

https://dspace.org/

# Gene editing of human CCR5 by an enhanced CRISPR /SaCas9 nickase system

**Hanson, Britt**

**2017**
https://hdl.handle.net/10539/24797

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

# Gene editing of human *CCR5* by an enhanced CRISPR/*Sa*Cas9 nickase system

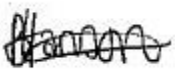
Britt Hanson

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine by Dissertation

**Johannesburg, November 2017**

**Declaration**

I, Britt Hanson, declare that this Dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in Medicine by Dissertation at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University. I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong. I have followed the required conventions in referencing the thoughts and ideas of others. I understand that the University of the Witwatersrand may take disciplinary action against me if there is belief that this is not my own unaided work or that I have failed to acknowledge the source of ideas or words in my writing.

Signature: 

On this 1st day of November 2017

## Publications and Presentations

*Publications*

- Hanson, B. (2016) Necroptosis: A new way of dying? *Cancer Biology and Therapy*. Sep;17(9):899-910. doi: 10.1080/15384047.2016.1210732. This review was initially submitted as an essay for my Honours course (graded 87%) and it was encouraged that I submit it for publication, as the sole author and with little external assistance.
- Kraberger, S., et al. (2017) Molecular diversity, geographic distribution and host range of monocot-infecting mastreviruses in Africa and surrounding islands. *Virus Research.* Jul 4;238:171-178. doi: 10.1016/j.virusres.2017.07.001. [Epub ahead of print]. I am a co-author on this paper for acknowledgement of the research contributions made as a component of the MCB Research Project in the final year of my Bachelor of Science degree at the University of Cape Town.

*Presentations*

- Hanson, B., Papathanasopoulos, M., Weinberg, M.S. Genetic Targeting of the Human *CCR5* Gene using a CRISPR/*Sa*Cas9 Nickase System. South African Society of Biochemistry and Molecular Biology (SASBMB) 25$^{th}$ Congress, 10 – 13 July 2016, East London, South Africa.
- Hanson, B., Papathanasopoulos, M., Weinberg, M.S. Genetic Targeting of the Human *CCR5* Gene by an Enhanced CRISPR/*Sa*Cas9 Nickase System. University of the Witwatersrand Faculty of Health Sciences Biennial Research Day, 1 September 2016, Johannesburg, South Africa.
- Hanson, B., Papathanasopoulos, M., Weinberg, M.S. & his researchers in the HPRU (University of the Witwatersrand) as well as Scripps Research Institute (USA) involved. Adapting the CRISPR/Cas9 Technology for HIV Cure Research. HIV Cure Research Workshop, 6 & 7 October 2016, Stellenbosch, South Africa.

**Abstract**

The human immunodeficiency virus (HIV) infects 37 million people and causes acquired immune deficiency syndrome (AIDS) that claims the lives of millions annually. Combination antiretroviral therapy has significantly reduced morbidity and mortality rates but latent viral reservoirs established during acute infection preclude drug- and immune-mediated clearance, necessitating lifelong treatment. HIV type 1 (HIV-1) entry is mediated by the interaction with human cluster of differentiation 4 and a co-receptor, mainly C-C chemokine receptor 5 (CCR5) or alternatively C-X-C chemokine receptor 4. A naturally occurring 32 base pair deletion in the *CCR5* gene (CCR5-Δ32) renders the receptor non-functional and sequesters native CCR5 in intracellular compartments. CCR5-Δ32 homozygotes have no known immune abnormalities and are resistant to R5-tropic (CCR5-utilising) HIV-1 infection while heterozygotes display reduced infectivity and prolonged progression to AIDS. CCR5 knock out has thus emerged as a promising model for HIV-1 functional cure development. Highly specific targeted genome engineering has provided the means to selectively and permanently manipulate the genetic code that governs all cellular processes. CRISPR/Cas9 is a novel versatile and powerful gene editing tool that employs a short programmable single guide RNA (sgRNA) for delivery of the Cas9 endonuclease to the desired DNA target site, inducing formation of a double stranded break and insertion, deletion and substitution mutations via the error-prone non-homologous end joining DNA repair pathway. The *Staphylococcus aureus* derived CRISPR/Cas9 system is AAV vector deliverable and has a high level of on target specificity. This system, as well as nickase and 'enhanced specificity' variants, were employed for disruption of the human *CCR5* gene in cell culture. The cleavage activity of ten sgRNAs with the nuclease and five sgRNA pairs with the nickase were assessed for cleavage activity using the T7EI, TIDE and ddPCR™ assays. Deep sequencing was carried out to characterise nuclease and nickase-mediated indels and assess off target activity of two sgRNAs at five highly similar genomic target sites each. The *Sa*Cas9 cleavage efficiency varied across sgRNA target sites and was reduced by the nickase and 'enhanced specificity' modifications, suggesting a trade-off between on target specificity and cleavage efficiency. Finally the effect of *CCR5* target site indels on mRNA levels and R5-tropic HIV-1 pseudovirus infectivity was assessed by qRT-PCR and a TZM-bl luciferase assay respectively to determine the functional effect of gene editing. The reason for the observed reduction in mRNA levels was inconclusive, however, the desired phenotypic effect of preclusion of R5-tropic HIV-1 infection was achieved using the WT *Sa*Cas9 nuclease,

nickase and ‘enhanced specificity’ nuclease systems. This study contributes towards understanding the functionality of novel highly specific CRISPR/Cas9 variants with the capacity to be delivered as a gene therapy both *in vivo* and *ex vivo* using AAV vectors for the functional cure of HIV-1.

**Acknowledgments**

I would like to express my gratitude to the following people and organisations:

- Assoc. Prof. Marco Weinberg, for your unending support from across the globe. You have guided me every step of the way and I have learnt more than I could ever have hoped from a Supervisor.
- Assoc. Prof. Maria Papathanasopoulos for the personal source of critical advice and oversight in Marco's physical absence. Thank you for keeping me on track and supporting my research in your laboratory.
- Dr Tristan Scott, for playing the incredible role of unofficial co-supervisor that you did. Thank you for your endless help, advice, patience and knowledge.
- Dr Stanford Kwenda, for your immeasurable assistance and guidance with my Bioinformatics analysis. You had no obligation to help and yet you spent hours of your free time willingly assisting me.
- The HIV Pathogenesis Research Laboratory, every one of you made me feel welcome and played an instrumental role in helping me through first my Honours and then my Masters research projects. Special thanks go to Dr Lorraine Mutsvunguma for the hands-on help and advice, Dr Mark Killick for the HIV infection assay, Melanie Moodie for looking after my cells when I was away and Dean Harris for the Next Generation Sequencing assistance.
- Prof. Patrick Arbuthnot and colleagues from the Antiviral Gene Therapy Research Unit for the assistance when in need of reagents or equipment and the unfaltering will to provide.
- The Wits Health Consortium (PTY) LTD RNA Therapeutics Research Group, Poliomyelitis Research Foundation and National Research Foundation for bursary funding.
- Lastly, to my family and friends for your endless support, sacrifice and love throughout my studies, without which I would not have grown to be the young and ambitious Scientist that I am.

*Funding for the research conducted in this project was provided by the Medical Research Council (MRC) Strategic Health Innovations Partnerships (SHIP) grant and the National Research Foundation (NRF).*

## Table of Contents

Declaration ..... i
Publications and Presentations ..... ii
Abstract ..... iii
Acknowledgments ..... v
Table of Contents ..... vi
List of Figures ..... xi
List of Tables ..... xii
Abbreviations ..... xiii
1. Introduction ..... 1
1.1. Epidemiology of the global HIV pandemic ..... 1
1.2. HIV-1 infective mechanisms ..... 2
1.3. Genome engineering for HIV-1 functional cure development ..... 5
1.3.1. Rational genome engineering technologies ..... 6
1.3.2. CRISPR/Cas9 as a natural defence mechanism ..... 7
1.3.3. A brief overview of CRISPR/Cas9 system functioning ..... 8
1.3.4. Molecular biology of the CRISPR/*Sa*Cas9 system ..... 10
1.4. Generation of highly specific CRISPR/*Sa*Cas9 variants ..... 13
1.5. CRISPR/Cas9 for HIV-1 functional cure development ..... 16
1.6. *CCR5* as a promising anti-HIV-1 gene therapy target ..... 17
1.6.1. Structure of the human *CCR5* gene and protein ..... 18
1.6.2. The CCR5-Δ32 resistance mutation ..... 21
1.6.3. Mimicking the CCR5-Δ32 phenotype by gene editing ..... 22
1.7. Aims and Objectives ..... 23
2. Materials and Methods ..... 25
2.1. Plasmids and cell lines used in this study ..... 25
2.1.1. Plasmid constructs ..... 25
2.1.2. Cell lines ..... 25
2.2. Cloning of *CCR5*-targeted *Sa*Cas9 sgRNAs; the *Sa*Cas9 D10A-mutated nickase and 'enhanced specificity' *Sa*Cas9 nuclease and nickase variants ..... 26
2.2.1. Generation of the *Sa*Cas9 D10A-mutated nickase construct by site-directed mutagenesis and Gibson Assembly cloning ..... 26
2.2.2. Production of a panel of *CCR5*-specific *Sa*Cas9 sgRNAs ..... 29

2.2.3. Generation of the 'enhanced specificity' *Sa*Cas9 nuclease and nickase constructs by Gibson Assembly cloning ..... 32
2.2.4. Production of the *CCR5*-targeting short hairpin RNA 1005 ..... 32
2.3. Transfections of HEK293T and TZM-bl cells ..... 35
2.3.1. HEK293T cell transfections ..... 36
2.3.2. TZM-bl cell transfections ..... 36
2.4. Quantification of indel frequency by the T7E1 assay ..... 37
2.4.1. Transfection and DNA extraction ..... 37
2.4.2. PCR and re-annealing of products for heteroduplex formation ..... 37
2.4.3. Heteroduplex cleavage and agarose gel electrophoresis ..... 38
2.5. Quantification of indel frequency by the TIDE assay ..... 39
2.6. Quantification of indel frequency by the ddPCR™ drop-off assay ..... 40
2.6.1. Transfection and gDNA extraction ..... 40
2.6.2. ddPCR™ probe and primer design ..... 40
2.6.3. Droplet generation and PCR ..... 42
2.6.4. Analysis of ddPCR™ drop-off assay results ..... 43
2.7. Indel quantification and detection of off target effects by deep sequencing ..... 45
2.7.1. Transfection and gDNA extraction ..... 45
2.7.2. Prediction and selection of potential off target sites ..... 46
2.7.3. PCR and pooling for deep sequencing ..... 48
2.7.4. Illumina deep sequencing and indel analysis ..... 51
2.8. Quantification of relative *CCR5* mRNA levels by qRT-PCR ..... 53
2.8.1. Transfection and RNA extraction ..... 53
2.8.2. gDNA wipeout and reverse transcription ..... 53
2.8.3. Real-time qRT-PCR ..... 54
2.8.4. Analysis of qRT-PCR results ..... 55
2.9. Assessment of R5-tropic HIV-1 infectivity by the TZM-bl luciferase assay ..... 56
2.9.1. Transfection and pseudovirus infection ..... 56
2.9.2. Luminescence acquisition and analysis of TZM-bl assay results ..... 57
2.10. Statistical analyses ..... 57
3. Results ..... 58
3.1. Assessment of CRISPR/*Sa*Cas9 nuclease and nickase activity with a panel of *CCR5*-specific sgRNAs ..... 58
3.1.1. Generation of the *Sa*Cas9-D10A nickase construct ..... 58

3.1.2. Production of a panel of human *CCR5*-specific *Sa*Cas9 sgRNAs ..... 60
3.1.3. Assessment of *CCR5*-specific U6-sgRNA cleavage activity with the *Sa*Cas9 nuclease and nickase ..... 62
3.2. Analysis of the efficiency and specificity of *CCR5*-specific gene editing mediated by CRISPR/*Sa*Cas9 nuclease and nickase systems ..... 68
3.2.1. High sensitivity indel quantification at the *CCR5*-specific CRISPR/*Sa*Cas9 nuclease and nickase target sites by the ddPCR™ drop-off assay ..... 68
3.2.2. High sensitivity indel analysis at the *CCR5*-specific CRISPR/*Sa*Cas9 nuclease and nickase on target sites by deep sequencing ..... 71
3.2.3. High sensitivity indel analysis at *CCR5*-specific CRISPR/*Sa*Cas9 off target sites by deep sequencing ..... 76
3.3. Generation and assessment of 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase activity at the *CCR5*-sgRNA 7 and 8 target sites ..... 78
3.3.1. Generation of 'enhanced specificity' *Sa*Cas9 nuclease and nickase constructs ..... 79
3.3.2. Assessment of cleavage activity of the original and 'enhanced specificity' *Sa*Cas9 nuclease and nickase systems ..... 80
3.4. Analysis of the effect of *CCR5*-specific gene editing on mRNA production and R5-tropic HIV-1 infectivity ..... 85
3.4.1. Production of shRNA 1005 as a positive control for analysis of the functional effects of *CCR5*-specific CRISPR/*Sa*Cas9 gene editing ..... 85
3.3.2. Functional analysis of CRISPR/*Sa*Cas9-mediated *CCR5* gene editing by qRT-PCR and TZM-bl luciferase assay ..... 86
3.3.3. Analysis of the effect of *CCR5*-specific gene editing on R5-tropic HIV-1 infectivity ..... 89
4. Discussion ..... 92
4.1. Rational target design for effective CRISPR/*Sa*Cas9 nuclease functioning ..... 93
4.1.1. *In silico* prediction of sgRNA cleavage activity ..... 93
4.1.2. Key determinants of sgRNA cleavage activity ..... 94
4.2. The trade-off between CRISPR/*Sa*Cas9 cleavage efficiency and specificity ..... 99
4.2.1. Reduced CRISPR/*Sa*Cas9 nickase activity with improved specificity ..... 99
4.2.2. Efficiency of the 'enhanced specificity' *Sa*Cas9 nuclease and nickase systems ..... 101
4.3. Downstream functional effects of targeted gene editing ..... 103
4.4. Future prospects of a CCR5-specific CRISPR-mediated anti-HIV-1 gene therapy ..... 107
5. Conclusions ..... 111
6. References ..... 112
6.1. Journals ..... 112
6.2. Websites ..... 126

7. Appendix ..... 128
7.1. Plasmid construct maps ..... 128
7.1.1. pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA vector (pAAV-CMV_SaCas9) ..... 128
7.1.2. pTZ57R/T_U6-BbsI-sgRNA-SaCas9_tracrRNA vector (U6-sgRNA) ..... 128
7.1.3. pTZ57R/T_U6-BbsI-sgRNA-SaCas9_modtracrRNA (U6-sgRNA_modtracr) ..... 128
7.2. Supplementary Sequences ..... 129
7.2.1. *Sa*Cas9 ..... 129
7.2.2. *Sa*Cas9-D10A ..... 130
7.2.3. e*Sa*Cas9 ..... 131
7.2.4. e*Sa*Cas9-D10A ..... 132
7.2.5. *Sa*Cas9 U6-sgRNA cloning vector ..... 134
7.2.6. *Sa*Cas9 U6-sgRNA_modtracr cloning vector ..... 134
7.2.7. pTZ57R/T-H1_sh1005 construct ..... 134
7.2.8. Amplicon Sequences for Deep Sequencing ..... 134
7.3. Supplementary reagents and recipes ..... 137
7.3.1. 1 X Phosphate Buffered Saline solution ..... 137
7.3.2. Freeze medium ..... 137
7.3.3. 1 X TAE Buffer ..... 137
7.3.4. Restriction enzyme digestion reactions and buffer components ..... 137
7.3.5. PIPES transformation buffer ..... 138
7.3.6. SOC outgrowth medium ..... 138
7.3.7. LB/Amp broth and plates ..... 138
7.3.8. Polyethylenimine “Max” Mw 40 000 solution ..... 139
7.4. Supplementary protocols ..... 140
7.4.1. Agarose gel electrophoresis ..... 140
7.4.2. GeneJET PCR Purification ..... 140
7.4.3. QIAquick® Gel Extraction ..... 141
7.4.4. Transformation of *E. coli* DH5α cells ..... 141
7.4.5. Generation of chemically competent *E. coli* DH5α cells ..... 142
7.4.6. QIAprep® Spin Miniprep ..... 142
7.4.7. QIAGEN® Plasmid Midi and Maxiprep ..... 143
7.4.8. KAPA Express Extract DNA Extraction ..... 144
7.4.9. QIAamp DNA Mini gDNA Extraction ..... 144

7.4.10. Qubit® dsDNA High Sensitivity (HS) Quantification ..... 145
7.4.11. TRIzol® Reagent RNA Extraction ..... 145
7.5. Supplementary Figures and Tables ..... 147
7.6. Ethics Waiver Letter ..... 169

## List of Figures


| Figure no. | Figure title | Page no. |
|---|---|---|
| **1.1.** | CRISPR/Cas9 nuclease system functioning for targeted genomic disruption. | 9 |
| **1.2.** | *Sa*Cas9 primary protein structure and sgRNA in complex with a target DNA locus. | 11 |
| **1.3.** | CRISPR/*Sa*Cas9-D10A nickase and dual-sgRNA system functioning. | 14 |
| **1.4.** | Structure of the *CCR5* gene, pre-mRNA transcripts and protein as expressed on the surface of human cells. | 19 |
| **2.1.** | Diagram of the PCR methodology followed for production of the *CCR5*-targeting shRNA 1005 and the hairpin loop structure formed. | 34 |
| **2.2.** | Design of *CCR5* target site indel quantification by the ddPCR™ drop-off assay. | 41 |
| **2.3.** | Outline of the ddPCR™ drop-off assay methodology and an example of a 2D fluorescence plot. | 44 |
| **3.1.** | Schematic of the cloning strategy employed for generation of the *Sa*Cas9-D10A nickase vector and Sanger sequencing confirmation. | 59 |
| **3.2.** | A schematic showing ten *Sa*Cas9 sgRNA target sites within the human *CCR5* gene and Sanger sequencing chromatograms of cloned U6-sgRNAs. | 60 |
| **3.3.** | T7EI assays performed on *Sa*Cas9/-D10A-treated HEK293T cells with a panel of ten *CCR5*-specific U6-sgRNAs. | 63 |
| **3.4.** | TIDE indel quantification from treatment of HEK293T cells with *Sa*Cas9 and U6-sgRNA 7. | 66 |
| **3.5.** | Sequence confirmation of *CCR5*-specific U6-sgRNAs cloned into the U6-sgRNA_modtracrRNA vector. | 67 |
| **3.6.** | Quantification of CRISPR/*Sa*Cas9-mediated *CCR5* target site indel formation by the ddPCR™ drop-off assay. | 69 |
| **3.7.** | Targeted amplicon deep sequencing of the *CCR5* target region following TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNAs 7 and 8 as well as *Sa*Cas9-D10A + U6-sgRNA 7/8. | 73 |
| **3.8.** | Alignment and deep sequencing results of the *CCR5*-specific sgRNA 7 and 8 exonic off target sites. | 77 |
| **3.9.** | Schematic of the cloning strategy used for generation of the e*Sa*Cas9/-D10A constructs and Sanger sequencing chromatograms showing successful modification. | 79 |
| **3.10.** | Visualisation and quantification of CRISPR/*Sa*Cas9-mediated *CCR5* target site indel formation by T7EI and ddPCR™ PCR drop-off assays. | 81 |
| **3.11.** | Sequence confirmation of *CCR5*-specific shRNA 1005 cloned into the pTZ57R/T vector. | 85 |
| **3.12.** | Measurement of *CCR5* mRNA levels by qRT-PCR following TZM-bl treatment with shRNA 1005 and CRISPR/*Sa*Cas9 systems. | 86 |
| **3.13.** | Measurement of R5-tropic HIV-1 pseudovirus infectivity by TZM-bl luciferase assay. | 90 |
| **4.1.** | Computationally predicted structures of *Sa*Cas9 *CCR5*-specific sgRNAs utilized in this study and by Friedland *et al*. (2015). | 98 |
| **4.2.** | CCR5 protein structure before and after CRISPR/*Sa*Cas9 nuclease treatment. | 104 |
| **7.5.1.** | Agarose gels showing PCR products from amplification of the *CCR5* target site. | 147 |
| **7.5.2.** | Quantification of *CCR5* target site indel formation in the no gDNA negative control by the ddPCR™ drop-off assay. | 148 |
| **7.5.3.** | Agarose gels showing PCR products from amplification of the *CCR5* target site and ten off target sites for targeted amplicon deep sequencing. | 149 |

## List of Tables


| Table no. | Table title | Page no. |
|---|---|---|
| **2.1.** | Sequences of the original and modified *Sa*Cas9 sgRNA scaffolds | 25 |
| **2.2.** | Primers used for SDM-PCR of the *Sa*Cas9-D10A codon mutation | 27 |
| **2.3.** | Cycling conditions used for KAPA2G Robust PCRs | 27 |
| **2.4.** | Primer used for sequence confirmation of the *Sa*Cas9-D10A codon mutation | 28 |
| **2.5.** | *CCR5*-specific *Sa*Cas9 ODN + and - strand sequences, the resulting U6-sgRNA variable sequences and target site PAM recognition motifs | 30 |
| **2.6.** | Primer used for sequence confirmation of the variable region within the cloned U6-sgRNA constructs | 31 |
| **2.7.** | Primers used to generate the *CCR5*-targeting shRNA 1005 | 33 |
| **2.8.** | Cycling conditions used for each round of PCR for production of the *CCR5*-targeting shRNA 1005 | 34 |
| **2.9.** | Primers used to amplify the region containing the ten *CCR5*-specific target sites for indel analysis by T7EI assay | 38 |
| **2.10.** | Denaturation and re-annealing of PCR products for T7EI heteroduplex formation | 38 |
| **2.11.** | Probes and primers used for indel quantification at two *CCR5* target sites by the ddPCR™ drop-off assay | 41 |
| **2.12.** | Cycling conditions used for *CCR5* target site indel quantification by the ddPCR™ drop-off assay | 43 |
| **2.13.** | Indexes added to the 5′ end of each primer to distinguish between CRISPR/*Sa*Cas9 treatments used | 48 |
| **2.14.** | Primers used for on and off target site amplification for deep sequencing | 49 |
| **2.15.** | Cycling conditions used for Q5 Hot Start Polymerase Chain Reactions | 51 |
| **2.16.** | TruSeq adapter index sequences used for targeted amplicon deep sequencing of three pools | 52 |
| **2.17.** | Primers used for qRT-PCR of *CCR5* and *β-actin* mRNA | 54 |
| **2.18.** | Cycling conditions for the relative quantification of *CCR5* and *β-actin* mRNA by Real-time qRT-PCR | 55 |
| **3.1.** | TIDE indel quantification of *Sa*Cas9-mediated U6-sgRNA cleavage activity | 65 |
| **3.2.** | ddPCR™ drop-off assay results obtained from TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNA 7 or 8 and *Sa*Cas9-D10A + U6-sgRNA 7/8 for three independent replicates | 70 |
| **3.3.** | CRISPRseek *Sa*Cas9-specific exonic off target prediction sites similar to the *CCR5*-specific sgRNAs 7 and 8 | 76 |
| **3.4.** | ddPCR™ drop-off assay results obtained from WT and ‘enhanced specificity’ CRISPR/*Sa*Cas9 nuclease and nickase treatment of TZM-bl cells | 83 |
| **3.5.** | qRT-PCR results obtained from *CCR5*-specific and non-specific (NS) shRNA treatment in TZM-bl cells | 87 |
| **3.6.** | qRT-PCR results obtained from CRISPR/*Sa*Cas9 treatment in TZM-bl cells | 88 |
| **3.7.** | HIV-1 infectivity assay results obtained from CRISPR/*Sa*Cas9 treatment and ZM53 infection in TZM-bl cells | 90 |
| **7.5.1.** | Results obtained from targeted amplicon deep sequencing of the *CCR5* target region following TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNAs 7 and 8 as well as *Sa*Cas9-D10A + U6-sgRNA 7/8 | 152 |
| **7.5.2.** | Technical duplicate ddPCR™ drop-off assay results obtained from WT and ‘enhanced specificity’ CRISPR/*Sa*Cas9 nuclease and nickase treatment of TZM-bl cells | 163 |
| **7.5.3.** | Biological triplicate qRT-PCR results for *CCR5* and the *β-actin* reference control obtained from Non-Specific (NS) shRNA and shRNA 1005 treatment of TZM-bl cells | 165 |
| **7.5.4.** | Technical duplicate qRT-PCR results for *CCR5* and the *β-actin* reference control obtained from CRISPR/*Sa*Cas9 as well as Non-Specific (NS) shRNA and shRNA 1005 treatment of TZM-bl cells | 166 |
| **7.5.5.** | Biological triplicate luciferase readings obtained from CRISPR/*Sa*Cas9; Non-Specific (NS) shRNA and shRNA 1005 treatment of TZM-bl cells | 168 |

## Abbreviations

| | | |
|---|---|---|
| **aa** | - | Amino acid/s |
| **AAV** | - | Adeno-associated virus |
| **AIDS** | - | Acquired immune deficiency syndrome |
| **APC** | - | Antigen presenting cell |
| **ART** | - | Antiretroviral therapy |
| **ARV** | - | Antiretroviral |
| **bp** | - | Base pair/s |
| **cART** | - | Combination antiretroviral therapy |
| **Cas9** | - | CRISPR-associated protein 9 |
| **CCR5** | - | C-C chemokine receptor type 5 |
| **CCR5-Δ32** | - | *CCR5* delta 32 mutation |
| **CD4** | - | Cluster of differentiation 4 |
| **cDNA** | - | Complementary DNA |
| **CMV** | - | Cytomegalovirus |
| **CRISPR** | - | Clustered regularly interspaced short palindromic repeat |
| **crRNA** | - | CRISPR-RNA |
| **C-terminus** | - | Carboxy terminus |
| **CXCR4** | - | C-X-C chemokine receptor 4 |
| **DC** | - | Dendritic cell |
| **DC-SIGN** | - | Dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin positive receptor |
| **DMD** | - | Duchenne muscular dystrophy |
| **dd$H_20$** | - | Double distilled water |
| **ddPCR** | - | Droplet Digital PCR |
| **DEAE** | - | Diethylaminoethyl |
| **DMEM** | - | Dulbecco's Modified Eagle Medium |
| **DMSO** | - | Dimethyl sulfoxide |
| **DNA** | - | Deoxyribonucleic acid |
| **dNTP** | - | Deoxyribonucleotide triphosphate |
| **DSB** | - | Double stranded break |
| **EDTA** | - | Ethylenediaminetetraacetic acid |
| **e*Sa*Cas9** | - | Enhanced specificity *Sa*Cas9 |
| **FBS** | - | Foetal Bovine Serum |
| **gDNA** | - | Genomic DNA |
| **GFP** | - | Green fluorescent protein |
| **gp120/41** | - | Glycoprotein 120/41 |
| **GPCR** | - | G-protein coupled receptor |
| **gRNA** | - | Guide RNA - refers to the short variable sequence of the sgRNA |
| **HAART** | - | Highly active antiretroviral treatment |
| **HDR** | - | Homology directed repair |
| **HEK** | - | Human embryonic kidney |
| **HIV-1** | - | Human immunodeficiency virus type 1 |

| | | |
|---|---|---|
| **HR** | - | Homologous recombination |
| **HSPC** | - | Haematopoietic stem/progenitor cell |
| **Indels** | - | Insertion/deletion/substitution mutations |
| **iPSC** | - | Induced pluripotent stem cell |
| **ITR** | - | Inverted terminal repeat |
| **LL** | - | Linker loop |
| **LRA** | - | Latency reversal agent |
| **LTR** | - | Long terminal repeat |
| **MCS** | - | Multiple cloning site |
| **MOI** | - | Multiplicity of infection |
| **mRNA** | - | Messenger RNA |
| **NHEJ** | - | Non-homologous end joining |
| **NMD** | - | Nonsense-mediated decay |
| **NS** | - | Non-specific |
| **nt** | - | Nucleotide/s |
| **N-terminus** | - | Amino terminus |
| **NUC** | - | Nuclease lobe |
| **ODN** | - | Oligodeoxynucleotide |
| **ORF** | - | Open reading frame |
| **PAM** | - | Protospacer adjacent motif |
| **PCR** | - | Polymerase chain reaction |
| **PEI** | - | Polyethylenimine |
| **pre-crRNA** | - | Precursor crRNA |
| **PSA** | - | Penicillin/Streptomycin/Amphotericin B |
| **qRT-PCR** | - | Quantitative reverse transcriptase PCR |
| **REC** | - | Recognition lobe |
| **RNA** | - | Ribonucleic acid |
| **RNAi** | - | RNA interference |
| **RNP** | - | Ribonucleoprotein |
| **RT** | - | Reverse transcriptase |
| ***Sa*Cas9** | - | *Staphylococcus aureus* Cas9 |
| **SCT** | - | Stem cell therapy |
| **SDM-PCR** | - | Site directed mutagenesis PCR |
| **sgRNA** | - | Single guide RNA |
| **shRNA** | - | Short hairpin RNA |
| **SIV** | - | Simian immunodeficiency virus |
| ***Sp*Cas9** | - | *Streptococcus pyogenes* Cas9 |
| **T/F** | - | Transmitted/founder virus |
| **T7EI** | - | T7 Endonuclease I |
| **TAE** | - | Tris-acetate EDTA |
| **TALEN** | - | Transcription activator-like effector nuclease |
| **TIDE** | - | Tracking of indels by decomposition |
| **$T_m$** | - | Melting temperature |
| **tracrRNA** | - | Trans-activating crRNA |

| | | |
|---|---|---|
| **U** | - | Units |
| **U6 Pol III** | - | U6 Polymerase III promoter |
| **UCSC** | - | University of California, Santa Cruz Genome Browser |
| **UTR** | - | Untranslated region |
| **VS** | - | Virological synapse |
| **WT** | - | Wild type |
| **ZFN** | - | Zinc finger nuclease |
| **-1 PRF** | - | Programmed -1 ribosomal frameshift |

# 1. Introduction

## 1.1. Epidemiology of the global HIV pandemic

The human immunodeficiency virus (HIV) is a lentivirus within the *retroviridae* family that hijacks human immune cell machinery for replication, culminating in immune cell decline characteristic of acquired immune deficiency syndrome (AIDS) (Barre-Sinoussi et al., 1983). At the end of 2015, 36.7 million people were living with HIV with an annual increase of 2.1 million and over a million deaths from AIDS-related causes (UNAIDS, 2015). HIV is classified into two major types, HIV-1 and -2, which are transmitted by equivalent mechanisms and both culminate in AIDS but differ epidemiologically (Reviewed in Campbell-Yesefu and Gandhi, 2011). HIV-1 has a higher infective capacity than HIV-2 and, as a result, has been responsible for the major global pandemic while the latter has mostly been contained to small pockets within West Africa, Europe, India and the USA (Guerreiro et al., 2012).

The emergence of the highly active, combination and single-dose antiretroviral therapies (HAART, cART and ART) and up-to-date treatment recommendations have substantially reduced the global transmission, morbidity and mortality rates of HIV-1 (Reviewed in Günthard et al., 2016). While these treatment regimens are pivotal in the quest to combatting HIV-1, they require lifelong administration and are insufficient to achieve full disease eradication. A major obstacle is the emergence of HIV-1 drug resistance as a result of selective pressure on the highly mutagenic genome replication mechanism of the virus, coupled with poor patient adherence to antiretroviral (ARV) treatment (Zagordi et al., 2010). Additionally, transcriptionally silent but replication competent reservoirs of latent HIV-1 proviruses are established during acute infection and persist within long-lived infected cells, precluded from clearance by ARVs as well as the human immune system (Reviewed in Abbas and Herbein, 2012). HIV/AIDS has thus become one of the greatest global human health challenges of the current age, placing an enormous burden on the livelihood of millions as well as incurring monumental costs to healthcare systems worldwide (UNAIDS, 2015). ARV drug resistance, toxicity and the necessity for lifelong administration warrant the discovery of more effective and permanent treatment strategies. Understanding the key HIV-1 lifecycle stages and nuances within these is paramount to designing and developing the most promising curative strategies for this devastating disease.

### 1.2. HIV-1 infective mechanisms

HIV-1 transmission occurs through the direct contact between body fluids of infected and healthy (or, in some cases, previously infected) people (Hansasuta and Rowland-Jones, 2001). In resource-limited countries where the prevalence and burden of HIV-1/AIDS is greatest, the major mode of transmission is sexual intercourse, despite several physical and innate immune obstacles at the mucosal interface impeding efficient transmission of the virus (Reviewed in Gouws et al., 2008). These barriers create a bottleneck, most often resulting in the transmission of a single variant, referred to as the 'transmitted/founder' (T/F) virus, that possesses the characteristics required for establishment of infection (Keele et al., 2008). In order to uncover a broad and highly effective cure for HIV-1, it is essential to unravel the intricacies of two major stages of HIV-1 infection – the various modes of infection as well as cellular demise leading to AIDS-related illness and patient death. A path towards HIV-1 cure development is forged if a treatment strategy is able to inhibit both of these processes permanently.

Cellular infection by HIV-1 can occur via cell-free and cell-associated mechanisms, each informing the mode of cell death that is incurred. Cell-free infection is initiated by the interaction between the HIV-1 envelope glycoprotein 120 (gp120) trimeric spike and the human CD4 (cluster of differentiation 4) cell surface receptor (Dragic et al., 1996). This is followed by an essential co-receptor binding step, most often mediated by CCR5 (C-C chemokine receptor 5) or CXCR4 (C-X-C chemokine receptor 4), required to induce conformational changes in the HIV-1 envelope glycoproteins for gp41-mediated host cell membrane penetration and viral attachment (Berger et al., 1999). HIV-1 particles that have an affinity for the CCR5 or CXCR4 co-receptor for cell entry are termed R5- or X4-tropic virus respectively, while those able to utilise either are referred to as dual-tropic. Following attachment of the HIV-1 particle to the target cell, the viral envelope and cell membrane fuse to allow for injection of the viral core antigen into the host cell cytoplasm. The HIV-1 ribonucleic acid (RNA) genome is exposed, reverse transcribed by the viral reverse transcriptase (RT) and the double stranded complementary deoxyribonucleic acid (cDNA) genomic copy integrated permanently within the $CD4^+$ host cell chromatin by HIV-1 integrase (Schroder et al., 2002). The integrated provirus can either remain transcriptionally active and productive infection ensues or, under certain cellular conditions and/or epigenetic modifications, the HIV-1 long terminal repeat (LTR) promoter can be conditionally silenced

(Chun et al., 1995). The former course sustains the circulating infectious viral pool with the propensity for concomitant evolution while the latter generates a permanent reservoir of latent provirus (Finzi et al., 1997, Chun et al., 1997). These latent HIV-1 genomes are refractory to drug and immune removal and have the capacity to reactivate and replenish the infective pool throughout the patient's lifespan necessitating lifelong treatment (Reviewed in Siliciano and Greene, 2011). Activation of the LTR promoter results in the expression of viral proteins and full length genomic RNA transcripts, the assembly of immature virions within the cytoplasm and release for further propagation by 'budding' at the plasma membrane (Reviewed in Sundquist et al., 2012). The HIV-1 lifecycle is completed upon maturation of the particle into a functionally infective virion, a process mediated by the HIV-1 protease (Reviewed in Sundquist et al., 2012).

Alternatively, the cell-associated HIV-1 mechanism of infection occurs via direct cell-cell contacts, commonly known as virological synapses (VSs) (Jolly et al., 2004). This has been shown to occur with greater efficiency than cell-free infection, particularly in immobile T cell compartments such as lymphoid tissue, and can constitute as much as 60 % of transmission events within an infected patient (Chen et al., 2007, Sourisseau et al., 2007, Iwami et al., 2015). In the context of sexual transmission, among the first clinically relevant cells to be encountered by HIV-1 at the mucosal interface are the antigen presenting cells (APCs), specifically dendritic cells (DCs) and macrophages (Shen et al., 2011). DCs and macrophages express CD4, CCR5 and CXCR4 on their surface, albeit at lower frequencies than the circulating and lymphoid $CD4^+$ T cell compartment (Lee et al., 1999). Reduced levels of receptor expression and the presence of antiviral restriction factors confer resistance to productive HIV-1 infection in these cells (Cavrois et al., 2006, Goldstone et al., 2011, Lahouassa et al., 2012). While HIV-1 has evolved strategies to circumvent the activity of such restriction factors to an extent, DCs and macrophages still display reduced susceptibility to productive infection (Hrecka et al., 2011, Hrecka et al., 2016). The clinical significance of these APCs is as a result of the role they play in systemic dissemination of HIV-1 to the highly permissive $CD4^+$ T cell compartment where latency is established, virus is propagated and immune cell death ensues.

The mechanism by which DCs and macrophages carry out the capture and delivery of HIV-1 to $CD4^+$ T cells is not identical but the outcome is similar. DCs express cell surface immunoreceptors, such as DC-SIGN (dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin positive), that have a high affinity for the heavily glycosylated HIV-1

envelope gp120 (Geijtenbeek et al., 2000, Lee et al., 2001). Following the combined interaction with DC-SIGN, CD4 and CCR5/CXCR4, the HIV-1 particle is internalised into cytoplasmic endosomes, protected from human protease-mediated degradation (Reviewed in Dutartre et al., 2016). Alternatively, macrophages can harbour HIV-1 within invaginations of the plasma membrane and as these cells are relatively long-lived they have a substantially high transmitting capacity to permissive $CD4^+$ T cells (Deneka et al., 2007, Welsch et al., 2007). Transmission of HIV-1 from non-productively infected DCs and macrophages to permissive $CD4^+$ T cells occurs when the activated APCs mature and migrate to lymph tissues for presentation of the viral proteins to T cells (Reviewed in Izquierdo-Useros et al., 2014). This results in the dissemination of HIV-1 and allows for infection of distally located $CD4^+$ T cells by a cell-associated mechanism.

The result of direct cell-cell interaction between non-productively infected APCs and $CD4^+$ T cells is the formation of actin-dependent VSs between the donor and acceptor cells (Jolly et al., 2004). VS formation is initiated by the interaction between donor cell HIV-1 envelope proteins and CD4 receptors on the acceptor cells and requires the participation of host adhesion molecules, such as ICAM1 and LFA1 (Jolly et al., 2007). Simply blocking the adhesion molecules involved in VS formation is not sufficient to prevent transmission of HIV-1 by cell-cell contact and therefore not a robust target for treatment strategies (Kondo et al., 2015). VSs form semi-permeable channels for the transport of immature HIV-1 particles from infected to uninfected cells at a high multiplicity of infection (MOI) (Jolly et al., 2004, Chen et al., 2007). Maturation of the HIV-1 particles is believed to occur within the endosomes of acceptor cells following transmission and thus immune responses directed towards mature envelope structures of circulating virus are ineffective under these conditions (Chen et al., 2007, Hubner et al., 2009, Dale et al., 2011). It has been shown that infectious particles are required for cell-cell transmission and it is not possible for HIV-1 genomic copies to be transferred via the VS to acceptor cells (Monel et al., 2012). Furthermore, while the VS allows entry of some molecules, for example the CCR5-specific entry inhibitor Maraviroc, antibodies are unable to penetrate the barrier and neutralise the virus.

The mechanism of $CD4^+$ immune cell cytotoxicity in HIV-1 infection is somewhat poorly understood but the mode of transmission is believed to determine the mode of cell death that ensues (Doitsh et al., 2010). Only a small proportion of $CD4^+$ T cells infected by cell-free HIV-1 are believed to engage in apoptotic cell death (Doitsh et al., 2014, Galloway et al., 2015). Conversely, the immobile and ordinarily non-permissive resting $CD4^+$ T cells in

lymphoid tissues are exposed to a high HIV-1 MOI during VS-mediated infection by productively infected $CD4^+$ T cells or non-productively infected APCs (Sloan et al., 2013, Agosto et al., 2015). The consequent cytoplasmic overload and accumulation of labile HIV-1 RNA genome transcripts triggers the highly inflammatory caspase-1-dependent pyroptotic cell death pathway (Doitsh et al., 2010). In addition to the inflammation induced by this mode of death, the healthy and ordinarily non-permissive immune cells can be eliminated, contributing significantly towards patient disease progression to AIDS. It is therefore important for therapeutic development to consider the effect of cell-free as well as cell-associated modes of HIV-1 transmission in order to prevent T cell death both by apoptosis as well as pyroptosis.

ART is undeniably a powerful immediate solution to controlling the spread and symptomatic effects of HIV-1 infection but in order to eradicate the disease in the foreseeable future, it is essential for permanent treatment strategies to be developed. A 'functional' cure for HIV-1 appears currently to be the most realistic therapeutic ambition, whereby the effects of the virus are no longer experienced by the patient following treatment cessation. Rapid advancements to gene therapy in the form of targetable genome engineering, the *in situ* manipulation of genetic material in order to achieve a desired phenotypic effect, have opened up a novel avenue for direct and permanent disease intervention. The selectable manipulation of human and HIV-1 genes provides a means to engineer a highly specific human response to HIV-1 infection by abolishing the infective, replicative and evolutionary capacity of the virus.

