The 3′ HA was amplified in the same way, using the HA3′ F2 and HA3′ R2 primers (10 µM stock concentration) instead. The HAs were then visualised on an agarose gel, column purified and quantified and (see section 2.2.2.1 - *eGFP fragment*).

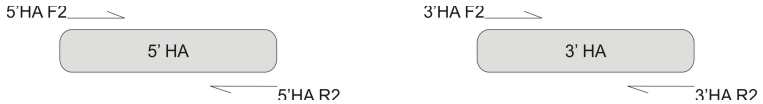


Figure 8: A schematic of the 5′ and 3′ HAs. Each HA was amplified, with the illustrated primer pairs, from gDNA extracted from HEK 293 cells.

2.2.2.2 dsDNA donor assembly

Three-way fusion of insertion DNA cassette (2 349 bp)

The column purified fragments of the insertion cassette (eGFP fragment, SV40-NeoR fragment and KRS fragment) were pooled into a single reaction for the three-way fusion of the gene fragments by overlap extension PCR. This involved the complementary binding and extension of corresponding and overlapping linker sequences, that were added to the terminals of each fragment, and the subsequent fusion of all the gene fragments in the designed configuration, into one seamless product (Figure 9). This reaction contained 10 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 1 µl of the 5′HA F and the KRAS R2 primers (10 µM stock concentration), equimolar amounts (28 fmol) of the eGFP fragment (16.5 ng), the SV40-NeoR fragment (25 ng) and the KRS fragment (3.5 ng) and ddH_2O to make up a final reaction volume of 20 µl. The amount of each fragment needed in the reaction was calculated by multiplying the molar mass of each fragment (length of fragment (bp) × 660 g/mol) by the molar concentration (28 fmol) needed. In order to ensure that the small *KRAS* reconstitution fragment (191 bp) was successfully incorporated into the construct, a control reaction was also setup. In this reaction, the KRS fragment was excluded from the reaction and the SV40-NeoR F primer (10 µM stock concentration) was used in place of the KRAS R2 primer. This produced a truncated fusion product that was compared to the full cassette. The reaction was subjected to the following cycle conditions: initial denaturation at 98°C for 10 seconds, followed by 35 cycles of denaturation at 98°C for 1 second, annealing at 56°C for 5 seconds, extension at 72°C for 30 seconds, and a final extension at 72°C for 1 minute. The PCRs were visualised on an agarose gel to ensure successful assembly, after which the products were column purified and quantified (see section 2.2.2.1 - *eGFP fragment*).

Homology arm extension (3 111 bp)

The 5′ and 3′ homology arms were finally added to the terminals of the insertion DNA cassette through a subsequent overlap-extension PCR (Figure 9). A reaction composed of

15 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 1 µl of the 5′HA F2 and the 3′HA R2 primers (10 µM stock concentration), equimolar amounts of the insertion DNA cassette (30 ng), 5′ HA (5.5 ng) and 3′ HA (5.4 ng) and ddH_2O to make up a final volume of 30 µl. The reaction was subjected to the following cycle conditions: initial denaturation at 98°C for 10 seconds, followed by 35 cycles of denaturation at 98°C for 1 second, annealing at 60°C for 5 seconds, extension at 72°C for 45 seconds, and a final extension at 72°C for 1 minute. The full dsDNA donor was gel purified as described in section 2.2.1.2.

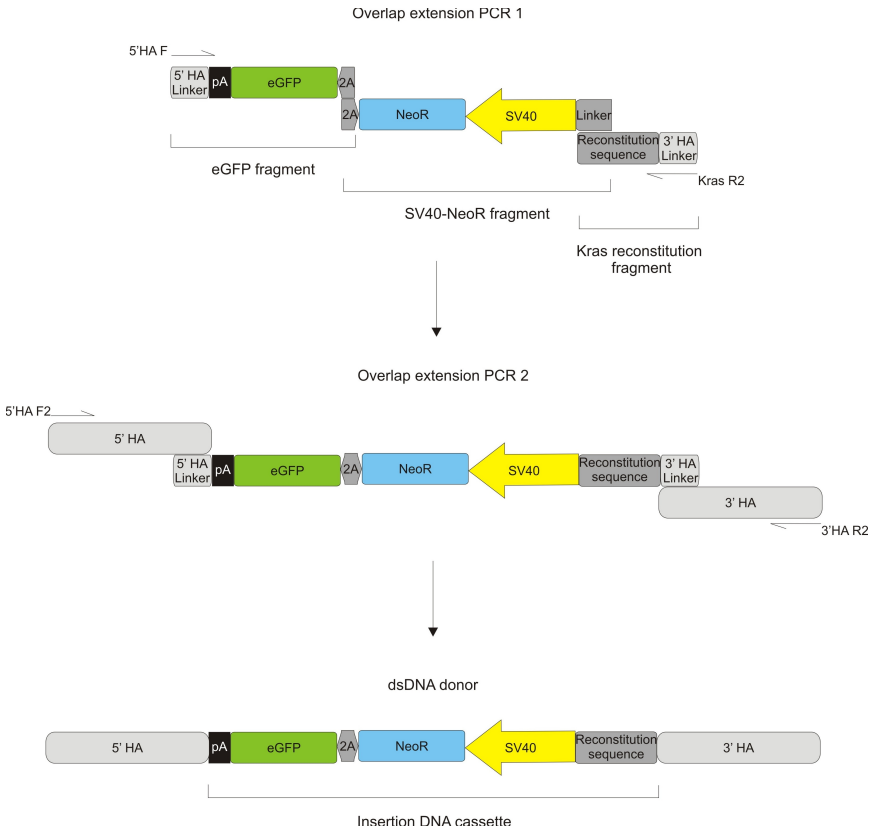


Figure 9: The overlap extension PCR assembly strategy for the dsDNA donor construct. Each of the fragments, illustrated previously, were fused together by terminal complementary linker sequences, in two sequential reactions. In the first reaction, the insertion DNA cassette, which comprises of the expression construct and the *KRAS* reconstitution sequence, was assembled and amplified. The 5′ and 3′ HAs were subsequently added to the ends of this, by a second overlap extension PCR.

2.2.2.3 Blunt-end ligation of the dsDNA donor cassette into pJet 1.2/blunt Cloning Vector

Ligation

The completed dsDNA donor cassette was ligated into the pJet1.2/blunt Cloning Vector (Appendix 7.5.3) by blunt-end ligation, using the CloneJet PCR Cloning Kit (Thermo Scientific, USA). 150 ng of the dsDNA donor was ligated into 50 ng of the plasmid vector, in a reaction containing 10 µl of 2X Reaction Buffer (which contains 80 mM Tris-HCl, 20 mM $MgCl_2$, 20 mM DTT and 1 mM ATP (pH 7.8 at 25°C), 2.5 U of T4 Ligase and ddH_2O for a final reaction volume of 20 µl. The reaction was incubated at 22.5°C for 1 hour in the thermal cycler (Bio-Rad, USA). 15 µl of the ligation reaction was used to transform and plate 100 µl of competent *E. coli* Stbl3 cells as described in section 2.2.1.1.

Colony PCR

Four colonies from the ampicillin-LB agar plate were picked, and inoculated into 10 ml of ampicillin-LB broth. Before these were cultured, each of them were screened for the newly ligated construct by colony PCR. This involved dipping the picked colony into a PCR mix prior to inoculating into the broth. Each colony PCR reaction was prepared with 1µl of the pJet 1.2 forward sequencing primer and 1 µl of the pJet 1.2 reverse sequencing primer from the CloneJet PCR Cloning Kit (Thermo Scientific, USA) (10 µM stock concentration) (which each primed to the outside of the cloning site of the plasmid vector) (Table 2), 5 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA) and 3 µl of ddH_2O. Thermal cycling conditions were set as follows: initial denaturation at 98°C for 10 seconds, followed by 30 cycles of denaturation at 98°C for 1 second, annealing at 60°C for 5 seconds, extension at 72°C for 45 seconds, and a final extension at 72°C for 1 minute. The PCRs were mixed with 2 µl 6X DNA Loading Dye (Thermo Scientific, USA) and subject to gel electrophoresis (see section 2.2.2.1 - *eGFP fragment*). Colonies that were positive for the dsDNA donor cassette product were cultured for 14-16 hours at 37°C in a shaking incubator set at 180 rpm. Following this, the plasmid DNA was extracted from these cultures and quantified as described in section 2.2.1.2.

Restriction digestion

To further confirm that the plasmid clones were positive for the dsDNA donor, a restriction digest was carried out. 500 ng of each plasmid clone was digested with 10 U of *Bgl*II in 2 µl of Buffer O (Thermo Scientific, USA), which contains 50 mM Tris-HCl (pH 7.5 at 37°C), 10 mM $MgCl_2$, 100 mM NaCl and 0.1 mg/ml BSA, and 16 µl of ddH_2O. This reaction was incubated for an hour at 37°C, after which the digested DNA was mixed with 4 µl 6X DNA Loading Dye (Thermo Scientific, USA) and resolved on an agarose gel (see section 2.2.2.1 - *eGFP fragment*). Plasmids containing the dsDNA donor (pJet-dsDNA donor) were digested such that the full dsDNA donor cassette was cut out of the plasmid. Positive clones were confirmed by Sanger sequencing, and prepared in bulk as described in section 2.2.1.1.

The functionality of the expression cassette was tested by transfecting 2 µg of the plasmid into HEK 293 cells (see section 2.3.1) and monitoring the expression of eGFP (excitation at 488 nm and emission at 509 nm) under an Olympus IX71 fluorescence microscope (Olympus, USA), 48 hours afterwards. pCI-NeoR-eGFP was used as a transfection control.

Table 2: List of primer sequences used for the production of the dsDNA donor

Primer Name	Usage	Sequence (5' - 3')
eGFP F	eGFP fragment R1	TCAAGTTGGCGGGAGACGTCGAGTCCAACCCTGGGCCCATGGTGAGCAAG GGCGAGGAGC
eGFP R	eGFP fragment R1	CACACAAAAAACCAACACACAGATGTAATGAAAATAAAGATATTTTATTTCTAG AGTATACGGACCGTTACTTG
5'HA F	eGFP fragment R2, Fusion PCR R1	GTTTGTATTAAAAGGTACTGGTGGAGTATTTGATAGTGTATTAACCTTATCACA CAAAAAACCAACACACAGATG
NeoR F	SV40-NeoR fragment	TAATGAAAATGTGACTATATTAGAACATGTCACACGCGCAGCACCATGGCCT GAAATAAC
NeoR R	SV40-NeoR fragment	ACGTCTCCCGCCAACTTGAGAAGGTCAAAATTCAAAGTCTGTTTCACGAAG AACTCGTCAAGAAGGCGATAGAAG
KRAS F	KRS fragment R1	GTGTGACATGTTCTAATATAGTCACATTTTCATTATTTTTATTATAAGGCCTGCT GAAAATGACTGAATATAAACTTGTGGTAGTTGGAG
KRAS R1	KRS fragment R1	GTCTACAAAATGATTCTGAATTAGCTGTATCGTCAAGGCACTCTTGCCTACGC CAACAGCTCCAACTACCACAAGTTT
KRAS R2	KRS fragment R1, Fusion PCR R1	GCATATTAAAACAAGATTTACCTCTATTGTTGGATCATATTCGTCTACAAAATG ATTCTG
5'HA F2	5' HA, Fusion R2	TAGCTGTTGCATATTGACTTCTAACACTTAGAGG
5'HA R2	5' HA	ATAAGGTTAATACACTATCAAATACTCCACCAGTACC
3'HA F2	3' HA	AATCATTTTGTGGACGAATATGATCCCACAAT
3'HA R2	3' HA, Fusion PCR R2	CCATCAAACAATTATATTTCACTAGTACAATTAAATCTAACCTTT
pJet 1.2 F	dsDNA donor colony PCR	CGACTCACTATAGGGAGAGCGGC
pJet 1.2 R	dsDNA donor colony PCR	AAGAACATCGATTTTCCATGGCAG

2.3 Surveyor Assay

The efficiency of each sgRNA to mediate a DNA DSB was assessed using the Surveyor Assay to measure the frequency of indel formation as a result of NHEJ-mediated DNA repair (Mussolino et al., 2011, Ran et al., 2013b). Briefly, the hybridisation of PCR amplicons from modified and unmodified *KRAS* alleles formed heteropduplexes that have base-pair mismatches at the DSB site as a result of indels that have formed from NHEJ repair. These heteroduplexes were digested by the mismatch-sensitive T7 endonuclease I (T7EI), and the ratio of digested products to undigested products was used to calculate the frequency of indel formation (Figure 11B).

2.3.1 Transfection

24 hours prior to transfection, 1.5 × 10^6 HEK 293 cells were harvested (see section 2.1) and seeded into a 6-well plate (250 000 cells/well), with 2 ml of complete DMEM per well. The total number of cells harvested was adjusted according to the number of transfection conditions being carried out. The next day, before commencing with the transfection procedure, the cell culture medium was renewed with 2 ml of fresh complete DMEM. All the plasmids used for transfection were prepared in bulk (see section 2.2.1.1) and normalised to a concentration of 1 µg/µl. In each transfection, 2 µg of the sgRNA expression plasmid (pcDNA-sgRNA KRAS 224 - 228) was mixed with 2 µg of the Cas9 expression plasmid (pX330 - Appendix 7.5.4) and 100 µl of serum-free DMEM lacking antibiotics in a 1.5 ml microcentrifuge tube. To this, 12 µl of branched polyethylenimine (PEI) (Sigma Aldrich, USA) (1 µg/ml stock concentration) was added, and each tube was immediately pulse-vortexed at a medium intensity 15 times, for 1 second each. This mixture was then incubated at room temperature (~25°C) for exactly 10 minutes. Following this, 600 µl of complete DMEM was added, and this transfection mixture was added, in a drop-wise manner, to a well containing HEK 293 cells. The 6-well plate was then gently rocked before it was placed back into the cell culture incubator. After 3 hours, an additional 2 ml of complete DMEM was added to each well of transfected cells. Negative controls for this assay included transfections with the pX330 plasmid only and a mock transfection with ddH_2O. In order to monitor transfection efficiency, a transfection with pCI-NeoR-eGFP was also carried out.

2.3.2 gDNA extraction

The transfection efficiency was assessed by visualising the expression of eGFP (at 488 nm) in the cells transfected with pCI-NeoR-eGFP. These cells were monitored 48 hours post-transfection using an Olympus IX71 fluorescence microscope (Olympus, USA). The transfected cells in each well were then harvested by washing the cells with 2 ml of 1X PBS (Appendix 7.3.10), followed by trypsinisation with 1 ml Trypsin-EDTA (Hyclone, USA). The cells were incubated for 2-3 minutes at 37°C, before 500 µl of complete DMEM was added to neutralise the trypsin. The cell suspension was then transferred to a 1.5 ml microcentrifuge

tube and the cells were collected by centrifugation at 260×g for 5 minutes (Eppendorf Centrifuge 5402, Germany). The supernatant was then discarded, and the pellet was resuspended in 1 ml of 1X PBS and centrifuged again at 260×g for 5 minutes. This wash step was repeated. The final cell pellet was resuspended in 200 µl of 1X PBS and the gDNA was extracted from these cells using the QIAamp DNA Mini Kit (Qiagen, Germany) and the manufacturer's recommended protocol (Appendix 7.4.4). The extracted gDNA samples were quantified using the Nanodrop ND-1000 (Nanodrop Technologies, USA) and each normalised to a working concentration of 20 ng/µl.

2.3.3 PCR and heteroduplex formation

The CRISPR/Cas9 cleavage regions of the *KRAS* gene were amplified using PCR. Two sets of primers were designed to amplify the two cleavage regions. The first set amplified the 5′ cleavage region (5′ Surveyor F and 5′ Surveyor R) (361 bp product), whereas the second set amplified the 3′ cleavage region (3′ Surveyor F and 3′ Surveyor R) (346 bp product) of the *KRAS* gene (see Table 3 for primer sequences). In each PCR, 1 µl of the appropriate forward and reverse Surveyor primer set (10 µM stock concentration) was added to 5 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 2.5 µl of gDNA and 0.5 µl of ddH_2O. Thermal cycling conditions were set as follows: initial denaturation at 98°C for 10 seconds, followed by 35 cycles of denaturation at 98°C for 1 second, annealing at 69.8°C for the 5′ cleavage primers) and 65.8°C for the 3′ cleavage primers, for 5 seconds, extension at 72°C for 6 seconds and a final extension at 72°C for 1 minute. For each transfection condition, the PCR was setup and carried out in triplicate. Each triplicate reaction set was then pooled together and 2 µl of this was visualised on an agarose gel (see section 2.2.2.1- *eGFP fragment*). The PCR products were then column purified and quantified as described in section 2.2.2.1 - eGFP fragment. Each of the eluted PCR products were then normalised to a concentration of 30 ng/µl.

For each PCR amplicon, 600 ng of DNA was added to 3 µl of 10X DreamTaq Buffer (Thermo Scientific, USA), which contains 750 mM Tris-HCl, 200 mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween 20 and 20 mM $MgCl_2$, and 7 µl of ddH_2O. The DNA duplexes were denatured by

heating to 95°C for 10 minutes in the thermal cycler (Bio-Rad, USA). After this, the machine was switched off and allowed to cool to room temperature, with the lid closed, to allow for the DNA to rehybridise and form heteroduplexes.

2.3.4 Heteropduplex digestion and polyacrylamide gel electrophoresis (PAGE)

150 ng of each heteroduplex product was digested by the mismatch-sensitive T7E1 (New England Biolabs, Germany). This was done in a reaction that contained 3 U of T7EI, 1 µl of NEB Buffer 2 (New England Biolabs, Germany), which was made up of 500 mM NaCl, 100 mM Tris-HCl, 100 mM $MgCl_2$ and 10 mM DTT (pH 7.9 at 25°C), and 1.2 µl of ddH_2O. The digestion reaction was incubated at 37°C for 15 minutes.

In the meantime, an 8% polyacrylamide gel was prepared. The 30% (w/v) polyacrylamide solution was prepared by dissolving 29 g of acrylamide (Bio-Rad, USA) and 1 g of bisacrylamide (Promega, USA) (29:1) in 1X Tris-borate EDTA (TBE) buffer (Appendix 7.3.8) to a final volume of 100 ml. The polyacrylamide gel was made by diluting 3.2 ml of the 30% polyacrylamide solution with 8.8 ml of 1X TBE, and catalysing its polymerisation by adding 200 µl of 10% ammonium persulphate (APS) (SaarChem, South Africa) and 10 µl of tetramethylethylenediamine (TEMED) (Sigma Aldrich, USA). This mixture was solidified between two glass plates, separated by a 15 well, 1.5 mm spacer comb (Bio-Rad Mini Protean Tetra system, USA). Before being loaded, the gel cast cassette was submerged in 1X TBE. The comb was then carefully removed and the wells were flushed out with a 1 ml pipette. Each T7EI digestion reaction was then mixed with 2 µl of 6X DNA Loading Dye and loaded into the gel. 2 µl of the GeneRuler 50 bp DNA Ladder (Thermo Scientific, USA) was used as a size marker. The gel was then subject to electrophoresis at 55 V (E-C Apparatus Corporation, USA) for 2.5 hours at 4°C. The polyacrylamide gel was then carefully removed from the gel cast and placed in 100 ml TBE buffer containing 10 µl of SYBR Green I Nucleic Acid Gel Stain (Invitrogen, USA). The gel was stained for 15 minutes while being gently rocked (40 rpm). The gel was then imaged using the Bio-Rad Gel doc EZ Imager system (Bio-Rad, USA).

2.3.5 Analysis

The gels of the Surveyor Assay were analysed using ImageLab (Bio-Rad, USA). A standard curve for each gel was created by the software, where the banding intensity of each band of the DNA ladder was correlated to its prescribed mass, as indicated by the manufacturer. This was used to quantify each of the bands, on the basis of band intensity, that resulted from the T7EI digestion of the heteroduplexes. The efficiency of target cleavage was then calculated using the following formula (Ran et al., 2013b):

$$indel\ (\%) = 100\times\left(1 - \sqrt{1 - fcut}\right)$$

$$where: fcut = \frac{b + c}{a + b + c}$$

$$and\ a = intensity\ of\ undigested\ PCR\ product$$

$$b\ and\ c = intensity\ of\ cleaved\ products$$

The two best sgRNAs (one that cleaved in the 5′ cleavage region and one that cleaved in the 3′ cleavage region) which showed the highest cleavage efficiency were chosen to be used in the subsequent experiments.