### 1.3. Genome engineering for HIV-1 functional cure development

Classical gene therapy has advanced rapidly since the emergence of DNA sequencing and genome targeting technology development. First generation methods relied on highly inefficient homologous recombination (HR) events for the introduction of exogenous DNA sequences into mammalian cells (Smithies et al., 1985, Thomas et al., 1986). At a frequency of around one successful HR event in $\sim 10^3 - 10^9$ human cells and with observations of integration events at undesired loci, this did not emerge as a practical approach for therapeutic development (Lin et al., 1985). The discovery that a DNA double stranded break (DSB) at an intended genetic target site enhances the endogenous homology directed repair (HDR) pathway, as well as allows for erroneous non-homologous end joining (NHEJ) activity, revolutionised the efficacy and applicability of human genome engineering,

particularly for gene therapy development (Rudin and Haber, 1988). Meganucleases were first employed for DSB induction and HDR- or NHEJ-mediated editing (Sussman et al., 2004, Rosen et al., 2006) but difficulties around the design and re-programming of these molecules drove development of simpler technologies, shaping the emergence of revolutionary designer endonucleases for *in situ* targeted genome engineering and gene therapy (Bibikova et al., 2002, Urnov et al., 2005). Rational genome engineering can be tailored towards any genome, diversifying therapeutic targets particularly in the context of diseases caused by exogenous agents, such as viruses, parasites and bacteria, or genetic diseases not amenable to classical gene replacement therapies, such as DMD (Duchenne muscular dystrophy) (Long et al., 2014). Over the past decade, various powerful gene editing technologies have been developed and manipulated to achieve safe and highly effective targeted genome engineering for advanced disease cure development.

### 1.3.1. Rational genome engineering technologies

Rational genome engineering for the introduction of controlled changes into selected genomic sites leverages the defined outcomes of endogenous host DNA repair mechanisms (Reviewed in Cox et al., 2015). The three major endonuclease-based gene editing tools commonly employed today are ZFNs (zinc finger nucleases), TALENs (transcription activator-like effector nucleases) and CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9). The premise of endonuclease-mediated gene editing is a two-component DNA recognition and cleavage system. A short region of the desired genomic target site is recognised by a customizable component which can be a DNA-binding protein, as with ZFNs and TALENs, or a complementary RNA sequence as with CRISPR/Cas9 (Urnov et al., 2005, Miller et al., 2011, Jinek et al., 2012). This designer navigation component is combined with a catalytic protein, such as FokI for ZFNs and TALENs or Cas9 for CRISPR/Cas9, able to induce DNA DSB formation and exploit the host cell repair machinery to introduce changes to the target locus. Provision of a homologous DNA template stimulates the high fidelity HDR pathway (Bibikova et al., 2001). Alternatively, the highly erroneous NHEJ DNA repair pathway is activated, resulting in the introduction of insertion, deletion or substitution mutations (indels) into the target site (Bibikova et al., 2002).

ZFNs and TALENs, while highly effective gene editing tools, are cumbersome to design and expensive to produce, requiring that a unique synthetic protein be engineered for each target

DNA site in order to induce the desired change (Reviewed in Cox et al., 2015). In contrast, the CRISPR/Cas9 system is modular, consisting of an invariant protein component (Cas9) and a small programmable RNA guide which can be generated with ease and low cost. The universal applicability and simplicity of CRISPR/Cas9 has led to its widespread adoption as a molecular biology tool and concurrently generated intense interest in both the scientific community and the mainstream media. CRISPR/Cas9 has the potential to revolutionise the field of molecular medicine by enabling treatment development for a plethora of currently incurable genetic and communicable diseases.

### 1.3.2. CRISPR/Cas9 as a natural defence mechanism

CRISPR/Cas9 is derived from a sequence-specific prokaryotic natural defence system to foreign invading DNA, or occasionally RNA, found in over 40 % of bacteria and 90 % of archaea (Ishino et al., 1987, Mojica et al., 2000, Barrangou et al., 2007). The most recent evolutionary classification report describes the division of the highly genetically and functionally diverse CRISPR/Cas systems into two major classes (1 and 2) and five types (I – V) (Makarova et al., 2015). CRISPR/Cas9 is a class 2, type II system with a single catalytic component while functionality of other systems in the group requires multi-domain interaction at the target site (Gasiunas et al., 2012).

The CRISPR genomic locus is comprised of a CRISPR array containing 'spacer' sequences interspersed between invariable precursor-CRISPR RNA (pre-crRNA) direct repeats; a non-coding trans-activating crRNA (tracrRNA) and the Cas9 gene (Mojica et al., 2000, Mojica et al., 2009, Fonfara et al., 2014). The first stage of CRISPR/Cas9-mediated immunity, as observed in the *Staphylococcus aureus*-derived system, is *adaptation* through the detection of invading DNA or RNA molecules and integration of short fragments, called 'protospacers', into the 'spacer' sites (Jinek et al., 2012). These exogenously-derived variable spacers are required to confer acquired 'immunity' towards future attack. The second stage involves *expression* of the CRISPR array, tracrRNA and Cas9 endonuclease. The pre-crRNAs within the CRISPR array are processed into mature crRNA targeting sequences by bacterial RNase III, facilitated by the tracrRNA. Thereafter, *interference* is carried out by Cas9:crRNA:tracrRNA complex formation and recognition by Watson-Crick base pairing of complementary DNA (or RNA) species with the variable region of the crRNA sequence. Target recognition is followed by Cas9 endonuclease-mediated cleavage, thereby destroying the invading nucleic acid (Mojica et al., 2009). A limiting factor of recognition is the

presence of a Cas9 orthologue-specific short protospacer adjacent motif (PAM) at the 3′ end of the target site in order to avoid self-recognition and genome destruction (Mojica et al., 2009).

Simplicity and ingenuity of this naturally occurring system has driven extensive research development over the past five years, providing researchers with a CRISPR toolbox from which to tailor a wide array of genome engineering applications (Deltcheva et al., 2011, Jinek et al., 2012). Critical to human disease treatment development, CRISPR/Cas9 has been optimised for targeted and multiplexed genomic modification in human cells (Mali et al., 2013b, Cong et al., 2013). Ultimately, a gene therapy based on gene editing should be easily delivered to target cells, efficiently induce the desired phenotypic outcome and avoid undesired 'off target' activity. Understanding the molecular biology of gene editing mediated by CRISPR/Cas9 and the nuances that set orthologous systems apart is crucial for the rational design of a potential therapeutic.

### 1.3.3. A brief overview of CRISPR/Cas9 system functioning

The general mechanism of action of CRISPR/Cas9 gene editing is conserved across orthologous systems from which they were derived but some structural and functional features are particular to the Cas9 orthologue in use. In general, CRISPR/Cas9 mediated gene editing is carried out by the Cas9 endonuclease in complex with a single guide RNA (sgRNA) – a short ~ 20 bp guide RNA (gRNA) transcriptionally fused to an invariable scaffold RNA (Jinek et al., 2012). Cas9-mediated DNA unwinding and sequence-specific genome scanning by the gRNA results in target site and Cas9-orthologue specific PAM motif recognition, target DNA:gRNA heteroduplex formation and activation of Cas9 cleavage activity. The two catalytic domains of the Cas9 endonuclease, HNH and RuvC, cleave the gRNA complementary and non-complementary strands respectively introducing a DSB at the PAM-proximal end of the target site (Ran et al., 2013). In the absence of a homologous DNA template, this DSB induces host NHEJ DNA repair and indel formation to carry out targeted gene editing (Figure 1.1) (Deltcheva et al., 2011, Jinek et al., 2012).

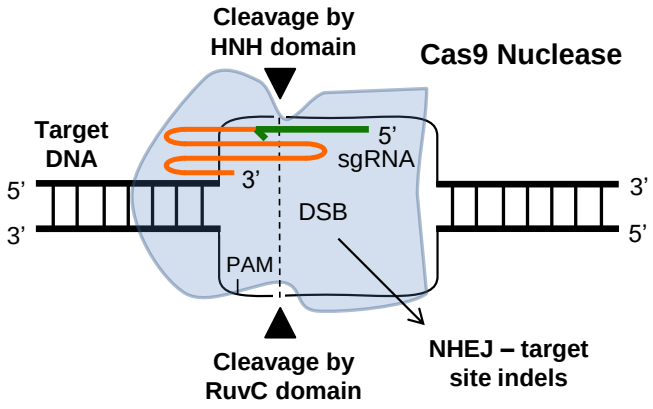



**Figure 1.1: CRISPR/Cas9 nuclease system functioning for targeted genomic disruption.** The Cas9 endonuclease recognises a cognate sgRNA by the particular folding of the scaffold sequence (orange). This forms a Cas9:sgRNA complex that scans the host genome for the desired target site and presence of non-target strand PAM motif. Full complementarity and heteroduplex formation between the variable gRNA region (green) and target site activates Cas9 nuclease activity at the RuvC and HNH domains. A DSB is introduced into the desired target site by cleavage at the gRNA complementary and non-complementary strands by HNH and RuvC respectively. In the absence of a homologous DNA template, erroneous NHEJ DNA repair is induced and potentially disruptive indels are introduced into the target site.

Major design considerations for CRISPR/Cas9-mediated gene therapy development are efficiency, safety and deliverability of the system. The CRISPR/Cas9 toolbox has advanced rapidly over the years to offer an array of ways in which to satisfy these key conditions. Of the class 2 type II CRISPR/Cas9 systems investigated, those derived from *Streptococcus pyogenes* and *Staphylococcus aureus* (CRISPR/*Sp*Cas9 and /*Sa*Cas9) have been shown to display the greatest *in vivo* cleavage activity (Mali et al., 2013b, Cong et al., 2013, Ran et al., 2015). These orthologous systems function through highly similar mechanisms despite sharing only 17 % sequence identity (Nishimasu et al., 2015). The *Sp*Cas9 system is employed in many studies for its high cleavage efficiency and tri-nucleotide 5′-NGG PAM recognition sequence which creates a broad targeting landscape within the human genome (Cong et al., 2013, Mali et al., 2013b). The undesirable consequence of this is a relatively high level of cleavage promiscuity, resulting from the abundance of canonical PAM sites, the potential for non-canonical 5′-NAG PAM recognition as well as a high level of tolerance for mismatches between the gRNA and similar off target genomic sequences (Jinek et al., 2012, Hsu et al., 2013, Cho et al., 2014). While not every PAM site leads to gRNA recognition, binding and cleavage, *Sp*Cas9 activity has been recorded at sites with up to 6 mismatches within the relatively short 20 bp gRNA region (Tsai et al., 2015). Even a low level of off target activity is unacceptable in most molecular biology applications but especially in the

context of gene therapy development as the downstream effects may be deleterious to treated cells and patients. This has stimulated important research into CRISPR/*Sa*Cas9 as an alternative system with improved safety profiles.

### 1.3.4. Molecular biology of the CRISPR/*Sa*Cas9 system

The *Sa*Cas9 endonuclease, with comparable cleavage activity to *Sp*Cas9, emerged as a highly promising candidate for minimised off target cleavage activity as the longer 5′-NNGRRT PAM recognition requirement narrows the number of potential cleavage sites within the human genome (Ran et al., 2015). In addition to a reduced targeting landscape, *Sa*Cas9 also has the potential for *in vivo* delivery by adeno-associated viral vectors (AAVs). AAVs are attractive gene therapy vehicles as they have broad and selectable tissue tropism, are able to transduce both dividing and non-dividing cells, have low immunogenicity and possess a non-integrative lifecycle that reduces the risk of insertional oncogenesis (Reviewed in Daya and Berns, 2008). A crucial drawback of AAV-mediated gene therapy delivery is the maximum genome packaging capacity of ~ 4.8 – 5.2 kb which limits the amount of genetic material that can be delivered using a single AAV (Wu et al., 2010). While *Sp*Cas9 is the most well-characterised CRISPR/Cas9 system to date, the combined size of the Cas9 gene and sgRNA, each driven by a separate promoter, exceed the AAV packaging limit. Dual-AAV delivery of CRISPR/Cas9 systems has been tested but, in the interest of efficiency optimisation, utilising CRISPR/*Sa*Cas9 instead circumvents this necessity. *Sa*Cas9 lacks non-essential recognition domains of *Sp*Cas9 and thus has a shorter gene sequence that is able to be packaged comfortably alongside the sgRNA component within a single AAV vector (Ran et al., 2015, Nishimasu et al., 2015). CRISPR/Cas9 gene editing technology has undergone extensive specificity enhancement over the recent years but, before these modifications can be discussed, it is important to understand the detailed molecular biology of CRISPR/*Sa*Cas9 system functioning.

From here on, the specifics of CRISPR/*Sa*Cas9 gene editing will be discussed for relevance to this study but the overall molecular biology is broadly applicable to the orthologous systems. The two major components of CRISPR/Cas9 are the Cas9 endonuclease and the sgRNA molecule that form a targetable genomic cleavage complex with the capacity to effect gene editing. The *Sa*Cas9 endonuclease is comprised of recognition (REC) and nuclease (NUC) lobes connected in the tertiary structure by a linker loop (LL) region (Figure 1.2A)

and an arginine-rich bridge helix (Jinek et al., 2013, Nishimasu et al., 2014, Nishimasu et al., 2015).

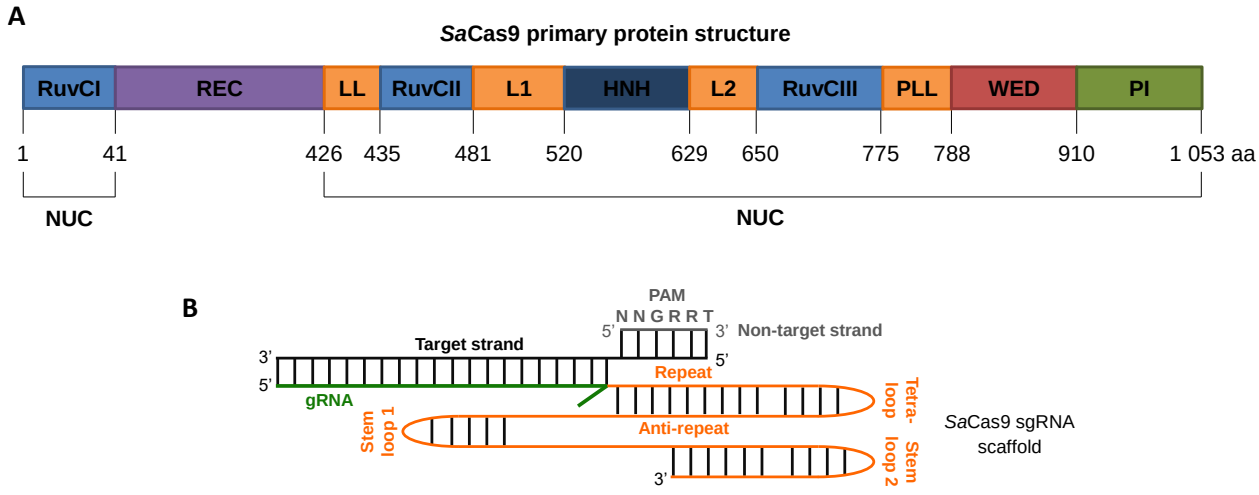



**Figure 1.2: *Sa*Cas9 primary protein structure and sgRNA in complex with a target DNA locus. A** – Primary structure of the 1 053 aa *Sa*Cas9 protein. The RuvC cleavage domain is divided into three subdomains, RuvCI – III. From left to right the primary protein structure consists of the RuvCI subdomain, recognition lobe (REC), linker loop sequence (LL), RuvCII subdomain, linker 1 (L1) region, HNH domain, linker 2 (L2) region, RuvCIII subdomain, phosphate lock loop (PLL), wedge (WED) domain and PAM interacting (PI) domain. **B** – *Sa*Cas9-specific sgRNA and target DNA heteroduplex. The short ~ 20 bp targeting guide RNA (gRNA) sequence (green) is fused to the scaffold sequence (orange), made up of the repeat and anti-repeat sequences as well as the tetraloop and stem loops 1 and 2. The structure formed by the scaffold is required for Cas9 recognition and complex formation prior to genomic targeting. The PAM sequence, 5′-NNGRRT for *Sa*Cas9 targeting, is located at the 3′ end of the gRNA on the non-target DNA strand (grey).

The two highly conserved Cas9 catalytic domains, namely HNH and RuvC, are contained within the NUC lobe and are responsible for targeted DNA cleavage of the gRNA complementary and non-complementary strands respectively (Nishimasu et al., 2014, Nishimasu et al., 2015). The RuvC domain is further divided into three subdomains (RuvCI – III), of which the second and third are connected to the HNH domain via the first and second linker regions (L1 and L2) respectively (Figure 1.2A) (Nishimasu et al., 2015). The REC lobe is required for recognition of a cognate sgRNA molecule with orthologue-specific structural features. The sgRNA is generated as a chimera of the target site-specific 21 – 24 bp gRNA sequence (the optimal length for *Sa*Cas9) and the invariable scaffold RNA, referred to in the native prokaryotic system description as the tracrRNA (Figure 1.2B) (Jinek et al., 2012, Ran et al., 2015, Friedland et al., 2015). The *Sa*Cas9 scaffold secondary structure contains a repeat:anti-repeat duplex, tetraloop region and two stem loops and is important for Cas9-

specific recognition and cleavage activity (Figure 1.2B). Extension of the tetraloop by five base pairs (bp) and the inclusion of an adenine to uracil (A-U) flip in the anti-repeat region are recent improvements to sgRNA structural feature stability (Nishimasu et al., 2015). This advancement has the propensity to enhance the efficiency of CRISPR/*Sa*Cas9 gene editing by improving recognition of the sgRNA by the wedge (WED) and REC domains (Figure 1.2A) of the Cas9 protein prior to genome scanning.

Cas9 structural rearrangement is induced upon complex formation with the unbound cognate sgRNA to facilitate genome scanning and target site cleavage (Nishimasu et al., 2015). The phosphate lock loop (PLL) Cas9 domain (Figure 1.2A) plays a role in ATP-independent strand separation and heteroduplex formation between the gRNA and target DNA sequences. The PAM interacting domain (PI) (Figure 1.2A) recognises the appropriate motif on the non-target strand which, for *Sa*Cas9, is 5′-NNGRR(T/N) (where N refers to any nucleotide and R a purine; Figure 1.2B) with a tolerance for another nucleotide at the 3′ end (Ran et al., 2015, Friedland et al., 2015). Cas9 unwinds the target DNA in a directional manner away from the PAM site with consecutive mismatches in the heteroduplex attenuating helicase activity (Sternberg et al., 2014). The N-terminus of the REC lobe plays a role in identifying full complementarity between the first 8 bp, otherwise known as the 'seed' region, of the gRNA:DNA heteroduplex (Nishimasu et al., 2015, Xu et al., 2017). This feature of the Cas9 protein is also important for recognition of the repeat:anti-repeat structure in the cognate sgRNA. The REC C-terminus is required for PAM-distal heteroduplex identification before the Cas9 catalytic domains become fully activated (Nishimasu et al., 2015). Cleavage by the HNH and RuvC domains results in DSB formation, following which the Cas9:sgRNA complex must dissociate from the target site to allow access and induce activity of host cell DNA repair machinery. Indels or specified recombinant sequences can be introduced into the target site by NHEJ or HDR respectively. Alternatively, non-catalytic CRISPR/Cas9 systems can be combined with activator or repressor complexes to alter gene expression levels without permanently disrupting the sequence.

Even with a narrowed genome targeting landscape, unwanted off target effects have been observed using the CRISPR/*Sa*Cas9 gene editing system in a number of studies (Cradick et al., 2013, Friedland et al., 2015, Kleinstiver et al., 2015a). It is therefore crucial to explore additional recent advancements for further improved specificity in the quest to uncovering a safe and AAV-deliverable anti-HIV-1 gene therapy. Such improvements to CRISPR/Cas9 specificity have largely been focused on *Sp*Cas9 and insights gained have informed the path

forward for other orthologous systems, particularly *Sa*Cas9. In studies conducted to determine optimal gRNA lengths for *Sp*Cas9 system functioning, truncated sequences of 16 – 17 bp, as opposed to the original 20 bp, were found to improve on target specificity contrary to expectation (Fu et al., 2014). Shorter recognition sequences, despite having a greater potential number of similar sequences in the genome, also have a decreased tolerance for mismatches within the gRNA:DNA heteroduplex thereby overall improving specificity. This was, however, found not to be applicable to *Sa*Cas9 as a longer gRNA sequence of 21 – 24 bp is required for efficient cleavage activity (Ran et al., 2015, Friedland et al., 2015). Alternatively, Cas9 protein engineering has been employed to abrogate cleavage activity at non-complementary off target sites. PAM recognition specificity requirements have been increased to reduce the targeting landscape of *Sp*Cas9 (Kleinstiver et al., 2015b) but the longer *Sa*Cas9 PAM motif already allows for this without the need for further system modification. Two effective and relatively simple specificity improvement alterations to the *Sp*Cas9 protein have successfully been extrapolated to *Sa*Cas9, specifically the creation of Cas9 nickase and 'enhanced specificity' variants.

### 1.4. Generation of highly specific CRISPR/*Sa*Cas9 variants

Targeted genome engineering of the complex human genome at the nucleotide level has provided an essential alternative approach to treatment development for currently incurable diseases. Owing to relatively short recognition sequence requirements within an infinitely larger landscape, one of the greatest challenges has been avoiding permanent modification at unintended genomic sites. Simple molecular biology manipulation of existing gene editing technologies provided a possible solution to this. Novel systems have been developed that require dual targeting for functional activity. This idea was stimulated by the generation of ZFNickases that, by introducing a single stranded DNA nick at the desired target site, enhanced gene editing by the HDR pathway (Wang et al., 2012, Kim et al., 2012, Ramirez et al., 2012). The added advantage is that only the intended target can be modified while nicks generated at non-homologous loci are efficiently repaired by the base excision DNA repair pathway. This stimulated the idea of Cas9 nuclease to nickase system conversion, requiring dual system targeting of opposite DNA strands for DSB induction and NHEJ-mediated indel formation (Ran et al., 2013, Mali et al., 2013a). Sequence complementarity at two gRNA target sites, instead of one, substantially limits the genomic targeting landscape thereby minimising potentially deleterious off target effects.

Inactivation of either the RuvCI or HNH cleavage domain of *Sp*Cas9 attenuated cleavage activity at one of the two catalytic domains (Ran et al., 2013, Mali et al., 2013a). This has recently been applied to *Sa*Cas9 with the finding that a RuvCI-inactivated *Sa*Cas9 nickase functions with greater cleavage efficiency than an HNH-defective nickase (Ran et al., 2015, Friedland et al., 2015). The RuvCI-defective *Sa*Cas9 nickase (*Sa*Cas9-D10A) harbours a glutamic acid to alanine codon mutation at amino acid position 10 within the primary protein sequence (Figure 1.3).

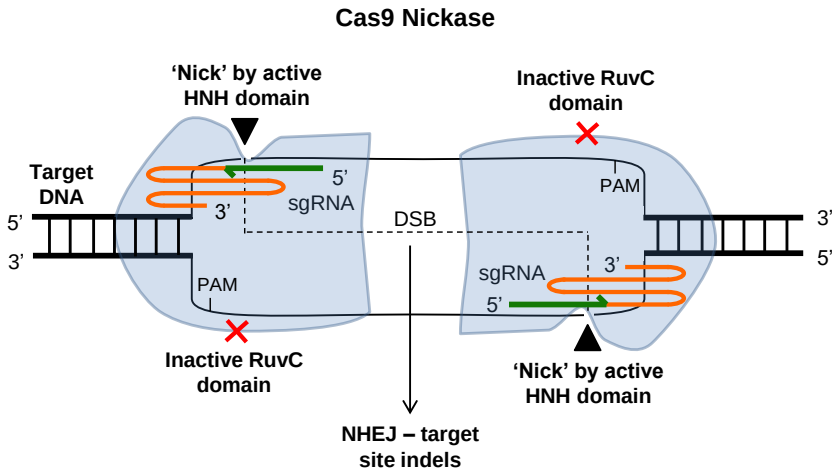



**Figure 1.3: CRISPR/*Sa*Cas9-D10A nickase and dual-sgRNA system functioning.** CRISPR/*Sa*Cas9 D10A nickase system functioning by tail-tail orientated dual-sgRNA targeting. Each *Sa*Cas9 D10A nickase:sgRNA complex scans the host genome for the desired target site and presence of non-target strand PAM motif. Full complementarity between the gRNA and target site activates *Sa*Cas9 D10A nickase activity of the HNH domains. A 5′ DSB overhang is generated and required for the induction of erroneous NHEJ DNA repair and target site indel formation.

The orientation of the two sgRNAs and offset distance between the 5′ ends of these is paramount to achieving effective DSB formation and target site indels (Ran et al., 2015, Friedland et al., 2015). Tail-tail orientated pairs are required for 5′ DSB overhang formation through the 'nicking' of opposite DNA target strands by active Cas9 HNH cleavage domains (Figure 1.3). Functional dual-sgRNA targeting with the nickase has been determined to occur at an offset distance range of -5 (a 5 bp overlap at the 5′ end of the sgRNAs) to 100 bp (Ran et al., 2015, Friedland et al., 2015). The specificity of *Sp*Cas9 nickases was determined to be as much as 100 – 1 000 fold greater than the corresponding nuclease systems but deeper investigation into nickase-induced off target activity is warranted (Ran et al., 2013, Satomura et al., 2017).

Measuring the safety profile of CRISPR/Cas9 nuclease systems has been carried out extensively (Cradick et al., 2013, Hsu et al., 2013, Fu et al., 2014, Frock et al., 2015, Tsai et al., 2015) but it appears that the methods developed for this purpose are not applicable or fully representative of nickase systems as few studies have reported on genome-wide nickase off target activity (Ran et al., 2013, Mali et al., 2013a). It is thus important to explore alternative means to improve CRISPR/Cas9 nuclease specificity to broaden the array of tools with improved safety profiles on offer.

An alternative approach to CRISPR/Cas9 gene editing specificity enhancement is structure-guided evolution of the Cas9 protein in order to destabilise electrostatic interactions at highly similar genomic off target sites (Slaymaker et al., 2016, Kleinstiver et al., 2016). This approach was focused largely on *Sp*Cas9 evolution but the study conducted by Slaymaker *et al*. (2016) extrapolated the three successful amino acid alterations within the *Sp*Cas9 nuclease to four pertinent to *Sa*Cas9 (namely R499A, Q500A, R654A and R655A). The quadruple codon-mutated *Sa*Cas9 nuclease has attenuated helicase activity such that unwinding and full length gRNA binding at mismatched off target sites is less energetically favourable (Slaymaker et al., 2016). DNA re-hybridisation is favoured over gRNA:DNA heteroduplex formation thus reducing the propensity for catalytic domain activation and Cas9-mediated cleavage. The R499A and Q500A codon mutations are located within the L1 region of *Sa*Cas9 and the R654A and R655A mutations within the RuvCIII subdomain (Nishimasu et al., 2015) (Figure 1.2A). Genome-wide detection of off target indel formation by a technique called BLESS was performed to confirm the substantial reduction in detectable off target effects compared to the wild type (WT) system (Slaymaker et al., 2016). The authors described the novel *Sa*Cas9 nuclease variant as an 'enhanced specificity' *Sa*Cas9 (e*Sa*Cas9) and this name was adopted in this study.

In the quest to uncovering a functional cure for HIV-1, various approaches have been taken to leverage the many available applications of CRISPR/Cas9-mediated genome engineering, including gene editing by both HDR and NHEJ as well as controlling gene expression levels. Assessing the advantages and disadvantages of each of these approaches informed the selection of a highly promising anti-HIV-1 target against which to test the disruptive performance of WT and 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase systems.

### 1.5. CRISPR/Cas9 for HIV-1 functional cure development

CRISPR/Cas9-mediated genome engineering has offered alternative safer and single-intervention approaches to current treatment strategies. For example, currently the "shock and kill" strategy entails the use of latency reversal agents (LRAs) for induced expression of dormant proviral genomes, rendering these vulnerable to drug and/or immune-mediated clearance (Reviewed in Deeks, 2012). This approach currently has many pitfalls, including that LRAs have broad cellular and genomic activity (Reviewed in Richon, 2006), effective treatment requires T cell activation (Bullen et al., 2014) and it has been demonstrated that LRA treatment does not result in sufficient removal of productive HIV-1 genomes (Passaes et al., 2017). The targeting capacity of CRISPR/Cas9 has been employed to circumvent these hurdles. For example, a catalytically inactive CRISPR/Cas9 system combined with four VP64 gene activator complexes induced highly specific and robust HIV-1 genome expression at the LTR without effecting permanent sequence alteration (Saayman et al., 2016). Fully catalytic CRISPR/Cas9 systems have been programmed to cleave within each of the flanking HIV-1 LTR sequences, or in two alternate regions of the genome, thereby excising the entire provirus or a portion thereof (Kaminski et al., 2016b, Kaminski et al., 2016a) and to induce the formation of permanently inactivating indels into essential HIV-1 genes (Hu et al., 2014). Importantly, the safety of anti-HIV-1 genome engineering has been demonstrated in humanised mouse models, showing great promise towards uncovering a functional cure for this devastating and currently incurable disease (Kaminski et al., 2016a, Yin et al., 2017).

The major disadvantage with direct targeting of the HIV-1 genome by CRISPR/Cas9 is that every HIV-1 genome must be effectively removed *in vivo* in order to eliminate the potential for viral rebound in the future. Current limited knowledge around the full complement of latent viral cellular and anatomical reservoirs, as well as underdeveloped gene therapy delivery modalities for *in vivo* treatment, makes this approach extremely challenging. Even if these hurdles are overcome, it was recently discovered that highly conserved regions of HIV-1 are able to circumvent the disruptive effect of gene editing because non-deleterious mutations have the paradoxical effect of rendering the target site resistant to subsequent rounds of recognition and modification (Wang et al., 2016a, Wang et al., 2016b). Alternatively, re-engineering of the human immune system for HIV-1 resistance has emerged as another functional cure approach that circumvents the challenges associated with direct HIV-1 genome targeting and does not solely rely on *in vivo* therapeutic activity.

Human elements that are essential for cell-free and cell-associated HIV-1 infection, but which are also dispensable to normal immune functioning, are highly promising anti-HIV-1 targets. The human CCR5 cell surface receptor is one such example, with no apparent deleterious immune effects if absent (Perez et al., 2008, Hutter et al., 2009). The other major HIV-1 co-receptor, CXCR4, has been knocked out by gene editing in a handful of studies (Schumann et al., 2015, Hou et al., 2015) but major deleterious immune effects have been identified in CXCR4-deficient mice making this an unattractive target (Ma et al., 1998, Cash-Padgett et al., 2016). CD4 function is also known to be essential for normal adaptive human immunity thus the possibility for genetic ablation of the two other key HIV-1 entry mediators is eliminated. The pivotal role that the human CCR5 receptor plays in HIV-1 infection as well as the putative dispensability of this protein has driven efforts to ablate expression by highly targeted gene editing. The versatility and specificity of genome editing paired with the preclusion of viral transmission by human cell engineering has emerged as a highly promising HIV-1 functional curative strategy.

### 1.6. *CCR5* as a promising anti-HIV-1 gene therapy target

There are a number of reasons for the attractiveness of CCR5-specific therapy for the functional cure of HIV-1 infection. CCR5 is the major co-receptor used by sexually transmitted T/F HIV-1 and it is necessary for both cell-free as well as cell-associated R5-tropic viral transmission (Salkowitz et al., 2003). An increased prevalence of R5- over X4-tropic HIV-1 has been observed and, although the mechanism remains to be fully elucidated, this phenomenon has been attributed in part to a lower replicative fitness of the X4- over the R5-targeting variants (Schweighardt et al., 2004). Another key motivator for the promise of CCR5-specific targeting is that certain human sub-populations display normal immune functioning despite a genetic mutation rendering the CCR5 co-receptor non-functional (Liu et al., 1996). This mutation is referred to as CCR5-delta 32 (CCR5-Δ32) and results in the production of a truncated, non-functional protein that is contained within intracellular compartments. A potent *CCR5*-specific short hairpin RNA (shRNA), shRNA 1005, has been developed for the safe and effective knock down of CCR5 messenger RNA (mRNA) and consequently cell surface protein expression, conferring R5-tropic HIV-1 resistance to a humanised mouse model of HIV-1 infection (Shimizu et al., 2009). This serves as a useful positive control for functional assessments of *CCR5*-specific gene editing but is less attractive than permanent gene editing as a gene therapy approach owing to the necessity for sustained shRNA expression to maintain the knock down phenotype.

A number of studies have successfully employed targeted gene editing for *CCR5* disruption, highlighting the potential of this approach for the functional cure of HIV-1. To describe but a few, TALENs were delivered to human T cell cultures using AAVs, resulting in efficient disruption of the *CCR5* gene through the high fidelity HDR DNA repair pathway (Sather et al., 2015). ZFNs have likewise been employed for *in vivo* editing of *CCR5* in mice (Didigu et al., 2014) as well as for *ex vivo* disruption in $CD4^+$ T cells followed by autologous T cell transplantation in mice (Perez et al., 2008, Tebas et al., 2014). A phase II clinical trial is currently underway to investigate the safety and efficacy of treatment with SB-728-T, *ex vivo* ZFN-modified patient-specific $CD4^+$ T cells with CCR5 cell surface receptor knock out (Sangamo Therapeutics, 2014). CRISPR/Cas9 has been employed for *CCR5* disruption within human-derived cell lines (Cho et al., 2013) as well as induced pluripotent stem cells (iPSCs) (Kang et al., 2015, Ye et al., 2014) thus demonstrating the potential for the application of gene editing in iPSC models and generation of a broad R5-tropic HIV-1-resistant cell repertoire. These studies employed the *Sp*Cas9 model for targeted disruption of the *CCR5* gene and more recently a study has investigated the use of the more compact *Sa*Cas9 variant for ablation of human *CCR5* (Friedland et al., 2015). The function of CCR5 and molecular basis for resistance conferred by one of the major mutant forms will be discussed so as to understand what may be required for successful preclusion of R5-, and potentially even X4-tropic, HIV-1 infection by gene editing. (Friedland et al., 2015)

### 1.6.1. Structure of the human *CCR5* gene and protein

The *CCR5* gene is located on chromosome 3, arm p21 and is 6 065 base pairs (bp) long with two promoters (Pr1 and 2), three exons and two introns (Liu et al., 1996, Mummidi et al., 1997) (Figure 1.4A).

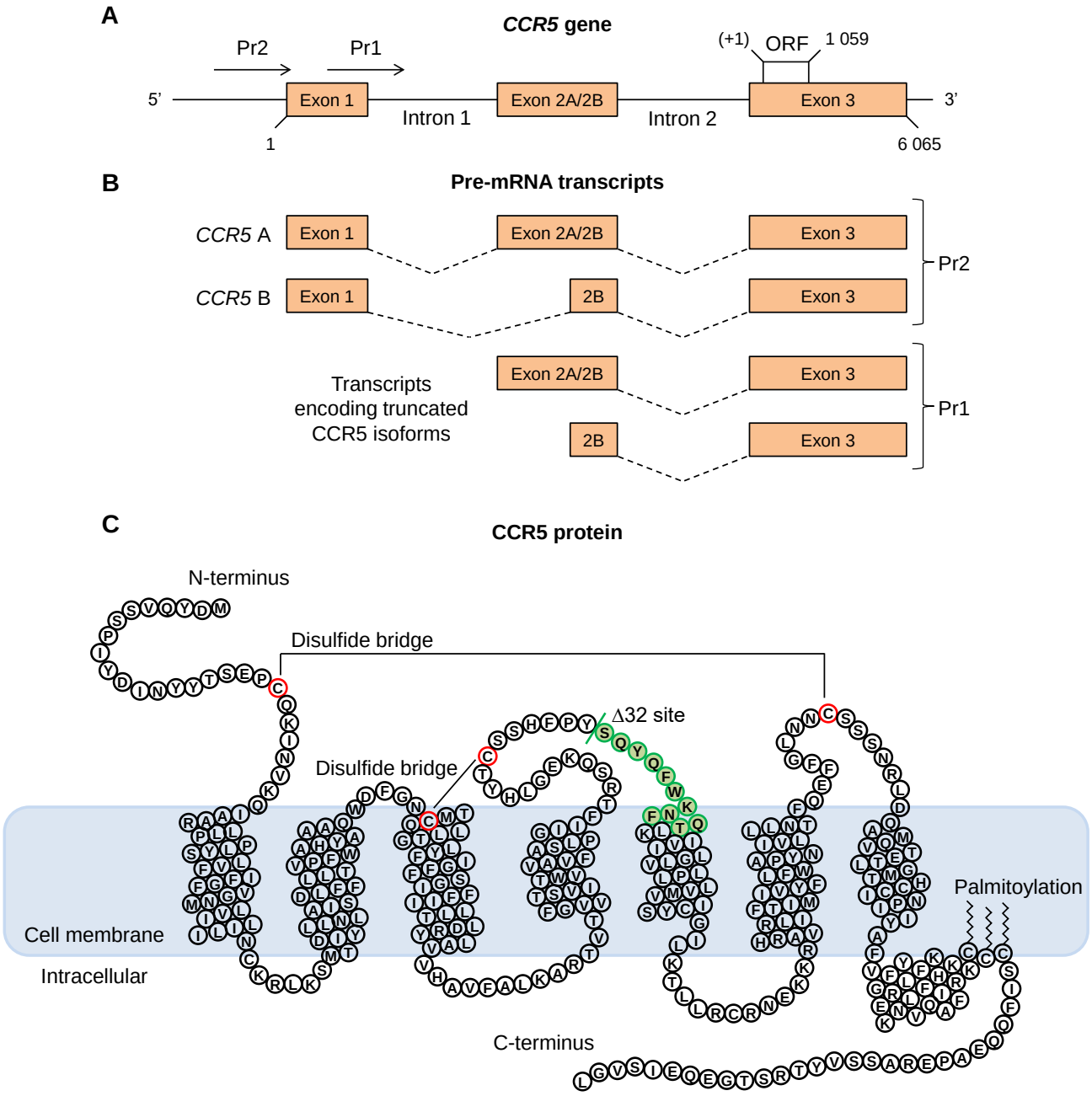



**Figure 1.4: Structure of the *CCR5* gene, pre-mRNA transcripts and protein as expressed on the surface of human cells. A** – Structure of the *CCR5* gene which consists of 6 065 bp, divided into three exons and two introns. Two promoters, Pr1 and 2, drive the expression of various *CCR5* pre-mRNA transcripts. Exon two is divided into two parts, A and B, that correlate with the two major mRNA transcripts produced by Pr2 expression. The ORF is contained within exon 3 and is 1059 bp in length. **B** – The pre-mRNA transcripts produced by Pr2 and 1 include those containing exon 1 (*CCR5*A and B) as well as those encoding various truncated protein isoforms respectively. **C** – The structure of the 352 aa GPCR CCR5 protein as expressed within the human cell plasma membrane*, consisting of an external amino terminus (N-terminus), seven transmembrane domains, three external loops, three internal loops and an intracellular signal transducing carboxy terminus (C-terminus). Two disulfide bridges stabilise the tertiary protein structure and palmitoylation moieties at the C-terminus are necessary for trafficking of the correctly folded receptor to the cell membrane.

The CCR5-Δ32 mutation site is indicated. This 32 bp deletion in the ORF eliminates the codons encoding the amino acids shown in green, generating a truncated, non-functional 215 aa protein. *The CCR5 protein structure diagram in **C** was adapted from Palmentier et al. (2015).

Exon 2 of the *CCR5* gene is divided into two parts, A and B. Pr2 expresses two different precursor messenger RNAs (pre-mRNAs) encoding exon 1-containing functional isoforms of CCR5, denoted *CCR5* A and B according to the length of exon 2 included in the transcript (Liu et al., 1998) (Figure 1.4B). Pr1 is situated downstream of Pr2 and expresses some functionally encoding, but mostly truncated, *CCR5* pre-mRNA transcripts owing to a number of ATG start sites within the 5′ end of the sequences as well as many premature stop codons (Mummidi et al., 1997, Mummidi et al., 2007). A programmed -1 ribosomal frameshift (-1 PRF) signal (UUUAAAA at nucleotide position 407 of the *CCR5* ORF) induces ribosomal slippage upon translation and reading of premature termination codons (PTCs) in the pre-mRNA molecule (Belew et al., 2008, Belew et al., 2014). Immediately downstream of this -1 PRF site the *CCR5* mRNA folds into a tandem stem loop, otherwise referred to as a pseudoknot. Furthermore, microRNA 1225 (miR-1224) has been shown to bind to the -1 PRF region, enhancing pseudoknot formation and ribosomal slippage at this site. These processes culminate in nonsense-mediated decay (NMD) of *CCR5* mRNA transcripts, a mechanism believed to allow cells to regulate CCR5 expression and fine-tune immune activity.