Table 3: List of primer sequences used for the testing of the sgRNAs

Primer Name	Usage	Sequence (5' - 3')
5' Surveyor F	Surveyor assay, micro-deletion detection, RFLP	CGCAGAACAGCAGTCTGGC
5' Surveyor R	Surveyor assay	CTACGCCACCAGCTCCAAC
3' Surveyor F	Surveyor assay	AAGGCCTGCTGAAAATGACTG
3' Surveyor R	Surveyor assay, micro-deletion detection, RFLP	GCACAGAGAGTGAACATCATGG

2.4 Micro-deletion detection

To detect the repair of two simultaneous DSBs (one 5′ cleavage and one 3′ cleavage), a PCR using primers that spanned across the repair junction was carried out (650 bp product). HEK 293 cells were transfected in the same way as described in section 2.3.1. However, the transfection for the dual cleavage contained 2 µg of the 5′ cleavage pcDNA-sgRNA KRAS plasmid, 2 µg of the 3′ cleavage pcDNA-sgRNA KRAS plasmid and 2 µg of pX330. Control transfections contained 2 µg of either the 5′ or 3′ sgRNA expression plasmid and 2 µg of pX330. A mock transfection was also included. After 48 hours, the transfected cells were harvested and the gDNA extracted (see section 2.3.1 and 2.3.2). The gDNA was normalised to 20 µg/µl, of which 2.5 µl was added to 5 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 1 µl of 5′ Surveyor F (10 µM stock concentration), 1 µl of 3′ Surveyor R (10 µM stock concentration) and 0.5 µl of ddH_2O. The PCRs underwent the following thermal cycling conditions: initial denaturation at 98°C for 10 seconds, followed by 35 cycles of denaturation at 98°C for 1 second, annealing at 65.4°C for 5 seconds, extension at 72°C for 10 seconds and a final extension at 72°C for 1 minute. The PCR products were then visualised on an agarose gel as described in section 2.2.2.1- *eGFP fragment*.

2.5 Assessing HDR efficiency with ssODNs

The CRISPR/Cas9-mediated DSBs and the micro-deletion were repaired using short 90-mer ssODNs synthesised by IDT. Each of these ssODN repair templates contained two homology arms of 42 nucleotides each. In between the two arms, was the sequence for the *EcoR*I restriction enzyme recognition site that was designed to be integrated into the repair site by HDR. Three ssODN repair templates were used to assess the HDR of a 5′ DSB, a 3′ DSB and a micro-deletion that resulted from simultaneous 5′ and 3′ cleavages. The ssODN sequences are as follows:

Homology arms are shown in black, the *EcoR*I restriction site in green.

> F1 ssODN
TTGAAATAATTTTTCATATAAAGGTGAGTTTGTATTAAAAGGGAATTCTACTGGTGGAGTATTTGATA
GTGTATTAACCTTATGTGTGAC

>R2 ssODN
CAGCTAATTCAGAATCATTTTGTGGACGAATATGATCCAACAGAATTCATAGAGGTAAATCTTGTTTT
AATATGCATATTACTGGTGCAG

>F1+R2 SSODN
TTGAAATAATTTTTCATATAAAGGTGAGTTTGTATTAAAAGGGAATTCATAGAGGTAAATCTTGTTTTA
ATATGCATATTACTGGTGCAG

HEK 293 cells were harvested and seeded into 6-well plates as described in section 2.3.1. The cells in each well were transfected with 2 µg of the chosen pcDNA-sgRNA KRAS expression plasmid, 2 µg of pX330 and 2 µl of the of the appropriate ssODN repair template (100 µM stock concentration) (see section 2.3.1) (see Supplementary figure 1 for ssODN dosage optimisation). For the micro-deletion, 2 µg of each pcDNA-sgRNA KRAS plasmid was used in a single transfection, for a total of 4 µg of sgRNA expression plasmids with 2 µg of pX330 and 2 µl of the of the appropriate ssODN repair template (100 µM stock concentration). Control transfections included the sgRNA expression plasmids and pX330 without the ssODN repair template, or the ssODN repair template only. After 72 hours, the gDNA was extracted from these transfected cells and adjusted to 20 ng/µl, as described in section 2.3.2. The repaired *KRAS* locus was amplified using PCR, exactly as described in section 2.4. The PCR products were visualised on an agarose gel, column purified and quantified as previously outlined in section 2.2.2.1- *eGFP fragment*. Each sample of purified PCR product was diluted to 30 ng/µl, and was subject to a restriction digestion reaction containing 300 ng of DNA, 2µl of Buffer O (Thermo Scientific, USA), which contains 50 mM Tris-HCl (pH 7.5 at 37°C), 10 mM $MgCl_2$, 100 mM NaCl and 0.1 mg/ml BSA, 10 U of *EcoR*I (10 U/µl stock concentration) (Thermo Scientific, USA) and 7 µl of ddH_2O. This was digested for 1 hour at 37°C before the products were resolved and visualised on an 8% polyacrylamide gel as described in section 2.3.4. HDR efficiency was calculated with the same formula that is described in section 2.3.5.

2.6 Testing the "dual cut" strategy for the production of G>T *KRAS* mutants

qPCR was used to quantify the efficiency of the CRISPR/Cas9-mediated integration of the dsDNA donor using the "dual cut" strategy. This involved the simultaneous induction of a 5′ and 3′ DSB as described in Figure 14.

2.6.1 Linear dsDNA donor preparation

A linearised form of the dsDNA donor was prepared for transfection for the dsDNA donor repair experiments. This was done by PCR amplification of the donor cassette from the pJet-dsDNA donor plasmid, which was constructed and validated in section 2.2.2. The linear dsDNA donor was generated in a PCR which contained 25 µl of 2X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, USA), 1 µl of 5′HA F2 (10 µM stock concentration), 1 µl of 3′HA R2 (10 µM stock concentration), 1 µl of pJet-dsDNA donor (25 ng/µl stock concentration) and 22 µl of ddH_2O. The thermal cycler (Bio-Rad, USA) was programmed to undergo an initial denaturation at 98°C for 10 seconds, followed by 35 cycles of denaturation at 98°C for 1 second, annealing at 62.7°C for 5 seconds, extension at 72°C for 45 seconds and a final extension at 72°C for 1 minute. Four identical reactions were prepared to generate enough linear dsDNA donor for a single experiment. The linear dsDNA donor product was column purified and quantified according to section 2.2.2.1 - *eGFP fragment*.

2.6.2 gDNA template preparation

HEK 293 cells were transfected in a 6-well plate (as described in section 2.3.1), with 2 µg of the appropriate pcDNA-sgRNA KRAS plasmid (2 µg of each pcDNA-sgRNA KRAS plasmid for the double cut), 2 µg of pX330 and 1 µg of the purified linear dsDNA donor PCR product. Two negative control transfections were included. The first of these contained 2 µg of the empty pcDNA.H1.sgRNA plasmid, 2 µg of pX330 and 1 µg of the linear dsDNA donor. The second control transfection contained 2 µg of each of the pcDNA-sgRNA KRAS plasmids, 2 µg of pX330 and 1 µg of a mock donor. 72 hours post-transfection, half of the transfected

cells in each well were harvested for gDNA extraction as described in section 2.3.2. The gDNA was quantified and normalised to 16.5 ng/µl.

The remaining cells were placed back into their wells, and supplemented with 2 ml of complete DMEM. These were cultured for another 11 days to dilute out exogenous episomal DNA, while maintaining recombinant cells. When the cells became confluent, the cell culture medium would be discarded and the cells were washed with 2 ml 1X PBS. Following this, the cells were detached by incubating them for 3 minutes at 37°C in 1 ml of Trypsin EDTA (Hyclone, USA). 1 ml of complete DMEM was added to neutralise the trypsin. The cell suspension was then thoroughly mixed by aspirating with a 1 ml pipette. 1 ml of this cell suspension was then discarded. The remaining cells were supplemented with 2 ml of complete DMEM and returned to the cell culture incubator. 14 days post-transfection, the HEK 293 cells were harvested and the gDNA was extracted from these cells as described above.

2.6.3 Quantitative real-time PCR (qPCR)

Four primer sets were designed for this experiment, using the NCBI Primer-BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The primer sequences are shown in Table 4. Primer set A spanned the 5′ cleavage-repair junction, with the forward primer binding to a flanking endogenous *KRAS* sequence and the reverse primer binding to the end of the eGFP coding sequence. This allowed for the quantification of dsDNA donor constructs that were successfully integrated into the *KRAS* gene target site. Primer set B primed to the NeoR coding sequence, which was present in the dsDNA donor and the sgRNA expression plasmids. The readout from this primer set served as an indication of the levels of exogenous episomal DNA over time. Primer set C primed to the KRS of the dsDNA donor. The forward primer is allele specific to the G>T point mutation, resulting in amplification from the dsDNA donor only, and not the WT allele. Primer set D was designed to measure the total *KRAS* DNA of the target region, as it primes to a region that is slightly downstream of the cleavage target sites. β-actin was used as a reference gene. Because the qPCR was performed on gDNA, reference gene considerations were not heavily scrutinised as it was assumed that

each cell will have the same number of β-actin gene copies. Each of these primer sets were optimised at different annealing temperatures (see Supplementary figure 2), and the reactions were all checked for product specificity by melt curve analysis and agarose gel electrophoresis (Supplementary figure 3).

Table 4: List of primer sequences used for qPCR

Primer Name	Usage	Sequence (5' - 3')
Primer set A F	qPCR	CCAAGAGAACTACTGCCATGATGC
Primer set A R	qPCR	GCATGGACGAGCTGTACAAGT
Primer set B F	qPCR	TGATATTCGGCAAGCAGGCA
Primer set B R	qPCR	GACCACCAAGCGAAACATCG
Primer set C F	qPCR	AACTTGTGGTAGTTGGAGCTGT
Primer set C R	qPCR	GGTCCTGCACCAGTAATATGCA
Primer set D F	qPCR	TTATTTGGGCGGAAGGCTGA
Primer set D R	qPCR	GTCAGGGACCGTCAGTTTCA
β-actin F	qPCR	ACCAACTGGGACGACATGGAGAAA
β-actin R	qPCR	TAGCACAGCCTGGATAGCAACGTA

To each qPCR, 5 µl of 2X SYBR Green PCR Mastermix (Applied Biosystems, USA), 1 µl of the forward primer (2 µM stock concentration), 1 µl of the reverse primer (2µM stock concentration) and 3 µl of the gDNA template (16.5 ng/µl stock concentration) was added. No template control reactions were also included. These were setup in a 96-well PCR plate, which was centrifuged at 865×g for 3 minutes (Boeco U-320, Germany), before being loaded into an Applied Biosystems 7500 RT PCR System, USA. These reactions then underwent an initial hold stage at 50°C for 2 minutes, followed by denaturation at 95°C for 10 minutes and 50 cycles of denaturation at 95°C for 15 seconds and annealing and extending at 55°C (for primer set A, B and Beta-actin) or 60°C (for primer set C and D) for 1 minute. An inter-plate calibrator (IPC) reaction was included for each primer set so that any technical variations between the 72 hours and 14 days samples could be controlled for. The IPC template mixture

was made up by adding 10 µl of each gDNA sample (16 ng/µl stock concentration) obtained 72 hours post-transfection. 3 µl of this was used as a template in each IPC control reaction.

2.6.4 Analysis

The data of each primer set were mathematically corrected according to the changes seen in the threshold (Cq) values of the corresponding IPC reactions by performing the following function:

$$Cq_i^{\text{corrected}} = Cq_i^{\text{uncorrected}} - Cq_i^{\text{IPC}} + \frac{1}{\text{no.plates}} \sum_{i=1}^{no.plates} Cq_i^{\text{IPC}}$$

The corrected data generated from the qPCR experiments were analysed and represented as a fold change relative to β-actin. Briefly, this involved normalising the Cq values for each sample to that of β-actin by subtracting the β-actin Cq from the Cq value of the corresponding sample, to obtain a ΔCq value. To calculate the fold change, the following transformation was performed: $2^{\Delta Cq}$. This was done according to "Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR" (Applied Biosystems, USA).

2.7 Statistical Analyses

Statistical analysis was carried out using the online QuickCalcs tool provided by GraphPad (http://www.graphpad.com/quickcalcs/). Student's t-tests were performed on the data from the Surveyor Assay and HDR efficiency experiment. Pearson's correlation coefficient was used to determine the strength of linear correlations. The F-test was used to determine the statistical significance of these linear regressions.

3. Results

3.1 Cloning of sgRNA expression plasmids

The online CRISPR tool identified 14 NGG PAM sites around the G>T *KRAS* mutation hotspot. Each of these potential sgRNA target sites were ranked according to a score that was inversely correlated to the number of predicted off-target sites within the human genome. The search area for these potential sgRNA target sites was limited to 250 bp around the mutation point (125 bp in each direction) so that the returned target sites were within very close proximity to each other. This ensured that variations associated with markedly different genomic target sites were eliminated, allowing for the effect of the guide sequence composition on sgRNA activity to be examined more accurately. Furthermore, in order to edit the point mutation, the HDR-stimulating DSBs needed to be sufficiently close to the mutation site for it to be effective. A total of five sgRNA guide sequences were chosen on the basis of the returned score of each target site identified by the online CRISPR tool, as well as the sequence requirements for the H1 Pol-III promoter transcriptional start site (Figure 10A). Target sequences that started (at the 5′ end) with a guanine or an adenosine were selected for optimal H1 Pol-III-mediated sgRNA expression. Therefore, the optimal sgRNA guide sequences that were chosen to target this region of the *KRAS* gene adhered to the formula G/A (19-20N) NGG. Three of these optimal targets sites were upstream of the G>T mutation (5′ cleavage) and the remaining two were downstream of the mutation (3′ cleavage). This allowed for different sites around the mutation to be targeted, and for potential gene deletions that span the mutation site to be induced. The guide sequence for each target site was ligated into the pcDNA.H1.sgRNA expression plasmid as an annealed pair of complementary oligonucleotides. This process resulted in the insertion of a 20 bp guide sequence immediately downstream of the H1 Pol-III promoter, which was designed to drive the expression of the inserted guide sequence and the RNA scaffold, as a single chimeric sgRNA molecule. Sanger sequencing was used to confirm the successful insertion of the guide sequences into the plasmid vector. This was done with a primer that bound upstream of the H1 promoter, allowing for the forward read into the ligated guide sequence. The sequence

readouts showed that all the *KRAS* target guide sequences were successfully ligated into the sgRNA expression vector (Figure 10B).

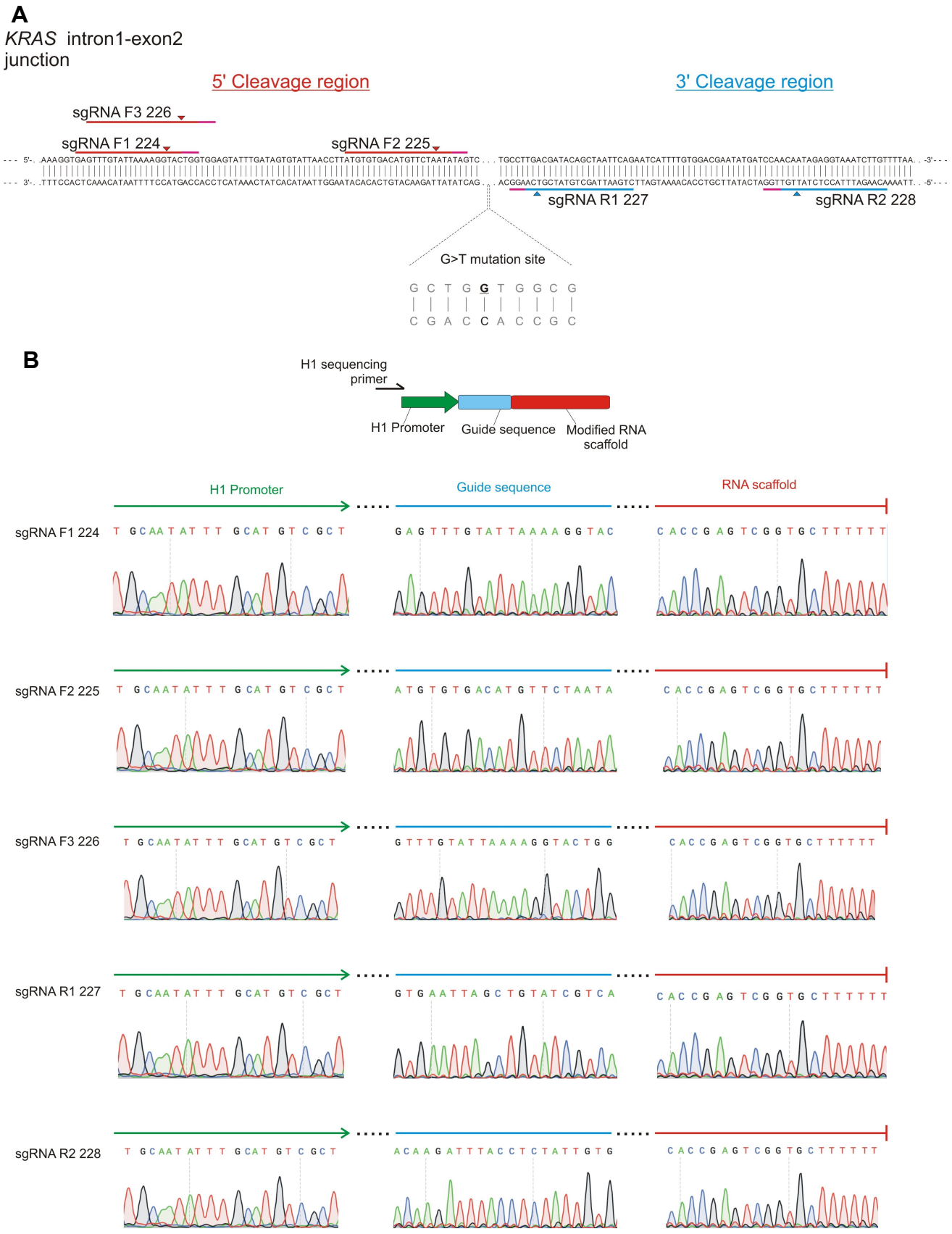


Figure 10: The design and screening of the sgRNA constructs. (**A**) A schematic showing the sgRNA target sites around the G>T mutation hotspot on the *KRAS* gene. Three of the sgRNAs targeted a region upstream of this hotspot (in the 5′ cleavage region - red) and the remaining two, a region downstream of the hotspot (in the 3′ cleavage region - blue). The coverage of these five sgRNAs was not more than 250 bp of the gene sequence.

The red and blue lines indicate the target sequences for the 5′ and 3′ sgRNAs, respectively. The pink line corresponds with the PAM tri-nucleotide recognition site. The arrowhead shows the cleavage site. The guide sequences of the three 5′ sgRNAs were complementary to the bottom strand, and the two of the 3′ sgRNAs to the top strand. (**B**) A schematic showing the sgRNA expression construct. Sanger sequencing was used to check the sequence of this construct in potentially positive clones. A forward primer that bound upstream of the H1 promoter was used. The chromatograms show the 5′ sequence of the H1 promoter (green), the full guide sequence (blue) and the 3′ end of the RNA scaffold (red) for each of the five sgRNA constructs, which all correspond to their target sites.

3.2 Surveyor Assay

In order to test the functionality of each sgRNA, the Surveyor Assay was used to measure the frequency of indel formation as a result of NHEJ repair of CRISPR/Cas9-mediated DSBs (Mussolino et al., 2011, Ran et al., 2013b). This was done in the absence of an exogenous homologous repair template. The Surveyor Assay provides a simple method to detect any indel formation as a result of NHEJ repair of a DSB and give a readout of whether the DNA was cleaved or not at each sgRNA target site (Figure 11A and B). It is able to do this from a population of diverse, stochastically-formed indels, compared to other well known polymorphism detection assays, such as single-strand conformation polymorphism analysis, which is better suited for the detection of and comparison between single polymorphic species (Orita et al., 1989). Furthermore, because the Surveyor Assay is able to clearly distinguish cleaved DNA products from uncleaved, irrespective of the type of indel formed, semi-quantitative data can be obtained from this assay, providing information on the cleavage efficiency mediated by each sgRNA.

The presence of digested PCR products suggested that the CRISPR/Cas9 system components were being successfully expressed and translated, and were functional in targeting the *KRAS* locus. Gel electrophoresis of the 5′ and 3′ cleavage region amplicons showed that the PCRs were successful, producing very specific products of 361 and 346 bp in size, respectively (Figure 11C). The specificity of the PCR products was particularly important for the Surveyor Assay, as any non-specific species would have formed heteroduplexes that contributed to background noise and possibly false positives. Subsequent digestion of these heteroduplexes with the T7EI nuclease revealed that all the sgRNAs were functional. This was seen on the

polyacrylamide gel as the typical banding pattern of a single restriction digest of PCR products. Digestion of the 5′ cleavage region heteroduplexes produced fragments of 224 and 137 bp for sgRNA F1 224, 274 and 87 bp for sgRNA F2 225 and 227 and 134 bp for sgRNA F3 226. The digestion of the 3′ cleavage region heteroduplexes generated fragments of 274 and 72 bp for sgRNA R1 227, and fragments of 227 and 119 bp for sgRNA R2 228. Controls in which pX330 was transfected into the cells without an accompanying sgRNA expression construct showed no targeted cleavage. Similarly, a transfection with pCI-NeoR-eGFP (to monitor transfection efficiency) and a mock transfection showed no specifically digested products either (Figure 11D). This confirmed that the CRISPR/Cas9 system was able to introduce targeted DSBs in the *KRAS* gene locus and stimulate the NHEJ DNA repair pathway. The cleavage efficiency of each sgRNA was calculated (see section 2.3.5) from three biologically independent repeats and showed that the sgRNAs functioned at varying efficiencies (Figure 11E). sgRNA F1 224 was the most efficient at cutting in the 5′ cleavage region, with an average efficiency of 6.01%. sgRNA R2 228 showed a cleavage efficiency of 2.74%, the highest out of the two 3′ cleavage region targeting sgRNAs. Interestingly, sgRNA F1 224 performed significantly better than the other sgRNAs, even though they all target sequences within the same region.

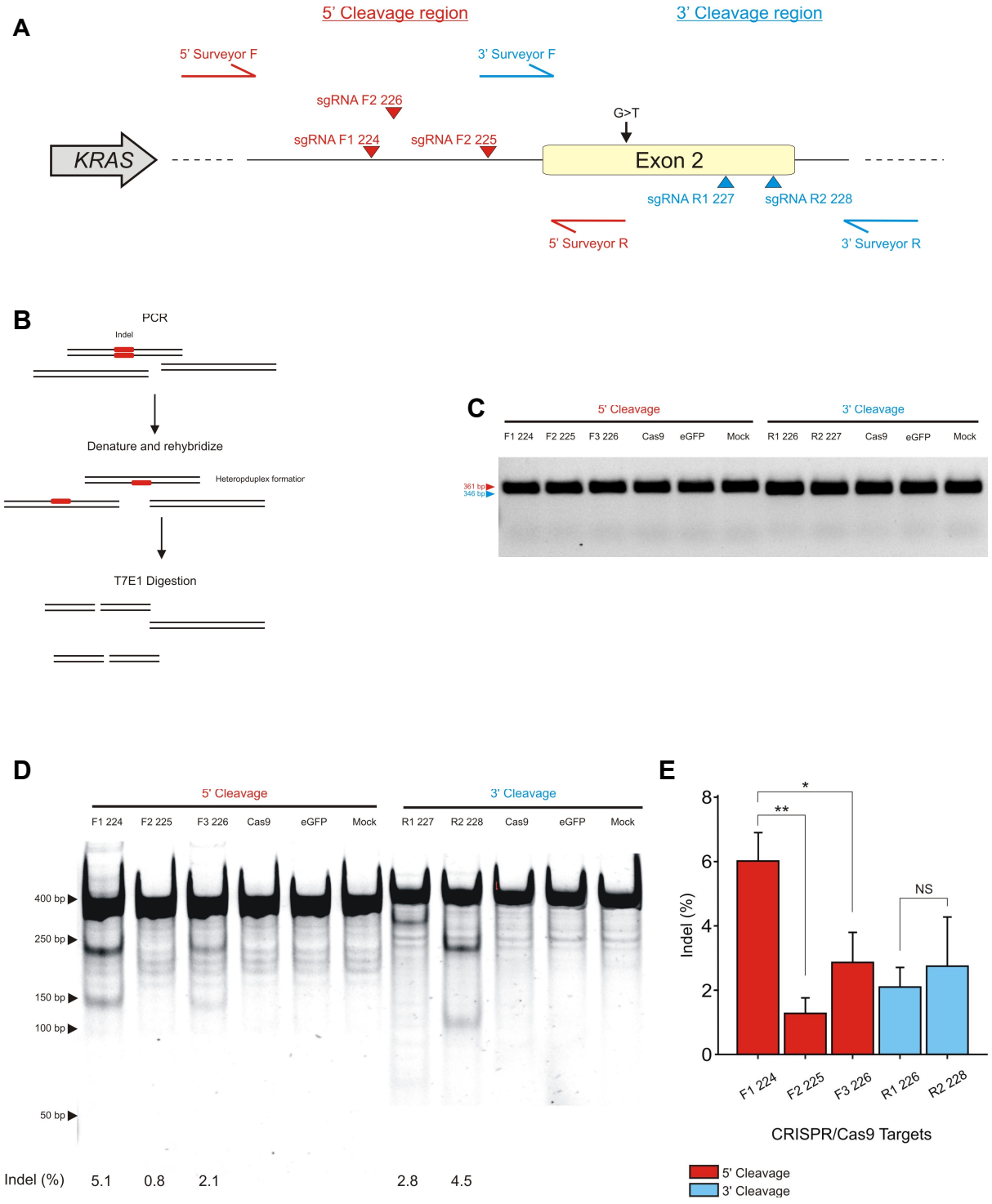


Figure 11: Functional assessment of the *KRAS* targeting sgRNAs with the Surveyor Assay. (**A**) A schematic of the *KRAS* gene locus chosen for CRISPR/Cas9 targeting. The cleavage site of each sgRNA is depicted as an arrowhead. The two primer sets used to amplify the 5′ and 3′ cleavage region are also shown, in red and blue, respectively. (**B**) An outline of the Surveyor Assay. (**C**) An agarose gel showing the PCR products amplified from the 5′ (361 bp) and the 3′ (346 bp) cleavage regions of the *KRAS* gene. (**D**) A polyacrylamide gel showing the heteroduplexes after being digested with the mismatch sensitive T7EI nuclease. The upper bands represents PCR products from unmodified alleles, with the lower, digested bands from alleles that have experienced a DSB and subsequent NHEJ repair. Cells transfected with pX330 only (Cas9), pCI-NeoR-eGFP (eGFP) or water

(mock) served as controls. For each sgRNA the percentage of indel formation was calculated after densitometric quantification of the bands (**E**) The percentage of indel formation mediated by each of the sgRNAs from a triplicate experiment (mean±SD, *p<0.05, **p<0.01, unpaired Student's t-test).

3.3 Micro-deletion detection

To demonstrate the use of the CRISPR/Cas9 system for gene deletions, sgRNA F1 224 and sgRNA R2 228 were used to introduce simultaneous DSBs in the 5′ and 3′ cleavage regions, respectively. It was expected that two concurrent DSBs at the target sites would result in the deletion of the region in between the two cuts. Because this was done in the absence of an exogenous repair template, the repair of the excised DNA should have occurred by NHEJ repair, whereby the outer ends of the broken DNA were re-ligated together. PCR across the deletion junction was used to detect alleles that had undergone the deletion event. Alleles that had experienced a micro-deletion would produce PCR amplicons that would reflect this, as they were predicted to be smaller in size compared to that of the WT alleles (Figure 12A).

Visualisation of the PCR products on an agarose gel showed that when sgRNA F1 224 and R2 228 were applied simultaneously, a micro-deletion of 199 bp in the *KRAS* gene was introduced. Amplification from the deletion alleles generated a 451 bp product. This was smaller than the 650 bp product, which was amplified from the WT allele of cells that underwent a mock transfection. A product of the same size was observed when single DSBs were introduced, indicating that the repair of individual DSBs was not sufficient to stimulate gene deletion. This demonstrated that sgRNA F1 224 and sgRNA R2 228 were able to function concurrently to introduce two simultaneous DSBs, which generated a specific deletion in the *KRAS* gene (Figure 12B). Furthermore, this highlights the ability of the CRISPR/Cas9 system to successfully mediate targeted gene deletions, which is a useful function for genome engineering applications.

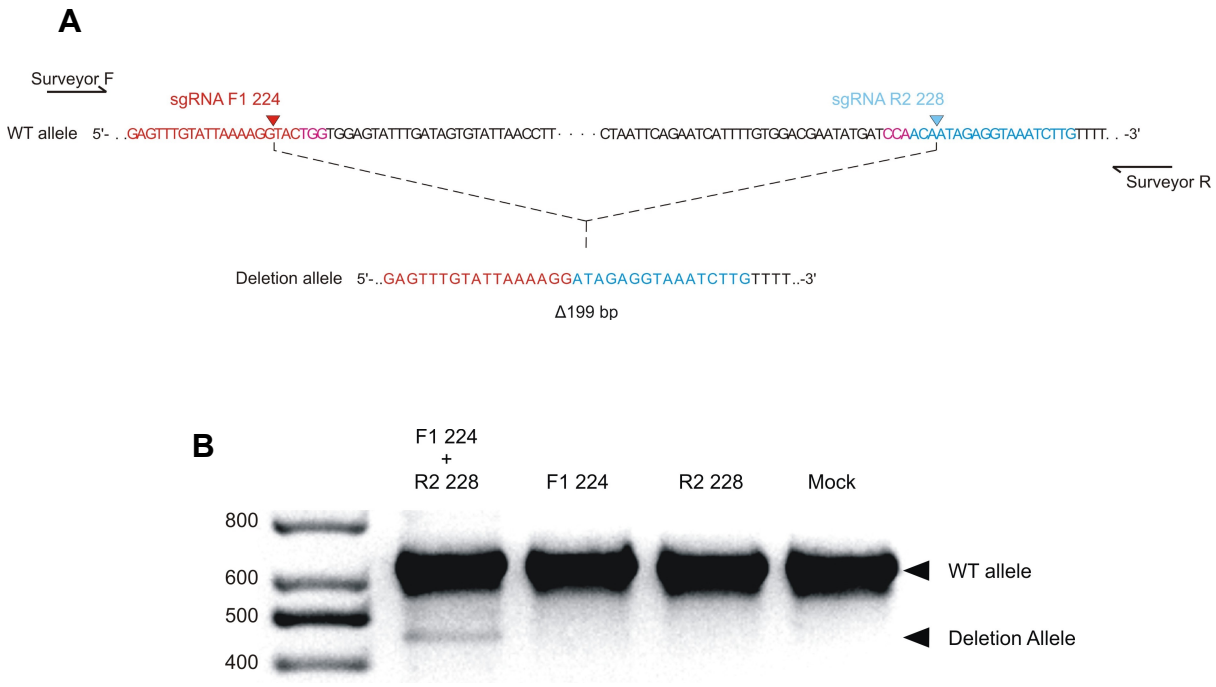


Figure 12: PCR detection of CRISPR/Cas9 mediated micro-deletions. (**A**) A schematic showing the deleted 199 bp sequence from the *KRAS* gene when sgRNA F1 224 and R2 228 were applied simultaneously to generate a double cut. This micro-deletion was detected by PCR amplification of this cleavage junction with the illustrated primers (**B**) An agarose gel showing the PCR products amplified from *KRAS* alleles that have undergone a micro-deletion, which was absent when only single DSBs are applied. Cells transfected with water (mock) produced PCR products from WT alleles only, which served as a comparative control.

3.4 HDR efficiency with ssODNs

The results above show that the CRISPR/Cas9 system is able to successfully stimulate the NHEJ repair pathway when DSBs are introduced in the absence of an exogenous homologous repair template. This is useful in mutating genes through the resultant indels or gene deletions formed. However, for the precise targeted modification of genes, the CRISPR/Cas9 system needs to be able to stimulate the HDR pathway for the repair of DSBs according to specific repair templates. ssODNs have been shown to be able to seamlessly repair and modify targeted single DSB sites via HDR. Based on the optimal architecture outlined by Church *et al.*, ssODNs harbouring the *EcoR*I restriction site were designed to be homologous to the top-strand of the *KRAS* target region and serve as repair templates of individual sgRNA F1 224 and R2 228 DSBs, as well as the previously described micro-deletion (Figure 13A) (Yang et al., 2013). The ability of the CRISPR/Cas9 system to stimulate the HDR of this targeted DNA damage, according to the ssODN templates, was assessed by RFLP analysis. It was expected that the individual DSBs and the micro-deletion would be repaired according to the homologous ssODN template, resulting in the integration of the *EcoR*I recognition sequence into the repair site. Digestion of PCR products amplified from this gene region, with *EcoR*I, would reveal the newly introduced restriction site, serving as an indicator for CRISPR/Cas9 stimulated site-specific HDR.

The repair region was successfully amplified in each transfection condition. WT alleles or alleles of repaired single DSB sites yielded a 650 bp long product. A 451 bp product was observed for the repair of the micro-deletion event. It is worth noting that the addition of the F1+R2 ssODN for the repair of the micro-deletion, seemed to stimulate gene deletion. This can be seen by the abundance of the deletion PCR product when the two sgRNAs were applied with the F1+R2 ssODN, compared to when the micro-deletion was induced in the absence of the F1+R2 ssODN repair template (Figure 13B). Densitometry of these PCR product bands showed that the micro-deletion occurred in approximately 53.2% of the amplified alleles when the F1+R2 ssODN was included. This only occurred in about 4.3% of the alleles when the F1+R2 ssODN was excluded. *EcoR*I digestion of these PCR amplicons revealed that both the individual DSBs as well as the micro-deletion were able to stimulate

HDR (Figure 13C). The repair of the micro-deletion produced fragments of 225 and 228 bp in length. However, quantification of these bands showed that HDR only occurred in 3.4% of the transfected cells, even though more than 50% of the amplified alleles experienced the micro-deletion. This indicated that a large proportion of the alleles that underwent this micro-deletion were not repaired according to the ssODN by the HDR pathway. The repair of the F1 224 target site produced digested products of 225 and 427 bp. The repair of the R2 228 target site generated products that were 228 and 423 bp long. The HDR of the individual sgRNA F1 224 and R2 228-mediated DSBs occurred at a frequency of 5.9 and 9.3%, respectively (Figure 13D). HDR of the *KRAS* locus was undetected in the controls, which either excluded the ssODN repair template, or the CRISPR/Cas9 components for DSB induction. This indicated that the observed HDR of the DSBs was not mediated by an endogenous homologous sequence that happened to also harbour an *Eco*RI restriction site (shown by cleavage only), or that the integration of the *Eco*RI restriction site was not from spontaneous HR with the ssODNs (shown by donor only). From this, it could be seen that the CRISPR/Cas9 system was able to stimulate HDR in the *KRAS* gene locus, through the introduction of single DSBs as well as a micro-deletion. Furthermore, this targeted DNA damage was able to be repaired according to small ssODN repair templates.

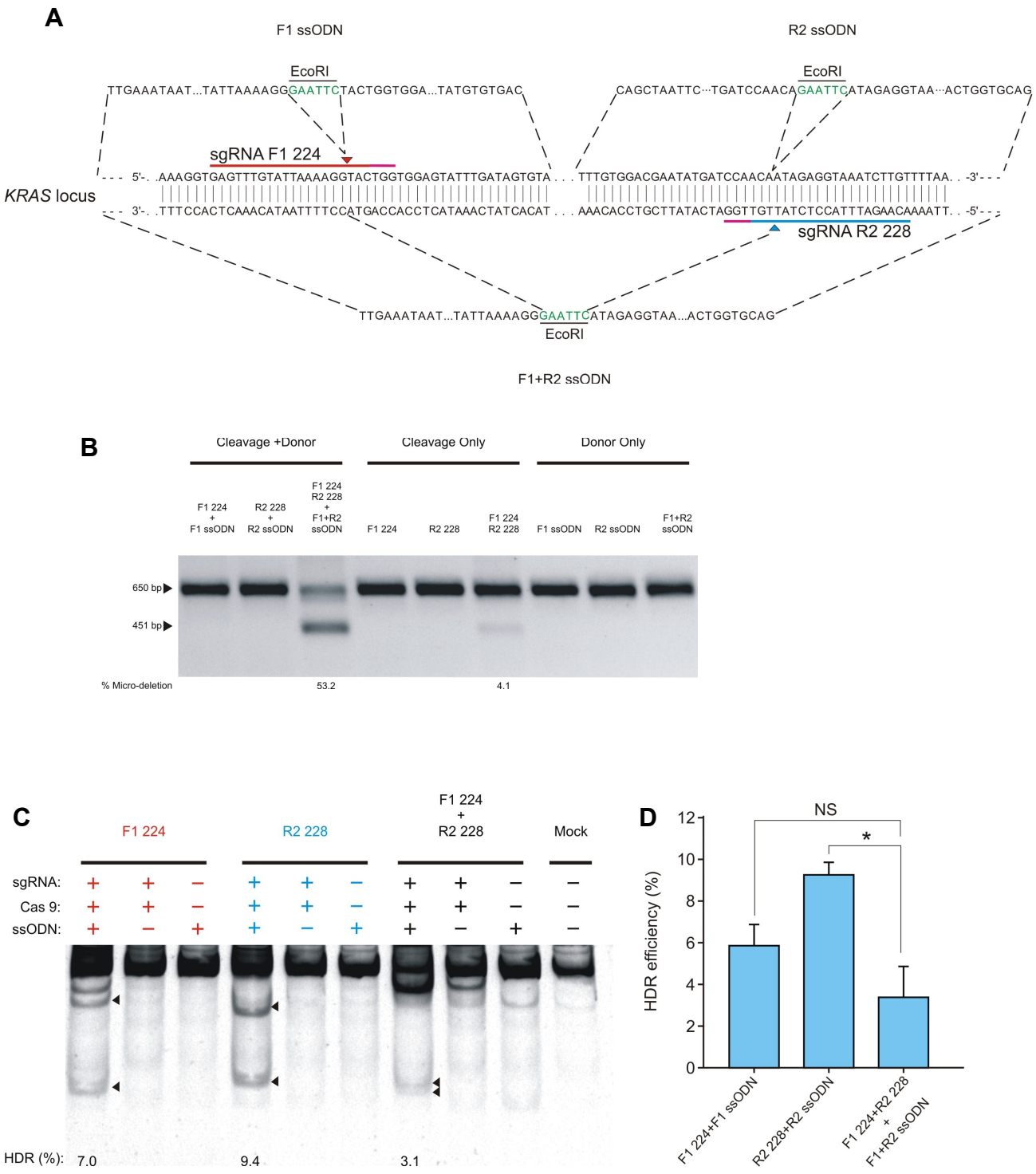


Figure 13: HDR of the individual DSBs and the micro-deletion with ssODNs. (**A**) A schematic showing the design of each 90-mer ssODN to repair the individual DSBs (top) and the micro-deletion (bottom) mediated by sgRNA F1 224 and R2 228. The homology arms of F1 ssODN and R2 ssODN flanked sequences directly adjacent to their respective DSBs. F1+R2 ssODN was designed to flank distal sequences that were outside of the F1 224 and R2 228 DSBs, for the repair of the micro-deletion. Each ssODN carried the *EcoR*I recognition site that was destined to be integrated into the *KRAS* gene via HDR. (**B**) PCR products of this repair region were amplified from gDNA. Controls for this experiment included cells that were transfected with the CRISPR/Cas9

components without the ssODN (cleavage only), or the ssODN without the CRISPR/Cas9 components (donor only). (**C**) A polyacrylamide gel showing the *EcoR*I digested PCR products. Successfully digested products were the result of the integration of the *EcoR*I recognition site into the *KRAS* gene, via HDR with the ssODNs. The density of the bands of digested products were quantified by densitometry to determine the efficiency of HDR. (**D**) The HDR efficiency was measured from triplicate experiments (mean±SD, *p<0.05, unpaired Student's t-test).

3.5 Testing the "dual cut" strategy for the production of G>T *KRAS* mutants

In order to generate selectable recombinant cells that harbour a desired point mutation, two genetic modifications have to be successfully introduced simultaneously into the gene region of interest. This commonly involves the insertion of a selection cassette into the nearest intronic sequence, where it can be integrated and expressed without affecting the gene coding sequence. The accompanying point mutation is often targeted to alter an exonic gene coding sequence. Traditional strategies employed to do this, involve introducing a targeted DSB to stimulate HDR according to a donor DNA construct that contains these two elements. This construct usually consists of two homology arms, which directly flank the DSB site. One of these homology arms will not completely be homologous to the target, with the point mutation variant embedded in its sequence. In between the two homology arms is the selection cassette. The repair of the DSB according to this donor sequence results in the integration of the expression cassette into the intronic sequence, and the point mutation into the targeted exon. However, due to the short gene conversion tracts in mammalian cells, the amount of DNA transferred from the donor construct to the gene locus diminishes greatly as the distance between the DSB site and the genetic modification increases (Elliott et al., 1998). Therefore, with the use of a single DSB, the successful co-transfer of both the expression cassette and the point mutation is often very inefficient when the two points of gene modification, that is, the insertion sites within the intron and exon, are far apart. This often results in the production of recombinant cells that are positive for the selection cassette and negative for the point mutation (Figure 14). This effect is further exacerbated when considerations into the preservation of intronic splicing elements are taken into account. In order to prevent the insertion of the expression cassette from interfering with gene splicing, the DSB site needs to be moved sufficiently far away from the intron-exon junction. Work done by Elliot *et al*. have demonstrated that gene conversion drops by a massive 60% when

the intended modification is a mere 46 bp away from the DSB site. From this, it was predicted that the success of producing selectable *KRAS* (G>T) mutants would be very limited with the single cut strategy, as the distance between the safe intronic insertion site and the exonic mutation point is 94 bp apart.