Exon 3 of *CCR5* encodes the full open reading frame (ORF) of CCR5 (which culminates at bp position 1 059), a portion of the 5′ untranslated region (UTR) and the entire 3′ UTR of functional *CCR5* mRNA transcripts. The human CCR5 protein is a 352 amino acid (aa) and 40.6 kDa G-protein coupled receptor (GPCR) (Combadiere et al., 1996, Samson et al., 1996a) (Figure 1.4C) utilised by many chemokines, including RANTES, the alpha and beta forms of the macrophage inflammatory factor one (MIP-1α and MIP-1β) and monocyte chemotactic protein two (MCP-2), which have been shown to inhibit HIV-1 infection when interacting with the receptor (Combadiere et al., 1996, Gong et al., 1998, Dragic et al., 2000). Like all GPCRs, it contains an external amino terminus (N-terminus), seven transmembrane domains, three external and three internal loops as well as an intracellular signal transducing carboxy terminus (C-terminus; Figure 1.4C) (Horuk, 1994). Two disulfide bonds, characteristic of human chemokine receptors, are required for correct folding and stability of the protein (Wu et al., 1997) and palmitoylation at the C-terminus is required for trafficking of the protein to, and anchorage within, the human cell plasma membrane (Blanpain et al., 2001). The CCR5-Δ32 mutation results in the deletion of the gene codons encoding the amino acids shown in

green (Figure 1.4C) and creates a truncated, 215 aa protein that is non-functional and not trafficked to the plasma membrane (Blanpain et al., 2001). The capacity to mimic the phenotypic effect of the CCR5-Δ32 mutation that confers resistance to R5-tropic HIV-1 infection is of intense interest in the quest to developing a functional cure for HIV-1 (Dean et al., 1996, Liu et al., 1996, Samson et al., 1996b).

### 1.6.2. The CCR5-Δ32 resistance mutation

Many naturally occurring deleterious mutations have been identified within the *CCR5* coding sequence but one has sparked intense interest in CCR5 knock out strategies for HIV-1 functional cure development (Reviewed in Barmania and Pepper, 2013). This variant is known as CCR5-Δ32 as there is a 32 bp deletion within the ORF of the *CCR5* gene (Dean et al., 1996, Liu et al., 1996, Samson et al., 1996b). The CCR5-Δ32 allele encodes a transcript with a premature stop codon, generating the truncated non-functional 215 aa protein. The evolutionary significance of CCR5-Δ32 is somewhat unknown but it is agreed that it did not emerge under HIV-1 selective pressures (Libert et al., 1998). This allele fortuitously confers resistance to R5-tropic HIV-1 infection in homozygous individuals and reduces transmission and prolongs progression to AIDS in heterozygotes (Eugen-Olsen et al., 1997, Marmor et al., 2001). The major reason for this protective effect is the lack of cell surface expression of the CCR5-Δ32 protein which thus limits R5-tropic infectivity. This phenomenon is supported by the finding that the density of CCR5 cell surface protein correlates with the level of R5-tropic HIV-1 infectivity and that even reduced levels of the co-receptor will have a protective effect against cellular infection (Platt et al., 1998, Reynes et al., 2000, Richardson et al., 2012).

In 2008 Timothy Ray Brown, otherwise known as the “Berlin patient”, was the first known human being to be cured of HIV-1 by homozygous CCR5-Δ32 haematopoietic stem cell transplantation (Hutter et al., 2009). This patient has not displayed viral rebound since 2009, despite having a total of 2.9 % X4- and dual-tropic HIV-1 prior to stem cell transplantation which he received primarily for the treatment of leukaemia. Unfortunately this cure event has not been recapitulated, the reason believed to be as a result of the extensive chemotherapy and fully ablative bone marrow therapy that the “Berlin patient” received prior to stem cell transplantation, potentially eliminating a substantial portion of the functional latent reservoir including those cells harbouring any X4- and dual-tropic HIV-1. This is not entirely deflating for CCR5-specific functional cure development as the molecular biology of CCR5-Δ32

confers additional benefits beyond the absence of co-receptor cell surface expression of this variant.

The lack of a palmitoylated C-terminus of the truncated CCR5-Δ32 protein results in containment of the defective protein in intracellular compartments (Blanpain et al., 2001). Furthermore, it has been shown that stable production of the CCR5-Δ32 protein exerts a *trans* dominant effect on native CCR5 as well as CXCR4 by dimerising with and sequestering these molecules within the intracellular compartments (Agrawal et al., 2007, Jin et al., 2008b). This reduces cell surface expression of both major HIV-1 co-receptors thereby potentially limiting R5-, X4- and dual-tropic viral infection. Importantly, mutations that prevent stable production of a truncated CCR5 protein do not allow for the *trans* dominant negative effect on the normal CCR5 and CXCR4 receptors (Jin et al., 2008a). The miraculous cure of the "Berlin patient", seemingly redundant nature of CCR5 in normal immune functioning and *trans* dominant negative effect of the truncated CCR5- Δ32 protein on HIV-1 infectivity have all stimulated immense interest in *CCR5* knock out as an approach to HIV-1 functional cure development. Owing to the difficulty, expense and risks associated with stem cell transplantation, alternative treatment strategies are being investigated to mimic the CCR5-Δ32 phenotype.

### 1.6.3. Mimicking the CCR5-Δ32 phenotype by gene editing

Gene editing of human *CCR5 in situ* has emerged as a highly promising avenue to achieving a CCR5-Δ32 phenotypic effect for the generation of functionally resistant human immune cells. The advantage of mimicking the CCR5-Δ32 phenotype in $CD4^+$ human immune cells is that the cell surface expression of CCR5 can be reduced and a *trans* dominant negative effect exerted on CCR5 in heterozygous edited cells as well as on CXCR4 in all cellular contexts. It is expected that gene editing by targeted endonucleases in close proximity to the CCR5-Δ32 mutation locus will be able to recapitulate the protective phenotypic effect. A CRISPR/Cas9-mediated gene therapy, whether employed *ex vivo* with the intention of autologous cell transplantation or directly *in vivo*, must be deliverable by a safe and highly efficient mechanism. Employing 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase systems for the disruption of the human *CCR5* gene ultimately contributes towards understanding the functioning and efficiencies of four different CRISPR/Cas9 gene editing systems with improved safety profiles as well as on-going efforts to uncover effective AAV-deliverable gene therapies for the functional cure of HIV-1. These research areas are of

paramount importance to uncovering deliverable, safe, effective and single intervention treatment strategies for currently devastating and incurable genetic and communicable diseases.

### 1.7. Aims and Objectives

Despite the undeniable successes of the ART era, there is an urgent need for a safe, effective and permanent functional cure for HIV-1 as the pandemic continues to intensify its global reach. The emergence of highly specific gene editing technologies has provided a promising avenue to thwart the infective and replicative capacity of HIV-1 permanently. The purpose of this study was to employ the 'enhanced specificity' CRISPR/Cas9 nuclease and D10A-mutated nickase gene editing systems derived from *Staphylococcus aureus* for the highly specific disruption of the human *CCR5* gene. The level of target site mutagenesis, effect on *CCR5* mRNA production and ultimately propensity for R5-tropic HIV-1 infectivity was measured to satisfy this ambition. In accordance with the theme of improved specificity, an auxiliary aim of the study was to characterise the types of on target indels generated by the WT *Sa*Cas9 nuclease and nickase systems. The goal was to select highly similar computationally predicted off target sites and compare cleavage activity between the nuclease and nickase variants by deep sequencing. In order to achieve these aims, the following detailed objectives were carried out:

1. *Assess the propensity for highly specific CRISPR/SaCas9 nuclease and nickase systems to induce CCR5 target site modification*

Two different CRISPR/*Sa*Cas9 gene modifications were created to improve on target specificity of the *CCR5*-specific gene editing system. Firstly, the *Sa*Cas9 nuclease was converted to a nickase by the introduction of a D10A codon mutation (Ran et al., 2015). Secondly, both the *Sa*Cas9 nuclease and nickase were converted to 'enhanced specificity' variants by the introduction of R499A, Q500A, R654A and R655A codon mutations into the *Sa*Cas9 gene sequence (Slaymaker et al., 2016). It has been demonstrated that the 'enhanced specificity' *Sa*Cas9 nuclease performs with comparable efficiency to the WT nuclease but no known study has investigated the effect on *Sa*Cas9-D10A nickase functioning. This study therefore aimed also to determine the effect of the 'enhanced specificity' *Sa*Cas9 mutations on both the nuclease as well as the D10A-mutated nickase. A panel of sgRNAs were designed to target the region of the *CCR5* ORF that harbours the well-characterised CCR5-Δ32 locus and the cleavage efficiency of each WT nuclease and nickase-based system tested

using indel sensitive endonuclease assay, online sequence deconvolution as well as digital polymerase chain reaction (PCR). The CRISPR/*Sa*Cas9 nuclease and nickase systems that introduced the highest percentage of indels with both the WT and 'enhanced specificity' variants were selected for use in downstream experiments for the functional analysis of *CCR5* gene targeting.

2. *Investigate the functional implications of CCR5 genetic modification*

Two functional effects of *CCR5* gene editing were measured. Firstly, the level of *CCR5* mRNA was quantified using reverse transcriptase PCR (qRT-PCR) following treatment of TZM-bl cells with the most efficient WT and 'enhanced specificity' nuclease and nickase CRISPR/*Sa*Cas9 systems identified to meet objective 1. Secondly, the propensity for R5-tropic HIV-1 pseudovirus to infect cells was determined by the TZM-bl luciferase assay. The short hairpin RNA 1005 (shRNA 1005), previously shown to knock down *CCR5* mRNA and cell surface expression in a potent and highly specific manner, was included in both the qRT-PCR and TZM-bl assays as a positive control (Shimizu et al., 2009, Shimizu et al., 2015).

3. *Characterise CCR5 on target indels and inspect off target activity of the WT SaCas9 nuclease and nickase systems at highly similar computationally derived genomic sites*

It has been demonstrated that *Sa*Cas9 nuclease mediated gene editing can result in off target activity and thus the specificity of this system was compared to that of the nickase to determine whether off target activity was detectable and, if so, reduced by dual-sgRNA nickase targeting. Five highly similar exonic sequences to the two most efficient *CCR5*-specific sgRNAs elsewhere in the human genome were selected for off target analysis. Both the *CCR5* on target indels as well as any off target activity at the selected genomic sites introduced by WT *Sa*Cas9 nuclease and nickase treatments were characterised by targeted amplicon deep sequencing.

## 2. Materials and Methods

### 2.1. Plasmids and cell lines used in this study

#### 2.1.1. Plasmid constructs

The pX601-AAV-CMV::NLS-*Sa*Cas9-NLS-3xHA-bGHpA;U6::*BsaI*-sgRNA packaging vector (pAAV-CMV_*Sa*Cas9; Appendix 7.1.1.) was a gift from Feng Zhang (Addgene plasmid #61591; Ran et al., 2015) and used to express the *Sa*Cas9 nuclease and to generate and express the D10A-mutated nickase and 'enhanced specificity' variants (gene sequences in Appendix 7.2.1. – 7.2.4.). The *CCR5*-specific sgRNAs were expressed from the pTZ57R/T_U6-*BbsI*-sgRNA-*Sa*Cas9_tracrRNA (U6-sgRNA; Appendix 7.1.2. and 7.2.5) and in later experiments from the pTZ57R/T_U6-*BbsI*-sgRNA-*Sa*Cas9_modtracrRNA (U6-sgRNA_modtracr) vectors (Appendix 7.1.3. and 7.2.6.) as an extended tetraloop and A-U flip modifications (Table 2.1) increase sgRNA stability and therefore potential cleavage activity.

**Table 2.1: Sequences of the original and modified *Sa*Cas9 sgRNA scaffolds**

| Type of Scaffold | Sequence (5′ to 3′) |
|---|---|
| **Original *Sa*Cas9** | $N_{(21-22)}$ – guuuuaguacucuggaaacagaaucuacuaaaacaaggcaaaaugccguguuuaucucgucaacuuguuggcgaga |
| **Modified *Sa*Cas9** | $N_{(21-22)}$ – guuuuaguacucug**UGCUG**gaaac**AGCAC**agaaucuacu**U**aaacaaggcaaaaugccguguuuaucucgucaacuuguuggcgaga |

Note: $N_{(21-22)}$ refers to the variable gRNA sequence

A plasmid containing the green fluorescent protein (GFP) gene within a pCI vector backbone was used to assess transfection efficiency and the CCR5 mRNA-specific shRNA 1005 was expressed from a pTZ57R/T vector backbone (Thermo Fisher Scientific, USA). For ZM53 pseudovirus production, the pSG3Δenv plasmid was obtained from Dr John C. Kappes & Xiaoyun Wu and the ZM53M.PB12 plasmid from Drs E. Hunter and C. Derdeyn.

#### 2.1.2. Cell lines

All cell culture reagents, flasks and plates were obtained from Thermo Fisher Scientific (USA) and serological pipettes from Merck (Germany), unless otherwise stated. Cells were grown in a humidified incubator at 37 °C with 5 % $CO_2$. Trypan Blue Solution (0.4 %) was mixed with cells in a 1:1 ratio and 10 µl added to a Countess Cell Counting Chamber Slide to determine cell number of viability using the Countess® Automated Cell Counter. Trypan Blue

dye permeates dead cell membranes thus the cell counter identifies the number of cells per ml able to exclude the dye.

Human embryonic kidney (HEK) 293T and TZM-bl (JC53-bl) adherent cells were maintained in High Glucose Gibco™ Dulbecco's Modified Eagle Medium (DMEM) to which 1X Gibco™ GlutaMAX™ Supplement, 10% Gibco™ Foetal Bovine Serum (FBS) and 1X Gibco™ Antibiotic-Antimycotic (PSA: Penicillin, Streptomycin and Amphotericin B) were added. Cells were maintained in T-25 flasks and passaged 1:10 at 70 – 90 % confluency by removing media, washing with 1X phosphate buffered saline (PBS; Appendix 7.3.1.) and detaching the cells with TrypLE™ Express reagent. Stocks of all cells were prepared by slow freezing to - 80°C of 1 ml aliquots in freeze medium (Appendix 7.3.2.) in order to maintain a low passage number for transfections. HEK293T cells were selected for screening of the panel of sgRNAs for cleavage activity owing to ease of transfection and robust growth characteristics. The TZM-bl cell model was selected for post-screening experimentation as these cells express a fully functional *CCR5* protein and are tractable to HIV-1 infection whereas HEK293T cells are heterozygous for the *CCR5*-Δ32 mutation and are thus partially resistant to infection (Qi et al., 2016).

### 2.2. Cloning of *CCR5*-targeted *Sa*Cas9 sgRNAs; the *Sa*Cas9 D10A-mutated nickase and 'enhanced specificity' *Sa*Cas9 nuclease and nickase variants

#### 2.2.1. Generation of the *Sa*Cas9 D10A-mutated nickase construct by site-directed mutagenesis and Gibson Assembly cloning

The *Sa*Cas9-D10A nickase construct (pAAV-CMV_*Sa*Cas9-D10A; Appendix 7.1.1.) was generated in the research project leading up to this study. Site-directed mutagenesis polymerase chain reaction (SDM-PCR) was carried out to introduce a GAC to GCC codon mutation, converting the aspartic acid residue at position ten of the *Sa*Cas9 protein to an alanine (D10A). This modification inactivates the RuvCI catalytic domain, generating an *Sa*Cas9-D10A nickase (Ran et al., 2015). Primers were designed and melting temperatures ($T_m$s) calculated using SnapGene® v3.0.3. The OligoAnalyzer 3.1 online programme (Integrated DNA Technologies, USA) was used to assess primer propensity for self- or heterodimer and hairpin formation. A change in Gibbs free energy (ΔG) of greater than - 9 was accepted for the primer. SDM-PCR was carried out using a 135 base pair (bp) Ultramer® forward primer (RuvCI F), harbouring the appropriate codon mutation and spanning the *AgeI*

restriction enzyme site, and a 31 bp reverse primer (RuvCI R), spanning the *HindIII* restriction site (Table 2.2; Appendix 7.1.1.).

**Table 2.2: Primers used for SDM-PCR of the *Sa*Cas9-D10A codon mutation**

| Primer name | Primer sequence (5′ – 3′) | Primer length (bp) | Primer $T_m$ (°C) |
|---|---|---|---|
| **RuvCI F** | TATAAGCAGAGCTCTCTGGCTAACTACCGGT GCCACCATGGCCCCAAAGAAGAAGCGGAAG GTCGGTATCCACGGAGTCCCAGCAGCCAAGC GGAACTACATCCTGGGCCTGGCCATCGGCAT CACCAGCGTGGG | 135 | 82 |
| **RuvCI R** | GTTGATCACTTTGATGCTCTGGATGAAGCTT | 31 | 62 |

The PCR reaction was carried out using 2 ng of pAAV-CMV_*Sa*Cas9 template, 0.5 µM RuvCI F and R primers and 1X KAPA2G Robust DNA Polymerase (containing 1.5 mM $MgC\ell_2$; KAPA Biosystems, South Africa) with the appropriate cycling conditions (Table 2.3).

**Table 2.3: Cycling conditions used for KAPA2G Robust PCRs**

| Step | Temperature (°C) | Hold (min:sec) | Cycles |
|---|---|---|---|
| **Initial denaturation** | 95 | 03:00 | 1 |
| **Denaturation** | 95 | 00:15 | |
| **Annealing** | 60 | 00:15 | 35 |
| **Extension** | 72 | 00:15 | |
| **Final extension** | 72 | 01:00 | 1 |

Approximately 1 – 2 µl of the PCR fragment was run on a 1 % agarose gel (Appendix 7.4.1.) to confirm the expected size of 1471 bp and thereafter purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA; Appendix 7.4.2). One microgram of the pAAV-CMV_*Sa*Cas9 vector was digested with 10 units (U) each of *AgeI* and *HindIII* and 1X R Buffer (Thermo Fisher Scientific, USA; Appendix 7.3.4.) at 37 °C for one hour to remove the wild-type (WT) RuvCI domain fragment. This digested backbone was purified using the QIAquick® Gel Extraction Kit (Qiagen, Germany; Appendix 7.4.3.) and Gibson Assembly cloning employed to generate the pAAV-CMV_*Sa*Cas9-D10A vector.

The RuvCI F and R primers were designed to introduce 20 bp of flanking homologous sequences to the digested vector backbone, allowing for recombination by Gibson Assembly

cloning. The concentration (ng/µl) of each fragment to be cloned was converted to femtomoles (fmol) according to the following formula:

$\text{no. fmol} = \frac{\text{ng/µl x 1000}}{\text{Molecular Weight}}$ x 1000 , where the dsDNA Molecular Weight was determined using the online Oligo Calculator programme version 3.27 (Kibbe, 2007).

A molar ratio of 1:3 of the digested backbone to insert was used to maximise the likelihood of vector-insert assembly. The reaction was carried out in a final volume of 10 µl containing backbone, insert and 1X Gibson Assembly® Master Mix (proprietary information, New England Biolabs, USA). A negative control with no insert DNA was included to determine the percentage of colonies formed as a result of undigested or re-ligated vector backbone. The reaction was allowed to proceed in 0.2 ml PCR tubes at 50 °C for one hour and 2 µl of assembly mixture was added to 25 µl of NEB Stable Competent *E. coli* (High Efficiency) cells (New England Biolabs, USA) for transformation (Appendix 7.4.4.) in 1.5 ml microcentrifuge tubes. Thereafter, 1 ml of pre-warmed (37 °C) SOC Outgrowth Medium (New England Biolabs, USA; Appendix 7.3.6.) was added to the cells and growth was allowed to proceed in an Orbital Shaker (Labotec, South Africa) set at 37 °C for one hour. The tubes were centrifuged at 4 000 rpm (Centrifuge 5 415; Eppendorf, Germany) for 5 minutes, 90 % of the supernatant removed and pellets suspended in the remaining solution for plating onto pre-warmed (37 °C) LB/Amp agar plates (Appendix 7.3.7.). Plates were left overnight at 37 °C in a humidified incubator to allow for colony formation. Colonies were screened by performing a miniprep with the QIAprep® Spin Miniprep Kit (Appendix 7.4.6.; Qiagen, Germany) and restriction enzyme digestion with *AgeI*/*HindIII* followed by 1 % agarose gel electrophoresis (Appendix 7.4.1.). DNA samples containing the insert were sent for Sanger sequencing (Inqaba Biotech, South Africa) to confirm successful introduction of the D10A-mutated site using the 18 bp RuvCI Seq primer (Table 2.4).

**Table 2.4: Primer used for sequence confirmation of the *Sa*Cas9-D10A codon mutation**

| Primer name | Primer sequence (5′ – 3′) | Primer length (bp) | Primer $T_m$ (°C) |
|---|---|---|---|
| **RuvCI Seq** | CGGCAGAGAACTCTTCCT | 18 | 54 |

A maxiprep was performed on the *Sa*Cas9 and *Sa*Cas9-D10A constructs using the QIAGEN® Plasmid Maxiprep Kit (Qiagen, Germany; Appendix 7.4.7.) to increase DNA purity and yield of pAAV-CMV_*Sa*Cas9-D10A for use in transfection experiments.

### 2.2.2. Production of a panel of *CCR5*-specific *Sa*Cas9 sgRNAs

A panel of ten sgRNAs with 21-22 nucleotide (nt) variable regions were designed to function individually with the *Sa*Cas9 nuclease or as tail-to-tail orientated pairs with the *Sa*Cas9-D10A. Potential target sites within the last 500 bp of the *CCR5* ORF (downloaded from GenBank, NCBI) were identified in combination with a 5′-NNGRR(T/N) PAM sequence (where the 3′ residue is preferably a thymine – T but can be any nucleotide – N) using the CRISPR MultiTargeter online tool (Prykhozhij et al., 2015). Ten sgRNAs were selected based on low homology with potential off target sites in the human genome using the Thomas Jefferson University 'Off-Spotter' online prediction programme (Pliatsika, 2015).

Double stranded oligodeoxynucleotides (dsODNs) encoding the plus (+) and minus (-) strands of each variable sgRNA region were synthesised (Inqaba Biotech, South Africa) and cloned into the U6-sgRNA expression vector (Appendix 7.1.2.). dsODNs were designed with 5′extensions facilitating cloning using *BbsI* restriction enzyme as this is a type IIS restriction enzyme with cleavage sites located externally to the recognition sequence. This allows for seamless insertion of the gRNA sequences between the promoter and sgRNA scaffolds and consequently loss of the restriction enzyme site which is useful for screening. The sequences of the single stranded ODN (ssODN) + and – strands, the resulting sgRNA variable sequences as well as the target site PAM recognition motifs are listed in Table 2.5 below.

**Table 2.5: *CCR5*-specific *Sa*Cas9 ODN + and - strand sequences, the resulting U6-sgRNA variable sequences and target site PAM recognition motifs**

| *Sa*Cas9 U6-sgRNA | ODN strand sequence (5′ – 3′) | Variable sgRNA sequence (5′ – 3′) | *Sa*Cas9 PAM motif (5′ – 3′) |
|---|---|---|---|
| **1** | (+) **CACC**G**C**TTTAATGTCTGGAAATTCTTC<br>(–) **AAAC**GAAGAATTTCCAGACATTAAAG**C** | CUUUAAUGUCUGGAAAUUCUUC | CAGAAT |
| **2** | (+) **CACC**G**T**GTCATGGTCATCTGCTACTC<br>(–) **AAAC**GAGTAGCAGATGACCATGACA**C** | UGUCAUGGUCAUCUGCUACUC | GGGAAT |
| **3** | (+) **CACC**GAAGCAGAGTTTTTAGGATTC<br>(–) **AAAC**GAATCCTAAAAACTCTGCTTC | GAAGCAGAGUUUUUAGGAUUC | CCGAGT |
| **4** | (+) **CACC**GAGAAGAAGAGGCACAGGGCT<br>(–) **AAAC**AGCCCTGTGCCTCTTCTTCTC | GAGAAGAAGAGGCACAGGGCU | GTGAGG |
| **5** | (+) **CACC**G**A**GAAAATAAACAATCATGATG<br>(–) **AAAC**CATCATGATTGTTTATTTTCT**C** | AGAAAAUAAACAAUCAUGAUG | GTGAAG |
| **6** | (+) **CACC**GTCCTTCTCCTGAACACCTTCC<br>(–) **AAAC**GGAAGGTGTTCAGGAGAAGGAC | GUCCUUCUCCUGAACACCUUCC | AGGAAT |
| **7** | (+) **CACC**G**C**TACTGCAATTATTCAGGCCA<br>(–) **AAAC**TGGCCTGAATAATTGCAGTAG**C** | CUACUGCAAUUAUUCAGGCCA | AAGAAT |
| **8** | (+) **CACC**G**T**ATGCAGGTGACAGAGACTCT<br>(–) **AAAC**AGAGTCTCTGTCACCTGCATA**C** | UAUGCAGGUGACAGAGACUCU | TGGGAT |
| **9** | (+) **CACC**GATGCAGCAGTGCGTCATCCC<br>(–) **AAAC**GGGATGACGCACTGCTGCATC | GAUGCAGCAGUGCGUCAUCCC | AAGAGT |
| **10** | (+) **CACC**G**T**GCCTTTGTCGGGGAGAAGTT<br>(–) **AAAC**AACTTCTCCCCGACAAAGGCA**C** | UGCCUUUGUCGGGGAGAAGUU | CAGAAA |

Note: A 5′-CACC extension was included for U6-sgRNA cloning with *BbsI*. An additional 5′ Guanine (red), if not already present in the target sequence, was included for improved promoter initiation.

Phosphorylation of 10 µM + and – strands for each ssODN was carried out using 5 U of T4 Polynucleotide Kinase (PNK; New England Biolabs, USA), 1X T4 PNK Reaction Buffer (70 mM Tris-HCl, 10 mM $MgCl_2$ and 5 mM DTT at pH 7.6; New England Biolabs, USA) and 1 mM adenosine triphosphate (ATP; Thermo Fisher Scientific, USA) for one hour at 37 °C. An equal ratio of phosphorylated + and – strands were mixed for each gRNA and T4 PNK inactivated at 75 °C for ten minutes in a heating block. The heating block was switched off and samples left to allow for slow annealing of complementary strands.

The vector backbones were linearised, resolved on an agarose gel and extracted for ligation with the insert. pTZ57R/T_U6-*BbsI*-sgRNA-*Sa*Cas9_tracrRNA and pTZ57R/T_U6-*BbsI*-sgRNA-*Sa*Cas9_modtracrRNA were digested with 10 U of *BbsI* in 1X NEB Buffer 2.1 (New England Biolabs, USA) according to the reaction times and temperatures described in

Appendix 7.3.4. The products were resolved on a 1 % agarose gel at 100 V for fifteen minutes. This short resolution time was used to reduce the length of exposure of the DNA to ethidium bromide, a dsDNA intercalating agent that can cause mutagenesis upon prolonged exposure. The linearised vector backbones were purified using the QIAquick® Gel Extraction Kit (Qiagen, Germany; Appendix 7.4.3.).

dsODNs were diluted from 5 µM to 200 nM and ligated with 50 ng of the linearised expression vector at 22 °C for an hour using 5 U of T4 DNA Ligase in a 1X T4 DNA Ligase Reaction Buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP and 10mM DTT at pH 7.5; New England Biolabs, USA). The T4 DNA Ligase was inactivated at 65 °C for ten minutes in a thermal cycler. A negative control was included containing no DNA insert in order to screen for the percentage of re-ligated vector backbones. As there are two restriction sites within each of the gRNA cloning sites, only the single-digested backbones will have compatible overhangs and these re-ligated vectors will result in colony formation. Following ligation of the gRNAs and vector backbone, chemically competent *E. coli* DH5α cells (Appendix 7.4.5.) were transformed (Appendix 7.4.4.), plated on LB/Amp agar (Appendix 7.3.7.) and grown at 37 °C overnight in a humidified incubator. Three colonies were picked from each plate, miniprepped using the QIAprep® Spin Miniprep Kit (Qiagen, Germany; Appendix 7.4.6.) and screened by restriction enzyme digestion with *BbsI* for the loss of the restriction site (Appendix 7.3.4.). DNA containing an insert was sent for Sanger sequencing (Inqaba Biotech, South Africa) using the 22 bp U6 Seq primer (Table 2.6) to confirm successful sgRNA cloning.

**Table 2.6: Primer used for sequence confirmation of the variable region within the cloned U6-sgRNA constructs**

| Primer name | Primer sequence (5′ – 3′) | Primer length (bp) | Primer $T_m$ (°C) |
|---|---|---|---|
| **U6 Seq** | ACTATCATATGCTTACCGTAAC | 22 | 51 |

A midiprep was performed using the QIAGEN® Plasmid Midiprep Kit (Appendix 5.7; Qiagen, Germany) following sequence confirmation to increase DNA purity and yield for use in transfection experiments. In later experiments, the two most efficient sgRNAs (7 and 8) were cloned into the U6-sgRNA_modtracr vector (Appendix 7.1.3.) as the modified scaffold sequence, including a tetraloop extension and an A-U flip (Section 2.1.1.), improves stability of sgRNA scaffold secondary structure.

### 2.2.3. Generation of the 'enhanced specificity' *Sa*Cas9 nuclease and nickase constructs by Gibson Assembly cloning

Four codon mutations were introduced into the pAAV-CMV_*Sa*Cas9 and pAAV-CMV_*Sa*Cas9-D10A constructs, generating the 'enhanced specificity' *Sa*Cas9 nuclease (e*Sa*Cas9) and nickase (e*Sa*Cas9-D10A) expression vectors respectively (Appendix 7.1.1.). These mutations were shown to improve on target specificity by reducing sgRNA:DNA heteroduplex stability at mismatched off target sites (Slaymaker et al., 2016). A gBlocks® Gene Fragment (Integrated DNA Technologies, USA) was ordered, containing the R499A, Q500A, R654A and G655A mutations, and spanning the two *XcmI* restriction sites (Appendix 7.1.1.). The mutated gene fragment was cloned into the appropriate vector using the Gibson Assembly technique (New England Biolabs, USA) as described in Section 2.2.1. The constructs were miniprepped using the QIAprep® Spin Miniprep Kit (Qiagen, Germany: Appendix 7.4.6.) and sent for Sanger sequencing (Inqaba Biotech, South Africa) to confirm successful cloning. A maxiprep was performed using the QIAGEN® Plasmid Maxiprep Kit (Qiagen, Germany; Appendix 7.4.7.) for each of the e*Sa*Cas9 and e*Sa*Cas9-D10A constructs to increase DNA purity and yield for use in transfection experiments.

### 2.2.4. Production of the *CCR5*-targeting short hairpin RNA 1005

The shRNA 1005 (referred to also as sh1005), previously shown to induce robust knock down of human *CCR5* mRNA, was included as a positive control in experiments aimed at measuring the effect of CRISPR/Cas9 gene editing on *CCR5* mRNA levels and R5-tropic HIV-1 infectivity (Shimizu et al., 2009). shRNA 1005 was expressed from an H1 promoter as this was shown to be non-toxic to transfected cells whereas a U6 Pol III promoter expresses toxic levels of the shRNA (Shimizu et al., 2009). shRNA1005 was generated by two rounds of PCR, both making use of the same 28 bp forward primer (sh1005 F; Inqaba Biotech, South Africa) which was designed to bind to the 5′ end of the H1 promoter and introduce an upstream *XhoI* restriction site to facilitate high fidelity and directional cloning into the vector backbone (Table 2.7).

**Table 2.7: Primers used to generate the *CCR5*-targeting shRNA 1005**

| Primer name | Primer sequence (5′ – 3′) | Primer length (bp) | Primer $T_m$ (°C) |
|---|---|---|---|
| **sh1005 F** | GATCCTCGAGTGCAATATTTGC ATGTCG | 28 | 59 |
| **sh1005 R1** | TACACCGTCGGACAAGGTGTAA ACTGAGCTTGCTCGGATCTGAG TGGTCTCATACAGAAC | 60 | 73 |
| **sh1005 R2** | CAGAGCGTGAGATCGGCGCGCC AAAAAAGAGCAAGCTCAGTTTA CACCGTCGGACAAGGT | 60 | 74 |

Note: Random sequences were incorporated into the final PCR product at each end (red) to allow for effective *XhoI* (sh1005 F) and *AscI* (sh1005R2) restriction enzyme recognition and cleavage at the target sites (green).

For the first round of PCR, a 60 bp Ultramer® reverse primer (sh1005 R1; Integrated DNA Technologies, USA) was designed to bind to the 3′ end of the H1 promoter and introduce a portion of the shRNA sequence (Table 2.7). The second round of PCR made use of sh1005 F and another 60 bp Ultramer® reverse primer (sh1005 R2; Integrated DNA Technologies, USA) able to bind to the 3′ end of the first PCR product and introduce the remaining shRNA sequence as well as an *AscI* restriction site (Table 2.7). PCR primer binding sites as well as the resulting sh1005 hairpin loop structure formed as a result of H1 promoter expression, as determined using the RNAstructure, Version 5.8.1 online tool (Mathews Lab, University of Rochester, USA), can be seen in Figure 2.1 A and B respectively.

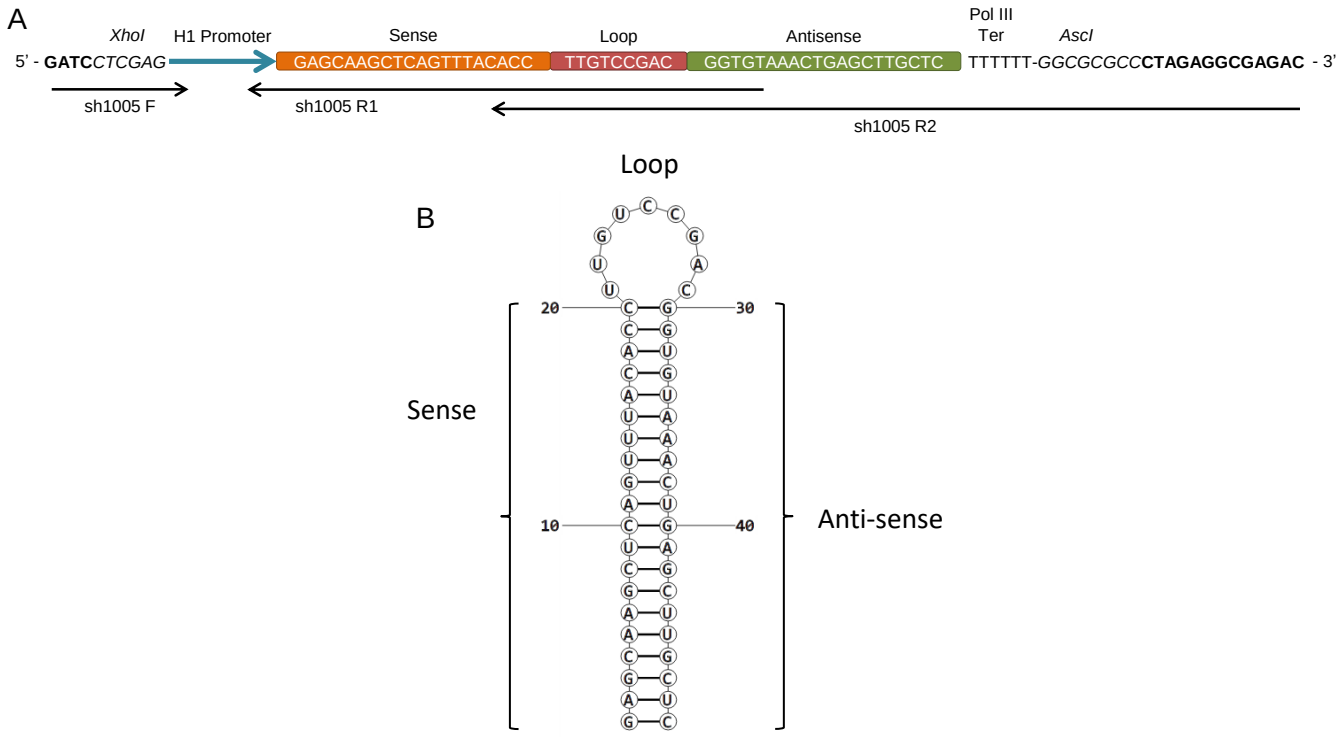



**Figure 2.1: Diagram of the PCR methodology followed for production of the *CCR5*-targeting shRNA 1005 and the hairpin loop structure formed. A** – Two rounds of PCR were carried out to generate the previously characterised *CCR5*-specific shRNA 1005 driven by an H1 promoter. The first round made use of primers sh1005 F and R1 and the second round with sh1005 F and R2 to introduce flanking *XhoI* and *AscI* sites and random 5′ and 3′ sequences to facilitate restriction enzyme target site recognition. The sense, loop, antisense and Pol III termination (Ter) signal sequences are shown in the diagram. **B** – Structure predicted by the online RNAstructure webserver of the H1-driven *CCR5*-specific shRNA showing the sense, loop and antisense sequences that fold to form the hairpin loop required to induce RNA interference within the target cell.

The PCR reaction was carried out using 2 ng of an H1 promoter-containing plasmid template, 2 mM $MgCl_2$, 5 mM dNTPs each, 10 U of the KAPA HiFi HotStart ReadyMix (HiFi HotStart) high-fidelity polymerase (KAPA Biosystems, South Africa) and 0.5 µM of the forward and appropriate reverse primer with the necessary cycling conditions (Table 2.8).

**Table 2.8: Cycling conditions used for each round of PCR for production of the *CCR5*-targeting shRNA 1005**

| Step | Temperature (°C) | Hold (min:sec) | Cycles |
|---|---|---|---|
| **Initial denaturation** | 95 | 3:00 | 1 |
| **Denaturation** | 98 | 00:15 | |
| **Annealing** | 55 | 00:15 | 35 |
| **Extension** | 72 | 00:30 | |
| **Final Extension** | 72 | 02:00 | 1 |
| **Hold** | 4 | ∞ | 1 |

Note: Ramp rate was set at 2 °C/sec.

The PCR products were run on a 1% agarose gel (Appendix 7.4.1.) to verify that amplicon sizes were as expected and that no contamination occurred during the reaction as a no DNA template was included as a blank control. The PCR products from the first round of PCR were diluted 100 times and 2 µl used in the second round with the same cycling conditions as before (Table 2.8). *XhoI* and *AscI* restriction enzyme sites were introduced using the sh1005 F and R2 primers (Table 2.7) to enable high fidelity PCR and cloning into the multiple cloning site (MCS) of the pTZ57R/T vector, instead of TA cloning which requires the use of an error-prone DNA polymerase for addition of 3′ adenosine triphosphate overhangs. The PCR product and pTZ57R/T backbone were digested with 10 U of *XhoI* and 20 U of *AscI* in 1X R Buffer (Thermo Fisher Scientific, USA) at 37 °C for one hour to create compatible overhangs according to the reaction conditions described in Appendix 7.3.4. The products were purified by gel extraction using the QIAquick® Gel Extraction kit (Qiagen, Germany; Appendix 7.4.3.) and fragments ligated in a 1:3 ratio of backbone to insert to generate the pTZ57R/T-H1_sh1005 construct (Appendix 7.2.7.). Ligation, transformation and colony miniprepping using the QIAprep® Spin Miniprep Kit (Appendix 5.6; Qiagen, Germany) was carried out in the same manner as that described for the sgRNA constructs (Section 2.2.2). Colonies were screened for the presence of the insert by *XhoI*/*AscI* digestion and agarose gel electrophoresis, and positive clones sent for Sanger sequencing (Inqaba Biotech, South Africa) to confirm the integrity of the product.

### 2.3. Transfections of HEK293T and TZM-bl cells

Prior to transfection, the constructs expressed from the pAAV-CMV_*Sa*Cas9 vector backbone (*Sa*Cas9/-D10A and e*Sa*Cas9/-D10A) were maxiprepped (Appendix 7.4.7.) owing to low plasmid copy number, and the U6-sgRNA as well as the pTZ57R/T-H1_sh1005 constructs were midiprepped (Appendix 7.4.7.), with an $A_{260}/A_{280}$ purity ratio of 1.8 – 2.0 determined using the NanoDrop® Spectrophotometer ND-1000 v3.8.1 (Thermo Fisher Scientific, USA). A plasmid containing the green fluorescent protein (GFP) gene within a pCI vector backbone was co-transfected with each sample treatment to assess transfection efficiency 24 and 48 hours later by fluorescence microscopy. Cells were detached for seeding prior to transfection and harvesting thereafter by adding TrypLE™ Express reagent to cover the base of the flask/well and incubation at 37 °C for two minutes. DMEM was added to neutralise the TrypLE™ Express and transferred to a microcentrifuge tube for downstream applications.

### 2.3.1. HEK293T cell transfections

HEK293T cell transfections were performed as a component of T7 Endonuclease I (T7EI; Section 2.4.), TIDE (Section 2.5.) and ddPCR™ (Section 2.6.) assays and were carried out in 24-well plates. Plates were seeded 24 hours prior to transfection with 1.2 x $10^5$ cells per well, at a viability of > 90 %, in 0.5 ml of DMEM lacking PSA (DMEM/$PSA^-$). HEK293T cells were transfected at 60 – 80 % confluency with 1 µg of total plasmid DNA. The pAAV-CMV_*Sa*Cas9 and U6-sgRNA expression plasmid concentrations were normalised to 450 ng/µl and the GFP construct to 100 ng/µl before setting up CRISPR/*Sa*Cas9 treatment combinations. For each treatment, 450 ng of the original or 'enhanced specificity' *Sa*Cas9 nuclease or nickase construct, 450 ng of the sgRNA expression vector or 225 ng of each sgRNA for nickase pair treatment and 100 ng of the GFP construct were mixed in 50 µl of Gibco™ Opti-MEM™ I Reduced Serum Media (Thermo Fisher Scientific, USA) in a 1.5 ml microcentrifuge tube. A 1:3 weight/weight ratio of DNA to Polyethylenimine 'Max' Mw 40 000 (PEI 'Max', Polysciences, USA) was used. Three microliters of a PEI 'Max' working solution of 1 mg/ml (Appendix 7.3.8.) in 50 µl of Opti-MEM™ I was added to each DNA sample and vortexed thoroughly before brief centrifugation. The transfection mixtures were incubated at room temperature for twenty minutes and then added drop-wise to the cells. Transfections were carried out in the absence of antibiotics as this has been found to be toxic when used in conjunction with the PEI 'Max' transfection reagent (Longo et al., 2013). Four hours post transfection the media in each well was replaced with 0.5 ml of fully supplemented DMEM ($PSA^+$) to prevent contamination.