In order to overcome this limitation, a "dual-cut" cleavage and repair strategy, using the CRISPR/Cas9 system, was tested. This involved introducing two simultaneous DSBs that flanked the intron-exon junction of interest, resulting in the deletion of the entire region between the DSBs. This region was then reconstituted according to a specially designed dsDNA donor, which functionally separated the intended gene modification elements (the selection cassette and the point mutation) from the homology arms. In this way, the homology arms, which are homologous to the outside of each DSB, function solely to direct the construct to the target region. The insertion DNA in between these two homology arms, serves to replace the deleted gene fragment with a modified gene sequence that contains the mutation, and introduces the selection cassette into the affected intron. This strategy should overcome the limitations of the single cut strategy, as the introduction of the two gene modifications are independent of gene conversion tract lengths. Also, by removing the point mutation from the homology arms, complete homology with the target sequence was maintained. This should further improve the efficiency of gene conversion, as small levels of homology divergence have been shown to greatly decrease the levels of recombination (Elliott et al., 1998). Furthermore, this strategy will allow for the introduction of multiple point mutations within the gene sequence, without affecting the rate of recombination. This strategy was tested by introducing two simultaneous DSBs into the exon 2 region of the *KRAS* gene, using sgRNA F1 224 and R2 228. The resultant micro-deletion, which was previously demonstrated, was repaired according to a donor DNA template which was designed to aid the successful co-insertion of both the oncogenic G>T transversion in exon 2 and the accompanying selection cassette into intron 1 (Figure 14).

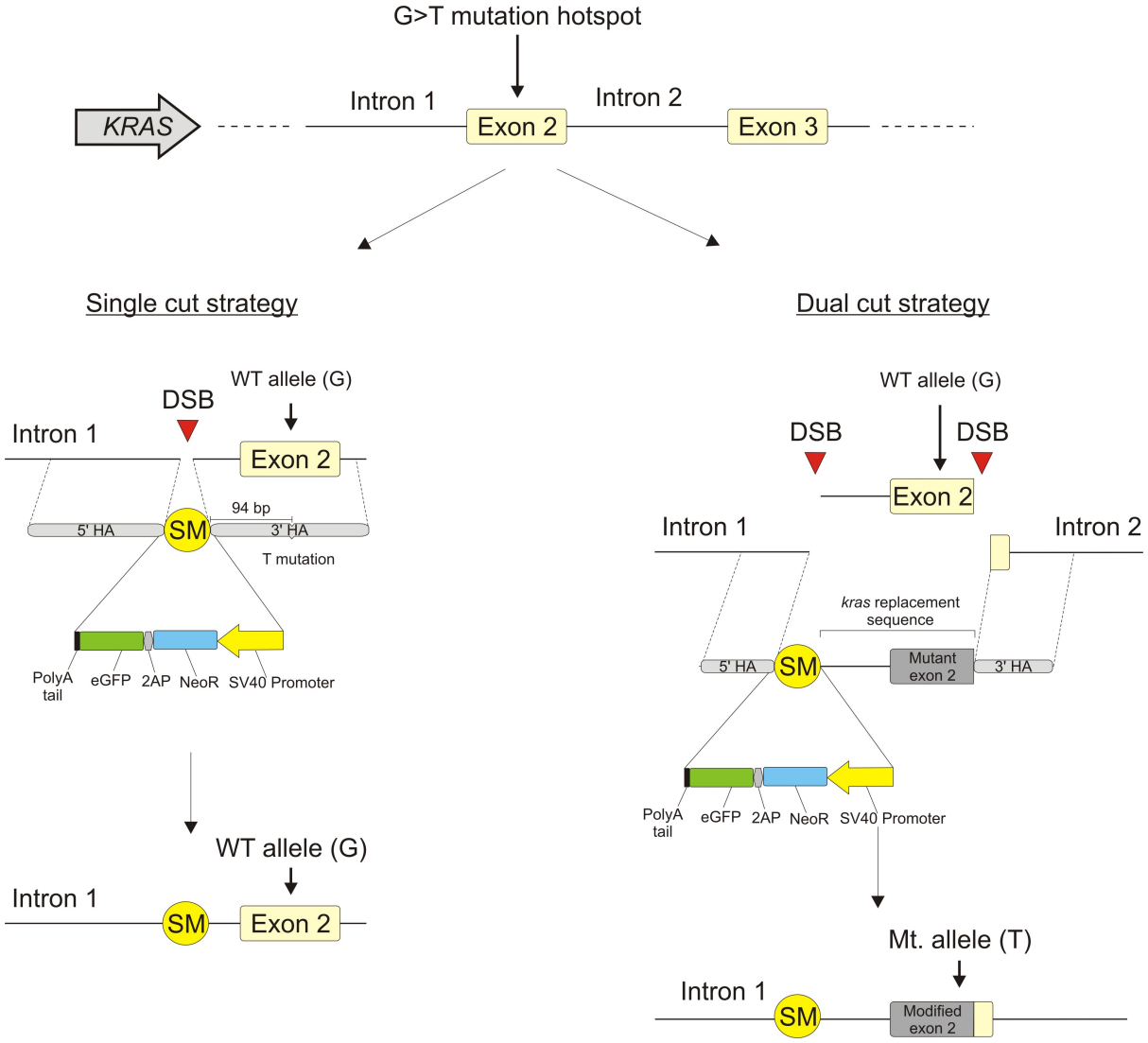


Figure 14: The envisioned "dual cut" strategy versus the tradition "single cut" strategy to modify the *KRAS* gene. The extent of gene conversion is limited with the traditional single cut strategy, which uses a donor design that has two homology arms (HAs) that directly flank the DSB site. The successful co-transfer of both the gene mutation and the selection marker cassette (SM) into sites that are 94 bp apart will be very inefficient as a result of short gene conversion tracts. The "dual cut" strategy should overcome this, by deleting the entire region between the two sites, and then reconstituting this with a modified gene sequence that contains the mutation and the selection marker.

3.5.1 dsDNA donor assembly

The dsDNA donor construct was designed to reconstitute the deleted *KRAS* gene sequence with the activating G>T mutation as well as insert a selection cassette that would allow for the isolation of recombinant cell clones. The dsDNA donor was composed of a 5′ and 3′ homology arm, which were designed to flank the sequences outside of each 5′ and 3′ DSB, respectively. In between the two homology arms is the insertion DNA sequence that is comprised of the selection cassette and the KRS. The expression cassette is driven by a constitutively active SV40 promoter, which expresses the genes for eGFP and NeoR (G418 resistance), allowing for a fluorescence and antibiotic-based approach for the selection of recombinant cells. Translation of the two selection markers occurs off a single transcript, after which the chimeric protein molecule is separated through the self-cleaving properties of the embedded 2A peptide linker. It is worth noting that the expression of this selection cassette occurs in the opposite orientation to that of the *KRAS* gene, preventing transcriptional read-through. In this way, the inserted DNA cassette is as isolated as possible from the endogenous *KRAS* gene function. Immediately downstream of this expression cassette is the KRS. This feature is inserted along with the selection cassette and is designed to replace the deleted *KRAS* DNA with a modified gene sequence that contains the oncogenic G>T transversion. With this repair template design, the desired gene modifications are completely independent of the homology arms, but are rather all located in between the two homology arms. This dedicates the homology arms for gene targeting purposes only, while the DNA in between these two arms is knocked-in to introduce the genetic modifications. This design should eliminate the loss of donor DNA features that result from short mammalian gene conversion tracts.

This dsDNA donor was created by fusing individual genetic parts together using overlap-extension PCRs. This assembly method allowed for the multiple DNA parts to be seamlessly assembled together in two consecutive PCRs, which would have otherwise required several steps if traditional cloning methods were used. Each genetic part was amplified using PCR and was visualised on an agarose gel and compared to a DNA ladder to ensure that the correct product was being amplified. All the DNA parts for the dsDNA donor were successfully

amplified. The eGFP (875 bp), KRS (191 bp) and the SV40-NeoR (1 332 bp) fragments were fused together via terminal overlapping linker sequences to generate the insertion DNA cassette (2 349 bp). Because fine resolution of larger DNA fragments is lost on agarose gels, direct comparison of this product to the DNA ladder to ensure complete assembly was not accurate. Therefore, a truncated product which lacked the addition of the small KRS fragment was generated as a comparative control. A small size shift can be clearly seen between the truncated insertion DNA cassette and the full insertion DNA cassette, indicating that the small KRS fragment was successfully assembled along with the eGFP and NeoR fragments (Figure 15A). A 5′ homology arm (437 bp) and a 3′ homology arm (432 bp) were then added to the terminals of the insertion DNA cassette in another overlap-extension PCR. An upward shift of 762 bps can be clearly seen between the insertion DNA cassette and the completed dsDNA donor (3 111 bp), indicating successfully fusion of the homology arms (Figure 15B).

Following the successful assembly of the dsDNA donor, Sanger sequencing confirmed the sequence integrity of the construct, particularly, the inclusion of the G>T mutation in the KRS (Figure 15C). Furthermore, florescent images of HEK 293 cells, 48 hours after being transfected with the dsDNA donor, showed that eGFP was successfully produced from the construct, albeit at a lower efficiency than pCI-NeoR-eGFP (Figure 15D). From the images, approximately 30% of the cells transfected with the dsDNA donor are fluorescing green. This is substantially lower than the 70% of fluorescent cells seen in the pCI-NeoR-eGFP transfection. Nevertheless, the production of eGFP is indicative of the co-expression of the upstream NeoR gene, based on the design of the construct. This shows that the expression cassette of the dsDNA donor is functional for fluorescence or antibiotic-based selection strategies.

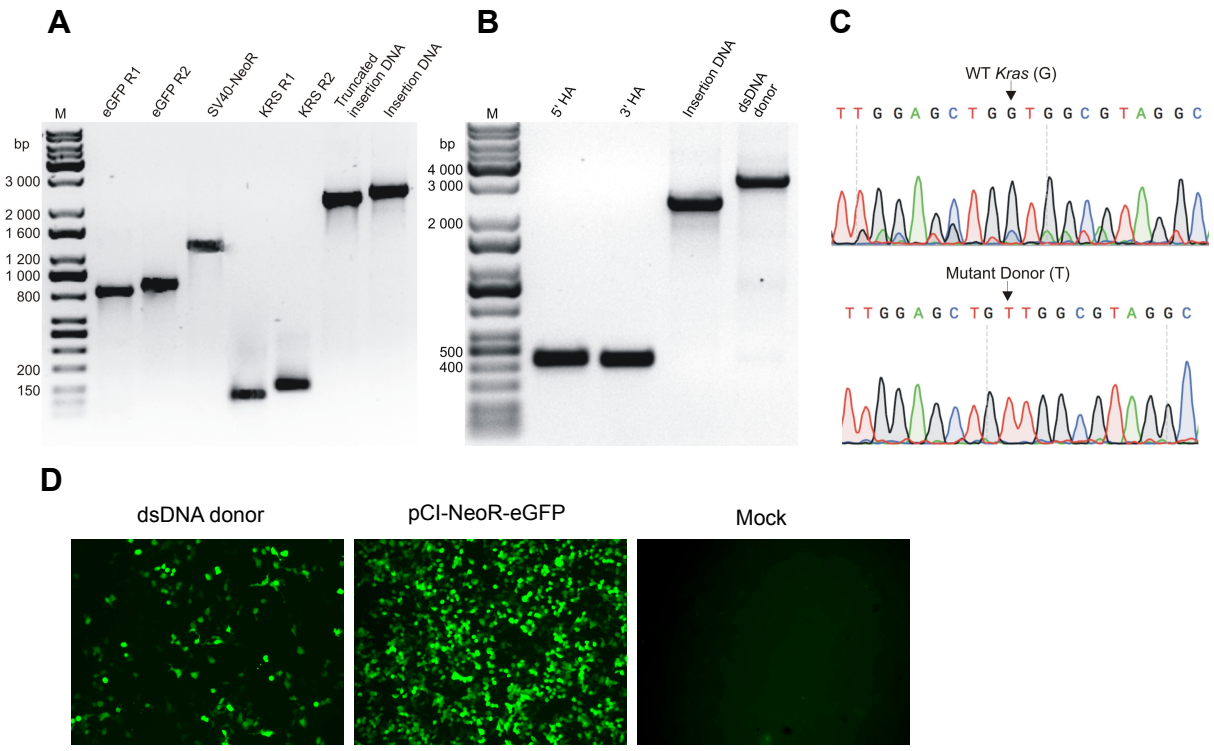


Figure 15: Assembly and verification of dsDNA donor construct. (**A**) An agarose gel showing each of the dsDNA donor fragments, which were generated with PCR. These products were fused together in an overlap extension PCR to produce the insertion DNA cassette, which comprised of the expression cassette and the *KRAS* reconstitution sequence (KRS). A truncated product was generated to serve as a size control to ensure the successful fusion of all the components. (**B**) The 5′ and 3′ HAs were amplified from gDNA and added to the terminals of the insertion DNA cassette through another overlap extension PCR. (**C**) Sanger sequencing was used to confirm the sequence integrity of the dsDNA donor. The chromatograms show a portion of this sequence that verifies the presence of the G>T mutation in the dsDNA donor construct. This was compared WT allele from the gDNA of HEK 293 cells. (**D**) Fluorescence images of HEK 293 cells 48 hours after being transfected with the dsDNA donor construct to test the functionality of the expression cassette. Cells transfected with pCI-NeoR-eGFP or water (mock) served as a positive and negative control, respectively.

3.5.2 HDR efficiency with the dsDNA donor using the "dual cut" strategy

To test the performance of the "dual cut" strategy, the amount of *KRAS* mutant DNA, as a result of successful donor integration, was measured. The detection and quantification of genome engineering outcomes is quite challenging, with conventional methods relying on the enrichment and isolation of potential recombinant cells and then screening clonal populations for the targeted genomic modification. This process is labour intensive and only provides empirical data collected from a sub-population of cells selected for screening. Furthermore, such enrichment techniques often result in the selection of illegitimate recombinants that have

arisen as a result of random donor integration. The advent of high-throughput sequencing technologies provide a more robust method for assessing gene modification outcomes across an entire cell population (Hendel et al., 2014). However, these methods are limited by their high cost and the need for intensive data analysis. In this study, a qPCR strategy was devised to quantify the efficiency of CRISPR/Cas9-mediated HDR with the dsDNA donor. This was performed on gDNA, which was extracted from the entire cell population grown in antibiotic-free conditions. This provided a simple way to detect and quantify the integration of the dsDNA donor across the entire cell population, without the need for clonal isolation of recombinants or reporter systems. Furthermore, with the lack of any selection pressure, the enrichment of illegitimate recombinants, and hence false positives, was also minimised. The detected amount of mutant *KRAS* DNA, relative to the total amount of *KRAS* DNA detected, gave an indication of how efficient the CRISPR/Cas9-mediated, "dual cut" knock-in process was. The use of qPCR allowed for the very specific and sensitive detection and quantification of the genomic integration of the complete dsDNA donor cassette (detected through the use of primer sets A-D) (Figure 16A) in the cell population 3 and 14 days post-transfection. This allowed for stable genomic integrants to be detected and monitored over time, as the residual episomal DNA from the transient transfection was diluted with cell proliferation.

Primer set A spanned the gene-donor junction, allowing for the detection of integrated donor DNA in the *KRAS* gene. From this, the insertion of the linearised dsDNA donor cassette was the most successful when sgRNA F1 224 and sgRNA R2 228 were employed together. This demonstrates that the "dual cut" strategy, which is designed to mitigate gene conversion limitations, is functional. However, interestingly, single DSBs were also able to mediate the insertion of this dsDNA donor, indicating that the asymmetric binding of one homology arm proximal to the respective DSB was sufficient for HDR. Transfections that were sgRNA deficient showed no detectable dsDNA donor integration. This confirms that the detected integration of the donor into the target sequence did not occur spontaneously. Transfections with sgRNA F1 224, sgRNA R2 228, pX330 and a mock donor PCR product also failed to show any amplification, indicating that the dsDNA donor was responsible for the detected HDR. Primer set B measured the amount of episomal DNA in the transiently transfected cells, over time. After 3 days, the levels of episomal DNA detected were similar across all the

samples, with the only exception being the transfection for the dual cleavage, which involved using twice as much sgRNA expression plasmid. This indicates that the amount of DNA entering the cells in each sample was the same, reflecting a constant transfection efficiency. The dilution of the episomal DNA from transient transfection (the linear dsDNA donor, pX330 and pcDNA.sgRNA-*KRAS*) was apparent after 14 days (for 11 passages), with a dramatic drop in the levels of detected episomal DNA by approximately 99% in each sample. Despite this, readily detectable amounts of episomal DNA were still present in the transfected cells after 14 days of sub-culture (600-1000-fold relative to β-actin). Primer set C, an allele specific primer, was designed to detect mutated *KRAS* alleles with the G>T transversion. Comparison of this to the quantified DNA of primer set A should have provided an idea of how successful the co-transfer of the expression cassette and the mutation was. Initially, the detected amount of DNA with primer set C was largely reflective of the dsDNA donor construct that existed episomally in the cells. It was expected that as the episomal DNA was diluted out of the transfected cells, the amount of modified (G>T) *KRAS* DNA, as a result of dsDNA donor integration into the *KRAS* gene, would become apparent. However, based on the observations made with primer set B, significant amounts of episomal dsDNA donor were still present in the cells after 14 days of sub-culture (approximately 150-fold of β-actin). The detection of the introduction of the G>T point mutation into the *KRAS* gene was overwhelmed by this background noise. The use of a mock donor for transfection yielded no detectable amplification with this primer set, which shows that the allele specific primer set is specific to the mutation which does not naturally exist in HEK 293 cells. Primer set D reflected the total amount of *KRAS* DNA in each sample, and remained relatively constant across all samples, throughout the 14 day period (Figure 16B).

The amount of detected dsDNA donor integrated into the *KRAS* gene (data from primer set A), taken as a percentage of the total *KRAS* DNA detected (data from primer set D), was used as a metric for site-specific donor integration efficiency (Figure 16D). The micro-deletion facilitated the integration of this large donor DNA sequence in 0.028 and 0.123% of the transfected cells, after 3 and 14 days, respectively. This is better by a factor of 1.6 and 1.8-fold, respective to the 3 and 14 days post-transfection, when compared to that of the single DSB generated by sgRNA F1 224. Similarly, it was found to be more efficient by 1.6 and 2.9-

fold compared to sgRNA R2 228, after 3 and 14 days, respectively. This showed that the application of both the sgRNAs to induce a micro-deletion facilitates the HDR process with this donor design. It is also worth noting that the detected amount of HDR for the three cleavage scenarios substantially increased, by a factor of 2.6 - 4.6-fold, after 14 days.

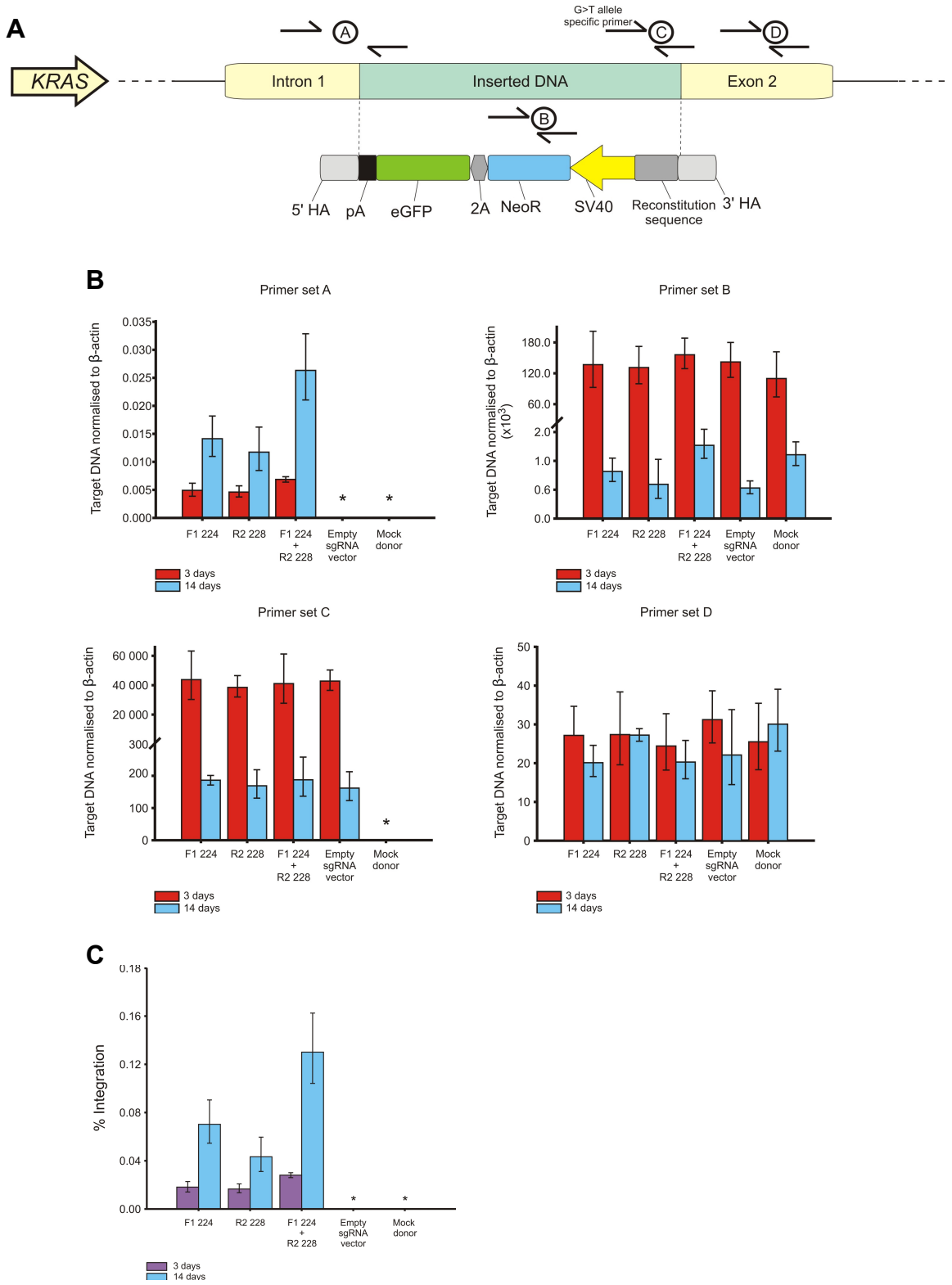


Figure 16: Integration efficiency of the dsDNA donor. (**A**) A schematic outlining the primer-target interactions for the qPCR strategy used to determine the integration efficiency of the dsDNA donor into the *KRAS* gene. Quantification was done on extracted gDNA 3 and 14 days post-transfection. (**B**) Relative amounts of detected target DNA of each primer set, normalised to β-actin. The error bars represent the SD calculated from the results of a triplicate experiment. Transfections with an empty sgRNA expression plasmid or a mock donor served as negative control. (**C**) The percentage of *KRAS* alleles that had successfully undergone dsDNA

donor integration at the *KRAS* gene locus, as calculated by taking the data from primer set A (integration events) as a percentage of primer set D (total *KRAS* DNA) (mean±SD, *denotes no detected amplification).

4. Discussion

The results demonstrate that the CRISPR/Cas9 system can modify the human *KRAS* gene in cultured HEK 293 cells in various, highly specific ways, which are each useful for genome engineering. The use of different sgRNA guide sequences, directed to the same region of the *KRAS* gene showed varying activity, possibly implicating a role of guide sequence composition on sgRNA functionality. Also, various strategies for the HDR of DSBs were simulated for the precise modification of the *KRAS* gene. From this, useful insights into sgRNA design and CRISPR/Cas9-mediated gene modification were gained, allowing for various considerations and strategies for the efficient and robust use of this system for genome engineering to be highlighted.

4.1 sgRNA guide sequence composition affects CRISPR/Cas9 indel forming activity

The Surveyor Assay showed that all the sgRNAs were functional and could successfully mediate the introduction of a DSB in the intron 1 - exon 2 region of the *KRAS* gene locus. The frequency of indel formation was variable, ranging from 1.3 - 6.0%. It has been demonstrated that the CRISPR/Cas9 system is able to introduce indels at an efficiency of greater than 50% (Hsu et al., 2013, Ran et al., 2013b, Fu et al., 2014). However, in these reports, different genes, which may be associated with variations in chromatin structure or repair pathway priority, were targeted. These factors are most likely to affect obtainable cleavage efficiencies. A pair of TALENs designed to cleave in exon 2 of the *KRAS* gene was shown to achieve an indel formation frequency of 9.4% (Reyon et al., 2012). In comparison to the efficiency observed with the CRISPR/Cas9 system here, the TALEN pair was 1.6 - 7.2-fold more efficient at indel formation than the tested sgRNAs. It is interesting that the CRISPR/Cas9 system performed poorly compared to the TALEN pair, and that the activity of each sgRNA was so variable, even though they all target sequences of very close proximity (less than 250 bp apart) within the *KRAS* gene. Because of this, these variations are not likely to be the result of differences in the genomic architecture of the sequences that the CRISPR/Cas9 system has to access, but rather due to factors that are intrinsic to the technology itself, which may limit its activity. It is possible that guide sequence composition may affect the

activity of the CRISPR/Cas9 system, and that these sequence variations may be tolerated differently to impact function.

It has been shown that the secondary structure of the RNA scaffold of the sgRNA is imperative for its interaction with the Cas9 protein. The repeat:anti-repeat duplex and stem loop 1 have been shown to be crucial for Cas9 function. Stem loop 2 and 3 have been found to play a supportive role, through stabilising the sgRNA and facilitating sgRNA-Cas9 complex formation (Chen et al., 2013, Hsu et al., 2013, Jinek et al., 2014, Nishimasu et al., 2014). The guide sequence introduced upstream of the RNA scaffold may potentially alter base-pair interactions that maintain the integrity of these functional secondary structures, resulting in the formation of dysfunctional sgRNA molecules that cannot efficiently assemble with the Cas9 endonuclease and activate it. In order to observe base-pair interactions between the guide sequence and the RNA scaffold, the online RNA folding simulation tool, RNAstructure (Reuter and Mathews, 2010), was used to predict the secondary structure of sgRNA F1 224, F3 226, R1 227 and R2 228 (Figure 17A). These were compared to the predicted secondary structure of the RNA scaffold in the absence of a guide sequence. The MaxExpect algorithm was used to generate structures that were based on the most probable base-pair interactions rather than the lowest free energy, which can be unrealistic under physiological conditions. sgRNA F2 225 was excluded from these analyses because it was designed to make use of the degenerate NAG PAM site, which has been shown to be far less efficient than when the canonical NGG PAM site is used (Hsu et al., 2013). In all of the predicted structures, the repeat:anti-repeat duplex is preserved exactly as seen in the native RNA scaffold. The structures predicted for sgRNA F3 226, R1 227 and R2 228 further show that stem loop 2 and 3 are also maintained. However, the critical stem loop 1 is disrupted in all of the sgRNA structures through direct base-pair interactions with the guide sequence, or resultant allosteric effects. This may have adversely affected the potential performance of the CRISPR/Cas9 system, resulting in its lower targeting activity of the *KRAS* gene locus compared to the TALEN pair. Interestingly, sgRNA F1 224, which was the most efficient, seems to lack the most structural features of the RNA scaffold. The predicted structure of this sgRNA displays an unwanted stem loop after the repeat:anti-repeat duplex and only maintains stem loop 3.

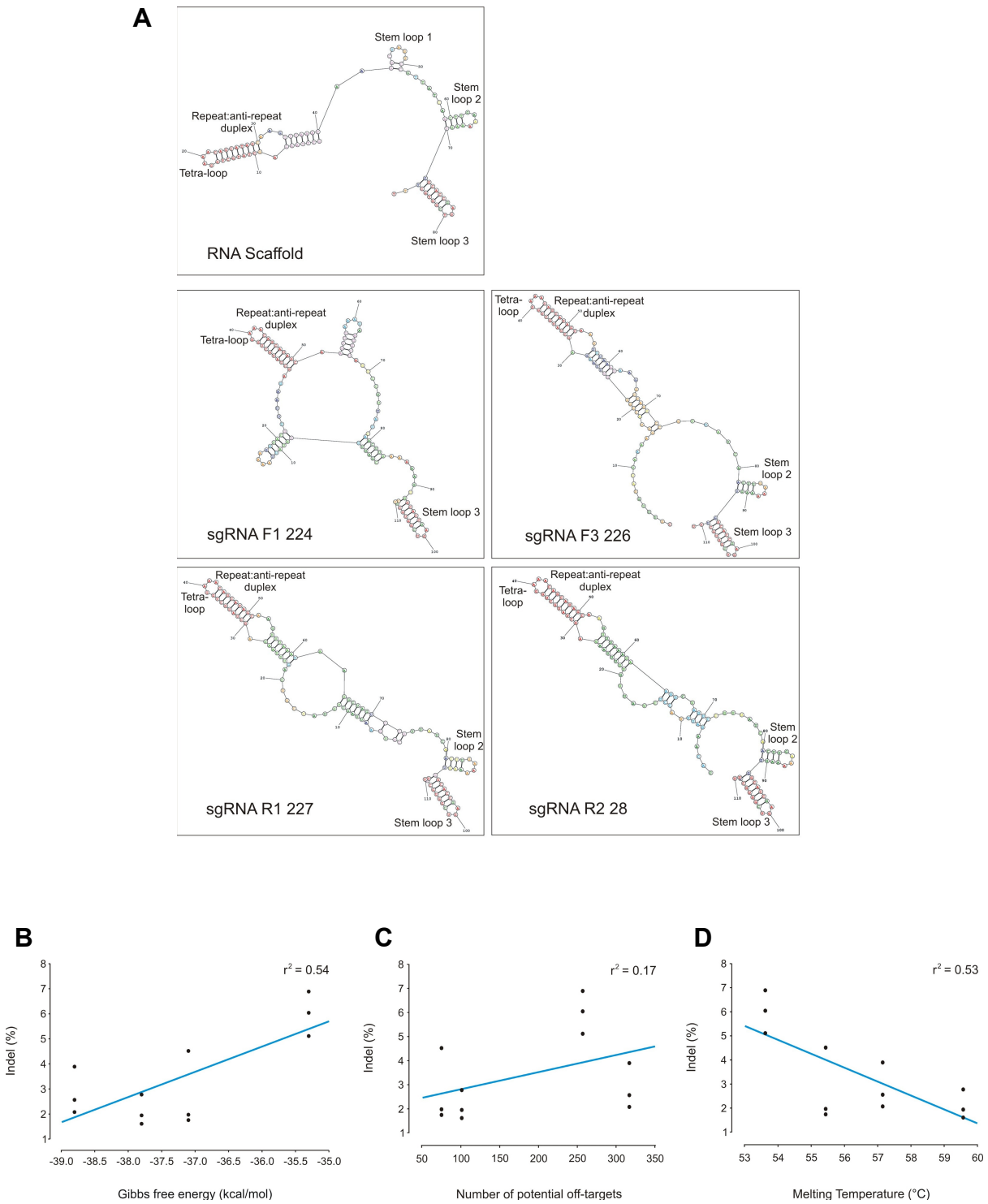


Figure 17: Analysis of guide sequence composition on sgRNA function. (**A**) Predicted RNA structures of four of the sgRNAs designed to target the *KRAS* gene. The native structure of the RNA scaffold without a guide sequence was used for comparison. The structures were predicted based on the most probable base-pair interactions. (**B**) Correlation between the stability of the sgRNA structures (as measured by Gibbs free energy) and the indel formation efficiency of the sgRNAs (p = 0.0062, F-test). (**C**) Correlation between the number of potential off-targets of each sgRNA and its indel forming activity (p = 0.18, F-test). (**D**) Correlation between the melting temperature of the guide sequence (calculated using the nearest-neighbour method) and sgRNA indel forming activity (p = 0.0072, F-test).

By inspecting these predicted secondary structures, it can be seen that the 20 nt guide sequence does interact with the RNA scaffold and disrupts various structural elements. But the maintenance of these stem loops, particularly stem loop 1, does not seem to account for the observed variations in sgRNA activity. It is worth noting, however, that the structure of sgRNA F1 224 looks significantly more "open", compared to the other three sgRNA structures, which adopt an almost completely double-stranded duplex formation. The stability of these structures is dependent on the strength and extent of base-pair interactions of the RNA scaffold and the guide sequence. Analysis of the thermodynamic properties of each of these structures, using mFold (Zuker, 2003), showed a very significant ($p = 0.0062$) negative correlation ($r^2 = 0.54$) between the sgRNA structure stability, as measured by Gibbs free energy (ΔG), and indel formation frequency (Figure 17B). Therefore, sgRNA F1 224, which has the least thermodynamically stable structure, displayed the most efficient activity. This can be explained by the ability of less stable sgRNA structures being more accessible to, and reactive with, the Cas9 endonuclease. Initial sgRNA-Cas9 interactions may be more energetically favourable with such structures and serve to catalyse a conformational change of the sgRNA to yield a structurally functional sgRNA. Such interactions may be less thermodynamically favourable with sgRNAs that are inherently stable as a result of strong intra-molecular base-pair interactions, inhibiting their efficient transformation into the functional conformation upon Cas9 interaction. Therefore, the process of sgRNA conformational change and its interaction with the Cas9 protein to produce an active complex, may be a cooperative process that is limited by the stability of the unbound sgRNA molecule.

The involvement of the guide sequence in sgRNA functionality prompted the investigation of other factors concerning the guide sequence that may also contribute to the observed sgRNA activity. It was hypothesised that the number of potential off-target sites of each sgRNA may affect the efficiency of the CRISPR/Cas9 system. Potential off-target substrates may possibly compete for the CRISPR/Cas9 components, resulting in the decrease of the frequency of *KRAS* targeted DSBs. Comparing the observed indel percentage with the number of potential off-target sites (determined with the CRISPR Design Tool), showed an insignificant ($p = 0.18$) and poor correlation ($r^2 = 0.17$) (Figure 17C). This indicates that guide sequence off-

targets do not significantly influence targeted sgRNA activity. The CRISPR/Cas9 components are probably sufficiently expressed to saturate all potential binding sites within the cell, eliminating any competitive effects. Interestingly, a very significant ($p = 0.0072$) negative correlation ($r^2 = 0.53$) was observed between the indel formation efficiency and the melting temperature of the sgRNA guide sequence (Figure 17D). This correlation shows that the increase in the binding affinity of the guide sequence with the genomic target DNA leads to a decrease in sgRNA activity. This can possibly be explained by the inefficient recycling of the CRISPR/Cas9 complex. A strong guide sequence target-binding affinity may result in the sgRNA-Cas9 complex sequestering the target DNA after DSB induction, affecting the accessibility of the target site for NHEJ repair. However, this is inconsistent with the trend previously observed (Yang et al., 2013), who found the opposite, but relatively poor, ($r^2 = 0.28$) positive correlation between sgRNA function and melting temperature. Nevertheless, this shows that the binding affinity of the sgRNA guide sequence to its cognate target plays a role in the efficiency of DSB repair. It would be useful to test a series of sgRNAs with a large range of nucleotide binding affinities to determine the optimal melting temperature for sgRNA function.

From this analysis, the guide sequence composition seems to be able to profoundly alter sgRNA function. Up to 54 and 53% of the variation observed in the efficiency of the CRISPR/Cas9 system to introduce indels in the *KRAS* gene, can be explained by sgRNA secondary structure stability and target sequence affinity, respectively. Both of these factors are affected by guide sequence composition. Therefore, guide sequence composition is an important factor that affects the robustness of CRISPR/Cas9 function and must be considered for the rational design of sgRNAs. However, because sgRNAs are so easy and cheap to generate, several guide sequences for a target region can be easily produced and tested in parallel to screen for the most efficient sgRNA.

4.2 CRISPR/Cas9-induced DSBs and micro-deletions can be predominantly repaired by different DNA repair pathways

The CRISPR/Cas9 system has been shown be able to successfully stimulate the NHEJ repair pathway to yield site-specific gene mutations and micro-deletions. The versatility of this system was further demonstrated through its ability to facilitate precise gene modifications, via the HDR pathway. Individual DSBs as well as a micro-deletion in the *KRAS* gene locus were able to be precisely repaired according to a short 90-mer ssODN repair template. RFLP analysis and densitometry showed that HDR of individual CRISPR/Cas9-mediated DSBs (sgRNA F1 224 and R2 228) occurred at an efficiency comparable to the ~8% reported by Zhang *et al.* (Ran et al., 2013a). Comparing the frequency of HDR and NHEJ showed that each of the two sgRNAs gave rise to quite different gene modification outcomes. DSBs mediated by sgRNA F1 224 were able to stimulate NHEJ (6.01%) and HDR (5.87%) at similar rates. This suggests that this target site is not preferentially repaired by either pathway. Interestingly, the repair of the DSBs mediated by sgRNA R2 228 were preferentially repaired by HDR, which occurred at an efficiency of 9.27%. This is significantly higher than the 2.74% detected for NHEJ repair of this target site. It is known that certain genetic elements are highly conserved, and preferentially repaired by either the NHEJ or HDR pathway (Wang et al., 2004). sgRNA R2 228 targets the coding sequence in exon 2 of the *KRAS* gene, which is an oncogene that needs to be carefully preserved. It is possible that the integrity of this coding sequence is maintained through the suppression of NHEJ DNA repair at the exons of this gene, and are preferentially repaired through the HDR pathway. Therefore, in the Surveyor Assay, DSBs mediated by sgRNA R2 228 may have been predominantly repaired by HDR through the recruitment of the endogenous homologous sister chromatid (Santiago et al., 2008), and less frequently by the mutagenic NHEJ pathway. In this way, the activity of sgRNA R2 228 was not able to be detected with the Surveyor Assay at this site. This explains the low percentage of indel formation observed. Furthermore, the recruitment of the homologous allele, after a targeted DSB, places the sister allele within close proximity to the CRISPR/Cas9 complex, increasing the chances of a second DSB. This facilitates the cleaving efficiency of sgRNA R2 228. Together, the increased frequency of DSB generation and the repair pathway preference of this site, enhanced the HDR of sgRNA

R2 228 DSBs with the exogenous ssODN when it was introduced. This hypothesis explains the high rate of HDR with the ssODN observed, relative to the indel formation efficiency detected with the Surveyor Assay for sgRNA R2 228. This phenomenon was not seen with sgRNA F1 224, which targets a less vital upstream intronic sequence. However, to confirm this, more robust and sensitive methods, such as high-throughput sequencing, are required to accurately quantify the gene modification outcomes at the different sites. Such sensitive characterisation would allow for repair pathway preferences associated with different genetic elements to be more accurately determined.

The simultaneous application of sgRNA F1 224 and sgRNA R2 228 was able to successful introduce a micro-deletion in the *KRAS* gene. It is expected that the observed micro-deletion only occurs with two concurrent DSBs, resulting in the excision of the DNA in between the two target sites. The agarose gels show that this deletion event was rare, with only a small fraction of transfected cells harbouring the deletion allele. This is possibly limited by the stochastic infrequency of the two sgRNAs binding to their cognate target sites, on the same copy of the *KRAS* gene, at the same time, and inducing DSBs simultaneously. However, the introduction of the F1+R2 ssODN, which was designed with homology arms that flank the outside of the micro-deletion, seemed to stimulate the frequency of this micro-deletion event quite profoundly. This is evident on the agarose gel from Figure 13B, which shows that the micro-deletion occurred 1 200-fold more frequently when the two sgRNAs were applied with the F1+R2 ssODN, than when only the two sgRNAs were applied without the repair template. Chen *et al*. showed in a study that specific gene deletions could be achieved efficiently by using ssODNs that were designed to have homology with distal chromosomal sequences to repair a single DSB (Chen et al., 2011). Therefore, it can be expected that single cuts introduced by the individual activity of sgRNA F1 224 and sgRNA R2 228 were sufficient to stimulate F1+R2 ssODN-mediated gene deletion. Furthermore, the occurrence of single cuts is stoichiometrically more favourable and likely to occur much more frequently than two simultaneous cuts, explaining why the gene deletion event occurred in only 4.1% of the transfected cells when the two sgRNAs were used in the absence of the F1+R2 ssODN, compared to the 53.2%, which was observed when the F1+R2 ssODN was included. Interestingly, however, RFLP analysis showed that only 3.4% of these deletion events were

repaired according to the F1+R2 ssODN, resulting in the successful integration of the *EcoR*I restriction site. This indicates that the remaining 50.1% of the observed micro-deletion events underwent NHEJ repair. It is uncertain as to how the F1+R2 ssODN so efficiently facilitates the introduction of this micro-deletion through the NHEJ pathway, without interacting with the homologous sequences around the DSB sites and recombining to introduce the *EcoR*I restriction site. Therefore, the predominant repair of the micro-deletion by the NHEJ pathway explains the poor HDR efficiency detected.