### 2.3.2. TZM-bl cell transfections

TZM-bl cell transfections were performed as a component of ddPCR™ (Section 2.6.), deep sequencing (Section 2.7.) and qRT-PCR (Section 2.8.) assays which were carried out in 24-well plates, as well as for the Luciferase assay (Section 2.9.) which was carried out in 96-well plates. The 24- and 96-well plates were seeded 24 hours prior to transfection with 1.2 x $10^5$ and 2 x $10^5$ cells per well respectively, at a viability > 90 %. Cells were seeded in 0.5 ml or 0.2 ml of DMEM respectively. TZM-bl cells were transfected at 60 – 80 % confluency with 1 µg of total plasmid DNA for 24-well plates as described for HEK293T cells (Section 2.3.1.), or 250 ng for 96-well plates. The *Sa*Cas9 and sgRNA expression plasmid concentrations were normalised to 450 ng/µl or 112.5 ng/ µl and the GFP construct to 25 ng/µl before setting up CRISPR/*Sa*Cas9 treatment combinations. For each transfection with 250 ng of total plasmid

DNA, 112.5 ng of the original or enhanced *Sa*Cas9 nuclease or nickase construct, 112.5 ng of the sgRNA expression vector or 56.25 ng of each sgRNA for nickase pair treatment and 25 ng of the GFP construct were mixed in 10 µl of Gibco™ Opti-MEM™ I Reduced Serum Media (Thermo Fisher Scientific, USA) in a 1.5 ml microcentrifuge tube. A 1:1 weight/volume ratio of DNA to Lipofectamine® 3000 (Thermo Fisher Scientific, USA), and 1:3 volume/volume ratio of Lipofectamine® 3000 to P3000 Reagent were used. For each sample, 0.75 µl of P3000 Reagent was added to 10 µl of Opti-MEM™ I and then mixed with the DNA in a 1.5 ml microcentrifuge tube. In a separate tube, 0.25 µl of Lipofectamine® 3000 and 10 µl of Opti-MEM™ was mixed, incubated at room temperature for five minutes and then added to the DNA containing P3000 Reagent. The transfection mixtures were mixed thoroughly, centrifuged briefly and incubated at room temperature for ten minutes. Samples were then added drop-wise to the cells.

### 2.4. Quantification of indel frequency by the T7E1 assay

The propensity for each CRISPR/*Sa*Cas9 system to introduce indels into the selected *CCR5* target sites as a result of NHEJ-mediated DNA repair was analysed using the T7EI assay (Mashal et al., 1995). Briefly, this method requires target region amplification by PCR followed by product denaturation and slow re-annealing (Figure 3.3). This process encourages the formation of DNA heteroduplexes containing single stranded bulges where indels mismatch with WT sequences. The T7EI enzyme specifically cleaves the single stranded DNA generating smaller digestion products that, when resolved by gel electrophoresis, are used to visualise and quantify the sample indel frequency.

#### 2.4.1. Transfection and DNA extraction

HEK293T or TZM-bl cells were transfected according to the protocol described in Sections 2.3.1 and 2.3.2 respectively with the appropriate *Sa*Cas9 and sgRNA constructs. Cells were harvested and pelleted at 4 000 rpm for five minutes (Centrifuge 5 415; Eppendorf, Germany). The supernatant was removed and the KAPA Express Extract DNA Extraction kit (KAPA Biosystems, South Africa; Appendix 7.4.8.) used for crude extraction of cellular DNA.

#### 2.4.2. PCR and re-annealing of products for heteroduplex formation

A 620 bp fragment spanning the ten selected *CCR5*-specific target sites was amplified by PCR using 25 bp *CCR5* F and 23 bp *CCR5* R primers (Table 2.9).

**Table 2.9: Primers used to amplify the region containing the ten *CCR5*-specific target sites for indel analysis by T7EI assay**

| Primer name | Primer sequence (5′ – 3′) | Primer length (bp) | Primer $T_m$ (°C) |
|---|---|---|---|
| ***CCR5* F** | AGAAGGTCTTCATTACACCTGCAGC | 25 | 60 |
| ***CCR5* R** | CAGCCCAGGCTGTGTATGAAAAC | 23 | 59 |

A 50 µl PCR reaction was carried out using 2 µl of the extracted DNA template, 0.5 µM *CCR5* F and R primers and 1X KAPA2G Robust DNA Polymerase (containing 1.5 mM $MgC\ell_2$; KAPA Biosystems, South Africa) with cycling conditions described in Table 2.2. A small volume (1 – 2 µl) of the PCR product was run on a 1 % agarose gel (Appendix 7.4.1.) to confirm the correct size and the remainder purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA; Appendix 7.4.2.). Each sample was normalised to 100 ng/µl and 4 µl (400 ng) added to 13.5 µl of Ambion® nuclease-free water (Thermo Fisher Scientific, USA) containing 2 µl of NEB Buffer 2.1 (50 mM of $NaC\ell$, 10 mM Tris-$HC\ell$, 10 mM $MgC\ell_2$ and 100 µg/ml BSA at pH 7.9; New England Biolabs, USA) in a 0.2 ml PCR tube. Heteroduplexes were formed by denaturation and re-annealing of PCR products in a thermal cycler according to the steps described in Table 2.10.

**Table 2.10: Denaturation and re-annealing of PCR products for T7EI heteroduplex formation**

| Step | Temperature (°C) | Hold (min:sec) | Ramp rate (°C/sec) |
|---|---|---|---|
| **Denaturation** | 95 | 10:00 | 0.2 |
| **Annealing** | 85 | 01:00 | 0.2 |
| | 25 | 01:00 | |

### 2.4.3. Heteroduplex cleavage and agarose gel electrophoresis

Heteroduplexes were cleaved by the addition of 5 U of T7EI enzyme (New England Biolabs, USA) and incubation in a thermal cycler at 37 °C for 30 minutes. The products were run on a 2 % agarose gel (Appendix 7.4.1.) with the presence of lower molecular weight digestion products indicative of indel formation at the target site. A negative control included in each experiment was created by transfection with the appropriate *Sa*Cas9 construct and a sgRNA expression vector lacking the dsODN insert within the cloning site. Densitometry (ImageJ

software) was used to measure the intensities of the bands on the T7EI agarose gels and the indel frequency calculated according to the formulae below (Schneider et al., 2012).

Indel % = 100 x ($1 - \sqrt{1 - \text{fcut}}$), where fcut = $\frac{b+c}{a+b+c}$, where a = intensity of undigested PCR product; b and c = intensity of digested PCR products.

The T7EI assay was employed to select the two sgRNAs with highest cleavage efficiency when complexed with the *Sa*Cas9 nickase to be used in downstream experiments.

### 2.5. Quantification of indel frequency by the TIDE assay

The TIDE (Tracking of Indels by DEcomposition) online indel quantification tool provides both quantitative and qualitative indel analysis by determining both the frequency and types of indels present in each sequenced sample (Brinkman et al., 2014). This method was employed to analyse indels formed within the *CCR5* target site of HEK293T cells as a result of *Sa*Cas9 nuclease treatment only as it was not found to be applicable to nickase analysis. The initial protocol followed was identical to that of the T7EI assay up to the point of PCR product size and integrity verification by gel electrophoresis (Section 2.4.). Thereafter, the raw PCR products were sent for Sanger sequencing (Inqaba Biotech, South Africa). The TIDE online tool was used to align the sequence file from the negative control (treatment with *Sa*Cas9 and an empty U6-sgRNA expression vector) with the gRNA sequence in order to predict the location of the CRISPR/Cas9 cleavage site. This was verified by manual inspection of the sequencing chromatogram. The test sample and negative control sequences were automatically compared to a computationally-derived consensus sequence by the algorithm in order to determine the relative contribution of aberrant signals from each sample. This is provided graphically in the form of a histogram indicating the position of the expected CRISPR/*Sa*Cas9 cleavage site and was used by the programme to determine the indel frequency and thus overall efficiency of the treatment.

In the advanced settings panel the alignment and decomposition windows were set at 1 – 100 bp and 115 – 500 bp, respectively so as to exclude decomposition of the poor quality sequence signal towards the ends of the read; the indel size range was set at a maximum of 10 bp and the p-value threshold significance cutoff for decomposition at 0.0001. Major indels and their frequencies were called from -10 to +1 bp with the statistical significance for each calculated by the programme. The $R^2$ value indicated the proportion of sequences that fit within the computational model while the remainder were classified as noise and indels greater than the selected threshold. The algorithm then calculated the overall cleavage

efficiency by subtracting the percentage of WT sequences from the $R^2$ value percentage. This method thus provided a simple yet sensitive means for indel quantification of CRISPR/*Sa*Cas9 nuclease-treated samples to aid in the selection of the most efficient sgRNAs for downstream experimentation.

### 2.6. Quantification of indel frequency by the ddPCR™ drop-off assay

A Droplet Digital™ PCR (ddPCR™) drop-off assay was performed for the highly sensitive quantification of *CCR5* on target indel formation (Hindson et al., 2011). Following the selection of two sgRNAs with highest cleavage activity, the ddPCR™ drop-off assay was employed to quantify the indel frequencies in the target site of each sgRNA separately and as a pair with the original and enhanced *Sa*Cas9 nuclease and nickase systems. Briefly, TZM-bl cells were treated with the CRISPR/*Sa*Cas9 systems and gDNA extracted for indel quantification. A PCR reaction mix was prepared for each sample, separated into tens of thousands of nanolitre-sized droplets and amplification carried out within each. Two TaqMan® hydrolysis probes, sensitive to the presence of target site indels, were used to quantify the percentage of target sites modified by CRISPR/Cas9-induced NHEJ DNA repair by measuring droplet fluorescence. All ddPCR™ reagents and consumables were compatible with the QX200™ Droplet Reader and obtained from Bio-Rad (USA), unless otherwise stated.

#### 2.6.1. Transfection and gDNA extraction

HEK293T or TZM-bl cells were transfected according to the protocol described in Sections 2.3.1 and 2.3.2 respectively with the appropriate *Sa*Cas9 and sgRNA constructs. Cells were harvested and pelleted at 4 000 rpm for five minutes (Centrifuge 5 415; Eppendorf, Germany). The supernatant was removed and genomic DNA (gDNA) extracted using the QIAamp DNA Mini Kit (Qiagen, Germany; Appendix 7.4.9.).

#### 2.6.2. ddPCR™ probe and primer design

The *CCR5* sgRNA 7 and 8 target site-specific ddPCR™ primers and probes were designed according to the recommendations in the protocol titled *Quantification of NHEJ events (nuclease activity) using ddPCR*™ (Bio-Rad, USA). $T_m$s were calculated using the OligoAnalyzer 3.1 online programme (Integrated DNA Technologies, USA). The 23 bp *CCR5* ddPCR F and 23 bp *CCR5* ddPCR R primers (Inqaba Biotech, South Africa) were

employed to amplify 179 bp centred around the desired sgRNA 7 and 8 target sites for indel quantification by the ddPCR™ drop-off assay (Table 2.11).

**Table 2.11: Probes and primers used for indel quantification at two *CCR5* target sites by the ddPCR™ drop-off assay**

| *CCR5* ddPCR probe/primer name | 5′ modification | Probe/primer sequence (5′ – 3′) | 3′ modification | Probe/primer length (bp) | Probe/primer $T_m$ (°C) |
|---|---|---|---|---|---|
| **F primer** | None | CTCCCTACAACAT TGTCCTTCTC | None | 23 | 54.9 |
| **R primer** | None | GAGGTAGTTTCTG AACTTCTCCC | None | 23 | 54.7 |
| **Reference (ref) probe** | 6-FAM | CATAGATGATGGG GTTGATGCAGCA | BHQ 1 | 25 | 59.2 |
| **NHEJ probe 1** | HEX | TTGGCCTGAATAA TTGCAGTAGC | BHQ 1 | 22 | 55.8 |
| **NHEJ probe 2** | HEX | ATGCAGGTGACAG AGACTCT | BHQ 1 | 20 | 55.4 |

Two TaqMan® hydrolysis probes (*CCR5* ddPCR reference/ref probe and NHEJ probes 1 and 2) were designed to bind to WT sequences thus allowing for indel quantification based on the loss of the NHEJ probe fluorescence signal. The reference probe was conjugated at the 5′ end to a 6-FAM fluorochrome, quenched by a 3′ Black Hole Quencher 1 (BHQ 1; Inqaba Biotech, South Africa) and designed to bind adjacent to, and on the same strand as, the forward primer (Figure 2.2).

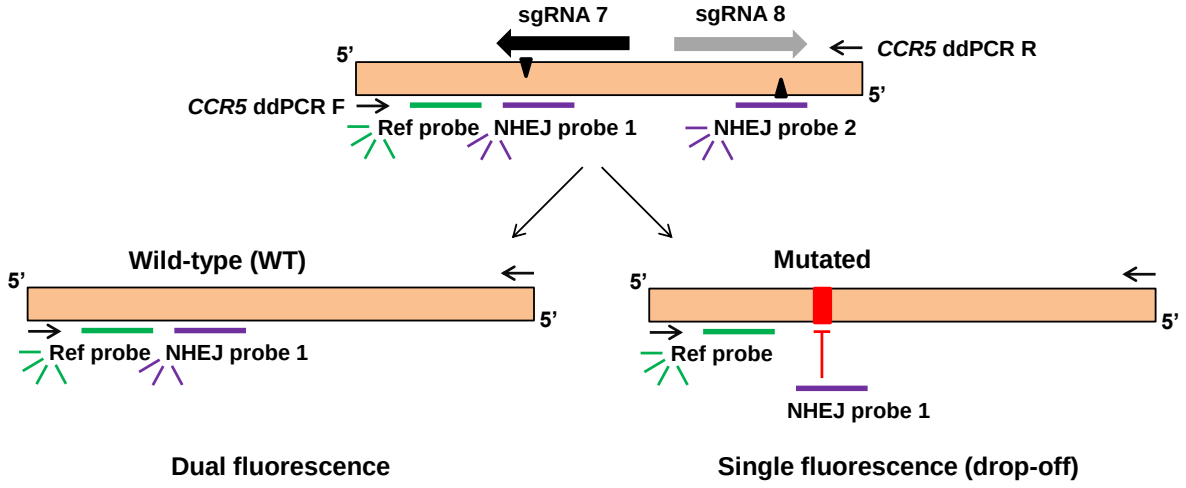



**Figure 2.2: Design of *CCR5* target site indel quantification by the ddPCR™ drop-off assay.**
A schematic showing the binding locations of probes and primers used for quantification of *CCR5* target site indels by the ddPCR™ drop-off assay. Primers *CCR5* ddPCR F and R were used to amplify a 179 bp region

containing the sgRNA 7 and 8 target sites. The TaqMan® hydrolysis probes were designed to bind on the same strand as the forward primer and only to WT sequences. The reference (ref) probe was used to quantify the number of target sites amplified and the NHEJ probes to show indel formation across the sgRNA cut site (black arrows). Each probe was conjugated with a fluorochrome at the 5′ end (the ref probe with 6-FAM and NHEJ probes with HEX) for quantification of binding by fluorescence. The Black Hole Quencher 1 (BHQ 1) was added to the 3′ end of each probe to prevent fluorescence unless hybridization occurs. Using the sgRNA 7 target site as an example, a WT sequence following treatment would result in binding of both the ref and NHEJ probes and thus dual fluorescence measured by the ddPCR™ droplet reader. Indels present in the target site would alternatively prevent NHEJ probe 1 binding and a loss of HEX fluorescence would be detected as compared to a negative (WT) control.

The ref probe design allowed for quantification of the number of *CCR5* target sites in the sample. The NHEJ probes (*CCR5* ddPCR probe 1 for sgRNA 7 or probe 2 for sgRNA 8 target site indel analysis; Figure 2.2) were conjugated to a 5′ HEX and quenched by a 3′ BHQ 1 molecule, designed to bind to the WT sequence spanning the sgRNA cut site (Inqaba Biotech, South Africa). These probes were used in separate PCR reactions to determine the number of NHEJ events in the sample as the HEX signal was expected to be absent at a mutated target site while the ref (FAM) signal is maintained. Lyophilised primers and probes were suspended in Ambion® nuclease-free water (Thermo Fisher Scientific, USA) at a concentration of 10 µM and stored in 5 µl single-use aliquots at -80 °C in the dark to protect from degradation and avoid repeated freeze-thaw cycles. According to the ddPCR™ drop-off assay guidelines, the NHEJ probe $T_m$ should be 56 – 57 °C, the primers 1 – 2 °C lower and the ref probe 60 $\pm$ 1 °C.

### 2.6.3. Droplet generation and PCR

For each sample, 100 ng of gDNA was amplified to determine the percentage of indels present by performing a ddPCR™ drop-off assay. A PCR reaction mix was set up for each sample in a 0.2 ml PCR tube containing the gDNA, 900 nM each *CCR5* ddPCR F and R primer (Table 2.11), 250 nM each of the ref probe and NHEJ probe 1 or 2 depending on the target site to be analysed (Table 2.11), 1X ddPCR™ SuperMix for Probes (No dUTP), 4 U of *HindIII*-HF (New England Biolabs, USA) and Ambion® nuclease-free water (Thermo Fisher Scientific, USA) to make up a final volume of 20 µl. Restriction enzyme digestion was carried out for fifteen minutes at room temperature in the dark, to prevent fluorophore degradation, using the *HindIII*-HF restriction enzyme as this has many target sites within the human genome to relieve DNA secondary structures that may affect assay efficiency but it does not cleave within the target site.

Each sample was mixed thoroughly to ensure even distribution of PCR reaction components for droplet generation. The 20 µl samples were loaded into the middle channel of wells in the DG8™ cartridge, air bubbles removed to prevent droplet shearing and 70 µl of Droplet Generation Oil for Probes added to the bottom channel of wells. A DG8™ gasket was secured over the cartridge and placed into the QX200™ Droplet Generator. Samples were automatically partitioned into up to 20 000 nanolitre-sized droplets which were collected in the top channel of wells in the cartridge. Droplets were then gently transferred to a clear 96-well plate (Eppendorf, Germany) which was sealed with a Pierceable Foil Heat Seal at 180 °C for five seconds in the PCR Plate Sealer. PCR was carried out in a deep well C1000 Touch™ Thermal Cycler with a heated lid set at 105 °C and according to the three-step protocol cycling conditions listed in Table 2.12.

**Table 2.12: Cycling conditions used for *CCR5* target site indel quantification by the ddPCR™ drop-off assay**

| Step | Temperature (°C) | Hold (min:sec) | Cycles |
|---|---|---|---|
| **Initial denaturation** | 95 | 10:00 | 1 |
| **Denaturation** | 94 | 00:30 | |
| **Annealing** | 58* | 01:00 | 40 |
| **Extension** | 72 | 02:00 | |
| **Final Denaturation** | 98 | 02:00 | 1 |
| **Hold** | 4 | ∞ | 1 |

Note: Ramp rate was set at 2 °C/sec. * The annealing temperature was determined empirically by a 60 – 50 °C thermal gradient experiment.

Two different negative controls were included in each ddPCR™ run. The first was a no template/water control to check for gDNA contamination as well as to set the FAM fluorescence threshold for indel quantification, discussed in Section 2.6.4. The second negative control included was samples treated with the *Sa*Cas9 variant and an empty U6-sgRNA vector in order to determine the background level of HEX signal drop-off and to set the HEX fluorescence threshold.

### 2.6.4. Analysis of ddPCR™ drop-off assay results

The TaqMan® hydrolysis probe design allowed for the ref and NHEJ probes to bind to the target site during the PCR denaturation step and fluorescence to accumulate as the hydrolysis probes become displaced during the extension step in this assay. At the end-point of PCR, the Droplet Reader was used to measure fluorescence of both FAM (number of droplets

containing a target site amplicon) and HEX (number of droplets containing a WT target sequence around the cut site), generating a 2D fluorescence plot (Figure 2.3).

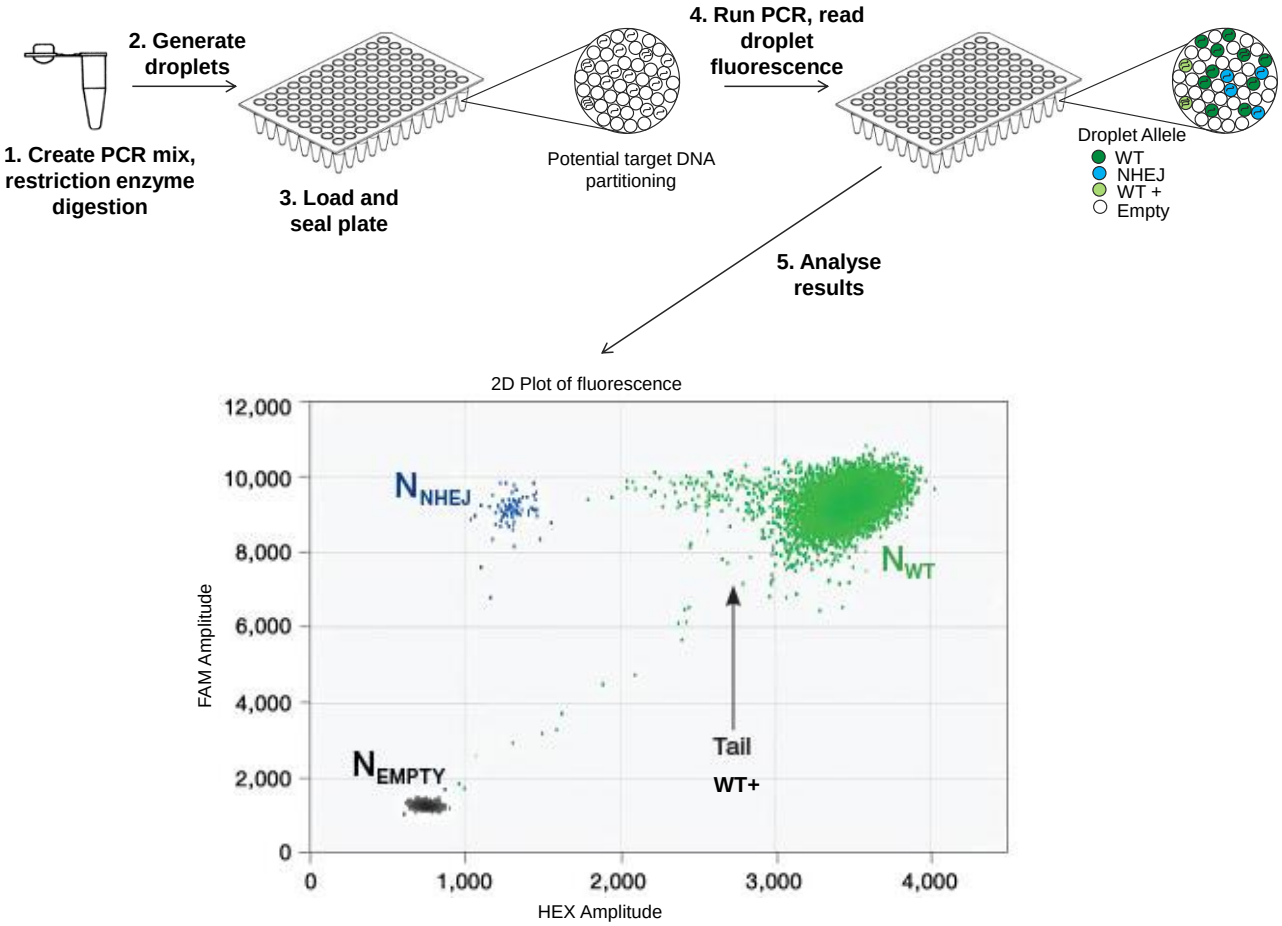



**Figure 2.3: Outline of the ddPCR™ drop-off assay methodology and an example of a 2D fluorescence plot.** The ddPCR™ method requires set up of the PCR reaction mix, containing gDNA, reference and NHEJ probes conjugated to 5′ 6-FAM and HEX respectively, forward and reverse primers, ddPCR™ Supermix for probes (no dUTPs) and a high fidelity restriction enzyme. The gDNA is fragmented by restriction digestion for fifteen minutes at room temperature before partitioning the PCR mix into droplets using the automated QX200™ Droplet Generator and Droplet Generation Oil for Probes. The droplets are gently loaded into the 96-well plate which is sealed with a Pierceable Foil Heat Seal at 180 °C for five seconds in the PCR Plate Sealer. PCR is carried out within each droplet by the three step protocol in a deep well thermal cycler and then fluorescence is read on the QX200™ Droplet Reader. Data is analysed using the QuantaSoft™ Analysis Pro software. The 2D fluorescence plot is the most visually informative output generated where clusters of fluorescence correspond to droplet contents: FAM + HEX (WT target site – 'WT'); FAM only (mutated target site – 'NHEJ'); FAM + HEX as well as FAM only (two or more target sites with both WT and mutated sequences – 'WT+' or the tail as indicated) and no FAM or HEX (no target site – 'Empty').

While it is most likely that either zero or one target site is captured per droplet, the possibility exists for the inclusion of more than one. The four potential single droplet fluorescence

outcomes include: FAM and HEX (wild-type target site/WT); FAM only (mutated target site/NHEJ); FAM and HEX as well as FAM only (two or more target sites with WT and mutated sequences/WT+) and neither FAM nor HEX (no target site/Empty; Figure 2.3). These major groupings are illustrated on the example 2D plot with FAM and HEX fluorescence amplitudes per droplet on the Y- and X-axis respectively. WT+ droplets (the 'tail') were grouped with the WT droplets during analysis.

The indel frequency per sample is calculated according to the following formulae:

$c = -\ln(\frac{\mathrm{N}neg}{\mathrm{N}total})/\mathrm{V}droplet$; where for WT quantification, $c$ = concentration of target copies amplified; ln is the natural logarithm; $N_{neg} = (N_{empty} + N_{NHEJ})$; $N_{total} = (N_{empty} + N_{NHEJ} + N_{WT+})$; $V_{droplet}$ = 1 nanolitre (nl) and for NHEJ quantification N*neg* = N*empty*; N*total* = $N_{empty} + N_{NHEJ;}$ $V_{droplet}$ = 1 nl.

Indel frequency (%) $= \frac{c(\mathrm{NHEJ})}{c(\mathrm{WT})} \mathrm{x}\ 100$

Analysis was performed using the newest version of the ddPCR software, QuantaSoft™ Analysis Pro (Bio-Rad, Sandton, SA), which automatically calculates the indel frequency (otherwise termed fractional abundance) based on the position of the FAM and HEX thresholds set manually using the position of droplets in the two negative controls.

### 2.7. Indel quantification and detection of off target effects by deep sequencing

The WT *Sa*Cas9 and *Sa*Cas9-D10A systems were analysed for on and off target indel formation using targeted deep sequencing to determine whether a nuclease to nickase conversion results in improved specificity of the *Sa*Cas9 at the *CCR5* target site. Briefly, TZM-bl cells were treated with the *Sa*Cas9 construct and an empty U6-sgRNA vector as a mock control, as well as with U6-sgRNA 7 and 8 individually to assess on and off target cleavage activity. TZM-bl cells were also treated with a combination of the *Sa*Cas9-D10A construct and the U6-sgRNA 7/8 pair. Following transfection, the gDNA was extracted from cells and regions that were identified as potential off target sites for sgRNA 7 and 8 were amplified along with the *CCR5* on target site for deep sequencing analysis using Illumina MiSeq technology.

#### 2.7.1. Transfection and gDNA extraction

TZM-bl cells were transfected in 24-well plates, according to the protocol described in Section 2.3.2, with the appropriate *Sa*Cas9 and sgRNA constructs. Cells were harvested and pelleted at 4 000 rpm for five minutes (Centrifuge 5 415; Eppendorf, Germany). The

supernatant was removed and gDNA extracted using the QIAamp DNA Mini Kit (Qiagen, Germany; Appendix 7.4.9.). Each transfection was carried out in triplicate wells and the cells from each combined before gDNA extraction to increase the yield. Three independent transfections were carried out for targeted amplicon deep sequencing of biological triplicates.

### 2.7.2. Prediction and selection of potential off target sites

In order to select the top five potential off target sites for each of the *Sa*Cas9 sgRNAs 7 and 8, the CRISPRseek prediction package was used (Zhu et al., 2014). This is a customisable tool for the prediction of off target gRNA sites based on experimentally-derived recognition and cleavage information. A single workflow is able to predict and score potential off target gRNAs, using the R programme platform with packages compatible with Bioconductor. R (http://www.r-project.org) and Bioconductor (http://www.bioconductor.org/install/) were installed before proceeding. In the R console, the CRISPRseek package was installed and loaded by using the following command-line:

*> source("http://bioconductor.org/biocLite.R")*
*> biocLite(CRISPRseek)*
*> library(CRISPRseek)*

The human genome sequence (hg19) obtained from UCSC (University of California, Santa Cruz Genome Browser) and the genome annotation packages were installed and loaded using the following command-line:

*> biocLite("BSgenome.Hsapiens.UCSC.hg19")*
*> biocLite("TxDb.Hsapiens.UCSC.hg19.knownGene")*
*> library("BSgenome.Hsapiens.UCSC.hg19")*
*> library("TxDb.Hsapiens.UCSC.hg19.knownGene")*

The "*offTargetAnalysis*" workflow function for sgRNA 7 was then run using the following command-line in the R console:

*> inputFilePath <-*
*DNAStringSet("GGAATTCTTTGGCCTGAATAATTGCAGTAGCTCTAACAGGTTGGACCAA*
*GCTATGCAGGTGACAGAGACTCTTGGGATGACGC")*
*> names(inputFilePath) <- "CCR5"*
*> outputDir = "CCR5CRISPRseekOutput"*

*> results <- offTargetAnalysis(inputFilePath, gRNAoutputName = "CCR5gRNAs", format="fasta", findgRNAs=TRUE, exportAllgRNAs=c("all", "fasta", "genbank", "no"), findgRNAsWithREcutOnly=FALSE, findPairedgRNAOnly=FALSE, gRNA.name.prefix="gRNA", PAM.size=6, gRNA.size=21, PAM="NNGRRT", BSgenomeName=Hsapiens, chromToSearch="all", max.mismatch=6, PAM.pattern = "NNG[A|G][A|G]N$", min.score=0, topN=1000, topN.OfftargetTotalScore=10, annotateExon=TRUE, txdb=TxDb.Hsapiens.UCSC.hg19.knownGene, outputDir=outputDir, fetchSequence=TRUE, upstream=300, downstream=300, weights=c(0, 0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583), overwrite=TRUE)*

For prediction of sgRNA efficacy once sgRNA target sites 1 – 10 were identified using MultiTargeter, the first command was changed to include the sequence of the last 500 bp of the *CCR5* ORF:

*>inputFilePath <-*
*DNAStringSet("ATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCT*
*GGTCCTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAATCCTAAAAACTCTGCTTC*
*GGTGTCGAAATGAGAAGAAGAGGCACAGGGCTGTGAGGCTTATCTTCACCATCATGAT*
*TGTTTATTTTCTCTTCTGGGCTCCCTACAACATTGTCCTTCTCCTGAACACCTTCCAGGA*
*ATTCTTTGGCCTGAATAATTGCAGTAGCTCTAACAGGTTGGACCAAGCTATGCAGGTGA*
*CAGAGACTCTTGGGATGACGCACTGCTGCATCAACCCCATCATCTATGCCTTTGTCGG*
*GGAGAAGTTCAGAAACTACCTCTTAGTCTTCTTCCAAAAGCACATTGCCAAACGCTTCT*
*GCAAATGCTGTTCTATTTTCCAGCAAGAGGCTCCCGAGCGAGCAAGCTCAGTTTACAC*
*CCGATCCACTGGGGAGCAGGAAATATCTGTGGGCTTGTGA")*

The region spanning the sgRNA 7 and 8 target sequences was given as the *DNAstringset* as these had been selected prior to analyses. The input file path was named “CCR5”. The *findgRNAsWithREcutOnly* parameter was set as “FALSE” because it was not necessary for a restriction enzyme site to be located at the cleavage site in this experiment. The *findPairedgRNAOnly* parameter was set as “FALSE” because when this was set as “TRUE” no gRNAs were identified. The *PAM.size* and *PAM* parameters were set as “6” nucleotides in the form “NNGRRT”, specific to *Sa*Cas9 targeting but the *PAM.pattern* was also set as “NNG[A|G][A|G]N$” (where N refers to any nucleotide, | denotes ‘or’ in pattern specification and $ indicates the end of the string) because this allowed flexibility within the

last base pair identified at off target sites. The *gRNA.size* parameter was set as "21" nucleotides and all chromosomes were searched for potential off target sites by setting *chromToSearch* as "all". The maximum mismatches parameter (*max.mismatch*) was set at 6 and the minimum score for cleavage (ranging from 0 as least similar to on target cleavage and 100 as most similar) set at 0. The *Genomicranges* function is employed by the programme to determine exonic sequences within the human genome thus the *annotateExon* parameter was set as "TRUE". Flanking sequences at off target sites predicted were retrieved 300 bp up and down stream. The weight scores were set based on experimentally determined penalty matrices for *Sa*Cas9 ranging from 0-1, 0 being most stringent and 1 being most lenient in terms of mismatch likelihood for each base pair in the gRNA from the PAM proximal to distal ends.

The top 1 000 off target sequence predictions for each sgRNA were narrowed down to those present within exons. Introns are AT-rich and have low complexity sequences thus it was found that primer $T_m$s in these regions were too low and primers did not bind specifically to target sites within the gene of interest. Furthermore, an off target effect detected within an exonic region is also more likely to be deleterious to gene function than intronic indel formation.

### 2.7.3. PCR and pooling for deep sequencing

Indexes of 6 bp in length were included at the 5′ end of each primer in order to distinguish between the four different CRISPR/*Sa*Cas9 treatments used (1 – *Sa*Cas9 + empty U6-sgRNA; 2 – *Sa*Cas9 + U6-sgRNA 7; 3 – *Sa*Cas9 + U6-sgRNA 8; 4 – *Sa*Cas9-D10A + U6-sgRNA 7/8), allowing for multiplexed sequencing within a single pool of PCR products (Table 2.13).

**Table 2.13: Indexes added to the 5′ end of each primer to distinguish between CRISPR/*Sa*Cas9 treatments used**

| Treatment* | Index Sequence (5′ to 3′) |
|---|---|
| **1** | CTAGAG |
| **2** | TTCGAC |
| **3** | GGTAAC |
| **4** | ATCTGC |

*Treatments: 1 – *Sa*Cas9 + empty-U6 sgRNA; 2 – *Sa*Cas9 + U6-sgRNA 7; 3 – *Sa*Cas9 + U6-sgRNA 8; 4 – *Sa*Cas9-D10A + U6-sgRNA 7/8

The five genes selected for sgRNA 7 off target analysis were *LCP1*, *CNTN3*, *SLITRK2*, *BMPR1B* and *FLNB* and those selected for sgRNA 8 were *CCR2*, *MCM3AP*, *SRL*, *TOMM5* and *POLL*. Primers were designed to amplify ~300 bp spanning each on and off target site identified (the product length including index sequences added to either end is shown in Table 2.14; Amplicon sequences are listed in Appendix 7.2.8.), the expected cut site from the 5′ end of the forward strand (excluding indexes) and the MW as well as the number of ng/µl of each amplicon required in order to prepare equimolar amounts for multiplexed deep sequencing.

**Table 2.14: Primers used for on and off target site amplification for deep sequencing**

| Treatment* | Primer Name | Primer Sequence (5′ – 3′) | $T_m$ (°C) | Amplicon Size (bp) | Amplicon MW (g/mol) | ng/µl |
|---|---|---|---|---|---|---|
| **1, 2, 3, 4** | *CCR5* F2 | CTGCCGCTGCTTGTCATGG | 60 | 313 | 193273.4 | 0.309 |
| | *CCR5* R2 | TGAACTTCTCCCCGACAAAGGC | 60 | | | |
| **1, 2, 4** | *LCP1* F | CCAGCCACTGCAGGAGTGAG | 61 | 312 | 192050.3 | 0.307 |
| | *LCP1* R | AGTCAGGAGTGAGTGCACCG | 59 | | | |
| **1, 2, 4** | *CNTN3* F | CACCCATATGTATGAATCATGTGCC | 57 | 316 | 195110 | 0.312 |
| | *CNTN3* R | GCACTACCACCACCACCTGC | 61 | | | |
| **1, 2, 4** | *SLITRK2* F | CAACAGTTAGCCTGCTCCAGC | 59 | 312 | 192657.9 | 0.308 |
| | *SLITRK2* R | CAGCCTCGATGGCACTGATG | 59 | | | |
| **1, 2, 4** | *BMPR1B* F | GCCTTAGAAAACCCAGACACATAGC | 59 | 341 | 210530.5 | 0.337 |
| | *BMPR1B* R | TTTCCAAGGGGAGAATGAAGCC | 58 | | | |
| **1, 2, 4** | *FLNB* F | CTCAAGTACCTGTCCCCTCGC | 60 | 313 | 193277.3 | 0.309 |
| | *FLNB* R | CATGGGACGGACCTTGGGATG | 60 | | | |
| **1, 3, 4** | *CCR2* F | GGCAGTGAGAGTCATCTTCACC | 58 | 313 | 193271.4 | 0.309 |
| | *CCR2* R | CACTGTCTCCCTGTAGAAAACTGG | 58 | | | |
| **1, 3, 4** | *MCM3AP* F | CTGTCCTGGAACTCTCATCTGTGG | 59 | 311 | 192045.4 | 0.307 |
| | *MCM3AP* R | CTGTGCCAGGCACTGTCC | 59 | | | |
| **1, 3, 4** | *SRL* F | GCCACCAAGGCTTAACATTGACC | 60 | 311 | 192049.3 | 0.307 |
| | *SRL* R | TCTCCTGGGCATGCACAGAC | 60 | | | |
| **1, 3, 4** | *TOMM5* F | ACAGGACATCACATATGAATGCACG | 59 | 323 | 199427.9 | 0.319 |
| | *TOMM5* R | CAGACACAGCTCCCTTAATACTTGC | 58 | | | |
| **1, 3, 4** | *POLL* F | GCTGACTCGGAAGCTATTCTGGC | 61 | 311 | 192085.6 | 0.307 |
| | *POLL* R | TCCGGGAATGGAGGAGTCTCG | 61 | | | |

*Treatments: 1 – *Sa*Cas9 + empty U6-sgRNA; 2 – *Sa*Cas9 + U6-sgRNA 7; 3 – *Sa*Cas9 + U6-sgRNA 8; 4 – *Sa*Cas9-D10A + U6-sgRNA 7/8

Full amplicon sequences for each gene can be found in Appendix 3.9, including the putative target site and PAM recognition motif. It is essential for the deep sequencing platform that all

PCR products are pooled and added to the flow cell in equimolar amounts. This was ensured by calculating the molecular weight (MW) of each dsDNA amplicon using the OligoCalc online tool (version 3.27) (Kibbe, 2007). The concentration of each PCR product required was determined by finding the molarity (nM) of each product that would result a final pool concentration of at least 10 ng/µl (see calculations below):

Total number of amplicons to be pooled:

11 genes (Treatment 1) + 6 genes (Treatment 2) + 6 genes (Treatment 3) + 11 genes (Treatment 4) = 34 amplicons in total to be pooled.

10 ng/µl total concentration / 34 amplicons = 0.294118 ng/µl per amplicon is required, assuming equal MW of each. Each amplicon does not have equal MW thus the molarity (nM) must be calculated so that the overall concentration of the pool is ≥ 10 ng/µl. To determine what molarity to use across all samples, the molarity of each when a concentration of 10 ng/µl is achieved was calculated:

$y$ nM = $\frac{x \text{ x } 1\text{x}10^6}{\text{MW}}$; where $x$ is the concentration of the purified PCR product in ng/µl and the MW is the molecular weight of the amplicon in g/mol (determined based on length and base composition using the OligoCalc online programme; Kibbe, 2007).

The range of molarities calculated was 1.397 nM – 1.532 nM (lowest being the BMPR1B amplicon and highest the MCM3AP amplicon) thus an overall molarity of 1.6 nM was selected for each amplicon and the concentration in ng/µl required calculated by rearranging the above formula to:

$x$ ng/µl = $\frac{y \text{ (nM) x MW}}{1\text{x}10^6}$ (as seen in Table 2.14).

Each target site was amplified by PCR using the primers listed in Table 2.14 (with indexes included). Reactions were carried out in a total volume of 25 µl with 100 ng of gDNA, 0.5 µM of each primer, 1X Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs, USA) and Ambion® nuclease-free water (Thermo Fisher Scientific, USA) with cycling conditions described in Table 2.15.