From this, it seems that single DSBs and micro-deletions in the *KRAS* gene can be predominantly repaired by either HDR or NHEJ. The factors that govern this repair pathway preference is currently unclear, and may be related to cellular mechanisms that prioritise repair pathways for specific genetic elements.

4.3 A "dual cut" CRISPR/Cas9 strategy can mediate the production of selectable G>T mutant *KRAS* cells

A dsDNA donor cassette was created to introduce an oncogenic G>T mutation in the *KRAS* gene along with a selection cassette that would allow for the fluorescence or antibiotic-based selection of mutant clones. This construct was designed to repair the micro-deletion generated by the simultaneous activity of sgRNA F1 224 and R2 228, so that common gene conversion limitations, intrinsic to the short gene conversion tracts of mammalian cells, could be overcome. This involved functionally separating the dsDNA donor cassette into two components: the homology arms and the insertion DNA. The homology arms were designed to perform the sole function of directing the donor construct to the gene target. The insertion DNA was responsible for the gene modifications. This was designed to reconstitute the deleted *KRAS* gene fragment with a modified sequence that contained the point mutation of interest, as well as insert the selection cassette into the adjacent intron. With this strategy, the point mutation and the expression cassette should always be successfully co-inserted into the into the *KRAS* gene, whenever an integration event occurred.

The fluorescence images of the HEK 293 cells transfected with the dsDNA donor construct showed that the expression cassette was functional. The observed eGFP indicated that the upstream NeoR gene must have also been concurrently expressed and translated. However, it can be seen that eGFP expression occurs less efficiently compared to that of the pCI-NeoR-eGFP plasmid. This is because the dsDNA donor makes use of a SV40 promoter, which has been shown to be approximately 2.5-fold less efficient, in HEK 293 cells, than the CMV promoter of pCI-NeoR-eGFP (Qin et al., 2010). Moreover, the *eGFP* transgene of pCI-NeoR-eGFP is preceded by a short intronic sequence, which is designed to stabilise and improve its expression (Xia et al., 2006). The folding kinetics and translation efficiency of the eGFP protein are also probably adversely affected by the production of an eGFP-NeoR chimera, which needs to be cleaved apart by the 2A-self cleaving peptide, before yielding individual functional proteins. The use of a stronger promoter might off-set the slower kinetics associated with the post-translational processing required for this chimeric protein. However, with the current dsDNA donor design, allowing sufficient time for the expression of these selection markers to occur, before the application of selective pressure, should overcome these limitations.

The data shows that the dual cleavage strategy was able to successfully facilitate the integration of the dsDNA donor into the targeted *KRAS* gene. HDR experiments with the ssODN and this dsDNA donor show that spontaneous HR (control transfections with the donor template only), without a DSB at the homologous target site, was undetectable. This suggests that random off-target integration events of the repair construct should be extremely rare as these sites lack the homology and the DSB to stimulate recombination. However, this needs to be confirmed by high-throughput sequencing to analyse the modified genome in more depth to get a better indication of HDR specificity.

The dual cleavage gene-editing strategy was able to generate up to 60 bi-allelic or 123 mono-allelic *KRAS* mutants out of one million cells. This is significantly higher than what can be achieved through gene targeting strategies that rely on spontaneous HR. Interestingly, the introduction of a single DSB by sgRNA F1 224 or sgRNA R2 228 was sufficient to facilitate the integration of the dsDNA donor sequence. However, this occurred at a significantly lower

efficiency than when both sgRNAs were employed simultaneously. This can be explained by the repair of the single cuts only being mediated by one of the homology arms of the dsDNA donor, whereas the micro-deletion makes use of both. It is also worth noting that the frequency of dsDNA donor integration improved over time, irrespective of the cleavage introduced. This is unlikely to be the result of slow CRISPR/Cas9-mediated HDR kinetics, as each cell is supplied with an abundance of CRISPR/Cas9 components and dsDNA donor to efficiently target the both *KRAS* alleles. Variations in cell numbers across the 14 day experiment period cannot explain this either, as these were kept relatively constant, as shown by the quantity of total *KRAS* DNA detected. Furthermore, this increase in the signal of integrated donor could also not be a result of spontaneous HR of the dsDNA donor with the *KRAS* gene over time. This is evident in the sgRNA-deficient control which showed no sign of spontaneous HR and targeted donor integration after 14 days. However, it is known that KRAS activity promotes cell growth and proliferation. The introduction of the activating point mutation into the *KRAS* gene results in the production of a constitutively active Kras protein mutant, which imbues the modified cells with a growth advantage (Loupakis et al., 2009). Therefore, *KRAS* mutant cells could have possibly been proliferating at a greater rate than the WT cells, resulting in the enrichment of the mutant cell population and the subsequent elevation in the detected levels of integrated dsDNA donor over time.

For the conclusive assessment of this dual cut strategy to successfully and efficiently facilitate the co-integration of both the point mutation and the expression cassette, the quantity of integrated expression cassette DNA (measured by primer set A) needs to be compared to that of the point mutation (measured by the allele specific primer set C). A ratio of 1 would indicate that there was a 100% efficiency in the co-transfer of these two genetic elements into the *KRAS* gene locus. However, this analysis could not be made, as vast amounts of residual episomal DNA from the transient transfection could still be detected in the cells, even after 14 days of sub-culture. Despite using the linearised form of the dsDNA donor for transfection, which was predicted to degrade much more efficiently episomally, compared to the plasmid format, approximately 20-30% of the episomal DNA detected was comprised of the dsDNA donor. This generated an overwhelming amount of signal noise that prevented the detection and quantification of successfully mutated *KRAS* sequences. In order

to successfully make these comparisons, less DNA should be used in the transfections so that it can be quickly diluted out after the modification. This may sacrifice the performance of this strategy, but qPCR should be sensitive enough to detect very rare events. Alternatively, the transfected cells could simply be cultured for a longer amount of time. Regardless of this, it has been demonstrated that the CRISPR/Cas9 system can mediate the introduction of a large DNA cassette into the *KRAS* gene locus. Furthermore, these recombinant cells can be isolated using fluorescent activated cell sorting or G418 selection and clonally expanded to generate cell lines for further study.

5. Conclusion

In this work, the use of the CRISPR/Cas9 system was explored to modify the human *KRAS* gene. The system was able to efficiently generate targeted DSBs and gene deletions, which were successfully repaired with small ssODNs or a custom designed dsDNA donor construct. The work highlights some important parameters that need to be considered when designing sgRNAs for the CRISPR/Cas9 system. The choice of guide sequence composition should aim to minimise base-pair interactions between the guide sequence and the RNA scaffold that stabilise the sgRNA structure. Furthermore, this sequence should not have a binding affinity to the target site that needs a temperature of greater than approximately 54ºC to be abolished. The repair of CRISPR/Cas9-mediated DSBs and micro-deletions can be easily and conveniently repaired using ssODNs for precise gene modifications. However, the gene target chosen may be preferentially repaired by NHEJ or HDR, affecting the efficiency of the desired outcome. The “dual cut” strategy was also functional to efficiently generate a high number of selectable mutant recombinant cells. The use of this strategy should overcome the gene conversion limitations in mammalian cells. Together, this work highlights some important considerations for the use of the CRISPR/Cas9 system for efficient and robust gene modification.

The CRISPR/Cas9 system is a novel genome engineering tool that has proven to be able efficiently facilitate the manipulation of DNA. The multiplex potential and the ease at which the CRISPR/Cas9 system can be applied is ideal for constructing biological models that more

accurately recapitulate the genetic elements of complex diseases such as cancer. However, the off-target effects of this system is concerning and must be considered, particularly for gene therapy applications. The CRISPR/Cas9 system is still a very new technology and will undoubtedly have to undergo substantial development to overcome current limitations. As our understanding of the CRISPR/Cas9 system improves, re-engineering of the system through rational design to improve target specificity of the system may be possible. Furthermore, new CRISPR/Cas systems are continuously being defined, providing a collection of systems with unique characteristics which may better satisfy the needs of genome engineering. Nevertheless, the functionality of the current system from *S. pyogenes*, for gene editing, has proven to be quite efficient and versatile. The expansion of this system to a general gene localisation platform, allows for the genome to be interrogated in even more ways. The CRISPR/Cas9 system currently provides a powerful, highly engineerable and cost-effective platform to efficiently interrogate and creatively explore the genome.

6. References

ANDERS, C., NIEWOEHNER, O., DUERST, A. & JINEK, M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature, 513, 569-73.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315, 1709-12.

BARRANGOU, R. & MARRAFFINI, L. A. 2014. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol Cell, 54, 234-44.

BOCH, J., SCHOLZE, H., SCHORNACK, S., LANDGRAF, A., HAHN, S., KAY, S., LAHAYE, T., NICKSTADT, A. & BONAS, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. Science, 326, 1509-12.

BROUNS, S. J., JORE, M. M., LUNDGREN, M., WESTRA, E. R., SLIJKHUIS, R. J., SNIJDERS, A. P., DICKMAN, M. J., MAKAROVA, K. S., KOONIN, E. V. & VAN DER OOST, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 321, 960-4.

CAPECCHI, M. R. 1989. Altering the genome by homologous recombination. Science, 244, 1288-92.

CARR, P. A. & CHURCH, G. M. 2009. Genome engineering. Nat Biotechnol, 27, 1151-62.

CARROLL, D. 2011. Genome engineering with zinc-finger nucleases. Genetics, 188, 773-82.

CARROLL, D. & BEUMER, K. J. 2014. Genome engineering with TALENs and ZFNs: repair pathways and donor design. Methods, 69, 137-41.

CHEN, B., GILBERT, L. A., CIMINI, B. A., SCHNITZBAUER, J., ZHANG, W., LI, G. W., PARK, J., BLACKBURN, E. H., WEISSMAN, J. S., QI, L. S. & HUANG, B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell, 155, 1479-91.

CHEN, F., PRUETT-MILLER, S. M., HUANG, Y., GJOKA, M., DUDA, K., TAUNTON, J., COLLINGWOOD, T. N., FRODIN, M. & DAVIS, G. D. 2011. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods, 8, 753-5.

CHEN, J. M., COOPER, D. N., CHUZHANOVA, N., FEREC, C. & PATRINOS, G. P. 2007. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet, 8, 762-75.

CHHABRA, A. 2014. Advances in genome engineering approaches. Adv Genet Eng, 3.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science, 339, 819-23.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature, 471, 602-7.

ELLIOTT, B., RICHARDSON, C., WINDERBAUM, J., NICKOLOFF, J. A. & JASIN, M. 1998. Gene conversion tracts from double-strand break repair in mammalian cells. Mol Cell Biol, 18, 93-101.

FONFARA, I., LE RHUN, A., CHYLINSKI, K., MAKAROVA, K. S., LECRIVAIN, A. L., BZDRENGA, J., KOONIN, E. V. & CHARPENTIER, E. 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res, 42, 2577-90.

FU, Y., SANDER, J. D., REYON, D., CASCIO, V. M. & JOUNG, J. K. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol, 32, 279-84.

GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A, 109, E2579-86.

GOLLOB, J. A., WILHELM, S., CARTER, C. & KELLEY, S. L. 2006. Role of Raf kinase in cancer: therapeutic potential of targeting the Raf/MEK/ERK signal transduction pathway. Semin Oncol, 33, 392-406.

HATOUM-ASLAN, A., SAMAI, P., MANIV, I., JIANG, W. & MARRAFFINI, L. A. 2013. A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. J Biol Chem, 288, 27888-97.

HEINEMANN, V., STINTZING, S., KIRCHNER, T., BOECK, S. & JUNG, A. 2009. Clinical relevance of EGFR- and KRAS-status in colorectal cancer patients treated with monoclonal antibodies directed against the EGFR. Cancer Treat Rev, 35, 262-71.

HENDEL, A., KILDEBECK, E. J., FINE, E. J., CLARK, J. T., PUNJYA, N., SEBASTIANO, V., BAO, G. & PORTEUS, M. H. 2014. Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. Cell Rep, 7, 293-305.

HOU, Z., ZHANG, Y., PROPSON, N. E., HOWDEN, S. E., CHU, L. F., SONTHEIMER, E. J. & THOMSON, J. A. 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci U S A, 110, 15644-9.

HSU, P. D., SCOTT, D. A., WEINSTEIN, J. A., RAN, F. A., KONERMANN, S., AGARWALA, V., LI, Y., FINE, E. J., WU, X., SHALEM, O., CRADICK, T. J., MARRAFFINI, L. A., BAO, G. & ZHANG, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol, 31, 827-32.

JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOULS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol, 43, 1565-75.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337, 816-21.

JINEK, M., JIANG, F., TAYLOR, D. W., STERNBERG, S. H., KAYA, E., MA, E., ANDERS, C., HAUER, M., ZHOU, K., LIN, S., KAPLAN, M., IAVARONE, A. T., CHARPENTIER, E., NOGALES, E. & DOUDNA, J. A. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science, 343, 1247997.

KHAN, I. F., HIRATA, R. K. & RUSSELL, D. W. 2011. AAV-mediated gene targeting methods for human cells. Nat Protoc, 6, 482-501.

KIM, H. & KIM, J. S. 2014. A guide to genome engineering with programmable nucleases. Nat Rev Genet, 15, 321-34.

LI, H. L., NAKANO T FAU - HOTTA, A. & HOTTA, A. 2014. Genetic correction using engineered nucleases for gene therapy applications. Dev Growth Differ.

LOUPAKIS, F., RUZZO, A., CREMOLINI, C., VINCENZI, B., SALVATORE, L., SANTINI, D., MASI, G., STASI, I., CANESTRARI, E., RULLI, E., FLORIANI, I., BENCARDINO, K., GALLUCCIO, N., CATALANO, V., TONINI, G., MAGNANI, M., FONTANINI, G., BASOLO, F., FALCONE, A. & GRAZIANO, F. 2009. KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. Br J Cancer, 101, 715-21.

MARRAFFINI, L. A. & SONTHEIMER, E. J. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science, 322, 1843-5.

MOJICA, F. J., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & ALMENDROS, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology, 155, 733-40.

MUSSOLINO, C. & CATHOMEN, T. 2012. TALE nucleases: tailored genome engineering made easy. Curr Opin Biotechnol, 23, 644-50.

MUSSOLINO, C., MORBITZER, R., LUTGE, F., DANNEMANN, N., LAHAYE, T. & CATHOMEN, T. 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res, 39, 9283-93.

NISHIMASU, H., RAN, F. A., HSU, P. D., KONERMANN, S., SHEHATA, S. I., DOHMAE, N., ISHITANI, R., ZHANG, F. & NUREKI, O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell, 156, 935-49.

ORITA, M., IWAHANA, H., KANAZAWA, H., HAYASHI, K. & SEKIYA, T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci U S A, 86, 2766-70.

PAEZ-ESPINO, D., MOROVIC, W., SUN, C. L., THOMAS, B. C., UEDA, K., STAHL, B., BARRANGOU, R. & BANFIELD, J. F. 2013. Strong bias in the bacterial CRISPR elements that confer immunity to phage. Nat Commun, 4, 1430.

PEREZ-PINERA, P., OUSTEROUT, D. G. & GERSBACH, C. A. 2012. Advances in targeted genome editing. Curr Opin Chem Biol, 16, 268-77.

PORTEUS, M. 2007. Using homologous recombination to manipulate the genome of human somatic cells. Biotechnol Genet Eng Rev, 24, 195-212.

PORTEUS, M. H. & CARROLL, D. 2005. Gene targeting using zinc finger nucleases. Nat Biotechnol, 23, 967-73.

QI, L. S., LARSON, M. H., GILBERT, L. A., DOUDNA, J. A., WEISSMAN, J. S., ARKIN, A. P. & LIM, W. A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 152, 1173-83.

QIN, J. Y., ZHANG, L., CLIFT, K. L., HULUR, I., XIANG, A. P., REN, B. Z. & LAHN, B. T. 2010. Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. PLoS One, 5, e10611.

RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013a. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell, 154, 1380-9.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013b. Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 8, 2281-308.

REUTER, J. S. & MATHEWS, D. H. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics, 11, 129.

REYON, D., TSAI, S. Q., KHAYTER, C., FODEN, J. A., SANDER, J. D. & JOUNG, J. K. 2012. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol, 30, 460-5.

RUSSELL, D. W. & HIRATA, R. K. 1998. Human gene targeting by viral vectors. Nat Genet, 18, 325-30.

SANTIAGO, Y., CHAN, E., LIU, P. Q., ORLANDO, S., ZHANG, L., URNOV, F. D., HOLMES, M. C., GUSCHIN, D., WAITE, A., MILLER, J. C., REBAR, E. J., GREGORY, P. D., KLUG, A. & COLLINGWOOD, T. N. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci U S A, 105, 5809-14.

SEDIVY, J. M. & DUTRIAUX, A. 1999. Gene targeting and somatic cell genetics--a rebirth or a coming of age? Trends Genet, 15, 88-90.

SHALEM, O., SANJANA, N. E., HARTENIAN, E., SHI, X., SCOTT, D. A., MIKKELSEN, T. S., HECKL, D., EBERT, B. L., ROOT, D. E., DOENCH, J. G. & ZHANG, F. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science, 343, 84-7.

SMITHIES, O., GREGG, R. G., BOGGS, S. S., KORALEWSKI, M. A. & KUCHERLAPATI, R. S. 1985. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature, 317, 230-4.

STERNBERG, S. H., REDDING, S., JINEK, M., GREENE, E. C. & DOUDNA, J. A. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature, 507, 62-7.

TOPALOGLU, O., HURLEY, P. J., YILDIRIM, O., CIVIN, C. I. & BUNZ, F. 2005. Improved methods for the generation of human gene knockout and knockin cell lines. Nucleic Acids Res, 33, e158.

VASILEVA, A., LINDEN, R. M. & JESSBERGER, R. 2006. Homologous recombination is required for AAV-mediated gene targeting. Nucleic Acids Res, 34, 3345-60.

WANG, H., YANG, H., SHIVALILA, C. S., DAWLATY, M. M., CHENG, A. W., ZHANG, F. & JAENISCH, R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell, 153, 910-8.

WANG, R. C., SMOGORZEWSKA, A. & DE LANGE, T. 2004. Homologous recombination generates T-loop-sized deletions at human telomeres. Cell, 119, 355-68.

WIRT, S. E. & PORTEUS, M. H. 2012. Development of nuclease-mediated site-specific genome modification. Curr Opin Immunol, 24, 609-16.

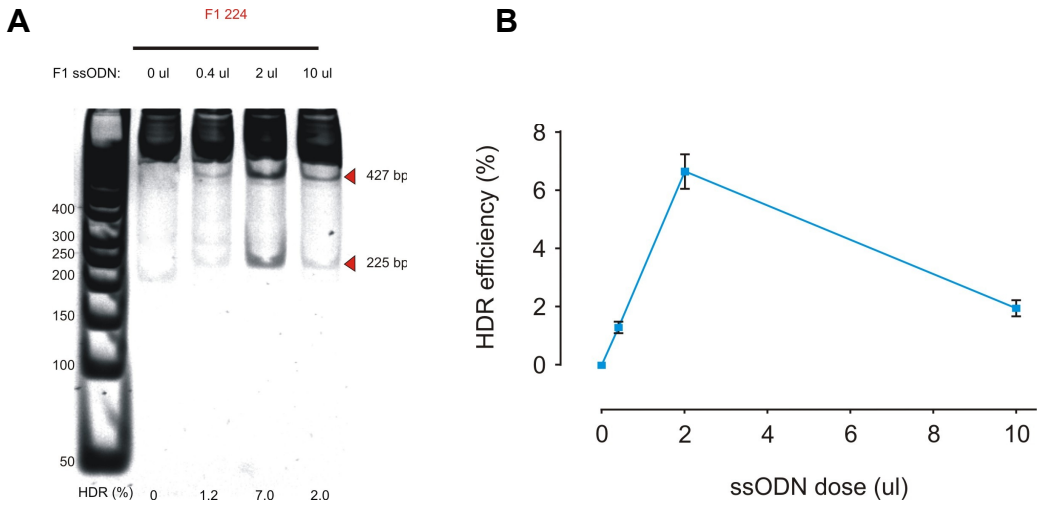
XIA, W., BRINGMANN, P., MCCLARY, J., JONES, P. P., MANZANA, W., ZHU, Y., WANG, S., LIU, Y., HARVEY, S., MADLANSACAY, M. R., MCLEAN, K., ROSSER, M. P., MACROBBIE, J., OLSEN, C. L. & COBB, R. R. 2006. High levels of protein expression using different mammalian CMV promoters in several cell lines. Protein Expr Purif, 45, 115-24.