**Table 2.15: Cycling conditions used for Q5 Hot Start Polymerase Chain Reactions**

| Step | Temperature (°C) | Hold (min:sec) | Cycles |
|---|---|---|---|
| **Initial denaturation** | 98 | 00:30 | 1 |
| **Denaturation** | 98 | 00:10 | |
| **Annealing** | 68* | 00:15 | 35 |
| **Extension** | 72 | 00:15 | |
| **Final extension** | 72 | 02:00 | 1 |

*Note: The annealing temperature was 65 °C for the BMPR1B target site PCR reactions

Approximately 1 – 2 µl of each PCR product was resolved on a 2 % agarose gel (Appendix 7.4.1.) to verify that the band in each sample was of the correct size and there were no non-specific products. Each amplicon was purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, USA; Appendix 7.4.2.) and the concentration determined by Qubit Fluorometric Quantification using the Qubit® dsDNA High Sensitivity (HS) Kit (Thermo Fisher Scientific, USA; Appendix 7.4.10.). This method allows for highly sensitive quantification using a dsDNA-specific fluorescent dye. This experiment was performed in triplicate thus three separate pools were prepared for deep sequencing. The volume of each amplicon required to achieve the desired concentration within the pool (Table 2.14) was determined and Ambion® nuclease-free water (Thermo Fisher Scientific, USA) added to a final volume of 170 µl in a 0.6 ml microcentrifuge tube. The calculation was performed according to the following formulae:

$$\text{Volume of purified amplicon} = \frac{x \text{ ng/µl x 170 µl}}{\text{concentration of purified amplicon (ng/µl)}}$$

$$\text{Total volume of water} = 170 \text{ µl} - \sum(\text{volume of each amplicon})$$

### 2.7.4. Illumina deep sequencing and indel analysis

Each replicate was mixed thoroughly and 20 µl sent to Inqaba Biotech (South Africa) for deep sequencing using the Illumina TruSeq® paired end chemistry system (New England Biolabs, USA) which generated 300 bp paired-end reads at a depth of ~ 20 000 X. Amplicons were designed to be ~ 310 bp in length to allow for sequencing to cover the majority of the amplicon in each direction. The TruSeq indexed adapters used for each of the three deep sequencing pools were D710 & D508, D711 & D508 and D712 & D508 (Table 2.16) and each sample within the pool had internal indexing to distinguish between the four different treatments (Table 2.13).

**Table 2.16: TruSeq adapter index sequences used for targeted amplicon deep sequencing of three pools**

| TruSeq adapter name | TruSeq adapter sequence (5′ to 3′) |
|---|---|
| **D710** | TCCGCGAA |
| **D711** | TCTCGCGC |
| **D712** | AGCGATAG |
| **D508** | GTACTGAC |

De-multiplexing was performed by Inqaba Biotech (South Africa) to remove the Nextera Adapter sequences and group data into six files containing .fastq formats of the forward and reverse sequencing reads (one file for each replicate and one for each read direction). Downstream analysis was performed with the help and expertise of a Postdoctoral researcher at the Sydney Brenner Institute for Molecular Bioscience, Dr Stanford Kwenda. Further de-multiplexing was performed to separate reads from each replicate based on the CRISPR/*Sa*Cas9 treatment using a programme called FASTX Barcode Splitter. Exact mapping of barcodes was performed as they are short 6 bp sequences. At this point, 24 files in .fastq format were obtained (the six files from before were further separated into four different treatments). Unmatched sequences were excluded from further analysis.
Quality control of reads within each file was determined using the FastQC programme (Babraham Bioinformatics). A general perception of the success of the sequencing run was performed based on the total number of reads obtained, overall read quality distribution, average length of reads, per base sequence quality and extent of sequence over-representation. This information was used to select sequence reads of at least 100 bp in length and with a quality score exceeding 20 when trimming index sequences using the Trim Galore! programme (Babraham Bioinformatics). Reads within the trimmed and filtered .fastq files were then mapped to the human chromosomes corresponding to the locus of the off target sites identified using the Bowtie 2 tool (John Hopkins University). The output .sam files were compressed to .bam files for variant calling. In order to determine whether the CRISPR/*Sa*Cas9 treatments induced indels in the *CCR5* on target and various identified off target sites compared to the mock treatment control, the Low Variant Detection Tool in the CLC Genomics Workbench was used for variant calling at a threshold of 0.1 % and a significance of 1 %. All other parameters were left as default except the base quality filter

settings under the noise filters tab were set as follows: neighbourhood radius at 5; minimum central quality at 20 and minimum neighbourhood quality at 10.

### 2.8. Quantification of relative *CCR5* mRNA levels by qRT-PCR

In order to determine whether indels generated in the *CCR5* target site resulted in a reduction of *CCR5* mRNA, qRT-PCR was carried out. Briefly, TZM-bl cells were treated with the CRISPR/*Sa*Cas9 systems and the total cellular RNA extracted for relative quantification of *CCR5* mRNA levels amongst treatments. Contaminating gDNA was removed during the two-step QuantiTect Reverse Transcription protocol (Qiagen, Germany) and the RNA reverse transcribed using Universal and gene-specific primers to *CCR5* and a reference gene, *β-actin*. Levels of *CCR5* mRNA were normalised to those of *β-actin* for each sample and adjusted relative to the negative control treatment in order to determine whether the CRISPR/*Sa*Cas9-induced indels had any effect on *CCR5* mRNA. A positive control was included by treatment with the *CCR5*-specific shRNA 1005. Two negative controls, the containing no reverse transcriptase and the other containing no gDNA were included in the final run for each primer set to verify that there was no gDNA or RNA contamination respectively.

#### 2.8.1. Transfection and RNA extraction

TZM-bl cells were seeded and transfected in 24-well plates as described in Section 2.3.2. Forty-eight hours post transfection, the media was removed, cells washed with 0.5 ml of ice cold 1 X PBS (Appendix 7.3.1.), and RNA extracted by the TRIzol® Reagent method (Thermo Fisher Scientific, USA; Appendix 7.4.11.).

#### 2.8.2. gDNA wipeout and reverse transcription

qRT-PCR for the quantification of *CCR5* mRNA was carried out by a two-step protocol, using the QuantiTect Reverse Transcription (RT) Kit (Qiagen, Germany) for cDNA generation and LightCycler® 480 SYBR Green I Master Mix (Roche, Switzerland) for real-time PCR. The QuantiTect RT Kit was used as it includes a preliminary gDNA 'wipeout' step thus reducing the potential for gDNA contamination. Another useful feature of the QuantiTect RT Kit is that it includes Universal primers for reverse transcription thus simulating nested PCR for target and reference gene amplification, enhancing the specificity of the reaction.

RNA samples were thawed and all reactions set up on ice to prevent degradation. gDNA wipeout was carried out by the addition of 2 µl of the 7 X gDNA Wipeout Buffer (Qiagen, Germany), 100 ng of RNA and Ambion® nuclease-free water (Thermo Fisher Scientific, USA) to a final volume of 14 µl in a 0.2 ml PCR tube. Reactions were mixed and incubated at 42 °C for precisely ten minutes in a thermal cycler and immediately placed on ice. A reverse transcription master mix was prepared in a 1.5 ml microcentrifuge tube by adding 1 µl of Quantiscript Reverse Transcriptase (containing an RNase inhibitor), 4 µl of the 5 X Quantiscript RT Buffer (containing $Mg^{2+}$ and dNTPs) and 1 µl of the RT Primer Mix (universal primers) per sample. The 14 µl gDNA elimination reaction was then added, the sample mixed and incubated at 42 °C for fifteen minutes for cDNA generation. The RT enzyme was inactivated by incubation at 95 °C for three minutes and samples were placed on ice. Real-time PCR was carried out immediately thereafter or samples were stored at -20 °C.

### 2.8.3. Real-time qRT-PCR

Real-time qRT-PCR for the relative quantification of *CCR5* mRNA following CRISPR/*Sa*Cas9 treatment was carried out in a LightCycler® 96 machine using the LightCycler® 480 SYBR Green I Master Mix (Roche, Switzerland). *CCR5*-specific primers were designed to amplify a 189 bp region within exon 3 (Lai et al., 2003) and *β-actin*-specific primers to amplify a 192 bp region spanning exons 3 and 4 (Batra et al., 2016) (Table 2.17).

**Table 2.17: Primers used for qRT-PCR of *CCR5* and *β-actin* mRNA**

| Primer name | Primer sequence (5′ – 3′) | Primer length (bp) | Primer $T_m$ (°C) |
|---|---|---|---|
| ***CCR5* qPCR F** | CAAAAAGAAGGTCTTCATTACACC | 24 | 53 |
| ***CCR5* qPCR R** | CCTGTGCCTCTTCTTCTCATTTCG | 24 | 59 |
| ***β-actin* qPCR F** | ACCAACTGGGACGACATGGAGAAA | 24 | 61 |
| ***β-actin* qPCR R** | TAGCACAGCCTGGATAGCAACGTA | 24 | 61 |

Real-time qRT-PCR reactions were set up to contain 2 µl of cDNA generated in the previous step, 0.5 µM each of the 24 bp F and R primers, 1 X LightCycler® 480 SYBR Green I Master Mix (Roche, Switzerland) and Ambion® nuclease-free water (Thermo Fisher Scientific, USA) to a final volume of 20 µl. Samples were transferred to wells within LightCycler® 8-tube white strips (Roche, Switzerland). With clear caps tightly in place, strips were centrifuged for 30 seconds to ensure removal of any bubbles. Samples were placed into the LightCycler® 96

real-time PCR machine and the reaction carried out with cycling conditions described in Table 2.18.

**Table 2.18: Cycling conditions for the relative quantification of *CCR5* and *β-actin* mRNA by Real-time qRT-PCR**

| Step | Temperature (°C) | Ramp rate (°C/sec) | Hold (min:sec) | Cycles |
|---|---|---|---|---|
| **Reverse Transcription** | 42 | 4.4 | 10:00 | 1 |
| | 95 | 2.2 | 03:00 | |
| **3 Step Amplification** | 95 | 4.4 | 00:10 | 35 |
| | 55 | 2.2 | 00:20 | |
| | 72* | 4.4 | 00:01 | |
| **High Resolution Melting** | 95 | 4.4 | 00:05 | 1 |
| | 65 | 2.2 | 01:00 | |
| | 97* | - | 00:0 1 | |
| **Cooling** | 40 | 2.2 | 00:10 | 1 |

*Acquisition of fluorescence reading (single at 72 °C and continuous at 97 °C)

Although RT had already been carried out on samples in the first step of the protocol, this was performed again as RT was present in the LightCycler® 480 SYBR Green I Master Mix and required deactivation prior to Real-time qRT-PCR. Fluorescence of the SYBR Green I was acquired during the amplification extension step at 72 °C. SYBR Green I is a cyanine dye that preferentially intercalates into dsDNA. The amount of cDNA in the Real-time qRT-PCR reaction was deduced by illumination as SYBR Green I absorbs light at a wavelength of 497 nm and emits at 520 nm. As the cDNA was amplified during the reaction, SYBR Green I was able to intercalate and thus was measured at the end of each extension step in real time, providing the cycle threshold value (Ct) of detection for each sample.

### 2.8.4. Analysis of qRT-PCR results

Analysis of the Real-time qRT-PCR results was carried out using the LightCycler® 96 Software. A reference gene was included as it is assumed that the mRNA levels of this should be unaltered across treatments but that external factors, such as the number of cells from which the mRNA was harvested and the total amount of starting material, could influence these levels. This gene is thus used as a reference for normalisation such that the differences in mRNA levels of the gene of interest can be attributed solely to the effects of the treatment. In this experiment, *β-actin* was used as a reference gene as it is important for cellular survival and thus is expected to be stably expressed across all cells. Cq values were determined by

qRT-PCR for each gene and each of the CRISPR/*Sa*Cas9 treatment samples and the amount of starting *CCR5* mRNA normalised to that of *β-actin* using the LightCycler® 96 software. Levels of *CCR5* mRNA for each treatment were adjusted relative to the negative control (transfection with the *Sa*Cas9 and an empty U6-sgRNA vector) which was set at 100 %.

### 2.9. Assessment of R5-tropic HIV-1 infectivity by the TZM-bl luciferase assay

In order to determine whether indels generated in the *CCR5* target site resulted in a reduction in R5-tropic HIV-1 pseudovirus infectivity, a TZM-bl luciferase assay was performed. TZM-bl cells harbour an integrated LTR-driven firefly luciferase gene activated by the HIV-1 Tat protein upon infection. The level of luciferase expression therefore correlates with the level of infection and, by deduction, the effects of CRISPR/*Sa*Cas9 treatment on *CCR5* protein expression and function (Sarzotti-Kelsoe et al., 2014). Briefly, TZM-bl cells were treated with the CRISPR/*Sa*Cas9 systems and infected with the R5-tropic ZM53 pseudovirus 48 hours later. The level of luciferase expression was measured after a further 48 hours to determine the propensity for ZM53 HIV-1 pseudovirus to infect the treated cells. A positive control was included by treatment with the *CCR5*-specific shRNA 1005 and a negative control with the *Sa*Cas9 and an empty U6-sgRNA vector.

#### 2.9.1. Transfection and pseudovirus infection

TZM-bl cells were seeded and transfected in 96-well plates, as described in Section 2.3.2., in biological triplicates. The media was removed from the cells and replaced with 200 µl of fully supplemented DMEM four hours prior to infection. Forty-eight hours post transfection, the cells were infected with R5-tropic ZM53 HIV-1 pseudovirus with a 50 % Tissue Culture Infectious Dose ($TCID_{50}$) of 4000 particles/ml, available for use in our laboratory and produced as follows:

The pSG3Δenv (the *env*-deficient backbone containing a defective *vpu* gene; Dr John C. Kappes & Xiaoyun Wu) and ZM53M.PB12 (Drs E. Hunter and C. Derdeyn) plasmids were co-transfected into HEK293T cells to produce ZM53 R5-tropic HIV-1 pseudovirus capable of a single round of infection. Briefly, 5 µg of plasmid DNA in a ratio of 1:2 env to backbone was incubated at room temperature with FuGENE® HD Transfection Reagent (Promega Corporation, USA) at a ratio of 1:6 of plasmid DNA to FuGENE for ten minutes in a final volume of 500 µl of Opti-MEM™ I Reduced Serum Media. The DNA:FuGENE complexes were then added drop-wise to HEK293T cells in a T25 tissue culture flask (Nunc, Denmark)

at ~ 60 – 65 % confluency and cells incubated at 37 °C. Cell media was replaced twelve hours post-transfection with 7 ml of DMEM. The supernatant containing pseudovirus was collected at 48 and 72 hours post-transfection and a final concentration of 20 % FBS added. Pseudovirus-containing supernatant was filtered using a 0.45 µm syringe filter and aliquots of 1 ml stored at -80 °C. A single aliquot was used to determine the $TCID_{50}$ in TZM-bl cells (Li et al., 2005). Five-fold serial dilutions of the pseudovirus were made in a volume of 100 µl DMEM in a 96-well plate, across eleven columns in quadruplicate. The final column contained cell-only controls. TZM-bl cells were trypsinised, counted and 10 000 cells in 100 µl of DMEM containing FBS, supplemented to a final concentration of 20 µg/ml of diethylaminoethyl-dextran (DEAE Dextran; Sigma-Aldrich, USA), added to each well. Luciferase readings were taken 48 hours later and the $TCID_{50}$ value calculated.

### 2.9.2. Luminescence acquisition and analysis of TZM-bl assay results

Infection of CRISPR/*Sa*Cas9-treated TZM-bl cells by the ZM53 pseudovirus was carried out for 48 hours before biological triplicate luminescence readings were obtained and averaged. A negative control for luciferase expression by not infecting cells and a positive control by infecting cells not treated with the CRISPR/*Sa*Cas9 systems were included in the assay. Infection of TZM-bl cells by the pseudovirus was facilitated by the addition of 20 µg/ml of DEAE Dextran as this reduces repulsive electrostatic forces between the virus and the cell surface. All media was removed from the plate and 80 µl of Bright-Glo® lysis buffer (Promega, USA) added to each well. Cells were lysed for five minutes at room temperature before 50 µl was transferred to a 96-well Luminometer plate (Promega, USA). In order to induce luminescence, 50 µl of the Bright-Glo® reagent (Promega, USA) was added and luciferase expression levels determined immediately using the GloMax® 20/20 Luminometer (Promega, USA). The relative level of infection was determined by comparison with *Sa*Cas9 treatment and an empty U6-sgRNA vector.

## 2.10. Statistical analyses

All bar graphs were generated using GraphPad Prism v5.03 and two-tailed unpaired t-tests with a 95% confidence interval carried out as a measure of the difference in sample means with equal variances assumed. The F-test was used to determine whether there was in fact equal variance between the two samples and if found to be significantly different, the Welch's correction was applied to the t-test.

## 3. Results

The CRISPR/Cas9 system derived from *Staphylococcus aureus* (*Sa*Cas9) was employed for targeted disruption of the human *CCR5* gene. The overarching aim was to determine whether gene editing could be achieved using four highly specific CRISPR/*Sa*Cas9 systems for the consequent reduction in R5-tropic HIV-1 infectivity in a TZM-bl cell model. Furthermore, the specificity of the *Sa*Cas9 nuclease and nickase systems was assessed at selected computationally predicted off target sites to determine the safety of the AAV-deliverable gene editing systems. Ultimately, this study contributes towards understanding the functioning of novel and highly specific AAV-deliverable CRISPR/*Sa*Cas9 systems as well as their potential applicability in the development of a gene therapy for the functional cure of R5-tropic HIV-1 by the preclusion of viral infection.

### 3.1. Assessment of CRISPR/*Sa*Cas9 nuclease and nickase activity with a panel of *CCR5*-specific sgRNAs

#### 3.1.1. Generation of the *Sa*Cas9-D10A nickase construct

The *Sa*Cas9 nuclease used in this study was expressed from the pAAV-CMV_*Sa*Cas9 packaging vector (Appendix 7.1.1.) which has been optimised for effective CRISPR/Cas9 gene editing in human cells (Ran et al., 2015). This plasmid includes a number of convenient features, some of which were relevant to this study, including 1) a cytomegalovirus (CMV) promoter and its cognate enhancer for constitutive and robust expression of the *Sa*Cas9 gene in a broad range of cell types; 2) 5′ Simian virus 40 (SV40) and 3′ nucleoplasmin nuclear localisation signals (NLSs) co-expressed with *Sa*Cas9 for efficient delivery of the *Sa*Cas9 protein to the nucleus where it is required for gene editing activity; 3) an ampicillin resistance (AmpR) gene for selective cloning and 4) a pBR322 origin of replication (pMB1 ori) to ensure plasmid survival and replication within transformed or transfected cells. Additional features of the pAAV-CMV_*Sa*Cas9 packaging vector include 1) three haemagglutinin (HA) tags at the 5′ end of the *Sa*Cas9 gene for detection of the expressed protein; 2) a sgRNA cloning site at the 3′ end of the *Sa*Cas9 gene for co-delivery of all components required for CRISPR/Cas9 functionality into target cells and 3) adeno-associated virus 2 (AAV-2) inverted terminal repeat (ITR) sequences (L-left and R-right) flanking the *Sa*Cas9-sgRNA cassette to allow for AAV-mediated CRISPR/*Sa*Cas9 system delivery.

The *Sa*Cas9-D10A nickase vector (pAAV-CMV_*Sa*Cas9 -D10A) was generated by the introduction of a codon mutation, GAC to GCC, within pAAV-CMV_*Sa*Cas9 (Appendix 7.1.1.) converting the aspartic acid to an alanine at position ten of the protein sequence. This was achieved by SDM-PCR and Gibson Assembly cloning. Sanger sequencing was carried out to confirm successful construct generation (Inqaba Biotech, South Africa; Figure 3.1).

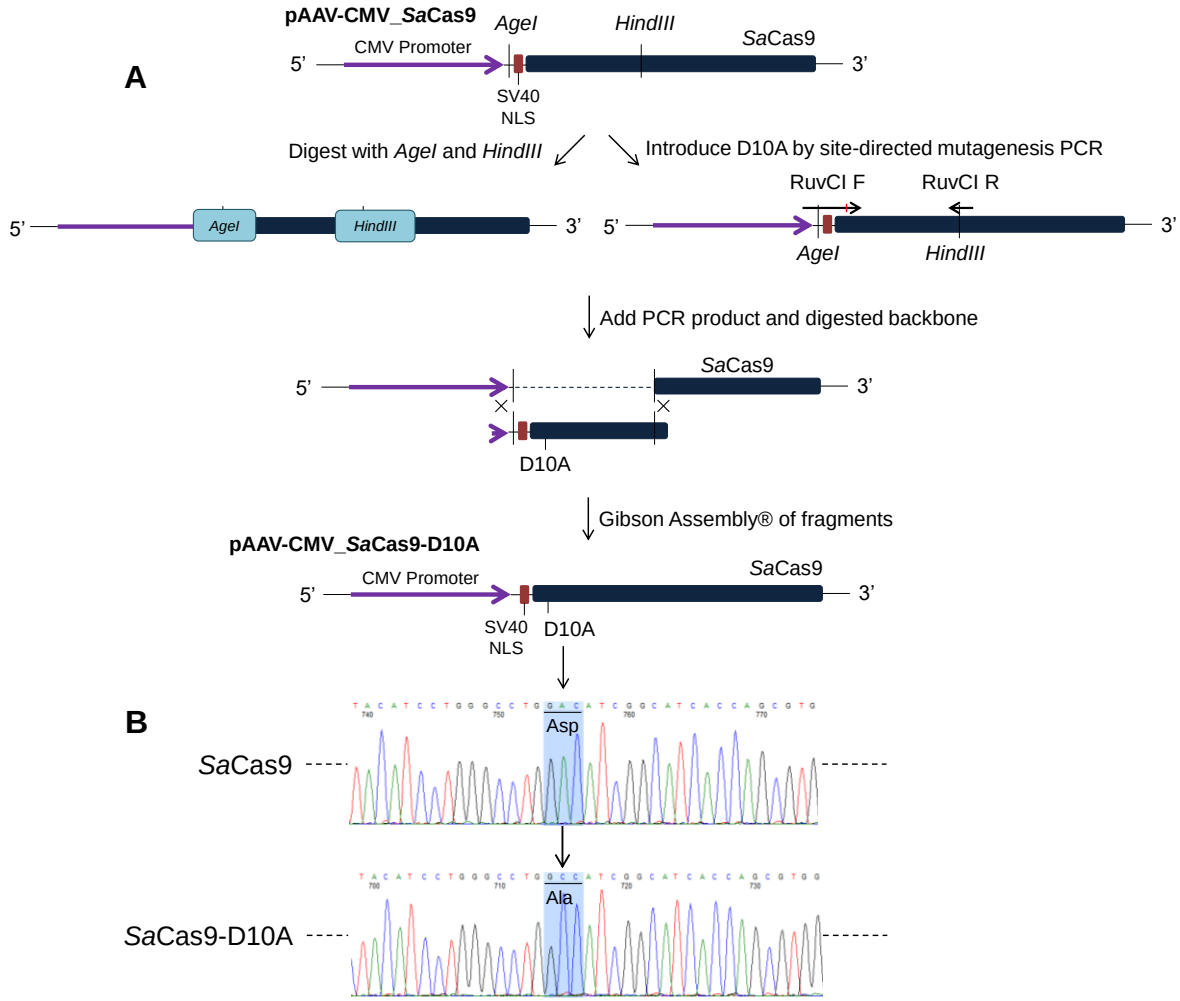


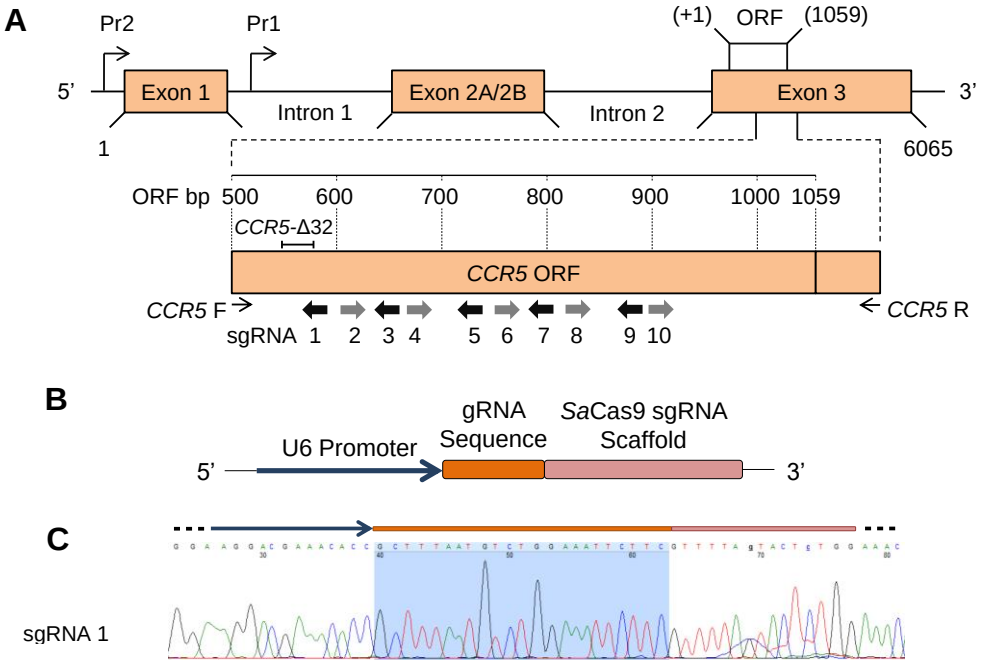

**Figure 3.1: Schematic of the cloning strategy employed for generation of the *Sa*Cas9-D10A nickase vector and Sanger sequencing confirmation. A** – A schematic of the cloning strategy employed for the generation of the pAAV-CMV_*Sa*Cas9-D10A vector by SDM-PCR and Gibson Assembly cloning. The aspartic acid to alanine (D10A) codon mutation, GAC to GCC, was introduced by the forward primer during a PCR reaction and the region spanning *AgeI* and *HindIII* restriction enzyme sites was amplified for incorporation of the modified PCR product into the digested pAAV-CMV_*Sa*Cas9 vector backbone. **B** – Sanger sequencing chromatograms of the *Sa*Cas9 nuclease and *Sa*Cas9-D10A nickase constructs showing the D10A aspartic acid to alanine codon mutation and integrity of the immediate flanking sequences.

Sanger sequencing both confirmed the successful modification of codon ten within *Sa*Cas9 and verified the integrity of the immediate flanking regions following PCR amplification

(Fig. 3.1B). The *Sa*Cas9-D10A was then assessed alongside the nuclease for cleavage activity within the *CCR5* gene. Many studies have demonstrated that cleavage activity across CRISPR/Cas9 target sites is variable (Cong et al., 2013, Mali et al., 2013b). In order to select the most efficient *CCR5*-targeting systems, ten *Sa*Cas9 sgRNAs within the ORF of the gene were designed and indel forming capacity measured for each individually with the *Sa*Cas9 nuclease or in pairs with the *Sa*Cas9-D10A nickase. This would allow for selection and application of the most efficient CRISPR/*Sa*Cas9 nuclease and nickase systems in downstream functional analyses with the greatest likelihood of realising the desired outcome.

### 3.1.2. Production of a panel of human *CCR5*-specific *Sa*Cas9 sgRNAs

In order to target the *Sa*Cas9 nuclease and *Sa*Cas9-D10A nickase (referred to collectively hereon as *Sa*Cas9/-D10A) to the human *CCR5* gene for disruption, a panel of ten guide RNAs (gRNAs – the variable region within the sgRNA) were designed, cloned into the U6-sgRNA expression vector (Appendix 7.1.2.) and screened for cleavage activity. Potential *Sa*Cas9 5′–NNGRR(T/N) PAM sequences were identified and gRNAs designed to target the opposite DNA strand upstream of this site (Ran et al., 2015). One of the design considerations was to allow for a 'PAM-out' (tail-to-tail) orientation for dual sgRNA nickase targeting (Friedland et al., 2015). Target site selection was confined to the last 500 bp of the *CCR5* open reading frame (ORF, Fig. 3.2A), as this region harbours the *CCR5*-Δ32 locus known to result in reduced infectivity of R5-tropic HIV-1.



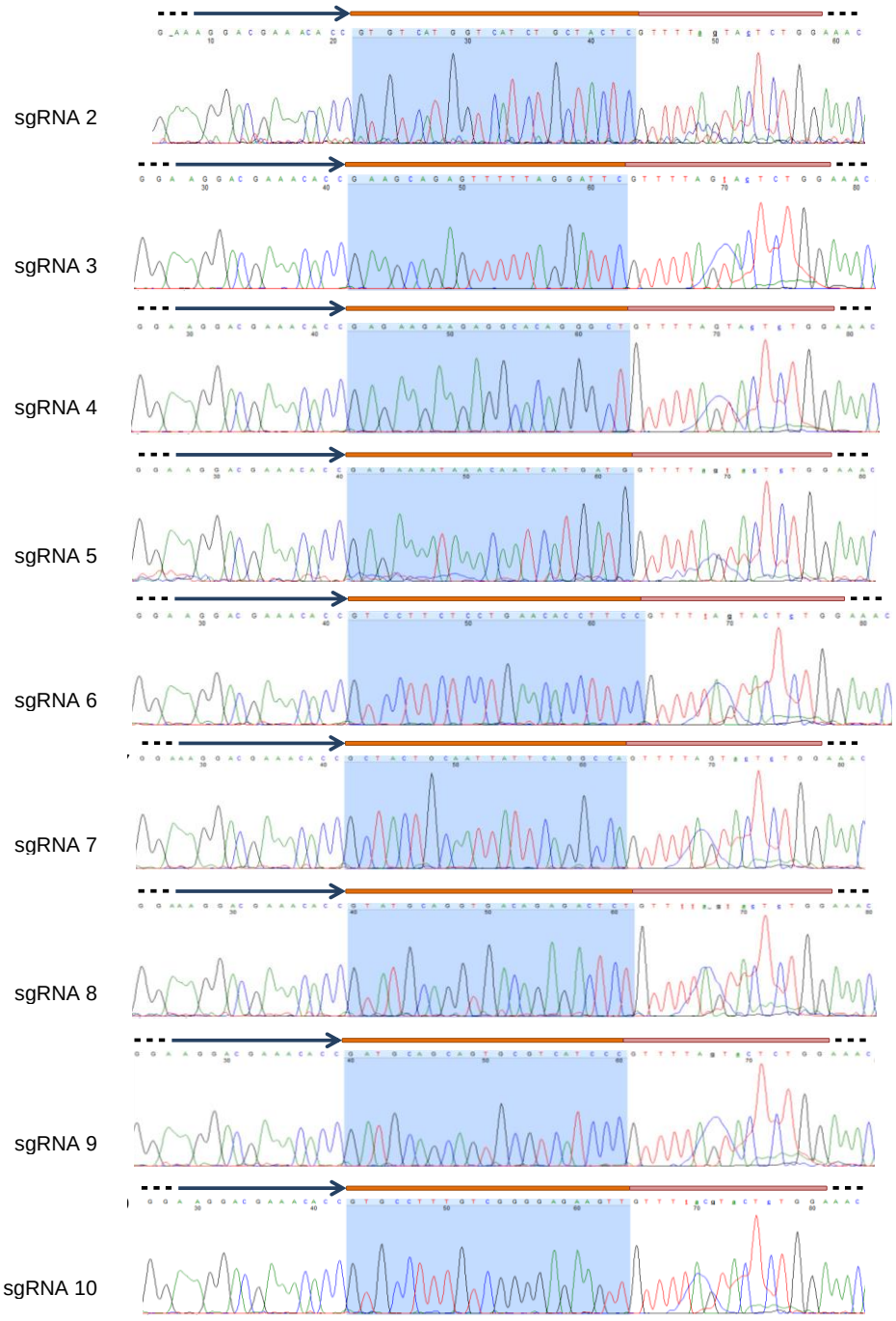



**Figure 3.2: A schematic showing ten *Sa*Cas9 sgRNA target sites within the human *CCR5* gene and Sanger sequencing chromatograms of cloned U6-sgRNAs. A** – A schematic of the *CCR5* gene showing exons 1, 2A, 2B and 3, introns 1 and 2 and the 1 059 bp ORF. The sgRNA 1 – 10 target sites and *CCR5*-Δ32 mutation locus are indicated within the last 500 bp of the *CCR5* ORF. Primers *CCR5* F and R were designed and utilised for target region PCR amplification and indel quantification. **B** – A schematic of the U6-sgRNA expression vector to illustrate the position of the U6 Pol III promoter, gRNA cloning region and *Sa*Cas9 sgRNA scaffold corresponding to the chromatogram sequences below. **C** - Sanger sequencing chromatograms of U6-sgRNAs 1 – 10 (gRNA sequences listed in Table 2.4.) showing seamless insertion of the gRNAs into the expression vector.

dsODNs encoding the variable gRNA sequences (listed in Table 2.4.) were annealed and cloned into the *BbsI* restriction sites of the U6-sgRNA expression vector (Appendix 7.1.2.) and successful cloning was confirmed by Sanger sequencing (Inqaba Biotech, South Africa; Fig. 3.2B). Either a mismatched or, if possible, a matched 5′ Guanine (G) was included in the design of the gRNA ODNs to enhance expression from the U6 Pol III promoter (Ran et al., 2015). In addition to this, the annealed dsODNs included 5′ end extensions to facilitate complementary binding and ligation to *BbsI* digested sites. This allowed for seamless insertion of the gRNA sequence between the promoter and tracrRNA as *BbsI* is a type IIS restriction enzyme with the cleavage site located externally of the recognition sequence. Sequencing chromatograms illustrate that each gRNA was successfully inserted within the *BbsI* cloning site of the expression vector (Figure 3.2C). The ten sgRNAs were then tested for activity individually with the *Sa*Cas9 or as 'PAM-out' pairs with *Sa*Cas9-D10A to assess cleavage activity and select the most efficient candidates for functional experimentation.

### 3.1.3. Assessment of *CCR5*-specific U6-sgRNA cleavage activity with the *Sa*Cas9 nuclease and nickase

HEK293T cells were selected for quick and effective screening of individual and dual-targeting sgRNA cleavage activity with *Sa*Cas9 and *Sa*Cas9-D10A respectively as these cells have a high propensity for transfection and robust growth characteristics. A mock negative control (absence of a gRNA target sequence within the U6-sgRNA vector, otherwise referred to as the empty U6-sgRNA) was included in the experiment as a measurement of baseline indels present within the target site. The propensity for each sgRNA and *Sa*Cas9 as well as dual sgRNAs and *Sa*Cas9-D10A to introduce indels into the target site was assessed by a T7EI assay (Figure 3.3A) and *Sa*Cas9-mediated indels were also quantified by TIDE, a novel online Sanger sequence deconvolution tool (Fig. 3.3B).

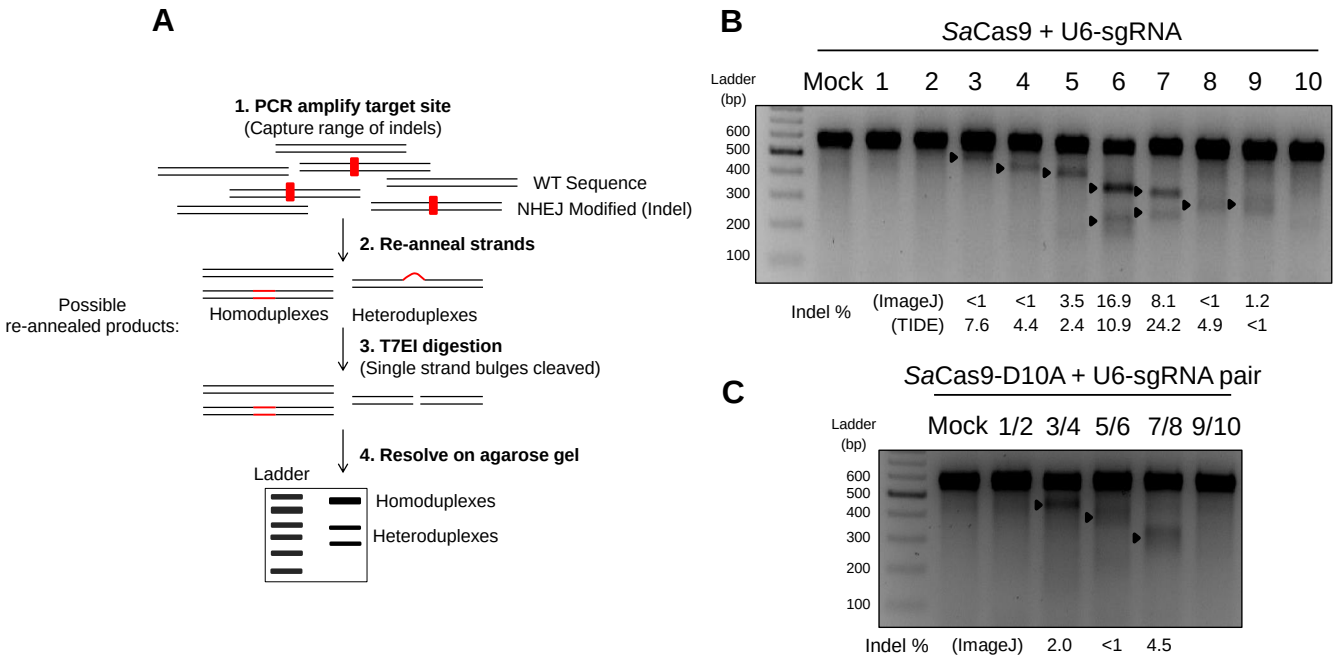



**Figure 3.3: T7EI assays performed on *Sa*Cas9/-D10A-treated HEK293T cells with a panel of ten *CCR5*-specific U6-sgRNAs. A** – An outline of the T7EI assay. HEK293T cells were transfected with *Sa*Cas9/-D10A constructs and sgRNAs 1 – 10 or pairs 1/2 – 9/10 respectively, gDNA was extracted 48 hours later and a T7EI assay was performed for target site indel visualisation. Briefly, the target region was PCR amplified and products denatured and slowly re-annealed. Target site indels (red) resulted in the formation of single stranded bulges (heteroduplexes) which were discriminated from the perfectly aligned sequences (homoduplexes) by T7EI cleavage activity. Products were then resolved on a 2 % agarose gel alongside a molecular weight ladder for analysis of indels resulting from HEK293T cell treatment with *Sa*Cas9 + U6-sgRNAs 1 – 10 (**B**) or *Sa*Cas9-D10A + U6-sgRNA pairs 1/2 – 9/10 (**C**). Black arrows point to the digestion products, indicative of indel formation. A mock negative control (absence of a targeting gRNA sequence) was also included in the experiment and indel frequencies calculated using densitometry (**B** and **C**) and TIDE (**B**).

The T7EI assay was employed as a simple, rapid and cost-effective method to screen the panel of sgRNAs for cleavage activity (Mashal et al., 1995). Selective cleavage by the T7EI enzyme of single stranded heteroduplex bulges, followed by agarose gel electrophoresis, provided direct evidence of NHEJ-mediated DNA repair. The efficiency of indel formation at each sgRNA target site was assessed by densitometric quantification of the digestion products relative to intact homoduplex bands. PCR products resolved on the agarose gels as single products of the expected 620 bp size (Supplementary Figure 7.5.1A) and the mock control samples did not induce target site indel formation as no digestion products were visible on the T7EI gels (Figure 3.3B and C). Indel formation was detected for samples treated with *Sa*Cas9 + U6-sgRNAs 3 – 9 and *Sa*Cas9-D10A + U6-sgRNA pairs 3/4 – 7/8. According to densitometric analysis, the cleavage efficiencies of successful treatments ranged from < 1 –

16.9 %. U6-sgRNAs 6 and 7 appeared to function most efficiently of the *Sa*Cas9 + U6-sgRNA treatments, generating indels at frequencies of 16.9 and 8.1 % respectively. The cleavage efficiencies of U6-sgRNAs 3, 4, 5 and 8 were relatively low, ranging from <1 – 3.5 % and U6-sgRNAs 1, 2 and 10 resulted in no T7EI digestion products, suggesting that these sgRNAs were non-functional. Of the five dual sgRNA nickase treatments, the U6-sgRNA 7/8 pair showed highest cleavage activity at 4.5 %, while U6-sgRNA 3/4 and 5/6 resulted in indels that were almost undetectable, despite the clear band seen for pair 3/4 on the gel. One of the pitfalls of the T7EI assay is that homoduplex-containing indels (Figure 3.3A) are potentially overlooked, thus underestimating true indel frequencies. The applicability of TIDE as an alternative to the T7EI assay was investigated in order to circumvent this drawback.

The novel TIDE computational tool was recently developed for rapid and cost-effective screening of CRISPR/Cas9 sgRNA cleavage activity (Brinkman et al., 2014). The TIDE online calculator deconvolutes Sanger sequencing trace signals within a single genetically modified population in order to quantify target site indel frequency as well as characterise major variants present. This alternative method was only applicable for *Sa*Cas9-mediated indel quantification (Figure 3.3B) and not for those introduced by *Sa*Cas9-D10A (Figure 3.3C). The variety of indels generated by CRISPR/Cas9 nickase systems (Shen et al., 2014) are most likely undetectable within a Sanger sequence trace by the TIDE algorithm which calls bases that are significantly different from the expected ‘consensus’ sequence derived from the negative control. The TIDE online analysis provided the total cleavage efficiency the U6-sgRNAs that resulted in T7EI digestion products (U6-sgRNAs 3 – 9; Table 3.1) and these sites were manually inspected within the Sanger sequencing chromatograms to confirm correspondence with the expected empirical cleavage site, three bp upstream of the PAM motif.

**Table 3.1: TIDE indel quantification of *Sa*Cas9-mediated U6-sgRNA cleavage activity**

| Nuclease treatment | $R^2$ Value | WT sequences (%) | Total efficiency (%) | Indels: % (no. bp deleted; p-value) |
|---|---|---|---|---|
| **U6-sgRNA 3** | 0.98 | 90.6 | 7.6 | 3.6 (-2; 6.2 E-05) |
| **U6-sgRNA 4** | 0.99 | 95.1 | 4.4 | 1.1 (-1; 0.015) |
| **U6-sgRNA 5** | 0.99 | 96.9 | 2.4 | - |
| **U6-sgRNA 6** | 0.99 | 87.8 | 10.9 | 5.1 (-1; 1 E-41) |
| **U6-sgRNA 7** | 0.93 | 68.4 | 24.2 | 10.5 (-2; 1.1E-31) and 6.6 (-1; 3.5 E-13) |
| **U6-sgRNA 8** | 0.99 | 94.2 | 4.9 | 2.2 (-2; 3.4 E-15) |
| **U6-sgRNA 9** | 0.99 | 98 | 0.5 | - |

The $R^2$ value obtained indicates the proportion of sequence traces that fit within the computational model while the remainder are categorised either as noise or as major indels above the 10 bp threshold (Section 2.5.). According to TIDE quantification, U6-sgRNA 7 functioned with the greatest efficiency through the introduction of 24.2 % indels, followed by sgRNAs 6, 3, 8, 4, 5 and 9 with 10.9, 7.6, 4.9, 4.4, 2.4 and <1 % indels respectively (p-values in Table 3.1). Based on the $R^2$ value of each result, at least 93 % of the sequence traces fitted within the computational model and thus the indels called and quantified by the online programme can be taken to be true with a high level of confidence. The most common types of indels predicted with the highest frequency were a one and two base pair deletion (-1 and -2) within the expected target site (Table 3.1). As a relatively large proportion of indels was introduced by the *Sa*Cas9 + U6-sgRNA 7 treatment, the region around the expected sgRNA cut site was inspected on the Sanger sequencing chromatogram to determine whether the aberrant signals detected by the programme were apparent above the background level or not (Figure 3.4).

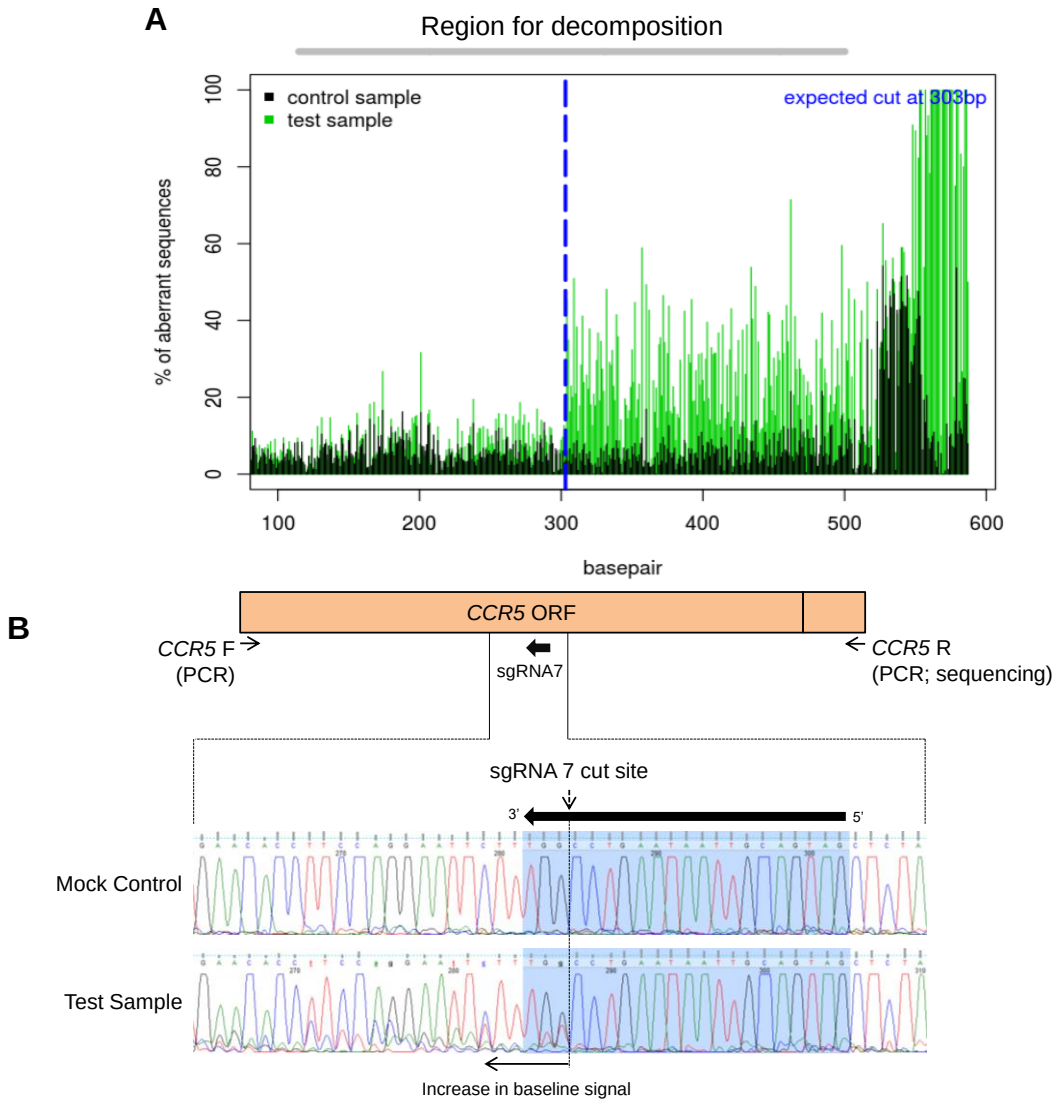



**Figure 3.4: TIDE indel quantification from treatment of HEK293T cells with *Sa*Cas9 and U6-sgRNA 7.** **A** – Output from the TIDE analysis tool showing the percentage of aberrant sequences, compared to a computationally-derived consensus sequence (grey bars), contributed by the control (dark green) and test (light green) sample sequences. A 100 bp window upstream of the expected cut site was set as the threshold for sequence alignment between the negative control and test sample and the decomposition window set from 115 – 500 bp so as to exclude poor Sanger sequence quality regions on either end. **B** – Schematic of the sgRNA 7 target site within the *CCR5* ORF and positions of *CCR5* primers used for PCR (F and R) and sequencing (R) before TIDE analysis was carried out and Sanger sequencing chromatograms following HEK293T cell treatment with the *Sa*Cas9 mock control and *Sa*Cas9 + U6-sgRNA 7. The predicted cut site is indicated showing the increase in the baseline signal downstream of the expected cut site as a result of indels introduced by the CRISPR/Cas9 system.

The output TIDE histogram shows the percentage of aberrant signals contributed by the mock control (dark green bars) and test (light green bars) sample sequences relative to the expected 'consensus' sequence (grey bars) (Figure 3.4A). To the right of the predicted cut site on the

histogram there is a clear increase in aberrant signal contributed by the test sample only. A similar effect can be seen on the Sanger sequencing chromatograms by the distinct increase in baseline signal downstream of the sgRNA cut site in the Test Sample (Figure 3.4B). It is important to note that sequencing was performed with the *CCR5* reverse (R) primer thus 'downstream' has been used to refer to the signal observed when moving towards the 3′ PAM-proximal end of the sgRNA. Five heterozygous peaks were corrected in the Test sample chromatogram (lower case) such that the two baseline sequences were identical and to confirm that indels were introduced at the expected target site. One and two bp positions downstream of the expected cleavage site on the Test sample chromatogram had strong background thymine (T) signals (Figure 3.4B). This shift in the baseline signal corresponds with the effect that the two major TIDE predicted indel types, one and two bp deletions at frequencies of 10.5 % and 6.6 % respectively (Table 3.1), would have on the background sequence. TIDE analysis thus provided a simple and cost-effective means to rank the efficiency of each sgRNA with confidence and affirmed that the digestion products observed on the T7EI gels were indicative of CRISPR/*Sa*Cas9-mediated DSB induction and NHEJ DNA repair. This experiment was performed for preliminary sgRNA activity screening and sgRNAs 7 and 8 were considered most appropriate for downstream analysis using both the *Sa*Cas9 and *Sa*Cas9-D10A systems owing to the highest level of cleavage activity both individually with the nuclease and as a pair with the nickase. In order to improve the stability, and consequently efficiency, of U6-sgRNAs 7 and 8 for downstream experimentation, the gRNA ODNs were cloned into the U6-*Sa*Cas9_modtracrRNA vector (Appendix 7.1.3.) (Tabebordbar et al., 2016). Sanger sequencing was carried out to confirm successful cloning of the modified scaffold sgRNAs (Inqaba Biotech, South Africa; Figure 3.5).

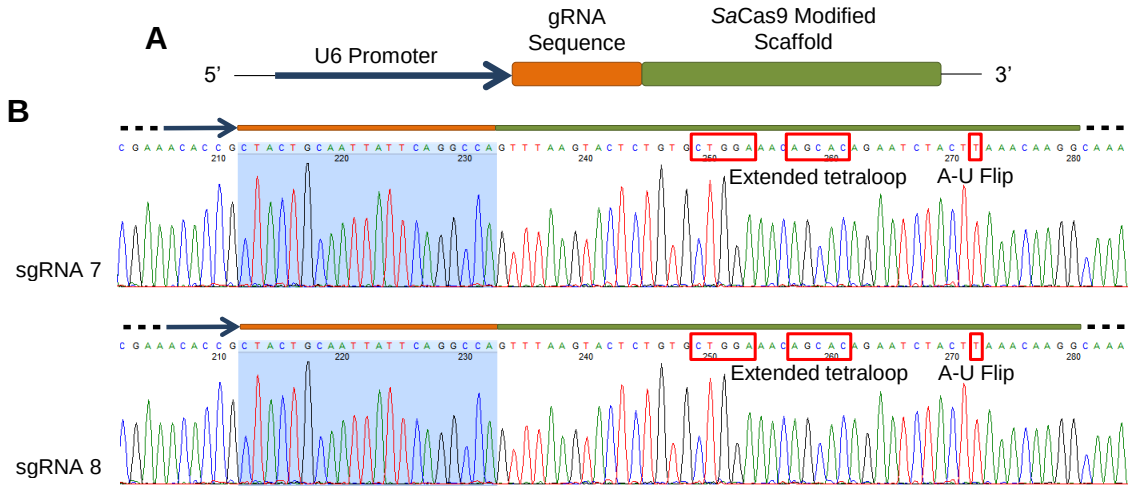


**Figure 3.5: Sequence confirmation of *CCR5*-specific U6-sgRNAs cloned into the U6-sgRNA_modtracrRNA vector. A** – A schematic of the U6-sgRNA_modtracrRNA vector to illustrate the position of the U6 promoter, gRNA sequence and modified *Sa*Cas9 scaffold corresponding to the chromatogram sequences. **B** – Sanger sequencing chromatograms of U6-sgRNAs 7 and 8 showing seamless insertion into the expression vector. The sequences of the extended tetraloop and A-U flip, introduced for improved sgRNA stability, are indicated within the scaffold sequence.

The two sgRNAs with improved scaffold stability were then tested for cleavage activity with *Sa*Cas9/-D10A using highly sensitive ddPCR™ and deep sequencing and the safety of each CRISPR/*Sa*Cas9 system assessed by targeted amplicon deep sequencing at selected computationally predicted off target sites.

## 3.2. Analysis of the efficiency and specificity of *CCR5*-specific gene editing mediated by CRISPR/*Sa*Cas9 nuclease and nickase systems

In order to develop an effective *CCR5*-specific CRISPR/Cas9 anti-HIV gene therapy, the system needs to generate functionally disruptive indels in the target site whilst avoiding potentially deleterious off target cleavage activity. The ddPCR™ drop-off assay and targeted amplicon deep sequencing were employed to measure the efficiency of the *Sa*Cas9 nuclease and nickase at the sgRNA 7 and 8 target sites with high sensitivity. It is important to quantify the propensity for indel formation prior to performing functional analyses of *CCR5* gene editing to determine whether the selected CRISPR/*Sa*Cas9 systems are likely to result in reduced infectivity of R5-tropic virus or not based on the extent of sequence modification.

### 3.2.1. High sensitivity indel quantification at the *CCR5*-specific CRISPR/*Sa*Cas9 nuclease and nickase target sites by the ddPCR™ drop-off assay

The ddPCR™ 'drop-off' assay allowed for the sensitive and absolute quantification of *CCR5*-specific target site indels without the necessity for reference gene standardisation (Section 2.6.). TZM-bl cells were treated with *Sa*Cas9 and U6-sgRNA 7 or 8 and with *Sa*Cas9-D10A and U6-sgRNA pair 7/8 in three independent transfections (replicates 1 – 3). *CCR5*-specific target site indels were quantified by the ddPCR™ drop-off assay using NHEJ probe 1, specific to the sgRNA 7 target site, and 2, specific to the sgRNA 8 target site as well as both for dual sgRNA targeting with *Sa*Cas9- D10A (Figure 3.6; Table 3.2).

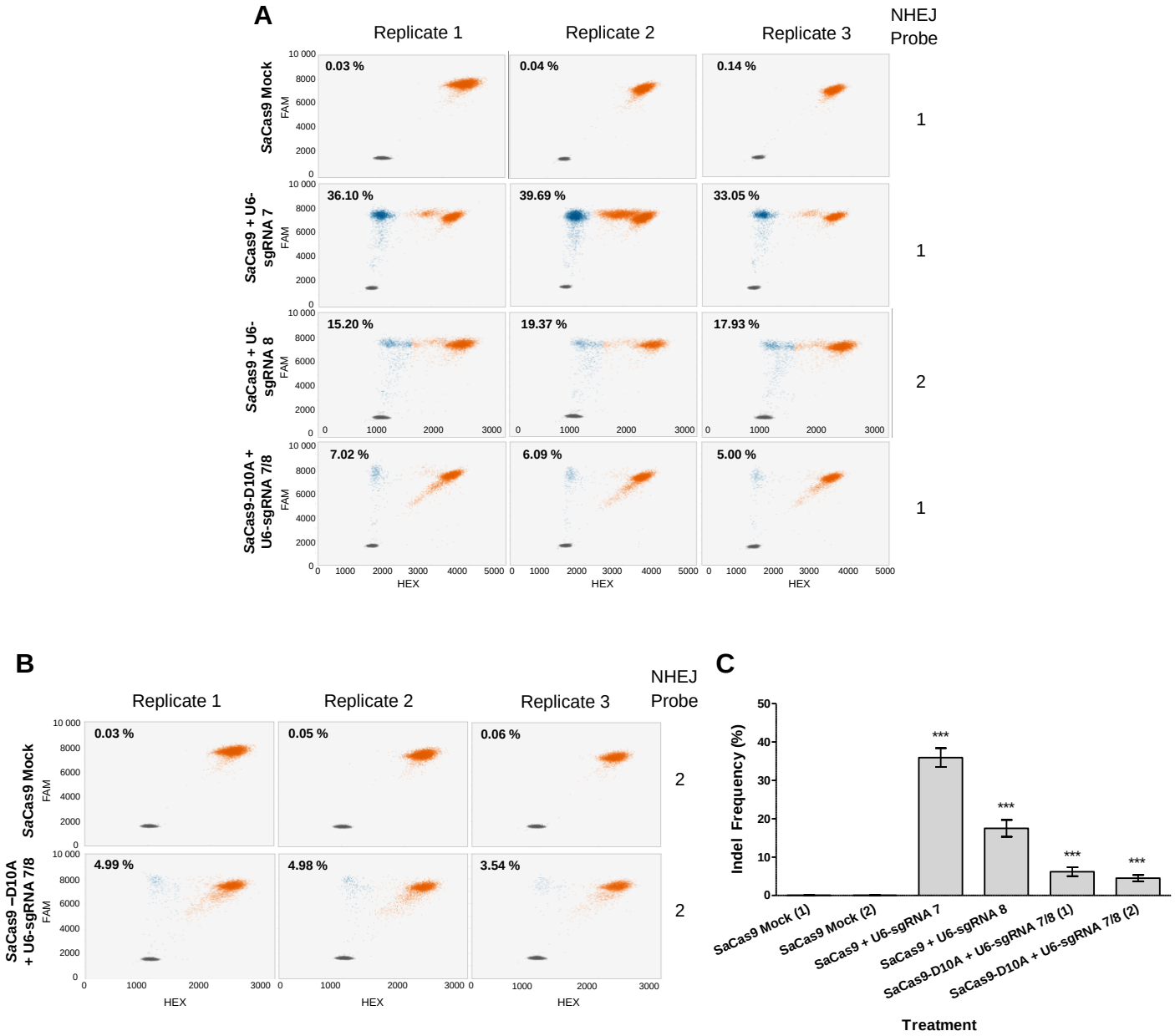



**Figure 3.6: Quantification of CRISPR/*Sa*Cas9-mediated *CCR5* target site indel formation by the ddPCR™ drop-off assay. A** and **B** – 2D-Plots of HEX (NHEJ probe 1 or 2) and FAM (ref probe) fluorescence measured by the ddPCR™ drop-off assay following TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNA 7 or 8 and *Sa*Cas9-D10A + U6-sgRNA 7/8 for three independent transfections (replicates 1 – 3). The gDNA was extracted 48 hours following transfection and a single ddPCR™ drop-off assay performed for target site indel quantification of each sample from each replicate. The threshold for positive FAM fluorescence was set at 2 250 using a no gDNA (water) negative control (Supplementary Figure 7.5.2A) and for HEX at 2 150 (NHEJ probe 1) using the mock control (treatment with the absence of a gRNA sequence). This allowed for clustering of the WT/WT+ droplets (orange), the NHEJ drop-off droplets (blue) and the negative droplets (black). QuantaSoft™ Analysis Pro software was used to calculate the indel frequencies as an average of technical duplicates. **C** – Bar graph showing indel frequencies calculated by the ddPCR™ drop-off assay averaged across the three biological replicates for NHEJ probe 1 and 2 (Table 3.2, replicates 1, 2 and 3). The significance of relative differences in

indel frequencies between treatments and the mock control are shown by asterisk/s above the bars. A two-tailed unpaired t-test with a 95% confidence interval was used to calculate p-values. Error bars show the sample standard deviations. * $p \leq 0.05$; ** $p \leq 0.005$; *** $p \leq 0.0005$.

**Table 3.2: ddPCR™ drop-off assay results obtained from TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNA 7 or 8 and *Sa*Cas9-D10A + U6-sgRNA 7/8 for three independent replicates**

| Replicate | Treatment | NHEJ Probe | Average Indel Frequency (%) | Standard Deviation | p-value |
|---|---|---|---|---|---|
| 1 | *Sa*Cas9 Mock Control | 1 | 0.04 | 0.007 | - |
| | | 2 | 0.03 | 0.021 | - |
| | *Sa*Cas9 + U6-sgRNA 7 | 1 | 36.10 | 0.622 | **0.0079** |
| | *Sa*Cas9 + U6-sgRNA 8 | 2 | 15.21 | 0.969 | **0.0287** |
| | *Sa*Cas9 + U6-sgRNA 7/8 | 1 | 6.54 | 0.686 | **0.0474** |
| | | 2 | 4.99 | 0.502 | 0.0051 |
| 2 | *Sa*Cas9 Mock Control | 1 | 0.05 | 0.007 | - |
| | | 2 | 0.05 | 0.007 | - |
| | *Sa*Cas9 + U6-sgRNA 7 | 1 | 39.69 | 1.442 | **0.0168** |
| | *Sa*Cas9 + U6-sgRNA 8 | 2 | 19.37 | 0.233 | **0.0054** |
| | *Sa*Cas9 + U6-sgRNA 7/8 | 1 | 7.02 | 1.315 | **0.0844** |
| | | 2 | 4.98 | 0.615 | **0.0560** |
| 3 | *Sa*Cas9 Mock Control | 1 | 0.14 | 0.057 | - |
| | | 2 | 0.06 | 0.042 | - |
| | *Sa*Cas9 + U6-sgRNA 7 | 1 | 33.68 | 0.636 | 0.0002 |
| | *Sa*Cas9 + U6-sgRNA 8 | 2 | 17.93 | 2.270 | **0.0571** |
| | *Sa*Cas9 + U6-sgRNA 7/8 | 1 | 5.00 | 0.170 | 0.0007 |
| | | 2 | 3.54 | 0.156 | 0.0011 |
| 1, 2 and 3 | *Sa*Cas9 Mock Control | 1 | 0.06 | 0.040 | - |
| | | 2 | 0.05 | 0.024 | - |
| | *Sa*Cas9 + U6-sgRNA 7 | 1 | 35.94 | 2.443 | **<0.0001** |
| | *Sa*Cas9 + U6-sgRNA 8 | 2 | 17.50 | 2.191 | **<0.0001** |
| | *Sa*Cas9 + U6-sgRNA 7/8 | 1 | 6.19 | 1.156 | **<0.0001** |
| | | 2 | 4.50 | 0.827 | **<0.0001** |

Note: a Welch's correction for unequal variance determined by the F test was applied to obtain p-values in bold.

The baseline indel frequencies from the triplicate mock treatments ranged from 0.03 – 0.14 % (0.06 % on average; Figure 3.6A and B, Table 3.2). The *Sa*Cas9 + U6-sgRNA 7 treatments introduced 33.05 – 39.69 % indels (35.94 % on average) as measured using NHEJ probe 1 and U6-sgRNA 8 introduced 15.20 – 19.37 % indels (17.50 % on average) as measured using NHEJ probe 2. Both NHEJ probes were used to quantify the *Sa*Cas9-D10A + U6-sgRNA 7/8 cleavage activity as indels could theoretically be introduced at either site (Figure 3.6A and B). The number of indels measured by NHEJ probe 1 was 5.00 – 7.02 % (6.19 % on average) and with NHEJ probe 2 lower per corresponding treatment, ranging from 3.54 – 4.99 % (4.50

% on average). As the exact types and locations of the nickase-induced indels has not been shown by this method, it is unknown whether addition of the two indel frequencies measured by NHEJ probes 1 and 2 is accurate to calculate the overall efficiency. The hypothesis is that indels could be introduced at either of the target sites, or both within a single sequence and thus the range of indels on average across the triplicate readings is expected to be from 6.19 – 10.69 % (Figure 3.6A and B, Table 3.2). To more accurately measure nickase-mediated indel formation using the ddPCR™ drop-off assay, an experiment should have been carried out using each NHEJ probe individually as well as together. A calculation can then be done to determine how many target regions contain indels at both sgRNA binding sites by subtracting the combination probe value from the sum of individual NHEJ probe indel frequencies. The slightly lower reading obtained using NHEJ probe 2 relative to NHEJ probe 1 for each of the *Sa*Cas9-D10A-treated samples could indicate that indel formation is favoured at the sgRNA 7 binding site over that of sgRNA 8. Alternatively, the binding efficiency of NHEJ probe 2 could be lower than NHEJ probe 1 resulting in a slight signal reduction for these measurements despite an even indel distribution between sgRNA 7 and 8 target sites. The latter scenario is possible as the WT amplitude of HEX fluorescence measured using NHEJ probe 1 was approximately a third higher than that of NHEJ 2 (Figure 3.6A and B).

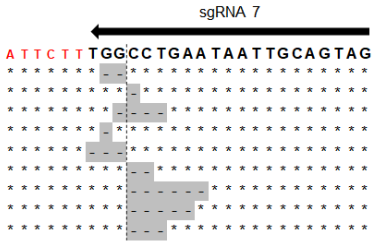
Overall, the ddPCR™ drop-off assay results demonstrate that the *Sa*Cas9 + U6-sgRNA 7 system has the highest capacity for indel formation while that of U6-sgRNA 8 is reduced by ~ 2-fold. The *Sa*Cas9 dual U6-sgRNA 7/8 system has further reduced cleavage activity of between ~ 3.4- and 5.8-fold compared to *Sa*Cas9 + U6-sgRNA 7 and between ~ 1.6- and 2.8-fold compared to *Sa*Cas9 + U6-sgRNA 8 treatment, however, the precise extent of this reduction cannot be ascertained by these results. In order to provide insight into the types and location of indels introduced by the *Sa*Cas9 with U6-sgRNAs 7 and 8 individually as well as *Sa*Cas9-D10A and U6-sgRNA 7/8, targeted amplicon deep sequencing was carried out on the *CCR5* target region. This method was employed in parallel to determine the safety of the *Sa*Cas9 and *Sa*Cas9-D10A treatments at a total of ten computationally predicted exonic off target sites within the human genome.

### 3.2.2. High sensitivity indel analysis at the *CCR5*-specific CRISPR/*Sa*Cas9 nuclease and nickase on target sites by deep sequencing

Targeted amplicon deep sequencing of the *CCR5* target region was carried out on TZM-bl cell samples treated with a mock control (*Sa*Cas9 with an empty U6-sgRNA vector), *Sa*Cas9

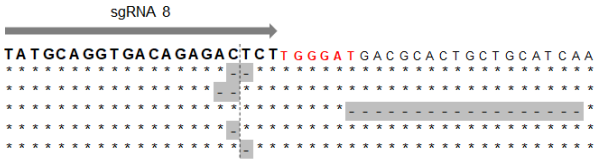
+ U6-sgRNA 7 and 8 individually and *Sa*Cas9-D10A + U6-sgRNA 7/8. This enabled high sensitivity quantification of *CCR5*-specific target site indels as well as characterisation of the major types. The *CCR5* target region was PCR amplified, the single ~ 300 bp products (Supplementary Figure 7.5.3) sequenced using the Illumina MiSeq platform (Inqaba Biotech, South Africa) and indels quantified using the Low Variant Detection Tool in the CLC Genomics Workbench at a threshold of 0.1 % and a significance of 1 %. The mock treatment control was used as a baseline for the level of CRISPR/Cas9-independent indels present from 20 bp up- and downstream of the combined sgRNA 7 and 8 target site. The types and frequencies of indels introduced by *Sa*Cas9 + U6-sgRNA 7 and 8 as well as by *Sa*Cas9-D10A + U6-sgRNA 7/8 for the three biological replicates at a threshold of ≥ 0.1 % were determined using the CLC Genomics Workbench Low Frequency Variant Detection Tool (Figure 3.7).

**A**




| Indel Type | Replicate 1 | Replicate 2 | Replicate 3 | Average |
|---|---|---|---|---|
| WT | Frequency (%) | Frequency (%) | Frequency (%) | Frequency (%) |
| -2 | 6.61 | 8.49 | 5.32 | 6.81 |
| -1 | 4.14 | 3.37 | 3.33 | 3.61 |
| -4 | 1.77 | 2.48 | 1.49 | 1.91 |
| -1 | 2.34 | 2.29 | 2.17 | 2.27 |
| -3 | 0.79 | 1.25 | 0.6 | 0.88 |
| -2 | 0.9 | 1.02 | 0.89 | 0.94 |
| -6 | 0.52 | 0.8 | 0.65 | 0.66 |
| -5 | 0.41 | 0.78 | 0.47 | 0.55 |
| -3 | 0.67 | 0.74 | 0.67 | 0.69 |
| Minor | 6.39 | 10.14 | 6.15 | 7.56 |
| Mock | 2.2 | 1.61 | 1.12 | 1.64 |
| **Total** | **22.34** | **29.75** | **20.62** | **24.24** |

**B**




| Indel Type | Replicate 1 | Replicate 2 | Replicate 3 | Average |
|---|---|---|---|---|
| WT | Frequency (%) | Frequency (%) | Frequency (%) | Frequency (%) |
| -2 | 2.05 | 2.03 | 1.8 | 1.96 |
| -2 | 1.62 | 1.59 | 1.14 | 1.45 |
| -18 | 0.72 | 1.53 | 1.22 | 1.16 |
| -1 | 0.7 | 0.9 | 0.58 | 0.73 |
| -1 | 0.65 | 0.76 | 0.71 | 0.71 |
| Minor | 5.92 | 10.23 | 6.59 | 7.58 |
| Mock | 2.2 | 1.61 | 1.12 | 1.64 |
| **Total** | **9.46** | **15.43** | **10.92** | **11.94** |

**C**




| Indel Type | Replicate 1 | Replicate 2 | Replicate 3 | Average |
|---|---|---|---|---|
| WT | Frequency (%) | Frequency (%) | Frequency (%) | Frequency (%) |
| +31 | 0.21 | 0.30 | 0.14 | 0.22 |
| -6 | 0.32 | 0.11 | 0.13 | 0.19 |
| -7 | - | 0.11 | 0.13 | 0.12 |
| Minor | 2.46 | 1.40 | 0.86 | 1.57 |
| Mock | 1.61 | 1.38 | 0.81 | 1.27 |
| **Total** | **1.38** | **0.54** | **0.45** | **0.79** |

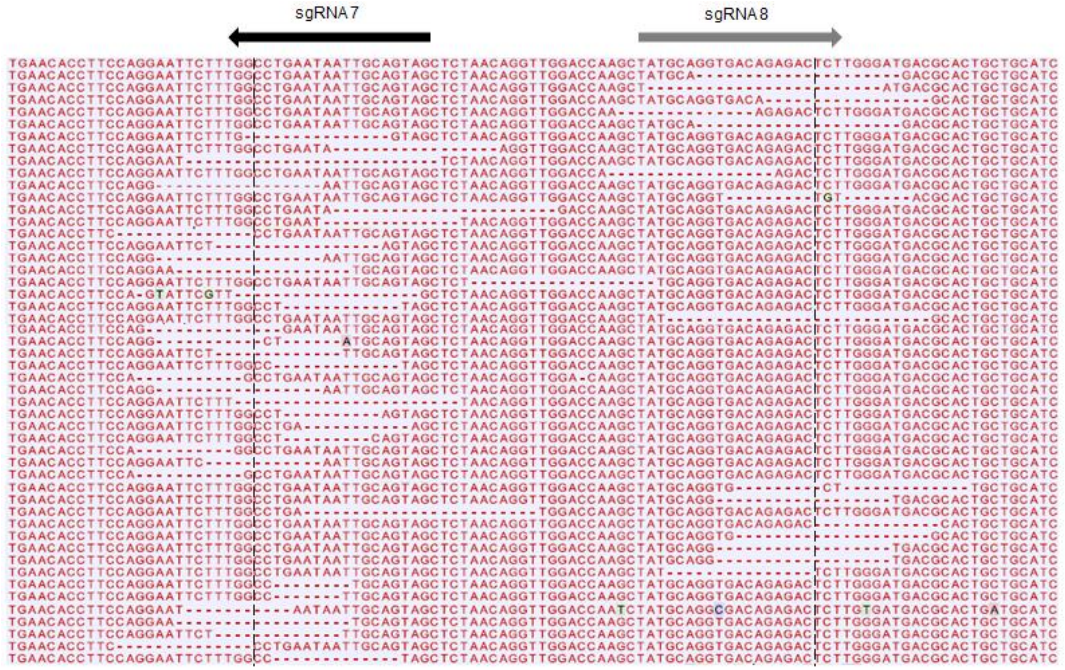


**Figure 3.7: Targeted amplicon deep sequencing of the *CCR5* target region following TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNAs 7 and 8 as well as *Sa*Cas9-D10A + U6-sgRNA 7/8.** Quantification of indels introduced into the TZM-bl cell *CCR5* target site following treatment with *Sa*Cas9 + U6-sgRNA 7 (**A**) and 8 (**B**) as well as *Sa*Cas9-D10A + U6-sgRNA 7/8 (**C**) by Illumina MiSeq targeted amplicon deep sequencing and analysis using the Low Variant Detection Tool in the CLC Genomics Workbench at a threshold of 0.1 % and a significance of 1 %. Most frequently identified indels are represented in the diagrams (deletions or insertions shaded) where '–' indicates a deletion and '+' an insertion of the number of bp indicated in the table alongside. Dotted lines represent the expected Cas9 cleavage site and asterisks correspond to perfect matches with the WT sequences. The total indel frequency for each biological replicate was calculated by addition of the major and minor variants and subtraction of the mock control background frequencies. The lower diagram in **C** is a representation of the indel alignment within the *Sa*Cas9-D10A + U6-sgRNA 7/8 target site obtained from the CLC Genomics Workbench.

The major indels introduced by *Sa*Cas9 and U6-sgRNAs 7 or 8 were situated mainly around the cleavage site, as would be expected (Figure 3.7A). One 18 bp deletion outside of the expected sgRNA 8 cleavage and indel forming region was found to be present in all three replicates, upstream and at the 5′ end of the PAM sequence. This indel was observed at baseline frequencies of 0.1, 0.15 and 0.17 % in the replicate 2 mock control and replicates 2 and 3 U6-sgRNA 7 samples respectively (Supplementary Table 7.5.1). The frequency of this type of indel was increased by at least 4-fold and as high as ~ 15-fold by the sgRNA 8 treatment, up to a frequency of 1.53 % in replicate 1 (Figure 3.7B). The four and two most frequent indels observed for the *Sa*Cas9 + U6-shRNA 7 and 8 treatments respectively were also observed in the mock control as well as the *Sa*Cas9-D10A + U6-sgRNA 7/8 treatments at low frequencies (Figure 3.7A and B; Supplementary Table 7.5.1). These were increased by ~ 18- to 29-fold and ~ 13- to 16-fold for each of these treatments respectively and thus the background levels of these indels could indicate that a small amount of cross-over contamination may have occurred during experimentation. This would be counteracted by subtracting the baseline level of indels from the sample treatments to determine the overall level of CRISPR/*Sa*Cas9-induced cleavage activity. The *Sa*Cas9 + U6-sgRNA 7 treatment resulted in the highest indel frequency of the three treatments ranging from 20.62 – 29.75 %, followed by *Sa*Cas9 + U6-sgRNA 8 ranging from 9.46 – 15.43 % and finally the *Sa*Cas9-D10A + U6-sgRNA 7/8 treatment which introduced indels ranging from 0.45 – 1.38 % (Figure 3.7A – C).

In the *Sa*Cas9-D10A + U6-sgRNA 7/8 sequence alignment obtained using the CLC Genomics Workbench, a number of indels were observed at both the sgRNA 7 and 8 target sites that are believed not to have been detected by the Low Frequency Variant Detector tool

at the cut-off of ≥ 0.1 % (Figure 3.7C). Additionally, few sequences appeared to contain indels at both the sgRNA 7 and 8 target sites with a biased distribution towards the sgRNA 7 target site. The indels quantified by the ddPCR™ drop-off assay with NHEJ probes 1 and 2 can thus most likely be added to determine the overall nickase-induced indel frequency in a treated sample. This would mean that, from the ddPCR™ quantification performed prior to deep sequencing, the frequency of indels would be ~ 12, 11 and 8.5 % for the three replicates respectively and ~ 10.5 % on average (Figure 3.6A and B, Table 3.2). The cleavage activity measured by deep sequencing for all three treatments was lower than those measured by the ddPCR™ drop-off assay. The indels below the 0.1 % cut-off in each of the samples are likely to account for the indel deficit between the ddPCR™-quantified indel frequencies and those detected by deep sequencing. The overall trend in replicate rankings for the three different on target treatments was found to be the same as that observed in the ddPCR™ results, with replicate 2 having the highest activity, 1 the second highest and 3 the lowest.

The deep sequencing analysis of the *CCR5* target site illustrates that each of the CRISPR/*Sa*Cas9 systems was capable of introducing indels into the desired *CCR5* target site. The two different nuclease treatments introduced short deletions at the expected cleavage site detectable above the indel frequency cut-off of 0.1 % while the nickase resulted in a range of larger deletions and one detectable insertion across either the sgRNA 7 or 8 cleavage site, a large proportion of which were undetectable using this analysis tool. While low variant detection of deep sequencing samples is highly sensitive to a threshold of 0.1 %, this may exclude a large proportion of indels that are highly variable and thus underrepresented within the nickase-treated samples. In general, the quantified indel frequencies for each CRISPR/*Sa*Cas9 treatment assessed were lower than those detected by the ddPCR™ drop-off assay (Figure 3.6 and 3.7). One of the drawbacks of the deep sequencing quantification method is therefore that highly variable indels are not cumulative to a detectable level and are overlooked during analysis. The advantage with deep sequencing over the ddPCR™ drop-off assay, however, is that a visual representation of the location and types of indels is attainable making it easier to predict the potential downstream functional impact of the gene editing events. Before functional assessment was carried out to determine the downstream effects of these modifications, the specificity and thus potential safety of the gene editing complexes was assessed. This was carried out by deep sequencing of various computationally predicted sgRNA 7 and 8 off target sites and searching for indels within the expected target regions.

### 3.2.3. High sensitivity indel analysis at *CCR5*-specific CRISPR/*Sa*Cas9 off target sites by deep sequencing

The specificity of the CRISPR/*Sa*Cas9 nuclease and nickase systems at sgRNA 7- and 8-specific potential off target sites in TZM-bl cells was assessed by targeted amplicon deep sequencing in parallel with on target indel quantification. Five of the top exonic off target sites as predicted by CRISPRseek with similarity to the sgRNA 7 and 8 gRNA sequences were selected for analysis as it was proposed that modification of exons would have a greater negative effect on gene function than editing within introns (Table 3.3).

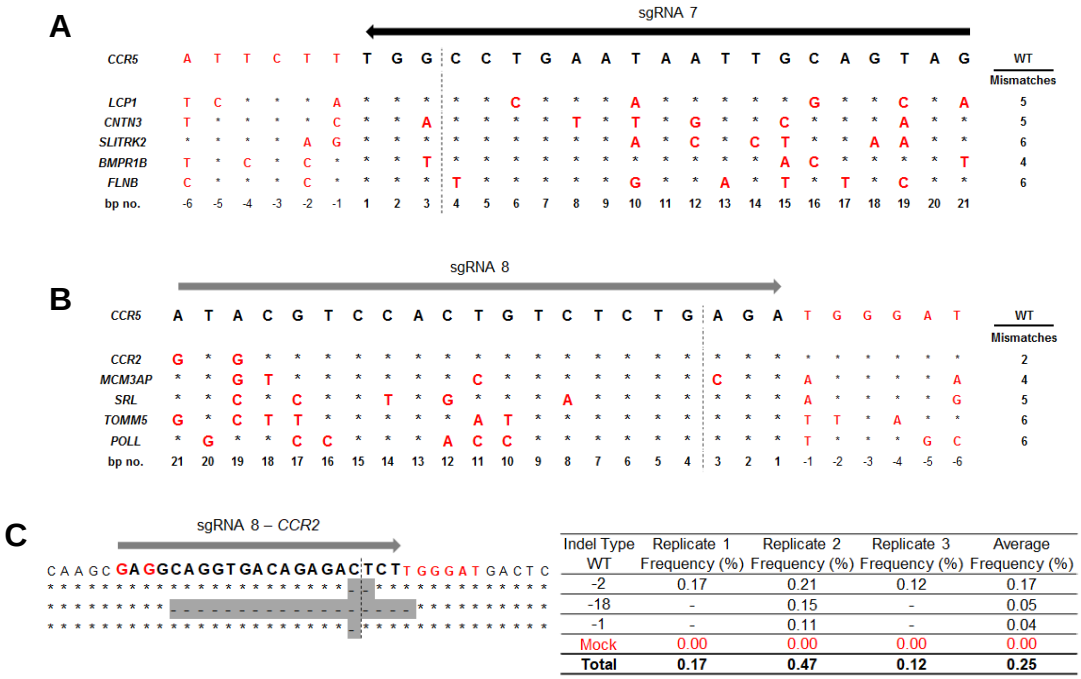
**Table 3.3: CRISPRseek *Sa*Cas9-specific exonic off target prediction sites similar to the *CCR5*-specific sgRNAs 7 and 8**

| Gene Name | Gene Entrez-ID | Similarity Score | Predicted gRNA and PAM Sequence (5′ – 3′) | UCSC Locus (chr bp start-end) |
|---|---|---|---|---|
| ***CCR5*-1** | 1234 | 100 | CTACTGCAATTATTCAGGCCAaagaat | chr3 46415179-46415205 |
| ***LCP1*** | 3936 | 0.1 | TTGCTCCAATTTTTCGGGCCAtagaga | chr13 46701824-46701850 |
| ***CNTN3*** | 5067 | 0.1 | CTTCTGGAACTATACAGGTCAgagaaa | chr3 74349081-74349107 |
| ***SLITRK2*** | 84631 | 0.1 | CTTTTGAGAGTTTTCAGGCCActgaat | chrX 144904258-144904284 |
| ***BMPR1B*** | 658 | 0.1 | ATACTGTAATTATTCAGGACTagggaa | chr4 96079480-96079506 |
| ***FLNB*** | 2317 | 0.1 | CTGCAGAATTTCTTCAGACCAaggaag | chr3 58157231-58157257 |
| ***CCR5*-2** | 1234 | 100 | TATGCAGGTGACAGAGACTCTtgggat | chr3 46415179-46415205 |
| ***CCR2*** | 729230 | 5.5 | CACGCAGGTGACAGAGACTCTtgggat | chr3 46399876-46399902 |
| ***MCM3AP*** | 8888 | 0.3 | TACACAGGTGGCAGAGACGCTagggaa | chr21 47669799-47669825 |
| ***TOMM5*** | 401505 | 0.3 | CAGAAAGGTGTAAGAGACTCTttgaat | chr9 37588790-37588816 |
| ***POLL*** | 27343 | 0.2 | TCTGGGGGTTGGAGAGACTCTtgggggc | chr10 103347863-103347889 |
| ***SRL*** | 6345 | 0.1 | TAGGGAGATCACATAGACTCTagggag | chr16 4240605-4240631 |

Note: *CCR5*-1 and -2 are the sgRNA 7 and 8 target sites respectively; non-canonical PAM sequence residues are underlined.