YANG, L., GUELL, M., BYRNE, S., YANG, J. L., DE LOS ANGELES, A., MALI, P., AACH, J., KIM-KISELAK, C., BRIGGS, A. W., RIOS, X., HUANG, P. Y., DALEY, G. & CHURCH, G. 2013. Optimization of scarless human stem cell genome editing. Nucleic Acids Res, 41, 9049-61.

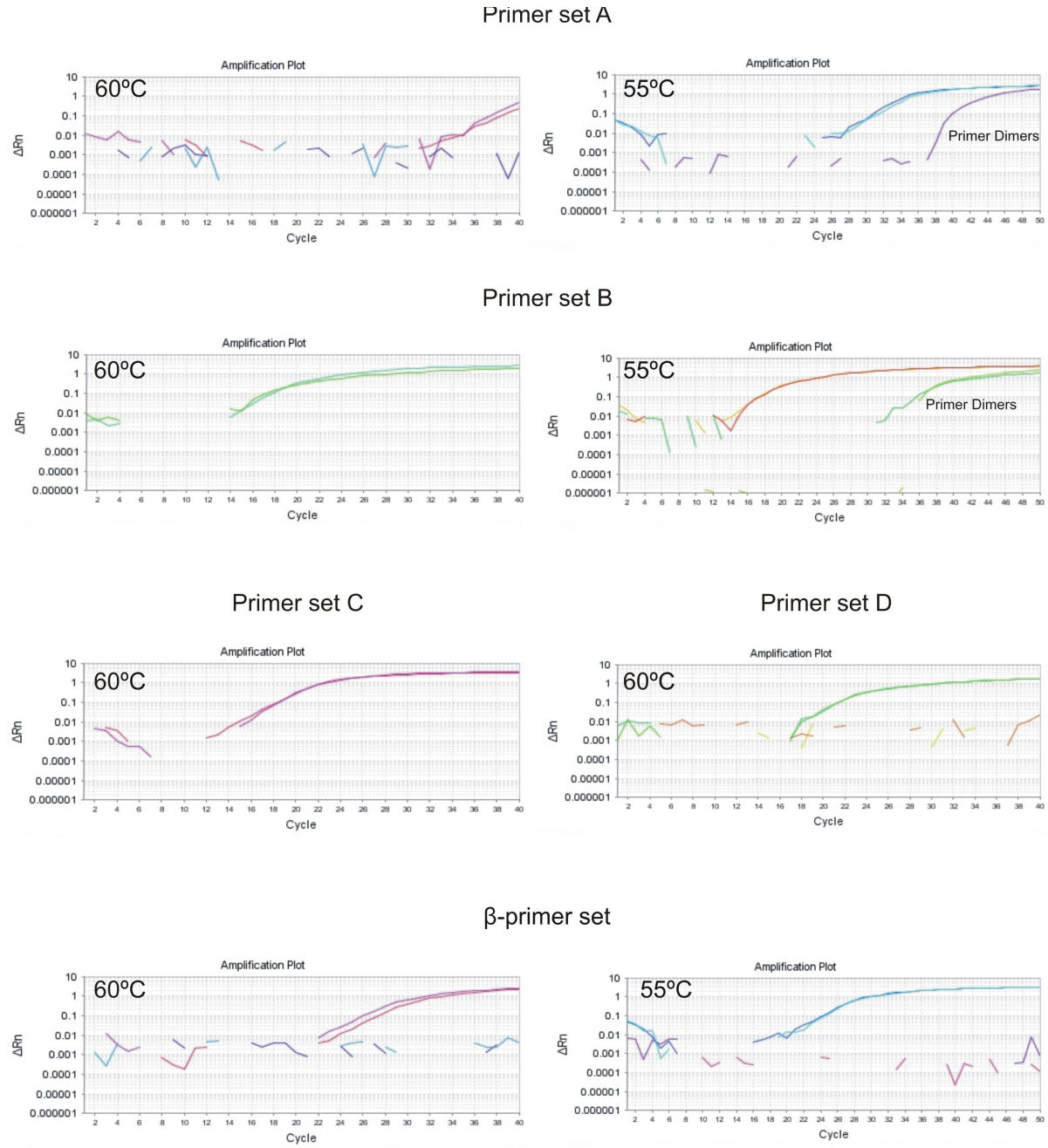
ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res, 31, 3406-15.

7. Appendix

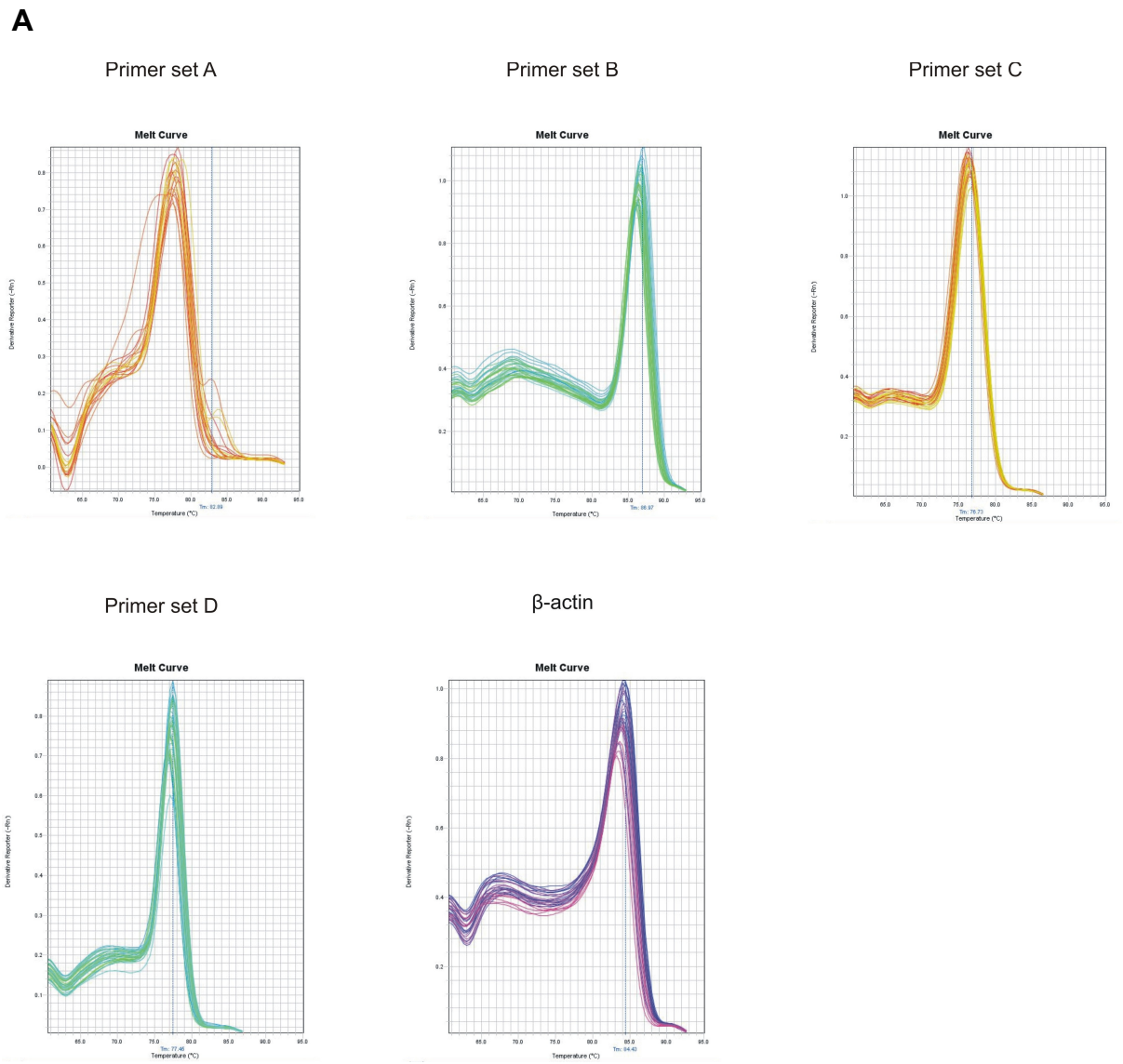
7.1 Supplementary data



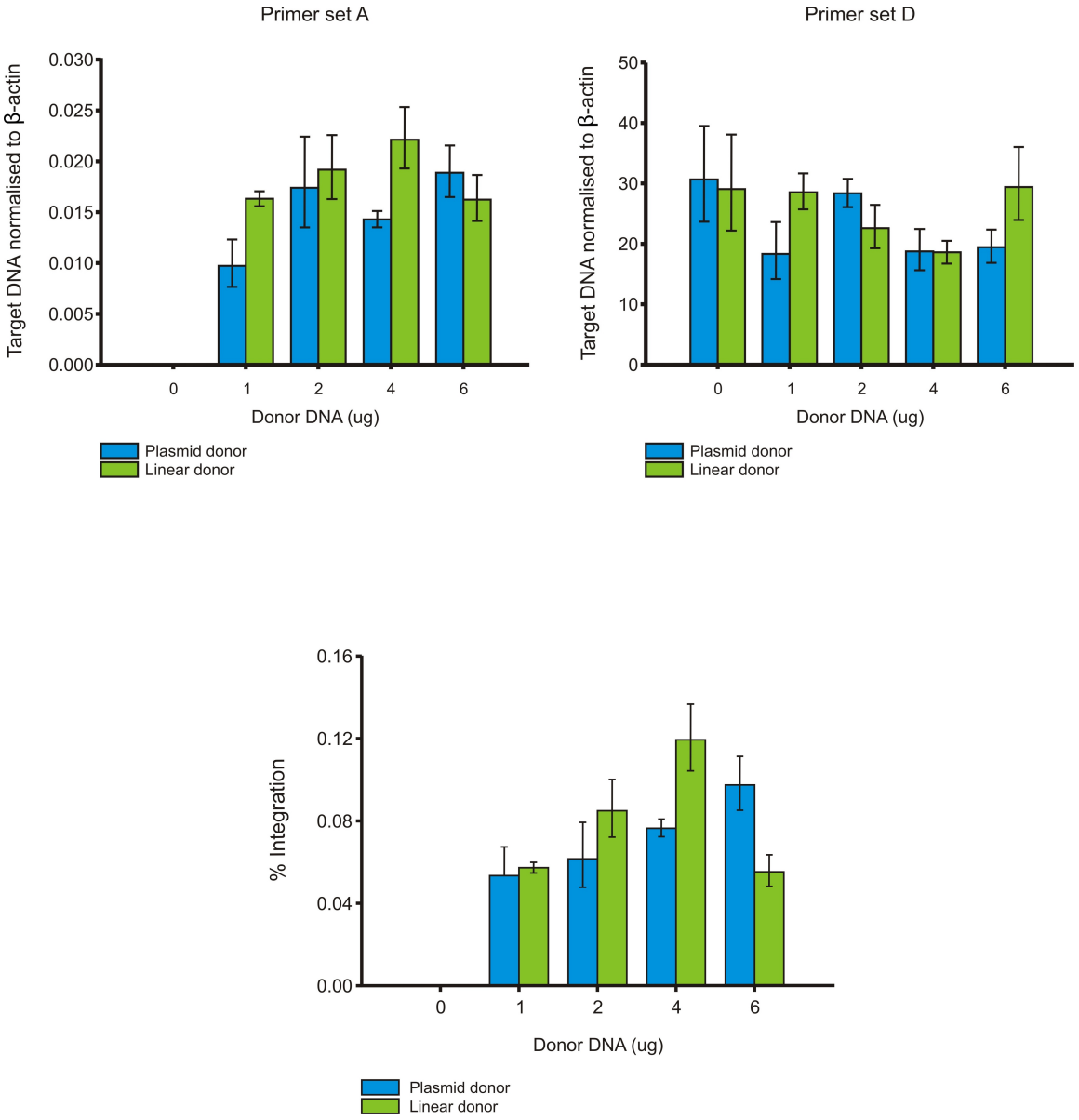
Supplementary figure 1: Optimisation of ssODN dosage for maximum HDR. A 5-fold dilution, ranging from 0.4 to 10 µl, of a 100 µM stock of F1 ssODN was tested for optimal HDR of sgRNA F1 224-mediated DSBs. 250 000 HEK 293 cells were transfected with the different doses and the gDNA was extracted from these cells 72 hours afterwards. PCR products amplified from this gene region were digested with the *EcoR*I restriction enzyme to detect HDR events. (**A**) A polyacrylamide gel showing the digested PCR products when different doses of ssODN are applied. Densitometry was used to quantify the digested bands and calculate the HDR percentage (**B**) Quantification of the digested bands show that HDR is linearly dose-dependent on the amount of ssODN used in the transfection between 0.4 and 2 µl. Increasing the amount of transfected ssODN to 10 µl led to a substantial decrease in HDR efficiency. From this 2 µl of ssODN (100 µM) was deemed to be optimal for HDR and was applied in subsequent experiments. The data here represents measurements made from a triplicate experiment (mean±SD).



Supplementary figure 2: Optimisation of the primer sets for qPCR. Amplification plots generated with the different primer sets at different annealing temperatures. This was done to determine the optimal annealing temperature for each primer set. Each reaction was carried out in duplicate and the optimal temperature was chosen on the basis of consistent amplification patterns from both reactions. The optimal annealing temperatures are as follows: Primer set A - 55ºC, primer set B - 55ºC, primer set C - 60ºC, primer set D - 60ºC , β-actin primer set - 55ºC. No template controls were set up for each reaction. No amplification was detected in these reactions for primer sets C, D and β-actin. However, primer set A and B showed some late amplification in these water controls, but these were regarded as primer dimers as they amplified more than five cycles later than the desired product.



Supplementary figure 3: Melt curve analysis to validate specific primer amplification. (A) At the end of each PCR, the generated DNA duplex products were melted and the SYBR green signal was monitored as it diminished. The melt curves show that each primer set was very specific, only amplifying a single specific product. Signs of the primer dimers can be clearly seen in the curves of primer set B and the β-actin as the small peak that preceded the major peak. (**B**) The products generated with each primer set were also visualised on an agarose gel to ensure that products of the correct size were being amplified.



Supplementary figure 4: The effect of different dsDNA donor concentrations and formats on HDR-mediated construct integration. For this experiment, 250 000 HEK 293 cells were transfected with different amounts of linear or plasmid dsDNA donor to repair the micro-deletion as a result of the simultaneous activity of sgRNA F1 224 and R2 228. The qPCR strategy described in Figure 16A was used to assess the effect of linear and plasmid dsDNA donor, introduced into the cells at different concentrations, on HDR and construct integration. Primer set A was used to measure the amount of integrated dsDNA donor and primer set B for the quantification of the total amount of *KRAS* DNA in each sample. The percentage integration, as calculated by taking the amount of integrated DNA as a percentage of total *KRAS* DNA, showed that HDR reaches a saturation point when 6 µg of the linear dsDNA donor was used for transfection. Transfections with the plasmid form of the dsDNA donor did not show any sign of HDR saturation, and seemed to follow a dose-dependent trend for the tested range. However, the percentage integration detected with equal amounts of linear and plasmid dsDNA donor showed that the linear format was always a slightly more efficient. This is probably as a result of a larger number of dsDNA molecules per microgram of DNA used, as the linear dsDNA construct is

significantly smaller that the plasmid form. This also explains the early saturation of HDR seen with the linear construct. Interestingly, the plasmid format is not significantly less efficient that the linear format, even though far fewer dsDNA donor molecules are introduced per microgram of transfected DNA. This could possibly be explained by the stability of circular DNA, allowing them to persist and function for a longer period of time compared to the linear dsDNA donor, which is probably degraded much more quickly. However, with this in mind, linear dsDNA constructs are more advantageous for gene editing experiments. A large number of molecules can be efficiently introduced in the cells using a relatively smaller amount of DNA for transfection. Furthermore, their activity as a donor construct is very transient, as they are degraded relatively quickly, which is desirable in applications in which their presence may interfere with the experiment.

7.2 Supplementary sequences

7.2.1 dsDNA donor components

Overlapping sequences are colour coded.

> 5′ HA

TAGCTGTTGCATATTGACTTCTAACACTTAGAGGTGGGGGTCCACTAGGAAAACTGTAACAATAAGA
GTGGAGATAGCTGTCAGCAACTTTTGTGAGGGTGTGCTACAGGGTGTAGAGCACTGTGAAGTCTCT
ACATGAGTGAAGTCATGATATGATCCTTTGAGAGCCTTTAGCCGCCGCAGAACAGCAGTCTGGCTAT
TTAGATAGAACAACTTGATTTTAAGATAAAAGAACTGTCTATGTAGCATTTATGCATTTTTCTTAAGCG
TCGATGGAGGAGTTTGTAAATGAAGTACAGTTCATTACGATACACGTCTGCAGTCAACTGGAATTTT
CATGATTGAATTTTGTAAGGTATTTTGAAATAATTTTTCATATAAAGGTGAGTTTGTATTAAAAGGTAC
TGGTGGAGTATTTGATAGTGTATTAACCTTAT

> eGFP fragment

G T T T G T A T T A A A A G G T A C T G G T G G A G T A T T T G A T A G T G T A T T A A C C T T A T
CACACAAAAAACCAACACACAGATGTAATGAAAATAAAGATATTTTATTTCTAGAGTATACGGACCG
TTACTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCAT
GTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGG
CAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCC
GTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATAT
AGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTC
GATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTG
TAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTG
AAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGGGTG
GTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTT
GCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCCGTCC
AGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCACCATGGGCCCAGGG
TTGGACTCGACGTCTCCCGCCAACTTGA

> SV40-NeoR fragment

ACGTCTCCCGCCAACTTGAGAAGGTCAAAATTCAAAGTCTGTTTCACGAAGAACTCGTCAAGAAGG
CGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTCAGC
CCATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCC
ACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAG
CAGGCATCGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGAAC
AGTTCGGCTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCA
TCCGAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAG
CGTATGCAGCCGCCGCATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGAC
AGGAGATCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACAACGTCG
AGCACAGCTGCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCCTGCAGT
TCATTCAGGGCACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCCTGCGCTGACAGCCG
GAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCATAGCCGAATAGCCTCTCCACC
CAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCATGGTGGCTCTAGCCTTAAGTTCGA
GACTGTTGTGTCAGAAGAATCAAGCTTTTTGCAAAAGCCTAGGCCTCCAAAAAAGCCTCCTCACTA
CTTCTGGAATAGCTCAGAGGCCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCAGCCA
TGGGGCGGAGAATGGGCGGAACTGGGCGGAGTTAGGGGCGGGATGGGCGGAGTTAGGGGCGGGA
CTATGGTTGCTGACTAATTGAGATGCATGCTTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTT
CCACACCTGGTTGCTGACTAATTGAGATGCATGCTTTGCATACTTCTGCCTGCTGGGGAGCCTGGGG
ACTTTCCACACCCTAACTGACACACATTCCACAGCTGGTTCTTTCCGCCTCAGAAGGTACCTAACCA
AGTTCCTCTTTCAGAGGTTATTTCAGGCCATGGTGCTGCGCGTGTGACATGTTCTAATATAGTCACAT
TTTCATTA

> KRS fragment

GTGTGACATGTTCTAATATAGTCACATTTTCATTATTTTTATTATAAGGCCTGCTGAAAATGACTGAATA
TAAACTTGTGGTAGTTGGAGCTGTTGGCGTAGGCAAGAGTGCCTTGACGATACAGCTAATTCAGAAT
CATTTTGTAGACGAATATGATCCCACAATAGAGGTAAATCTTGTTTTAATATGC

> 3′ HA

AATCATTTTGTAGACGAATATGATCCCACAATAGAGGTAAATCTTGTTTTAATATGCATATTACTGGTG
CAGGACCATTCTTTGATACAGATAAAGGTTTCTCTGACCATTTTCATGAGTACTTATTACAAGATAATT
ATGCTGAAAGTTAAGTTATCTGAAATGTACCTTGGGTTTCAAGTTATATGTAACCATTAATATGGGAA
CTTTACTTTCCTTGGGAGTATGTCAGGGTCCATGATGTTCACTCTCTGTGCATTTTGATTGGAAGTGT
ATTTCAGAGTTTCGTGAGAGGGTAGAAATTTGTATCCTATCTGGACCTAAAAGACAATCTTTTTATTG
TAACTTTTATTTTTATGGGTTTCTTGGTATTGTGACATCATATGTAAAGGTTAGATTTAATTGTACTAGT
GAAATATAATTGTTTGATGG

7.3 Reagents and recipes

7.3.1. Freeze medium

7.5 ml DMSO (Sigma Aldrich, USA)
32.5 ml DMEM (Lonza, USA)
10 ml FBS (Hyclone, USA)
100 µl Pen/Strep (Lonza, USA)

The DMSO, DMEM, FBS and Pen/Strep in a 50 ml tube. The mixture was filter sterilised with a 0.4 µm syringe filter and stored in the dark at 4ºC.