The selected off target sites were sequenced using the Illumina MiSeq platform (Inqaba Biotech, South Africa) in parallel with the *CCR5* on target site and indels quantified using the Low Variant Detection Tool in the CLC Genomics Workbench at a threshold of 0.1 % and a

significance of 1 %. An alignment of the sgRNA 7 or 8 on and off target sequences provided insight into the likelihood for CRISPR/Cas9-mediated recognition and cleavage as the position and identity of mismatches influences this process (Figure 3.8A and B).




| Indel Type | Replicate 1 | Replicate 2 | Replicate 3 | Average |
|---|---|---|---|---|
| WT | Frequency (%) | Frequency (%) | Frequency (%) | Frequency (%) |
| -2 | 0.17 | 0.21 | 0.12 | 0.17 |
| -18 | - | 0.15 | - | 0.05 |
| -1 | - | 0.11 | - | 0.04 |
| Mock | 0.00 | 0.00 | 0.00 | 0.00 |
| **Total** | **0.17** | **0.47** | **0.12** | **0.25** |

**Figure 3.8: Alignment and deep sequencing results of the *CCR5*-specific sgRNA 7 and 8 exonic off target sites.** The *CCR5*-specific sgRNA 7 (**A**) and 8 (**B**) WT target sequences and predicted off target gene names and sequences for sgRNA 7 and 8 that were assessed by Illumina MiSeq targeted amplicon deep sequencing and analysis using the Low Variant Detection Tool in the CLC Genomics Workbench at a threshold of 0.1 % and a significance of 1 %. PAM sequences as well as mismatched residues between the *CCR5* target site and off target sequence are shown in red. **C** – The indels detected above the 0.1 % threshold in the predicted *CCR2* off target site from *Sa*Cas9 + U6-sgRNA 8 treatment.

The top exonic off target sites predicted by CRISPRseek had between 4 and 6 mismatches within the sgRNA 7 or 8 variable sequences. One of the sgRNA 7 (*SLITRK2*) and two of the sgRNA 8 (*CCR2* and *TOMM5*) off target sites contained the canonical PAM recognition sequence while the remainder lacked the preferred 5′ T (thymine) residue. Only two of the sgRNA 8 off target sequences while four of the sgRNA 7 sites had mismatches within the first 8 bp ('seed region') and tandem mismatches were located in three and two of these sequences respectively. Consecutive mismatches have been shown to further attenuate cleavage efficiency (Jiang et al., 2015). The off target sequencing analysis results revealed that the only indels introduced with a frequency ≥ 0.1 % were those mediated by *Sa*Cas9 + U6-sgRNA 8 within the *CCR2* site (Figure 3.8C). These indels were not observed in the mock control, or the *Sa*Cas9 + U6-sgRNA 7 and *Sa*Cas9-D10A + U6-sgRNA 7/8 treatments,

indicating that they are most likely sgRNA 8-specific. The *CCR2* target site has only two mismatches towards the 5′ end of the sgRNA which would be expected to have a lower impact on cleavage efficiency than mismatches towards the PAM proximal 'seed region' end. Conversely, with only two mismatches in the 5′ gRNA region and an identical PAM recognition motif, it would be expected that a greater percentage of indels be introduced at this target site, however, a truncation effect may be experienced by the mismatched sgRNA with full complementarity of only 18 instead of 21 bp. This length of *Sa*Cas9-specific sgRNA has been shown to attenuate cleavage activity of the *Sa*Cas9 nuclease (Friedland et al., 2015). Even so, only a small number of deleterious mutations are needed to negate the therapeutic benefits of CRISPR/Cas9-mediated gene editing *in vivo* and thus a drive towards perfectly specific systems is paramount.

As only five potential exonic off target sites were investigated in this study, additional off target effects may be experienced in highly similar intronic regions, especially with the relatively efficient *Sa*Cas9 and U6-sgRNA 7 system. The highly specific *Sa*Cas9-D10A nickase is expected to be able to circumvent off target effects, as was observed at the *CCR2* target site. The cleavage efficiency of this system is, however, significantly reduced compared to the nuclease systems at the selected *CCR5* on target site. An effective gene therapy should balance efforts for improved specificity with on target editing efficiency and thus alternative methods to improve the specificity of the *Sa*Cas9 nuclease, while maintaining on target cleavage efficiency, were investigated for activity at the *CCR5* sgRNA 7 and 8 target sites.

### 3.3. Generation and assessment of 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase activity at the *CCR5*-sgRNA 7 and 8 target sites

In an effort to improve both specificity and maintain the cleavage potential of CRISPR/*Sa*Cas9-mediated gene editing, the recently developed 'enhanced specificity' *Sa*Cas9 (e*Sa*Cas9) system was generated and utilised for the targeted disruption of the human *CCR5* gene (Slaymaker et al., 2016). Based on results obtained by Slaymaker *et al*. (2016), it was hypothesised that gene editing could be achieved by the WT *Sa*Cas9 and e*Sa*Cas9 systems with comparable efficiencies. As no known study has investigated the effect of combining the nickase as well as specificity enhancing mutations into a single system, the indel forming propensity of the e*Sa*Cas9-D10A nickase was also assessed for functionality in this experiment.

### 3.3.1. Generation of 'enhanced specificity' *Sa*Cas9 nuclease and nickase constructs

The four previously characterised *Sa*Cas9 specificity-enhancing codon mutations (R499A, Q500A, R654A and G655A) (Slaymaker et al., 2016) were introduced into the *Sa*Cas9 nuclease and nickase constructs by Gibson Assembly cloning and Sanger sequencing (Inqaba Biotech, South Africa) was carried out to confirm successful introduction as well as maintenance of flanking sequence integrity (Figure 3.9).

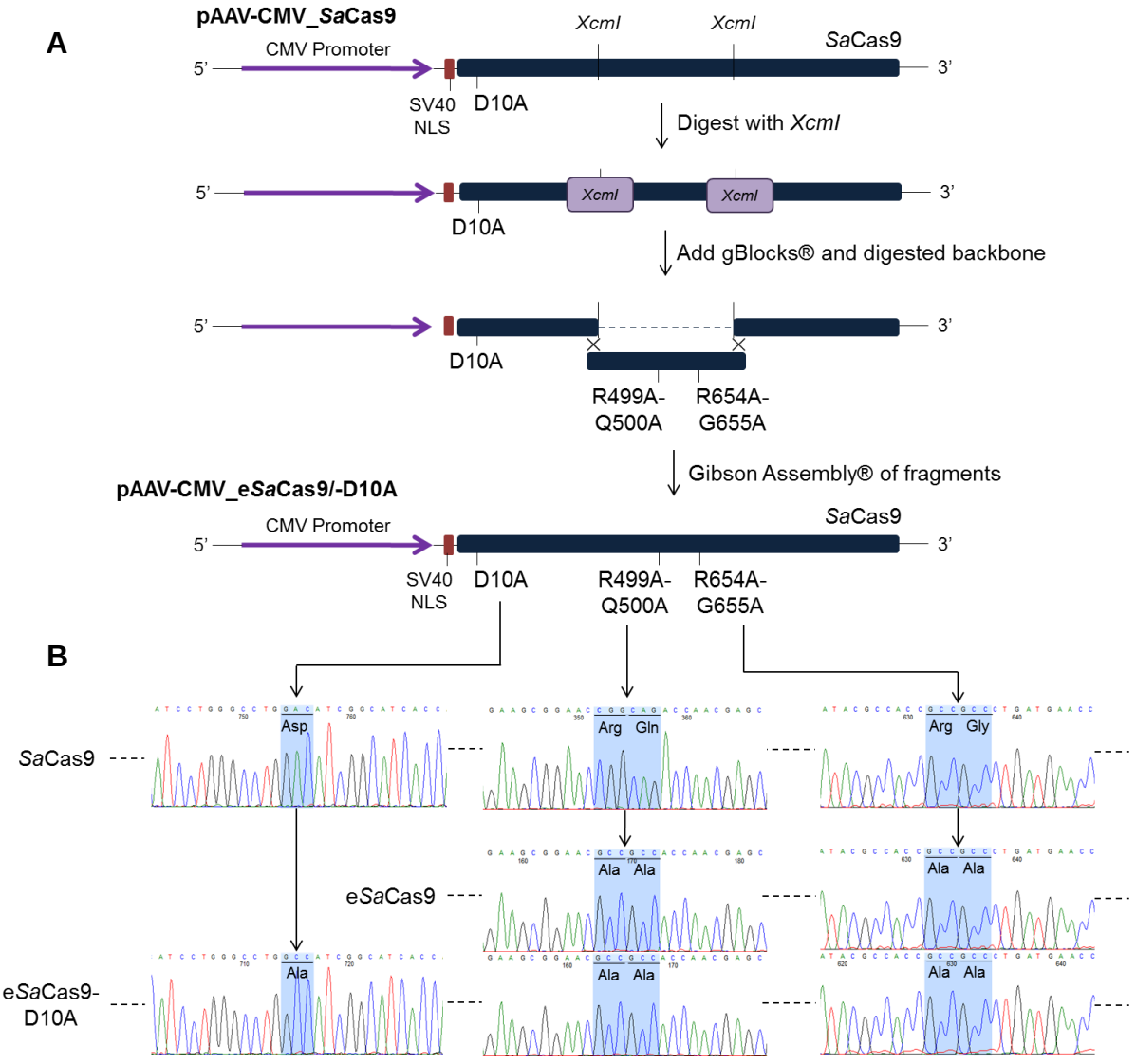



**Figure 3.9: Schematic of the cloning strategy used for generation of the e*Sa*Cas9/-D10A constructs and Sanger sequencing chromatograms showing successful modification. A** – A schematic of the cloning strategy employed for the generation of the 'enhanced specificity' *Sa*Cas9/-D10A (e*Sa*Cas9/-D10A) constructs by Gibson Assembly cloning. A synthetic fragment containing the four previously characterised *Sa*Cas9 specificity-enhancing codon mutations (R499A, Q500A, R654A and G655A) was cloned into *XcmI*-digested pAAV-CMV_*Sa*Cas9 nuclease and nickase vectors using the Gibson Assembly method. The fragment was

designed to overlap with 20 bp of each exposed end of the digested vector backbone for homologous recombination and generation of the e*Sa*Cas9 and e*Sa*Cas9-D10A constructs. **B** – Sanger sequencing chromatograms confirming the introduction of correct codon mutations at the indicated positions within the *Sa*Cas9 gene and integrity of immediate flanking sequences (Asp – aspartic acid; Arg – arginine; Gln – glutamine; Gly – glycine and Ala – alanine).

Sanger sequencing confirmed that the codons at position 499, 500, 654 and 655 within the *Sa*Cas9 gene were all successfully converted to the alanine-encoding GCC tri-nucleotide. Following successful generation of the e*Sa*Cas9/-D10A constructs, the cleavage activity of the native and enhanced systems was measured for comparison before functional analysis of *CCR5*-mediated gene editing was carried out.

### 3.3.2. Assessment of cleavage activity of the original and 'enhanced specificity' *Sa*Cas9 nuclease and nickase systems

Based on results obtained by Slaymaker *et al*. (2016), it was expected that the efficiencies of the original and enhanced *Sa*Cas9 systems would not be significantly different. In order to verify this, and determine whether the same phenomenon applies to the nickase systems, TZM-bl cells were treated with the original or 'enhanced specificity' *Sa*Cas9 nuclease or nickase constructs and U6-sgRNAs 7, 8 or 7/8. T7EI and ddPCR™ drop-off assays were carried out to visualise and quantify indels introduced by the different *Sa*Cas9 systems into the *CCR5* target site to assess indel forming potential and compare cleavage efficiencies across the various systems (Figure 3.10).

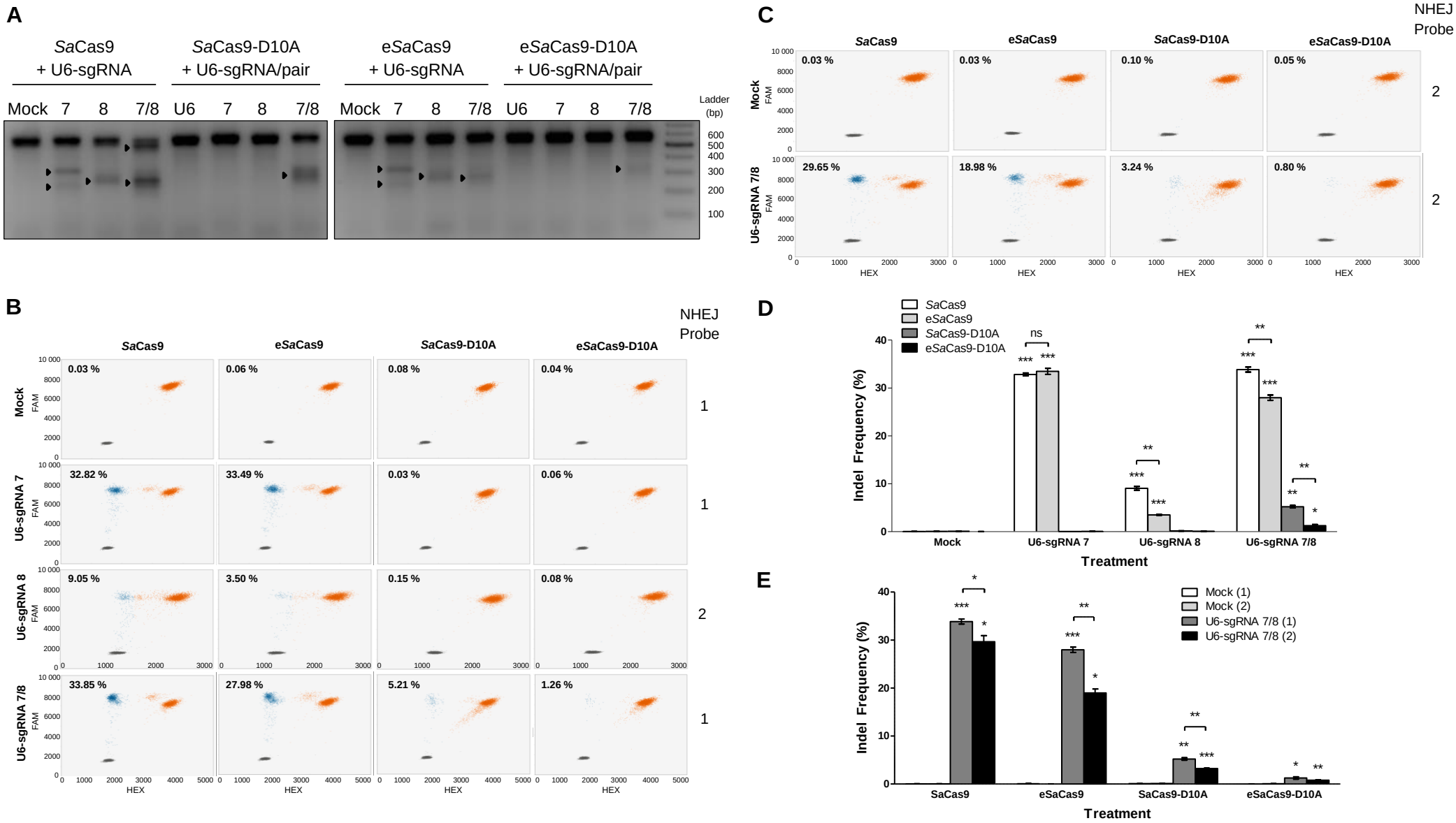

A
SaCas9 + U6-sgRNA
SaCas9-D10A + U6-sgRNA/pair
eSaCas9 + U6-sgRNA
eSaCas9-D10A + U6-sgRNA/pair
Mock 7 8 7/8 U6 7 8 7/8 Mock 7 8 7/8 U6 7 8 7/8
Ladder (bp)
600
500
400
300
200
100
B
NHEJ Probe
SaCas9
eSaCas9
SaCas9-D10A
eSaCas9-D10A
Mock
U6-sgRNA 7
U6-sgRNA 8
U6-sgRNA 7/8
FAM
HEX
0.03 %
0.06 %
0.08 %
0.04 %
32.82 %
33.49 %
0.03 %
0.06 %
9.05 %
3.50 %
0.15 %
0.08 %
33.85 %
27.98 %
5.21 %
1.26 %
1
1
2
1
C
NHEJ Probe
SaCas9
eSaCas9
SaCas9-D10A
eSaCas9-D10A
Mock
U6-sgRNA 7/8
FAM
HEX
0.03 %
0.03 %
0.10 %
0.05 %
29.65 %
18.98 %
3.24 %
0.80 %
2
2
D
SaCas9
eSaCas9
SaCas9-D10A
eSaCas9-D10A
ns
Indel Frequency (%)
Mock
U6-sgRNA 7
U6-sgRNA 8
U6-sgRNA 7/8
Treatment
E
Mock (1)
Mock (2)
U6-sgRNA 7/8 (1)
U6-sgRNA 7/8 (2)
Indel Frequency (%)
SaCas9
eSaCas9
SaCas9-D10A
eSaCas9-D10A
Treatment

**Figure 3.10: Visualisation and quantification of CRISPR/*Sa*Cas9-mediated *CCR5* target site indel formation by T7EI and ddPCR™ PCR drop-off assays. A** – Resolved products from T7EI assays run on 2 % agarose gels alongside a molecular weight ladder for analysis of indels resulting from TZM-bl cell treatment with *Sa*Cas9/-D10A or e*Sa*Cas9/-D10A + U6-sgRNA 7, 8 or 7/8. The gDNA was extracted 48 hours following transfection using the Qiamp DNA Mini Kit and a T7EI assay carried out to visualise target site indel formation. Black arrows point to the digestion products, indicative of indel formation. Mock negative controls (the absence of a targeting gRNA sequence) were also included in the experiments and the observed indels on the T7EI gels were quantified by the ddPCR™ drop-off assay. **B** and **C** – 2D-Plots of HEX (NHEJ probe 1 or 2) and FAM (reference probe) fluorescence measured by the ddPCR™ drop-off assay. The threshold for positive FAM fluorescence was set at 2 250 using a no gDNA negative control (Supplementary Figure 7.5.2B) and for HEX at 2 500 (NHEJ probe 1) or 1 450 (NHEJ probe 2) using the mock controls. This allowed for clustering of the WT/WT+ droplets (orange), the NHEJ drop-off droplets (blue) and the negative droplets (black). QuantaSoft™ Analysis Pro software was used to calculate the indel frequencies as an average of technical duplicates. **D** and **E** – Bar graphs showing indel frequencies calculated by the ddPCR™ drop-off assay (**B** and **C**). The significance of relative differences in indel frequencies between treatments and the corresponding mock control are shown by asterisk/s above the bars. A two-tailed unpaired t-test with a 95% confidence interval was used to calculate p-values. Error bars show the sample standard deviations. * $p \leq 0.05$; ** $p \leq 0.005$; *** $p \leq 0.0005$.

PCR products resolved on the agarose gels as single products of the expected 620 bp size (Supplementary Figure 7.5.1B) and the mock control samples did not induce target site indel formation as no digestion products were visible on the T7EI gels (Figure 3.10A). Indels were detected in each of the three *Sa*Cas9 and e*Sa*Cas9 samples but only for the *Sa*Cas9-D10A and e*Sa*Cas9-D10A when complexed with the U6-sgRNA 7/8 pair and not with individual sgRNAs, as expected. Quantification of these indels was carried out by the ddPCR™ drop-off assay as this was previously found to be more sensitive and applicable to both nuclease- and nickase-treated samples as opposed to densitometry and TIDE online analysis (Sections 3.1.3. and 3.2.1.). The 2D fluorescence amplitude plots provided by the QuantaSoft™ Analysis Pro software show HEX and FAM fluorescence patterns using NHEJ probe 1 for the mock, U6-sgRNA 7 and U6-sgRNA 7/8 treatments (Figure 3.10B) and NHEJ probe 2 for U6-sgRNA 8 (Figure 3.10B) and U6-sgRNA 7/8 treatments (Figure 3.10C). The background levels of indels determined by mock control treatment ranged from 0.03 – 0.10 % (Figure 3.10B and C). These indels were then used to determine whether or not the CRISPR/*Sa*Cas9 treatments resulted in significant target site modification as an average of technical duplicates (Table 3.4; Supplementary Table 7.5.2).

**Table 3.4: ddPCR™ drop-off assay results obtained from WT and 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase treatment of TZM-bl cells**

| Treatment | NHEJ Probe | Average Indel Frequency (%) | Standard Deviation | p-value |
|---|---|---|---|---|
| ***Sa*Cas9 Mock Control** | 1 | 0.03 | 0.04 | - |
| | 2 | 0.03 | 0.04 | - |
| ***Sa*Cas9 + U6-sgRNA 7** | 1 | 32.82 | 0.31 | <0.0001 |
| ***Sa*Cas9 + U6-sgRNA 8** | 2 | 9.05 | 0.39 | 0.0009 |
| ***Sa*Cas9 + U6-sgRNA 7/8** | 1 | 33.85 | 0.52 | 0.0001 |
| | 2 | 29.65 | 1.26 | **0.0191** |
| **e*Sa*Cas9 Mock Control** | 1 | 0.06 | 0.04 | - |
| | 2 | 0.01 | 0.02 | - |
| **e*Sa*Cas9 + U6-sgRNA 7** | 1 | 33.49 | 0.63 | 0.0002 |
| **e*Sa*Cas9 + U6-sgRNA 8** | 2 | 3.50 | 0.10 | 0.0004 |
| **e*Sa*Cas9 + U6-sgRNA 7/8** | 1 | 27.98 | 0.56 | 0.0002 |
| | 2 | 18.98 | 0.82 | **0.0196** |
| ***Sa*Cas9-D10A Mock Control** | 1 | 0.08 | 0.02 | - |
| | 2 | 0.10 | 0.01 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 7** | 1 | 0.03 | 0.00 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 8** | 2 | 0.15 | 0.04 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | 1 | 5.21 | 0.26 | 0.0013 |
| | 2 | 3.24 | 0.06 | 0.0002 |
| **e*Sa*Cas9-D10A Mock Control** | 1 | 0.02 | 0.03 | - |
| | 2 | 0.05 | 0.02 | - |
| **e*Sa*Cas9-D10A + U6-sgRNA 7** | 1 | 0.06 | 0.03 | - |
| **e*Sa*Cas9-D10A + U6-sgRNA 8** | 2 | 0.08 | 0.07 | - |
| **e*Sa*Cas9-D10A + U6-sgRNA 7/8** | 1 | 1.26 | 0.24 | 0.0185 |
| | 2 | 0.80 | 0.04 | 0.0023 |

Note: a Welch's correction for unequal variance determined by the F test was applied to obtain p-values in bold text.

The individual activity of U6-sgRNAs 7 and 8 was assessed with the *Sa*Cas9 nickase constructs to verify that single nicks introduced at the target site were corrected by high fidelity DNA repair pathways. This was observed through the lack of T7EI gel digestion products (Figure 3.10A) as well as no significant increase in the background indel level with the ddPCR™ drop-off assay (Figure 3.10B). *Sa*Cas9 and e*Sa*Cas9 + U6-sgRNA 7 induced formation of 32.82 and 33.49 % indels within the *CCR5* target site respectively, with no significant difference between the two treatments (Figure 3.10D, $p = 0.3134$), in agreement with the findings of comparable cleavage activity between WT and e*Sa*Cas9 as previously reported (Slaymaker et al., 2016). This was not true for treatments with *Sa*Cas9/e*Sa*Cas9 + U6-sgRNA 8 or 7/8. Indel frequency was reduced by 2.6-fold for the former (from 9.05 to 3.5 %, $p = 0.0026$) and 1.2-fold for the latter (from 33.85 to 27.98 %, $p = 0.0084$). When U6-

sgRNA 7/8 treatment was combined with the *Sa*Cas9-D10A and the e*Sa*Cas9-D10A constructs, there was just over a 4-fold decrease in efficiency (5.21 to 1.26 %, p = 0.0005) and thus these two observations indicate that the efficiency of the WT and enhanced nuclease systems could be sgRNA sequence- and Cas9-context dependent.

When comparing the *Sa*Cas9 nuclease and nickase treatments, there is a 6.5-fold decrease in indel frequency from the U6-sgRNA 7 treatment (33.85 to 5.21%, p = 0.0001) and a 1.7-fold reduction with U6-sgRNA 8 (9.05 to 5.21 %, p = 0.0010). These results are similar to those observed by deep sequencing, however, the overall indel frequencies are lower than those detected by ddPCR™ in the previous experiment (Figure 3.6). The use of NHEJ probes 1 and 2 individually for U6-sgRNA 7/8 treatment indel quantification showed that modifications occurred at both the U6-sgRNA 7 as well as 8 cut sites (Figure 3.10B and C). The frequencies determined by each of the NHEJ probes for the U6-sgRNA 7/8 treatments were interleaved in a single graph (Figure 3.10E) showing that the indels at each of the sgRNA 8 target sites (NHEJ probe 2) was reduced compared to the sgRNA 7 site (NHEJ probe 1). As it was determined in the deep sequencing experiment that few sequences appear to experience modification at both the sgRNA 7 and 8 target sites with *Sa*Cas9-D10A and dual sgRNA targeting (Figure 3.7C), the combined nickase indel frequencies for the *Sa*Cas9-D10A and e*Sa*Cas9-D10A systems should be at most 8.45 and 2.06 % respectively. There is a 4-fold decrease in cleavage activity when 'enhanced specificity' modifications are applied to the *Sa*Cas9-D10A at the selected dual sgRNA target site.

From these results it has emerged that the WT *Sa*Cas9 system functioned with similar efficiency to the 'enhanced specificity' system at the sgRNA 7 target site but exceeded it at that of sgRNA 8. Furthermore, there was a substantial reduction in the activity of the *Sa*Cas9 nickase compared to the nuclease, as well as with the further enhanced e*Sa*Cas9 nickase compared to the original nickase. The maximal ~ 2 % indel frequency introduced by the e*Sa*Cas9-D10A and dual sgRNA 7/8 system is believed to be too low to exert an observable functionally disruptive effect on the *CCR5* gene and downstream processes. This construct was thus excluded from further experimentation and the effect of *CCR5* gene editing on mRNA production and functional cell surface expression, in the context of R5-tropic HIV-1 infection, was assessed using sgRNAs 7 and 8 with the *Sa*Cas9, e*Sa*Cas9 and *Sa*Cas9-D10A systems only.

### 3.4. Analysis of the effect of *CCR5*-specific gene editing on mRNA production and R5-tropic HIV-1 infectivity

It was demonstrated by the ddPCR™ drop-off assay that the WT and 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase systems introduced varying levels of indels in the *CCR5* target site (Figure 3.10). The next goal was to determine whether the systems with the highest cleavage capacities could exert a negative effect on downstream processes, including mRNA production and cell surface protein expression and functioning. These experiments were carried out in order to realise the major objective of the study – to determine whether *CCR5*-specific gene editing precludes R5-tropic HIV-1 infection. A positive control was included in these functional experiments in the form of a previously characterised, non-cytotoxic *CCR5*-specific short hairpin RNA (shRNA 1005) (Shimizu et al., 2009).

#### 3.4.1. Production of shRNA 1005 as a positive control for analysis of the functional effects of *CCR5*-specific CRISPR/*Sa*Cas9 gene editing

shRNA 1005 targets the 3′ end of the *CCR5* mRNA, within the same region as was selected for CRISPR/*Sa*Cas9-mediated gene editing. The H1-promoter driven shRNA 1005 was cloned with flanking *XhoI* and *AscI* restriction sites into the MCS of the pTZ57R/T vector and Sanger sequencing (Inqaba Biotech, South Africa) carried out to confirm successful pTZ57R/T-H1_sh1005 construct generation (Figure 3.11).

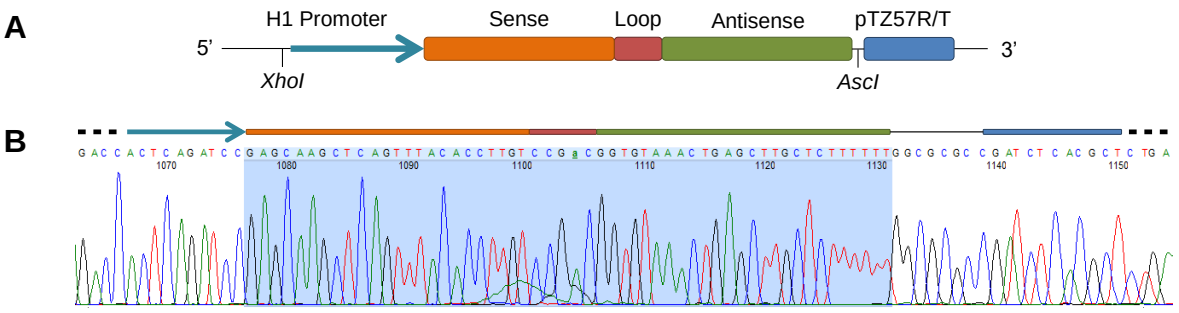



**Figure 3.11: Sequence confirmation of *CCR5*-specific shRNA 1005 cloned into the pTZ57R/T vector. A** – A schematic of the pTZ57R/T-H1_sh1005 vector to illustrate the position of the H1 promoter, sense/loop/antisense sequences of shRNA 1005 and *XhoI*/*AscI* restriction sites corresponding to the chromatogram below. **B** – Sanger sequencing chromatogram showing the successful cloning of shRNA 1005 into the expression vector.

A non-specific (NS) shRNA also expressed from a pTZ57R/T vector backbone was included as a negative control to validate that effects observed by the use of shRNA 1005 were

specific to *CCR5* mRNA knock down and not as a result of cellular toxicity or broad spectrum shRNA activity. TZM-bl cells were then treated with the selected CRISPR/*Sa*Cas9 systems and two shRNA controls to determine the effect of *CCR5*-specific gene editing on mRNA expression and R5-tropic HIV-1 infectivity.

### 3.3.2. Functional analysis of CRISPR/*Sa*Cas9-mediated *CCR5* gene editing by qRT-PCR and TZM-bl luciferase assay

In order to assess whether the indels introduced into the *CCR5* target site by the *Sa*Cas9 and e*Sa*Cas9 + U6-sgRNA 7, 8 and 7/8 as well as *Sa*Cas9-D10A + U6-sgRNA 7/8 treatments resulted in reduction in mRNA expression, *CCR5*-specific qRT-PCR was carried out. The human *β-actin* gene was used as an endogenous reference control as this gene product plays an important role in maintenance of cellular integrity and was expected to be constitutively and stably expressed across all treated cells. Two qRT-PCR experiments were performed in TZM-bl cells in order to determine the level of *CCR5* mRNA knock down induced firstly by the *CCR5*-specific shRNA positive control compared to a NS shRNA and secondly by CRISPR/*Sa*Cas9 treatment compared to the shRNA controls (Figure 3.12).

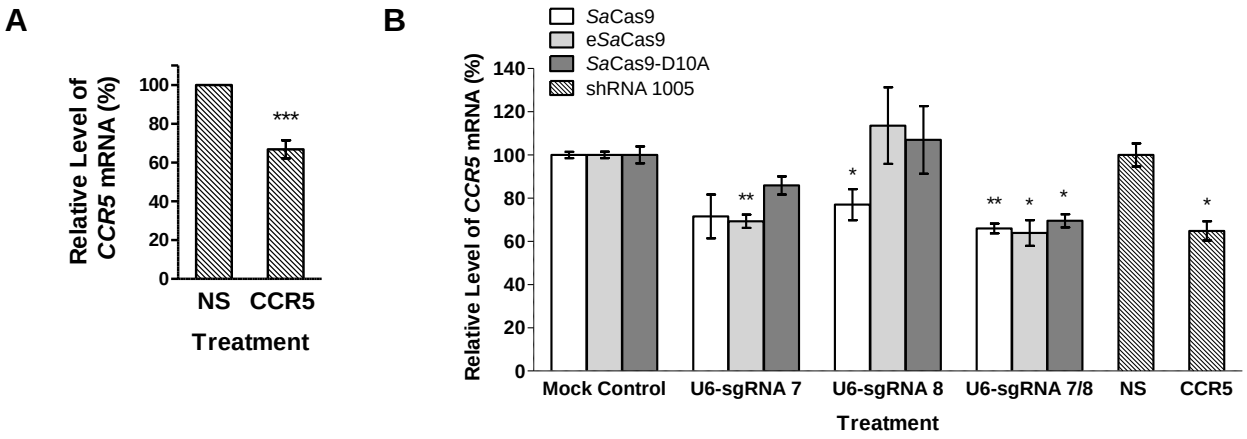



**Figure 3.12: Measurement of *CCR5* mRNA levels by qRT-PCR following TZM-bl treatment with shRNA 1005 and CRISPR/*Sa*Cas9 systems.** TZM-bl cells were transfected with a non-specific (NS) shRNA and *CCR5*-specific shRNA 1005 only (**A**) or in addition to *Sa*Cas9, e*Sa*Cas9 and *Sa*Cas9-D10A with U6-sgRNAs 7, 8 or 7/8TZM-bl cells (**B**). Total cellular RNA was extracted 48 hours following transfection and *CCR5* mRNA-specific qRT-PCR, normalised to *β-actin*, carried out. **A** – *CCR5* mRNA levels resulting from TZM-bl treatment with shRNA 1005 adjusted relative to the non-specific (NS) shRNA control as an average of biological triplicate readings. **B** – *CCR5* mRNA levels resulting from TZM-bl treatment with CRISPR/*Sa*Cas9 systems as described above, as well as with the NS and *CCR5*-specific shRNAs adjusted relative to corresponding mock controls (treatment with the *Sa*Cas9 variant and an empty U6-sgRNA vector) as an average of technical duplicates. The significance of relative *CCR5* mRNA level differences is shown by asterisk/s above the bars. A two-tailed

unpaired t-test with a 95% confidence interval was used to calculate p-values. Error bars show the sample standard deviations. * $p \leq 0.05$; ** $p \leq 0.005$; *** $p \leq 0.0005$.

*CCR5* mRNA knock down by shRNA 1005 (Figure 3.12A) was conducted for inclusion as a positive control in the CRISPR/*Sa*Cas9 treatment experiment (Figure 3.12B) as well as for the HIV-1 infectivity assay described later. The first qRT-PCR experiment was carried out on three biological replicates of RNA extracted from shRNA 1005 and non-specific (NS) shRNA treated TZM-bl cells to determine the level of *CCR5* mRNA knock down induced. The threshold (Ct) values of *CCR5* mRNA for each were normalised to those of *β-actin* using the LightCycler® 96 software (Supplementary Table 7.5.3) and the average percentage knock down, as well as the significance, calculated using a two-tailed unpaired t-test with a 95% confidence interval (Table 3.5).

**Table 3.5: qRT-PCR results obtained from *CCR5*-specific and non-specific (NS) shRNA treatment in TZM-bl cells**

| Treatment | Relative level of *CCR5* mRNA (%) | Standard Deviation | p-value |
|---|---|---|---|
| **NS shRNA** | 100.00 | - | - |
| ***CCR5* shRNA** | 66.85 | 4.65 | 0.0002 |

shRNA 1005 treatment resulted in a 33.15 % reduction in *CCR5* mRNA, calculated as a relative percentage of the NS shRNA negative control. This treatment was then included as a positive control in the following qRT-PCR experiment in order to validate results obtained from CRISPR/*Sa*Cas9 treatment in TZM-bl cells (Figure 3.12B; Table 3.6 and Supplementary Table 7.5.4).

**Table 3.6: qRT-PCR results obtained from CRISPR/*Sa*Cas9 treatment in TZM-bl cells**

| Treatment | Relative level of *CCR5* mRNA (%) | Standard Deviation | p-value |
|---|---|---|---|
| ***Sa*Cas9 Mock Control** | 100.00 | 1.462 | - |
| ***Sa*Cas9 + U6-sgRNA 7** | 71.55 | 10.138 | 0.0581 |
| ***Sa*Cas9 + U6-sgRNA 8** | 77.00 | 7.176 | 0.0475 |
| ***Sa*Cas9 + U6-sgRNA 7/8** | 65.98 | 2.259 | 0.0030 |
| **e*Sa*Cas9 Mock Control** | 100.00 | 1.478 | - |
| **e*Sa*Cas9 + U6-sgRNA 7** | 69.28 | 3.060 | 0.0063 |
| **e*Sa*Cas9 + U6-sgRNA 8** | 113.59 | 17.739 | 0.3916 |
| **e*Sa*Cas9 + U6-sgRNA 7/8** | 63.87 | 5.939 | 0.0142 |
| ***Sa*Cas9-D10A Mock Control** | 100.00 | 3.900 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 7** | 85.87 | 4.219 | 0.0747 |
| ***Sa*Cas9-D10A + U6-sgRNA 8** | 106.97 | 15.668 | 0.6144 |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | 69.51 | 3.056 | 0.0125 |
| **NS shRNA** | 100.00 | 5.370 | - |
| ***CCR5* shRNA** | 64.85 | 4.447 | 0.0191 |

The resulting knock down of *CCR5* mRNA from shRNA 1005 treatment was 35.15 %, similar to the 33.15 % reduction obtained in the previous experiment. The two negative controls included in this experiment were a no reverse transcriptase (no RT) and no template control, verifying that neither gDNA nor RNA contamination occurred respectively (Supplementary Table 7.5.4). Based on the cleavage potential of the *Sa*Cas9 and e*Sa*Cas9 + U6-sgRNA 7 and 7/8 treatments, determined by ddPCR™ to range from 27.98 – 33.85 % (Figure 3.10B), it was expected that these too would exert a significant negative effect on *CCR5* mRNA production. The observed relative reduction in mRNA was found to be 28.45 – 36.13 % for these treatments, affirming this hypothesis, however the 28.45 % reduction in mRNA was found not to be significant for the *Sa*Cas9 + U6-sgRNA 7 treatment at a 95 % CI ($p = 0.0581$). At face value, the reduction does fit within the expected trend and a large standard deviation for the test sample is most likely the cause for reduced confidence in the result observed.

As expected, no significant difference in *CCR5* mRNA levels was observed between the negative controls and nickase treatments with U6-sgRNAs 7 or 8 individually. When cells were treated with *Sa*Cas9-D10A + U6sgRNA 7/8, there was ~ 30 % reduction in *CCR5* mRNA ($p = 0.0125$) but based on the ~ 5.2 – 8.5 % (Figure 3.10B and C) or ~ 7 – 12 % (Figure 3.6A and B) indel frequencies measured by ddPCR™, the effective reduction would

be expected to be far lower. Treatment with e*Sa*Cas9 + U6sgRNA 8 showed no decrease in *CCR5* mRNA levels, which was expected with a measured indel frequency of 3.50 % (Figure 3.10B). *Sa*Cas9 + U6-sgRNA 8 resulted in a 23 % reduction and while this was not in accordance with the 9.05 % indels detected previously by ddPCR™ (Figure 3.10B), the deep sequencing results revealed that the treatment is able to introduce up to ~ 19 % indels in the target site (Figure 3.6A) which could result in a larger mRNA reduction.

This experiment highlights that *CCR5* mRNA can be reduced by CRISPR/Cas9 treatment, albeit with generally higher efficiencies than would be expected based on indel quantification. Crucially, the treatments that were anticipated not to cause any or, at most, a small reduction in *CCR5* mRNA expression were found to uphold this expectation. Furthermore, the shRNA 1005 positive control induced the expected level of *CCR5* mRNA reduction, further validating the CRISPR/Cas9-mediated knock down results obtained. This experiment was conducted in parallel with the HIV infectivity assay in order to assess the functional implications of *CCR5* gene editing in TZM-bl cells. A TZM-bl luciferase assay was carried out to illustrate whether the CRISPR/*Sa*Cas9-mediated indel formation in *CCR5*, and consequent reduction in mRNA expression, had the capacity also to exert a negative effect on CCR5 cell surface co-receptor function and preclude R5-tropic HIV-1 pseudovirus infection.

### 3.3.3. Analysis of the effect of *CCR5*-specific gene editing on R5-tropic HIV-1 infectivity

It was determined that certain CRISPR/*Sa*Cas9 system combinations could inflict both genetic modification as well as a reduction in transcription of the *CCR5* gene. The ultimate goal is for this to impede R5-tropic HIV-1 infectivity of treated cells and thus a TZM-bl luciferase assay was carried out to assess whether this effect was attainable (Montefiori, 2009). TZM-bl cells are susceptible to HIV-1 infection owing to the high level of *CCR5* and *CD4* expression and are an attractive research model in this study as they harbour an HIV-1 Tat-inducible firefly luciferase gene driven by the HIV-1 LTR promoter. TZM-bl cells were transfected with identical combinations of CRISPR/*Sa*Cas9 systems as were used to measure *CCR5* mRNA knock down (Figure 3.12), following assessment of cleavage efficiency (Figure 3.10). Cells were then infected with the R5-tropic HIV-1 Subtype C ZM53 pseuodvirus and the level of luciferase expression measured as an indicator of infectivity. The shRNA 1005 positive control was included in this experiment as well as the three different mock controls (Figure 3.13).