7.3.2 Transformation buffer

15 ml Glycerol (SaarChem, South Africa)
1.47 g $CaCl_2.2H_2O$ (Sigma Aldrich, USA)
302.4 mg PIPES.HCl (Boehringer Manheim, Germany)
NaOH (Sigma Aldrich, USA)
ddH_2O

The glycerol, $CaCl_2.2H_2O$ and PIPES.HCl were added together and made up to 80 ml with ddH_2O. The solution was mixed with a magnetic stirrer. Once the $CaCl_2.2H_2O$ and PIPES.HCl had dissolved and the solution was homogeneous, NaOH was added dropwise until the pH was exactly 7.0. The solution was made up to 100 ml with ddH_2O, autoclaved and stored at -20ºC.

7.3.3 Luria Bertani (LB) broth

10 g Bacto-tryptone (Sigma Aldrich, USA)
10 g NaCl (SaarChem, South Africa)
5 g Yeast extract (Sigma Aldrich, USA)

ddH_2O

These were dissolved in 500 ml of ddH_2O and then made up to 1 L. The broth was then autoclaved and stored at room temperature.

7.3.4 Ampicillin 1 000X stock (100 mg/ml)

5 g Ampicillin (Sigma, Aldrich, USA)
25 ml of 100% Ethanol
25 ml ddH_2O

The ampicillin was completely dissolved in a mixture of 25 ml of 100% ethanol and 25 ml ddH_2O by vortexing. The ampicillin stock was stored at -20ºC.

7.3.5 Ampicillin agar plates

10 g Bacto-tryptone (Sigma Aldrich, USA)
10 g NaCl (SaarChem, South Africa)
5 g Yeast extract (Sigma Aldrich, USA)
12.5 g Agar (Sigma Aldrich, USA)
1 ml Ampicillin 1 000X stock (100mg/ml)
ddH_2O

The bacto-tryptone, NaCl, yeast extract and agar were dissolved in 500 ml of ddH_2O, made up to 1 L and autoclaved. The sterilised mixture was allowed to cool and 1 ml of ampicillin 1 000X stock (100mg/ml) was added for a final concentration of 10 µl/ml. This was poured into bacterial culture petri dishes and allowed to solidify. The plates were then stored at 4ºC.

7.3.6 Ampicillin-LB broth

1 L LB broth
1 ml Ampicillin 1 000X stock (100 mg/ml)

1 ml of ampicillin 1 000 X stock (100 mg/ml) was added to 1 L of cooled LB broth (for a final concentration of 10 µl/ml). This was stored at 4ºC.

7.3.7 TAE buffer (50X)

242 g Tris-base (Sigma Aldrich, USA)
57.1 ml of 100% acetic acid (Merck, Germany)
100 ml of 0.5 M sodium EDTA (pH 8.0)
ddH_2O

The components were mixed together, dissolved in ddH_2O and made up to 1 L. For a 1X TAE working stock, 20 ml of the 50X TAE solution was mixed with 980 ml of ddH_2O.

7.3.8 TBE buffer (5X)

108 g Tris-base (Sigma Aldrich, USA)
55 g of Boric acid (Merck, Germany)
40 ml of 0.5 M sodium EDTA (pH 8.0)
ddH_2O

The components were mixed together, dissolved in ddH_2O and made up to 1 L. For a 1X TBE working stock, 100 ml of the 5X TBE solution was mixed with 900 ml of ddH_2O.

7.3.9 EDTA solution (0.5M, pH 8.0)

93.06 g EDTA (Merck, Germany)
ddH_2O
NaOH (Sigma Aldrich, USA)

The EDTA was dissolved in 400 ml of ddH_2O using a magnetic stirrer. The pH of the solution was adjusted to 8.0 with NaOH. The solution was then made up to 500 ml with ddH_2O.

7.3.10 PBS (1X)

PBS tablets (Sigma Aldrich, USA)
ddH_2O

1 PBS tablet was dissolved in 250 ml ddH_2O. The solution was autoclaved and store at 4ºC.

7.3.11 Tris-HCl (1M, pH 8.0)

60.55 g Tris-base (Sigma Aldrich, USA)
HCl (Merck, Germany)
ddH_2O

Dissolve Tris-base in 400 ml of ddH_2O. Add HCl until the pH reaches 8.0. Make up the solution to 500 ml with ddH_2O.

7.3.12 TE buffer (10X)

100 ml of 1M Tris-HCl (pH 8.0)
20 ml of 0.5M EDTA (pH 8.0)
ddH_2O

The Tris-HCl and EDTA solutions were mixed and made up to 1 L with ddH_2O. The solution was filter sterilised with a 0.4 µm syringe filter and then autoclaved.

7.3.13 Autoclaving conditions

All autoclaving was performed at 121ºC at approximately 104 kPa for 20 minutes.

7.4 Protocols

7.4.1 Competent *E. coli* Stbl3 cells

8 ml of LB broth was inoculated with *E. coli* Stbl3 cells. These cultures were grown in a shaking incubator (180 rmp) for 16 hours at 37ºC. 2 ml of this culture was added to 48 ml of LB broth and placed back into the incubator until the cells reached the logarithmic growth phase. This was estimated using a spectrophotometer (OD_{600} = 0.4 - 0.6). The 50 ml culture was then removed and placed immediately on ice. The cells were then centrifuged at 4ºC at 355×g (Boeco U-320R, Germany) for 10 minutes, after which the supernatant was poured off, and the pellet was very gently re-suspended in 1 ml of transformation buffer. Another 4 ml of transformation buffer was added and the cells were placed on ice for 30 minutes. The cells were pelleted by centrifugation for 10 minutes at 160×g rpm (Boeco U-320R, Germany) at 4ºC. The supernatant was poured off and the pellet was gently resuspended in 1 ml of transformation buffer. 100µl aliquots of competent of cells were stored at -80°C.

7.4.2 QIAprep Spin Miniprep Kit

A 10 ml culture of transformed bacteria was cultured for 14-16 hours at 37ºC in a shaking incubator (180 rpm). The cells were harvested by centrifugation at 3 900×g (Boeco U-320R, Germany) for 10 minutes at 4ºC. The supernatant was discarded and the pellet was resuspended in 250 µl Buffer P1 and transferred to a microcentrifuge tube. 250 µl of Buffer P2 was added and the tubes were inverted 24 times and incubated at room temperature for 5 minutes exactly. 350 µl of Buffer N3 was added and mixed by immediately inverting the tube 24 times. The precipitant was collected by centrifuging the mixture for 10 minutes at 14 000×g (Eppendorf 5402). The supernatant was then applied to a QIAprep spin column. The column was centrifuged for 30 seconds at 14 000×g (Eppendorf 5402) and the flow-through was discarded. The spin column was washed with 500 µl of Buffer PB by centrifugation at 14 000×g (Eppendorf 5402) for 30 seconds and discarding the flow-through. The column was washed again with 750 µl of Buffer PE by centrifugation at 14 000×g (Eppendorf 5402) for 30 seconds and discarding the flow-through. Residual wash buffer was removed by

centrifuging the tubes again for 1 minute. The QIAprep column was placed in a clean microcentrifuge tube and the DNA was eluted by adding 100 µl of Buffer EB to the centre of the column and incubating at room temperature for 1 minute. The DNA was collected by centrifugation at 16 000×g (Eppendorf Centrifuge 5402, Germany) for 1 minute.

7.4.3 Qiagen Plasmid Midi Kit

A 100 ml culture of transformed bacteria was culture for 14-16 hours at 37ºC in a shaking incubator (180 rpm). The cells were harvested by centrifuging 50 ml of the culture at a time, at 3 900×g (Boeco U-320R, Germany) for 15 minutes at 4ºC. The supernatant was discarded and the remaining 50 ml of the culture was added to the pellet and centrifuged again. The supernatant was then discarded and 4 ml of Buffer P1 was used to resuspend the pellet. 4 ml of Buffer P2 was added and mixed by vigorously inverting the mixture 24 times. The cells were incubated for 5 minutes at room temperature. 4 ml of pre-chilled Buffer P3 was then added and mixed by inverting the tube 24 times again. The mixture was then incubated on ice for 15 minutes before being centrifuged at 3 900×g (Boeco U-320R, Germany) for 30 minutes at 4ºC. The supernatant was then carefully pipetted into a clean 15 ml tube and centrifuged again at 3 900×g (Boeco U-320R, Germany) for 15 minutes at 4ºC. In this time, an Qiagen-tip 100 was equilibrated by applying 4 ml of Buffer QBT and allowing for the column to empty by gravity flow. The supernatant from the centrifuged samples were added to the column and allowed to run through the resin by gravity flow. The Qiagen-tip 100 was then washed twice with 10 ml of Buffer QC each time. The DNA was eluted from the column by adding 5 ml of Buffer QF and allowing it to run into a clean 15 ml tube. The DNA was precipitated by adding 3.5 ml of room-temperature isopropanol and vortexing. The precipitated the DNA was collected by centrifugation at 3 900×g (Boeco U-320R, Germany) for 30 minutes at 4 ºC. The supernatant was carefully decanted and the remaining pellet was washed with 2 ml of 70% ethanol. The DNA was recollected by centrifugation at 3 900×g (Boeco U-320R, Germany) for 10 minutes at 4ºC. The supernatant was carefully decanted and the pellet was dried by placing the tube in a 50ºC heating block for 10 minutes. The pellet was then resuspended in 100 µl TE buffer.

7.4.4 QIAamp DNA Mini Kit

Cells grown in a 6-well plate were harvested by washing the cells in each well with 2 ml of 1X PBS and trypsinised with 1 ml Trypsin-EDTA (Hyclone, Thermo Scientific). The cells were incubated for 2-3 minutes at 37ºC, before 500 µl of complete DMEM was added to neutralise the trypsin. The cell suspension was then transferred to a 1.5 ml microcentrifuge tube and the cells were collected by centrifugation at 260×g for 5 minutes (Eppendorf Centrifuge 5402, Germany). The supernatant was then discarded, and the pellet was resuspended in 1 ml of 1X PBS and centrifuged again at 260×g for 5 minutes. This wash step was repeated. The final cell pellet was resuspended in 200 µl of 1X PBS. 20 µl of Proteinase K was added followed by 200 µl of Buffer AL. This was mixed by pulse vortexing the microcentrifuge tube for 15 seconds and incubated at 56ºC for 10 minutes. Droplets on the side of the tube were collected by brief centrifugation. 200 µl of 100% ethanol pre-chilled in a -80ºC freezer was added and mixed by pulse vortexing for 15 seconds. The microcentrifuge tube was briefly centrifuged again to collect droplets on the side of the tube. The mixture was then applied to a QIAamp Spin Column and centrifuged at 8 000×g for 1 minutes (Eppendorf Centrifuge 5402, Germany). The column was placed in a clean 2 ml collection tube and 500 µl of Buffer AW1 was added and centrifuged at 8 000×g for 1 minute (Eppendorf Centrifuge 5402, Germany). The column was placed in a clean 2 ml collection tube and 500 µl of Buffer AW2 was added and centrifuged at 16 000×g for 3 minutes (Eppendorf Centrifuge 5402, Germany). The column was then carefully placed in a clean 1.5 ml microcentrifuge tube and the DNA was eluted by adding 200 µl of Buffer AE and incubating at room temperature for 5 minutes. The DNA was collected by centrifugation at 8 000×g for 1 minute.

7.4.5 GeneJet Gel Extraction Kit

The DNA band of interest was excised from the agarose gel using a clean scalpel and placed into a pre-weighed 1.5 ml microcentrifuge tube. The gel slice was then weighed and a 1:1 volume of Binding Buffer was added to the gel slice (volume:weight). The gel slice was melted by incubating the mixture for 10 minutes at 50-60ºC and vortexing periodically. The mixture was then transferred to a GeneJet purification column and centrifuged for 1 minute at

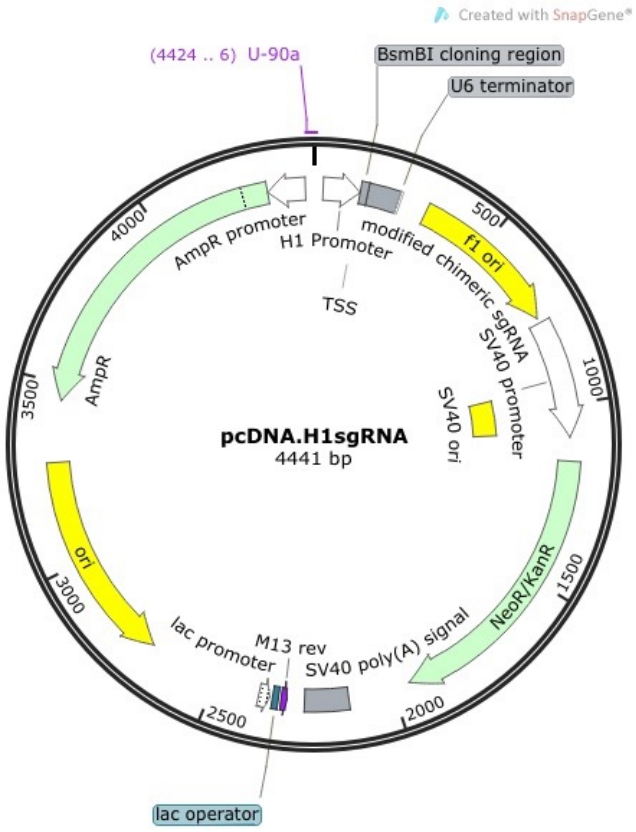
8 000×g (Eppendorf Centrifuge 5402, Germany). The flow-through was discarded and 700 µl of Wash Buffer was added to the column and it was centrifuged for another minute at 8 000×g (Eppendorf Centrifuge 5402, Germany). The flow-through was discarded and the residual buffer was removed by centrifugation at 8 000×g (Eppendorf Centrifuge 5402, Germany) for 1 minute again. The DNA was then eluted by transferring the GeneJet purification column into a clean 1.5 ml microcentrifuge tube and adding 50 µl of pre-warmed (65ºC) Elution Buffer to the centre of the purification column membrane. This was incubated for 1 minute at room temperature before being centrifuged for 1 minute at 16 000×g (Eppendorf Centrifuge 5402, Germany).

7.4.6 GeneJet PCR Purification Kit

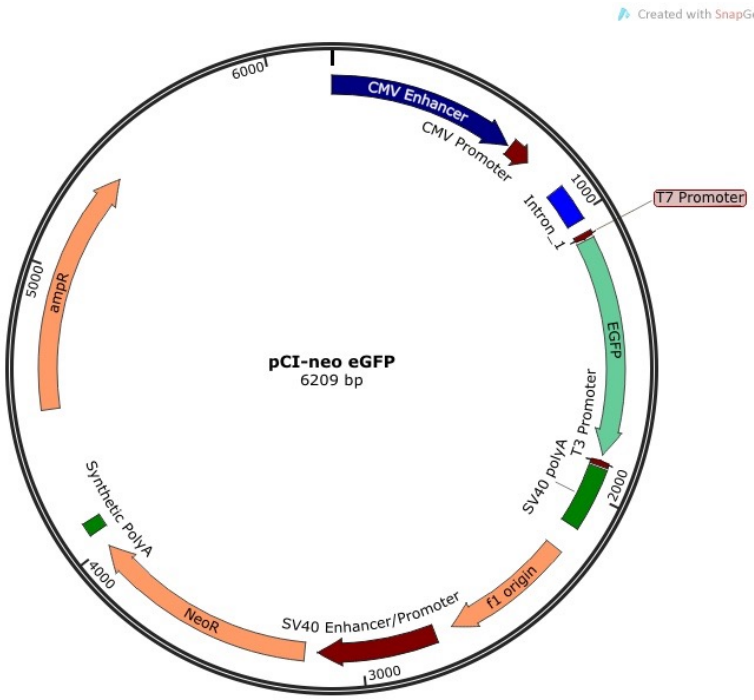
This was carried out as described in section 7.4.5, however a 1:1 ratio of PCR and Binding Buffer volume was used instead.

7.5 Plasmid maps

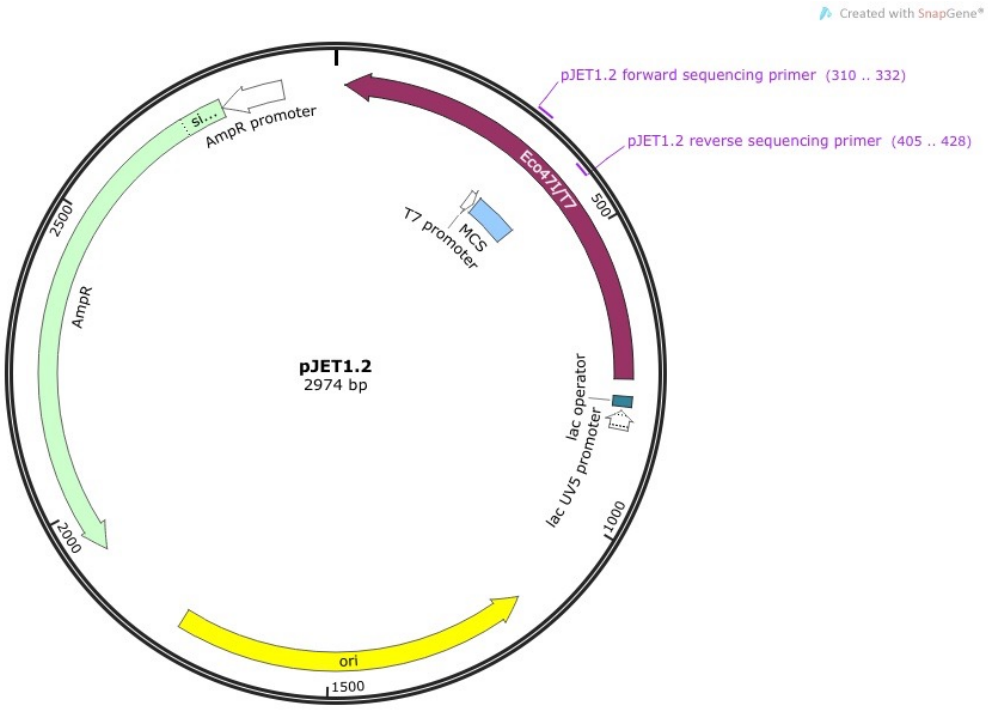
7.5.1 pcDNA.H1.sgRNA



7.5.2 pCI-neo-eGFP



7.5.3 pJet 1.2



7.5.4 pX330



7.6 Ethics approval waiver

University
of the Witwatersrand,
Johannesburg

Human Research Ethics Committee (Medical)
(formerly Committee for Research on Human Subjects (Medical))
Secretariat: Research Office, Room SH10005, 10th floor, Senate House • Telephone: +27 11 717-1234 • Fax: +27 11 339-5708
Private Bag 3, Wits 2050, South Africa

Ref: W-CJ-13-2

30/01/2013

TO WHOM IT MAY CONCERN:

Waiver: This certifies that the following research does not require clearance from the Human Research Ethics Committee (Medical).

Investigator: **Dr C Penny, Mr E Fok (student no 300204).**

Project title: **Discovering novel microRNA signatures and their potential use in the mutational analysis of *KRas* and *Braf* genes in cancer.**

Reason: This is a wholly laboratory study using commercial cell lines including HEK, HEF, HT29, SW480 and DLD1. No humans are involved.

UNIVERSITY OF THE WITWATERSRAND
PROF PE CLEATON-JONES
HREC (MEDICAL)
2013 -01- 30
JOHANNESBURG

Professor Peter Cleaton-Jones
Chair: Human Research Ethics Committee (Medical)

copy: Anisa Keshav, Research Office, Senate House, Wits