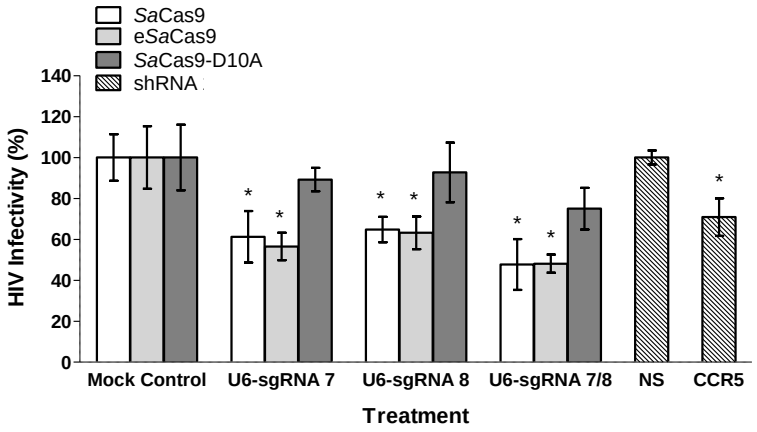



**Figure 3.13: Measurement of R5-tropic HIV-1 pseudovirus infectivity by TZM-bl luciferase assay.** TZM-bl cells were transfected with a non-specific (NS) shRNA and *CCR5*-specific shRNA 1005 as well as with *Sa*Cas9, e*Sa*Cas9 and *Sa*Cas9-D10A with U6-sgRNAs 7, 8 or 7/8. TZM-bl cells were infected with the R5-tropic HIV-1 ZM53 pseudovirus ($TCID_{50}$ = 4000) 48 hours following transfection and a further 48 hours later luciferase activity was measured. HIV-1 pseudovirus infectivity was adjusted relative to the corresponding mock control treatments. Values shown were obtained from averaged biological triplicate readings. The significance of changes in relative infectivity levels is indicated by asterisks placed above each bar. A two-tailed unpaired t-test with a 95% confidence interval was used to calculate p-values. Error bars show the sample standard deviations. * $p \leq 0.05$; ** $p \leq 0.005$; *** $p \leq 0.0005$.

Luciferase readings were taken across biological triplicate transfection wells and averaged to determine the level of expression and, by deduction, infection (Table 3.7 and Supplementary Table 7.5.5).

**Table 3.7: HIV-1 infectivity assay results obtained from CRISPR/*Sa*Cas9 treatment and ZM53 infection in TZM-bl cells**

| Treatment | Relative Level of HIV-1 Infectivity (%) | Standard Deviation | p-value |
|---|---|---|---|
| ***Sa*Cas9 Mock Control** | 100 | 13.99 | - |
| ***Sa*Cas9 + U6-sgRNA 7** | 61.18 | 15.42 | 0.0321 |
| ***Sa*Cas9 + U6-sgRNA 8** | 64.81 | 7.56 | 0.0188 |
| ***Sa*Cas9 + U6-sgRNA 7/8** | 47.68 | 15.20 | 0.0116 |
| **e*Sa*Cas9 Mock Control** | 100 | 18.75 | - |
| **e*Sa*Cas9 + U6-sgRNA 7** | 56.48 | 8.23 | 0.0212 |
| **e*Sa*Cas9 + U6-sgRNA 8** | 63.22 | 9.75 | 0.039 |
| **e*Sa*Cas9 + U6-sgRNA 7/8** | 48.12 | 5.41 | 0.01 |
| ***Sa*Cas9-D10A Mock Control** | 100 | 19.60 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 7** | 89.23 | 6.99 | 0.4176 |
| ***Sa*Cas9-D10A + U6-sgRNA 8** | 92.69 | 17.90 | 0.6678 |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | 75 | 12.50 | 0.137 |
| **NS shRNA** | 100 | 4.17 | - |
| ***CCR5* shRNA** | 70.89 | 11.26 | 0.0137 |

Luminescence readings obtained were adjusted relative to the corresponding mock control treatment to show the level of infectivity of the pseudovirus following CRISPR/*Sa*Cas9 treatment of TZM-bl cells. The positive shRNA 1005 control incurred ~ 30 % reduction in infectivity which was slightly lower than would be predicted based on the level of *CCR5* mRNA knock down, determined to be 33 – 35 % (Figure 3.12A and B). A significant 38.82/43.52 %, 35.19/36.78 % and 52.32/51.88 % reduction in infectivity of the pseudovirus was observed for *Sa*Cas9/e*Sa*Cas9 + U6-sgRNA 7, 8 and 7/8 treatments respectively (Figure 3.13). These values are higher than would be expected from the observed indel frequency and mRNA knock down trends. In the case of *Sa*Cas9/e*Sa*Cas9 + U6-sgRNA 7 treatment, there was only a marginal difference in expectation but this was slightly more pronounced for U6-sgRNA 7/8. Treatments with U6-sgRNA 8 showed a substantial inhibitory effect at ~ 35 % where the level of indels were below 10 % for both *Sa*Cas9 and e*Sa*Cas9, and where mRNA knock down for e*Sa*Cas9 + U6sgRNA 8 was not observed. The *Sa*Cas9-D10A treatments with each sgRNA individually showed no inhibition of HIV-1 pseudovirus infection, as was expected, and treatment with the U6-sgRNA 7/8 pair resulted in a non-significant reduction of 25 %. This reduction at face-value does mirror the trend of results obtained from mRNA knock down.

While the level of R5-tropic HIV-1 pseudovirus entry inhibition measured following the treatment of TZM-bl cells with the various *CCR5*-specific CRISPR/*Sa*Cas9 systems was, in general, higher than expected, the overall trends observed do appear to mirror those obtained from the mRNA knock down studies. Furthermore, the shRNA positive control results provide validity to those obtained in the experiments performed to assess the functional impact of *CCR5* gene editing on mRNA production and R5-tropic HIV-1 infectivity. Overall, the experiments carried out in this study reveal that targeting the human *CCR5* gene with the original as well as 'enhanced specificity' *Sa*Cas9 nuclease and nickase systems has the propensity to induce indels with frequencies ranging from < 5 – 39 % as well as induce mRNA knock down and preclude infection of R5-tropic HIV-1 pseudovirus by up to 36 and 52 % respectively. Importantly, the CRISPR/Cas9 systems utilised in this study are highly specific as well as AAV-deliverable which is desirable for future experimentation into the application of these systems as anti-HIV gene therapies in the quest to uncovering a functional cure for this devastating and unrelenting global disease.

## 4. Discussion

This study employed four variants of the novel and powerful CRISPR/Cas9 gene editing tool with various specificity enhancing characteristics for the targeted disruption of the human *CCR5* gene. The overarching research aim was to provide insight into the functioning of highly specific CRISPR/Cas9 gene editing systems and the development of safe and effective gene therapies, particularly for the functional cure of HIV-1. An effective gene therapy should be deliverable to target cells, efficient, safe and exert the desired functional effect. The CRISPR/Cas9 gene editing system derived from *Staphylococcus aureus* was selected for AAV vector deliverability, robust cleavage activity and a high level of on target specificity. Owing to significant sequence similarity of *CCR5* with other chemokine receptors, particularly *CCR2*, the nickase and 'enhanced specificity' CRISPR/*Sa*Cas9 variants were generated and employed for improved safety over the WT system.

This study revealed a number of important findings relating to the functioning of highly efficient CRISPR/Cas9 systems as well as the applicability of these for an effective gene therapy-based HIV-1 functional cure. The CRISPR/*Sa*Cas9 systems tested functioned with variable cleavage efficiencies depending on the sgRNA targeting sequence and Cas9 variant employed. Cleavage activity decreased with the inclusion of further specificity enhancing mutations to the point where the 'enhanced specificity' nickase was ineffective at the selected *CCR5* target site, suggestive of a trade-off between specificity improvement and modification efficiency. Importantly, while the *Sa*Cas9 nickase functioned with significantly reduced efficiency compared to the nuclease, specificity was improved as demonstrated at one selected off target site within the *CCR2* gene containing a single PAM-distal end nucleotide mismatch. The WT nuclease and nickase as well as the 'enhanced specificity' nuclease systems all successfully introduced indels into the *CCR5* target site, reduced mRNA expression and precluded R5-tropic HIV-1 pseudovirus infectivity in cell culture. Furthermore, the protein structure inferred from deep sequencing of the major mutations introduced by the WT CRISPR/Cas9 systems illustrated that a Δ32-like phenotype is achievable through NHEJ-mediated gene editing of the *CCR5* gene. These findings enable a direct path to developing a safe and effective anti-HIV-1 gene therapy by CRISPR/Cas9-mediated *CCR5* knock out using smaller vehicle delivery technologies such as AAVs.

### 4.1. Rational target design for effective CRISPR/*Sa*Cas9 nuclease functioning

A panel of ten CRISPR/*Sa*Cas9 sgRNAs were targeted to a small region of the human *CCR5* ORF adjacent to the CCR5-Δ32 locus with the intent to mimic the end-point disruptive phenotype of this naturally occurring mutation. The cleavage activity of the various *Sa*Cas9 systems was inconsistent across the selected target sites, with indel frequencies ranging from non-functional to ~ 24.2 % as determined by TIDE online sequencing analysis. Friedland *et al*. (2015) also employed CRISPR/*Sa*Cas9 for targeted disruption of the human *CCR5* gene, instead at intronic sites upstream of the region targeted in this study. The authors described a large variation in indel frequencies at the thirteen target sites, ranging from ~ 11.8 – 46.4 %, indicating that this is a common characteristic of CRISPR/*Sa*Cas9-mediated gene editing. The sgRNA component of a CRISPR/Cas9 system performs two pivotal tasks – locating the target site and activating Cas9 endonuclease activity – thus it was inferred that the difference in sequence composition likely contributes to variations in cleavage activity (Liu et al., 2016a). *In silico* prediction of potentially highly functioning sgRNAs using online tools has enabled reduced time and cost associated with pre-screening of large sgRNA panels prior to empirical analysis. While this step is ideal in the path to designing an effective gene therapy, a fully reliable model is not yet available, particularly in the context of novel and less commonly employed CRISPR/Cas9 systems such as those utilised in this study.

#### 4.1.1. *In silico* prediction of sgRNA cleavage activity

Since the emergence of CRISPR/Cas9 as a powerful and versatile gene editing tool, a major collaborative effort has been made to identify the contributing factors of sgRNA cleavage activity and using these to develop accurate, broadly applicable models for *in silico* prediction (Hsu et al., 2013, Doench et al., 2014, Moreno-Mateos et al., 2015). *In silico* prediction tools have the potential to revolutionise rational target design by reducing the cost and time associated with screening for potentially efficient sgRNAs (Reviewed in Lee et al., 2016). The contribution of gRNA sequence and base position to cleavage activity has been studied extensively using *Sp*Cas9 (Hsu et al., 2013, Wang et al., 2014, Kuscu et al., 2014, Xu et al., 2015, Doench et al., 2016). These studies have scrutinised the activity of hundreds or even thousands of sgRNA sequences to identify a nucleotide enrichment profiles upon which parts of the various prediction models discussed earlier have been based. There appears to be no overarching consensus and very limited agreement between the findings from these studies, contributing to inconsistencies in the prediction models. Two studies found the

presence of a cytosine at the cleavage site to correlate with improved cleavage activity (Wang et al., 2014, Xu et al., 2015) and two others have shown that a 3′ PAM proximal guanine is enriched amongst highly efficient sgRNAs (Doench et al., 2016, Xu et al., 2017). An important general finding is that residues within the seed region contribute most to cleavage potential and this tapers towards the 5′ end, but those at the extreme PAM distal end are also important for activity (Prykhozhij et al., 2015, Liu et al., 2016a). As this study investigated the gene editing potential of various novel *Sa*Cas9 systems, the question arose as to whether the available computational prediction models could be used both for the design of non-*Sp*Cas9 system target sites as well as for the accurate prediction of sgRNA cleavage activity.

The two tools capable of projecting sgRNA cleavage efficiency scores at degenerate PAM sites for *Sa*Cas9 are sgRNA Scorer 2.0 and CRISPRseek. The former tool runs according to a putative 'generalised' sgRNA predictive score algorithm, was trained using the combination of *Sp*Cas9 and *St*Cas9 experimental data and was said to be relevant to other CRISPR/Cas9 systems, including *Sa*Cas9 (Chari et al., 2017). The same assumption about broad applicability is true of CRISPRseek but this tool was derived solely from *Sp*Cas9-specific experimental data (Zhu et al., 2014). Prediction scores obtained using these two models did not correlate with experimental measurements of *Sa*Cas9 sgRNA activity in this study or in a meta-analysis performed on the indel frequency data acquired by Friedland *et al.* (2015) for thirteen other *Sa*Cas9 *CCR5*-specific sgRNAs (data not shown). Large numbers of gRNA sequences are required for assessment of correlations between predicted and empirical cleavage activities but ultimately these tools need to be able to predict the functioning of individual sgRNAs with accuracy for this to be of benefit to rational target design in small-scale studies such as this. Ultimately, the sgRNA cleavage activity determinants need to be elucidated in a Cas9-orthologue specific manner and employed to train separate *in silico* prediction models.

### 4.1.2. Key determinants of sgRNA cleavage activity

As prediction tools were found not to be applicable to the *Sa*Cas9-based systems employed in this study, it is of greater value to investigate key individual features that have been determined to influence cleavage activity in order to explain the indel frequency variation amongst the *Sa*Cas9 sgRNAs. The most likely determinant of cleavage activity is the sgRNA and target site sequence composition but the effect of sequence-independent factors within the cellular environment have also been shown to influence this process. Some of the major

findings from CRISPR/Cas9 gene editing mechanistic studies over the past three years will be discussed in the context of features pertinent to the ten *CCR5*-specific sgRNAs used in this study.

- *CRISPR/Cas9 system delivery modalities and experimental conditions*

The delivery modality of sgRNA and Cas9 components is an important early mediator of the end-point gene editing efficiency, however, this is a relatively under-researched area. The *in vitro* sgRNA delivery form, when introduced as a Cas9 ribonucleoprotein (RNP) complex, has the benefit of reducing the time from treatment to genetic targeting and consequently the accumulation of off target editing events (Kim et al., 2014). A potential caveat with this approach is the increased error rate of *in vitro* over *in vivo* transcription (Reviewed in Tycko et al., 2016). Mutations introduced into the sgRNAs can reduce on target efficiency and alter the off target site landscape. Plasmid DNA transfection was used in this study as an alternative sgRNA and Cas9 delivery approach thereby circumventing the effects of erroneous *in vitro* transcription. Conversely, relative sgRNA cleavage activity can be affected by variability in transfection efficiency and nuclear co-delivery of the CRISPR components. AAV-mediated co-delivery of sgRNA and Cas9 components is an elegant solution to overcoming these hurdles (Friedland et al., 2015, Tabebordbar et al., 2016). This approach would be most applicable to post-screening experiments owing to the labour and time associated with viral vector production, purification and quantification. Experimental delivery conditions should be titrated for optimal activity using a small panel of sgRNAs with known cleavage activities prior to study-specific screening in order to optimise cleavage activity.

- *PAM recognition motif*

Two studies have demonstrated that the *Sa*Cas9 protein has the highest affinity for the 5′-NNGRRT PAM sequence but that the presence of the 3′-T is not essential for CRISPR/*Sa*Cas9 activity (Friedland et al., 2015, Ran et al., 2015). Three sgRNAs from this study with relatively poor or no cleavage activity, U6-sgRNAs 4, 5 and 10, also recognise target sites accompanied by the degenerate 5′-NNGRRN PAM sequence. An additional PAM recognition preference was found to be the 5′-NNGRGT sequence over the 5′-NNGRAT (Ran et al., 2015). The former sequence was most abundant at the thirteen sgRNA target sites in the Friedland *et al*. (2015) study and, in the three instances where the latter was used, the sgRNAs performed more poorly. Recent studies have investigated relaxing the PAM recognition requirements for both the *Sp*Cas9 and *Sa*Cas9 orthologues, each time illustrating

that this approach broadens the target range of sgRNAs but the effect on cleavage activity, as well as on target specificity, appears to be unaltered (Kleinstiver et al., 2015b, Kleinstiver et al., 2015a). A potential avenue for future development is therefore the modification of PAM recognition domains within the Cas9 proteins for more stringent recognition, specific to the desired target site in order to enhance sgRNA activity at this locus.

- *Pre-Cas9-bound sgRNA secondary structures*

Once the CRISPR DNA plasmids have been delivered successfully into the target cell nuclei, productive transcription is required following which the sgRNA folds into an ensemble of pre-Cas9 bound structures (Xu et al., 2017). The variable gRNA sequence portion can potentially interact with the invariable scaffold residues resulting in the disruption of key features, such as the stem loops, required for recognition by, and complex formation with, the cognate Cas9 protein. The *Sa*Cas9 scaffold sequence forms a repeat:anti-repeat duplex with an internal loop, a tetraloop and two stem loops, all of which interact specifically with the tertiary *Sa*Cas9 protein structure (Nishimasu et al., 2015). The modified scaffold sequence showed improvement to sgRNA 7 and 8 activity in this study as this improves stability of the sgRNA and recognition of the key structural features by the Cas9 protein (Tabebordbar et al., 2016). An active sgRNA-Cas9 complex is then able to scan the genome for the target site recognition, R-loop formation, strand invasion and cleavage. The pre-Cas9-bound sgRNA structure is therefore pivotal in the overall functioning of the CRISPR/Cas9 system. The most likely secondary structure of each *CCR5*-specific U6-sgRNA was predicted computationally using the RNAstructure tool (Reuter and Mathews, 2010), as well as for the three highest and lowest performing sgRNAs (L12, L11, R20 and L13, R18, R12 respectively) from the Friedland *et al*. (2015) study for comparison (Figure 4.1).

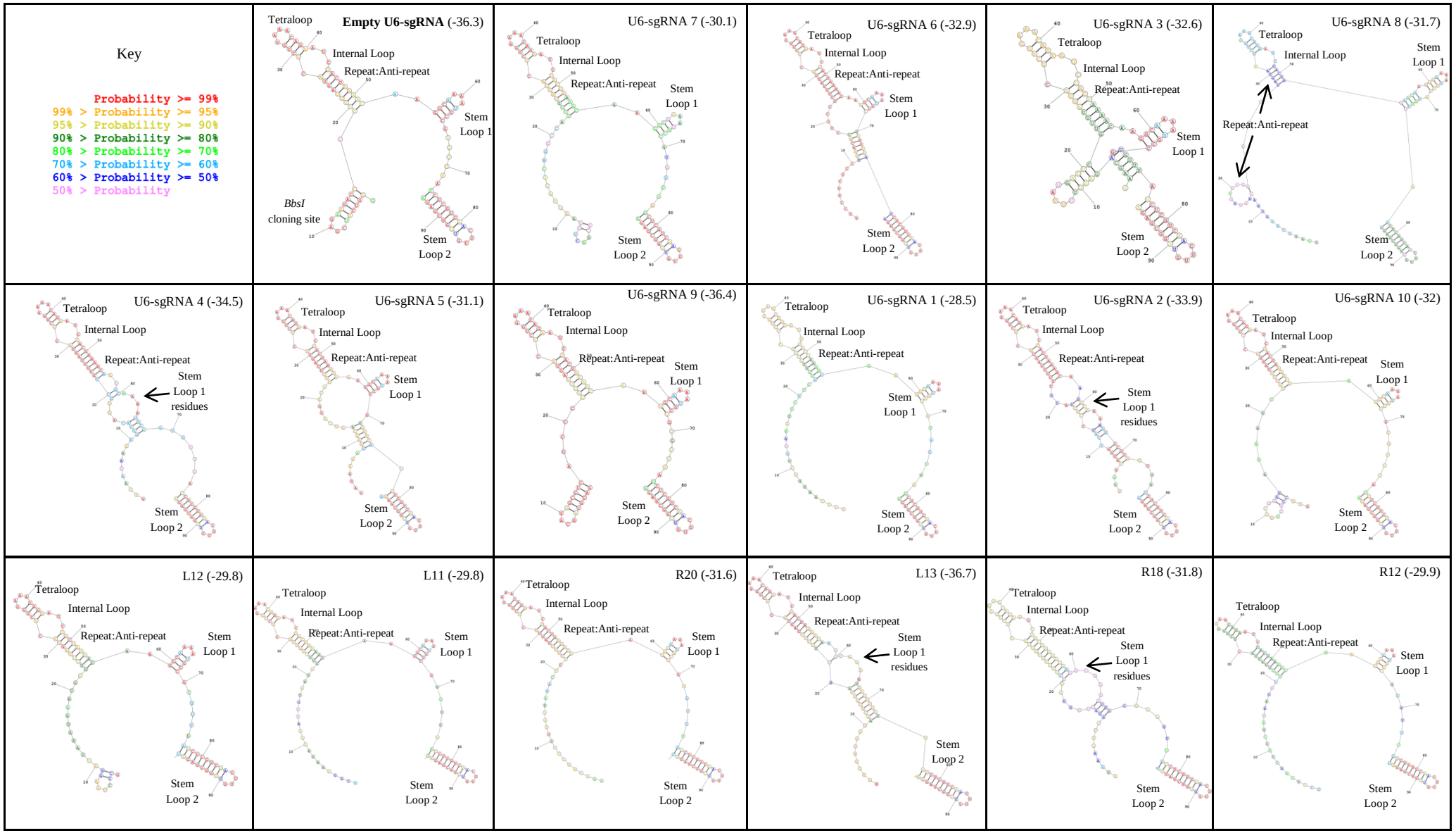

Key
Probability >= 99%
99% > Probability >= 95%
95% > Probability >= 90%
90% > Probability >= 80%
80% > Probability >= 70%
70% > Probability >= 60%
60% > Probability >= 50%
50% > Probability
Tetraloop
Empty U6-sgRNA (-36.3)
Internal Loop
Repeat:Anti-repeat
Stem Loop 1
BbsI cloning site
Stem Loop 2
U6-sgRNA 7 (-30.1)
U6-sgRNA 6 (-32.9)
U6-sgRNA 3 (-32.6)
U6-sgRNA 8 (-31.7)
U6-sgRNA 4 (-34.5)
Stem Loop 1 residues
U6-sgRNA 5 (-31.1)
U6-sgRNA 9 (-36.4)
U6-sgRNA 1 (-28.5)
U6-sgRNA 2 (-33.9)
U6-sgRNA 10 (-32)
L12 (-29.8)
L11 (-29.8)
R20 (-31.6)
L13 (-36.7)
R18 (-31.8)
R12 (-29.9)

**Figure 4.1: Computationally predicted structures of *Sa*Cas9 *CCR5*-specific sgRNAs utilized in this study and by Friedland *et al*. (2015).** The RNAstructure online tool was used to predict the secondary structure of each *CCR5*-specific *Sa*Cas9 sgRNA (U6-sgRNAs 1 – 10) and the three highest and lowest performing sgRNAs (L12, L11, R20 and L13, R18, R12 respectively) from the Friedland *et al*. (2015) study, in descending order of empirically determined cleavage activity. Gibbs free energy values are shown in brackets. The MaxExpect algorithm was applied to determine the most probable base pair interactions and the probabilities of predicted structures is described in the key. The empty sgRNA vector (pTZ57R/T_U6-*BbsI*-sgRNA-*Sa*Cas9_tracrRNA) is included to show the unaffected scaffold structural features. Specifically, the native *Sa*Cas9 sgRNA structure has a 5′ variable gRNA sequence, repeat:anti-repeat duplex with internal loop, tetraloop (extended in the modified version) and stem loops 1 and 2 at the 3′ end.

The secondary structures predicted by the online tool show the presence or absence of the hairpin loop features required for recognition of the sgRNA by the Cas9 protein. The most efficient sgRNAs (U6-sgRNA 7, L12, L11 and R20) have an 'open'-like structure with an exposed gRNA component and distinct hairpin features matching those of the Empty U6-sgRNA. This is expected to both enhance recognition by the Cas9 and allow for efficient binding of the exposed gRNA seed region to the target site for R-loop formation. The less efficient sgRNAs show varied non-canonical hairpin loop structures through the interaction with the gRNA sequence and backbone scaffold. U6-sgRNA 4, L13 and R18 appear not to contain the first stem loop and the repeat:anti-repeat duplex is split between two different hairpins in the U6-sgRNA 8 structure. The loss of these structural features would reduce the efficacy of Cas9 recognition and thus overall cleavage activity, relative to the other sgRNAs. The sgRNAs undoubtedly form alternative structures, some possibly containing the correct structural components, as folding is a dynamic process. Because these are the most probable base pair interactions predicted by the computational algorithm, the majority of species are expected to exist in this form. U6-sgRNAs 9, 1 and 10 as well as R12 have similar 'open' state structures to the most efficient sgRNAs and this highlights that, in general, no one characteristic is expected to be responsible for the efficiency of sgRNA cleavage activity.

The CRISPR/Cas9 system delivery modality, sgRNA structure and target site sequence features, including the PAM motif, each influence the cleavage activity of individual sgRNAs. While some sequence-dependent features are certainly more influential than others, none can be considered in isolation and thus continued improvement of *in silico* design and prediction tools for individual systems is a necessity. In the interim, this study has found the computationally predicted secondary structure of sgRNAs to be particularly informative of the relative potential for Cas9 recognition and complex formation. The more 'open' variable

region structures with native-like backbone features should be selected during the pre-screening process over those with more divergent forms. Some key research areas to be explored further in the context of this study are nucleotide enrichment profiles for *Sa*Cas9 sgRNA sequences and the optimal experimental and cellular conditions for efficient CRISPR/Cas9 activity.

### 4.2. The trade-off between CRISPR/*Sa*Cas9 cleavage efficiency and specificity

As research delves deeper into the immense potential of the CRISPR/Cas9 technology, the expanse of undesired off target effects are becoming increasingly apparent and concerning, particularly for human therapeutic development. The *Sa*Cas9 nickase and 'enhanced specificity' systems were employed in this study to investigate the effects of two different specificity improvement modifications on the efficiency of CRISPR/*Sa*Cas9 system functioning (Ran et al., 2015, Slaymaker et al., 2016). Furthermore, the specificity of the nickase system at selected potential off target sites in the human genome was compared to that of the nuclease. From indel quantification analysis of the four different CRISPR/*Sa*Cas9 systems it emerged that there could exist a significant trade-off between cleavage efficiency and specificity. Compromising on safety is not an option and thus it may be necessary either to uncover ways in which the reduced efficiency effects are mitigated, such as by *ex vivo* gene editing and selection of successfully modified cells for gene therapy, or alternative safety enhancing approaches need to be developed.

#### 4.2.1. Reduced CRISPR/*Sa*Cas9 nickase activity with improved specificity

Thus far, no computational tools have been developed to predict either the activity or specificity of *Sa*Cas9 nickase-dual sgRNA systems. It was therefore essential to test the *CCR5*-specific sgRNA pairs empirically and, following the selection of sgRNA 7/8 as the most effective *Sa*Cas9-D10A nickase pair, the efficiency of cleavage as well as specificity of the system was investigated by targeted amplicon deep sequencing. Results from this method as well as ddPCR™ analysis revealed that the activity of the *Sa*Cas9 nuclease with individual sgRNAs 7 and 8 surpassed that of the dual sgRNA nickase system but these were found not to be predictive of nickase cleavage activity. Likewise, for all *CCR5*-specific *Sa*Cas9 nickase sgRNA combinations investigated by Friedland *et al*. (2015), cleavage efficiency was lower than at least one, if not both, of the individual sgRNAs and the nuclease. This phenomenon has also been demonstrated using Zinc Finger Nickases (Ramirez et al., 2012). Lower relative activity of one of the sgRNAs with the nuclease system could account for low nickase

efficiency even when paired with a highly efficient sgRNA, however, Friedland *et al*. (2015) also demonstrated that selecting two efficient individual *Sa*Cas9 sgRNAs for dual sgRNA nickase targeting does not necessarily result in a high level of indel formation. A major reason for reduced nickase versus nuclease efficiency is possibly the necessity for temporal synchronicity in dual sgRNA binding and nicking in order to generate a DSB and induce NHEJ-mediated DNA repair. Alternatively, competition between the two sgRNAs for recognition by the Cas9 could result in an imbalance of functional complex formation as it has been demonstrated that the formation of cleavage incompetent Cas9:sgRNA structures can interfere with cleavage activity of functional complexes (Thyme et al., 2016).

The types of indels introduced into the target sites as detected by deep sequencing of the *CCR5* target site for the *Sa*Cas9 nuclease were mainly one and two bp deletions at the expected cleavage site with both sgRNA 7 and 8. The *Sa*Cas9 nickase treatment with sgRNA pair 7/8, however, introduced an array of larger indels, each suspected to exist at a frequency below the 0.1 % detection limit resulting in underestimation of true indel frequencies. Recent studies have explored the effect of DNA repair type on the outcome of both nuclease- and nickase-mediated cleavage (Vriend et al., 2016, Bothmer et al., 2017). The initial hypothesis was that a nickase-induced 5′ overhang DSB could stimulate the canonical error-prone NHEJ DNA repair pathway but it has now been suggested that alternative NHEJ and both high fidelity as well as erroneous HDR pathways contribute to indel formation (Friedland et al., 2015, Bothmer et al., 2017). The extent of HDR on nickase-mediated 5′ overhang DSB repair has been overlooked in the past because a 3′ overhang is required for induction of this process. Recently it has been shown that the 5′ overhangs are converted to 3′ intermediates by endo- and exonucleases providing the template necessary for HDR (Vriend et al., 2016). The participation of multiple DNA repair pathways in nickase-induced DSB resolution is likely to cause inconsistent indel formation in the target site, to a greater extent than with nuclease-based systems.

Highly similar sites to each of the individual sgRNAs were selected for off target analysis but proving that CRISPR/Cas9 nickase systems have overall improved specificity over the wild type system is difficult without performing unbiased genome-wide off target analysis. GUIDE-Seq (Tsai et al., 2015) and BLESS (Crosetto et al., 2013) are two such methods that have revolutionised nuclease-specific off target analysis but the nature of the DSB induced by the nickase system is incompatible with these approaches. Genome-wide deep sequencing is thus the only truly unbiased approach currently available for nickase-specific off target

analysis but the cost and time associated with this method was beyond the scope of this study. Targeted amplicon deep sequencing revealed that the WT *Sa*Cas9 nuclease system resulted in a low level of indels at the sgRNA 8-specific *CCR2* off target site while this was overcome by nickase and dual sgRNA treatment. This off target site has an identical PAM motif and only a single mismatch at the PAM-distal end of the variable region where mismatches are tolerated to a certain extent. The nickase system therefore displayed improved specificity over the nuclease system at this particular site but an unbiased genome-wide approach would be beneficial to assess broader specificity profiles. Additionally, it is important to determine whether the putative enhancing mutations introduced into the *Sa*Cas9 construct generating e*Sa*Cas9 do in fact improve overall specificity of the system.

A potential way in which CRISPR/Cas9 nickase activity could be improved is by screening a far greater panel of sgRNA pairs with varying offset distances should be assessed for cleavage activity to select for the most efficient system. Thereafter, AAV-mediated delivery of an all-in-one system will further enhance activity by ensuring delivery of all required components to each targeted cell. Another highly promising approach to circumventing the reduced cleavage activity of CRISPR/Cas9 nickases is Cas9-guided evolution for maintained on target activity with reduced off target mutagenesis.

### 4.2.2. Efficiency of the 'enhanced specificity' *Sa*Cas9 nuclease and nickase systems

A second approach to specificity enhancement is structure-guided evolution of the Cas9 protein to destabilise mismatched off target site interactions (Slaymaker et al., 2016). The e*Sa*Cas9 system was reported to have comparable efficiency, but improved specificity, to the native *Sa*Cas9 nuclease (Slaymaker et al., 2016). Based on the relatively low indel frequency induced by the *Sa*Cas9-D10A nickase, this alternative specificity enhancing approach was taken in an effort to improve the efficiency of highly specific *CCR5* gene editing. Additionally, because no known study has investigated the capacity for a combined 'enhanced specificity' *Sa*Cas9 nickase system to induce target site indels, the e*Sa*Cas9-D10A construct was generated and included in the analysis. ddPCR™ was employed to measure the absolute indel frequencies introduced by the native and 'enhanced specificity' *Sa*Cas9 nuclease and nickase systems. The individual *Sa*Cas9 sgRNA treatments did not both fit with the hypothesis that the efficiency of the *Sa*Cas9 and e*Sa*Cas9 is comparable. The U6-sgRNA 7 treatment did function with similar efficiencies between the two *Sa*Cas9 constructs but U6-sgRNA 8 treatment resulted in a significant ~ 2.6 fold lower activity with e*Sa*Cas9 than the

native system (Figure 3.10). A significant 4.1 fold reduction in efficiency was also observed for the dual U6-sgRNA 7/8 treatment with the WT and e*Sa*Cas9 nickases respectively. The additional specificity enhancing modification appeared to reduce nickase cleavage activity to highly inefficient levels and the nature of the changes to the Cas9 protein could compound the already reduced nickase system activity.

The four specificity enhancing mutations are located within the *Sa*Cas9 non-target strand groove – two within the linker 1 region (L1) and two in the RuvC III domain, both within the *Sa*Cas9 nuclease lobe (Nishimasu et al., 2015). From *Sa*Cas9-sgRNA complex crystal structures it is understood that L1 undergoes a conformational change upon binding of the RuvC nuclease domain to the non-target DNA strand in order to bring the HNH domain into contact with the target DNA strand for cleavage (Nishimasu et al., 2015). The arginine at position 654 (R654) was shown to contact the phosphate group of base position 12 from the PAM proximal end of the target DNA strand for stabilisation. The effect of the 'enhanced specificity' mutations is believed to destabilise strand separation, particularly where mismatches occur between the gRNA and target DNA sequence. With a fully complementary on target site, this was expected to have minimal impact on cleavage efficiency as sgRNA:DNA hybridisation would outcompete DNA:DNA re-hybridisation. These four mutations may thus not significantly affect the activity at thermodynamically favourable sgRNA target sites, such as was observed with sgRNA 7, but cleavage could be impaired by destabilisation at sites which are already less favourable with the native system, such as that observed with sgRNA 8. Additionally, perhaps the L1 modification does not allow for sufficient time wherein the HNH domain is able to cleave the target DNA strand in order to generate the required DSB for indel formation. It has also been demonstrated that truncated *Sp*Cas9 sgRNAs do not tolerate non-target strand groove modifications, pointing to potential additional effects of the 'enhanced specificity' mutations that are yet to be determined (Slaymaker et al., 2016).

In summary, the destabilisation of non-target strand binding within the *Sa*Cas9 nuclease was found not to affect cleavage activity of the more efficient sgRNA 7 but activity was reduced at the less efficient sgRNA 8 target site. The reduction in activity of the e*Sa*Cas9 nickase also raises an important unanswered question that is to what extent specificity enhancing modifications can be combined before the compromise in efficiency is too high. A broader range of individual e*Sa*Cas9 and dual e*Sa*Cas9-D10A target sites needs to be investigated and

compared to the original systems to draw more definitive conclusions about the specificity enhancing mutations and the trade-off between safety and successful genome engineering.

### 4.3. Downstream functional effects of targeted gene editing

To elucidate whether *CCR5* ORF sequence disruption by gene editing resulted in a simple phenotypic knock out effect, downstream processes of mRNA production and cell surface protein function were assessed. A logical hypothesis is that the level of *CCR5*-specific target site indels would induce the NMD pathway, directly correlate with mRNA production and consequently cell surface receptor expression and functioning as measured by R5-tropic HIV-1 infectivity. The NMD pathway is known to be responsible for the removal of mutated mRNA transcripts encoding non-functional and potentially deleterious *trans* dominant proteins by the recognition of PTCs located at least 50 – 55 nt upstream of the final exon-exon junction in a pre-mRNA transcript (Le Hir et al., 2000a, Le Hir et al., 2000b). As the *CCR5* ORF is located within solely within the third and final exon of the gene, it is unlikely that canonical NMD would be stimulated upon the introduction of gene editing-mediated PTCs, however this may be a possibility and remains to be elucidated. The *trans*-dominant negative effect of the naturally occurring CCR5-Δ32 mutated protein that is expressed despite a large deletion suggests that the relationship between *CCR5* DNA, mRNA and protein is able to circumvent NMD or a form thereof, at least to a certain extent. Measurement of the *CCR5* mRNA and R5-tropic HIV-1 infectivity following indel quantification with the *Sa*Cas9 nuclease, nickase and 'enhanced specificity' nuclease systems as well as analysis of likely protein structures inferred from deep sequencing results provided insight into the influences of gene editing on downstream molecular biology.

Relative levels of mRNA knock down as measured by qRT-PCR were comparable to the indel frequencies at the corresponding target sites for all CRISPR/*Sa*Cas9 systems but found to be ~ 2.5 and at least 3 fold higher for the *Sa*Cas9 with sgRNA 8 and *Sa*Cas9-D10A with dual sgRNA 7/8 treatments respectively. In general, the level of preclusion of R5-tropic HIV-1 infectivity as a measurement of CCR5 protein cell surface expression and function was slightly higher than the corresponding mRNA and indel frequencies for each sample. Before delving into the likely explanations for the observed results, the major mutations introduced into the sgRNA 7, 8 and 7/8 target sites with the *Sa*Cas9 and *Sa*Cas9-D10A systems as determined by deep sequencing were assessed and the likely primary protein sequence inferred from this. The two major single and dinucleotide mutations from *Sa*Cas9 and sgRNA

7 treatment in TZM-bl cells (Figure 3.7) were determined to introduce stop codons a short distance downstream of the modified target site. If these mutations are able to avoid activation of the NMD pathway and are transcribed as well as translated, as is the case with the CCR5-Δ32 mutation, this would generate two similar truncated proteins (Figure 4.2).

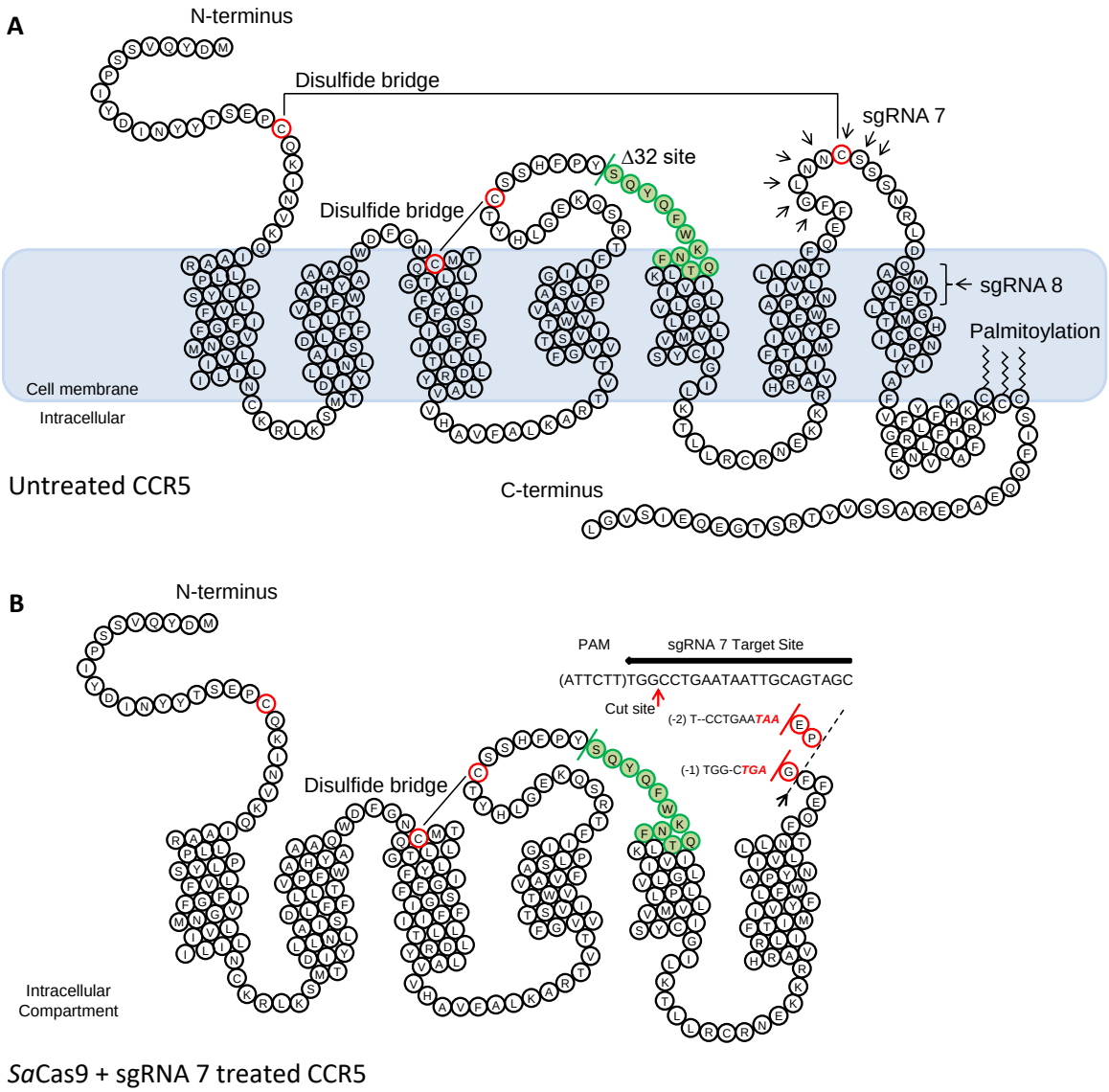



**Figure 4.2: CCR5 protein structure before and after CRISPR/*Sa*Cas9 nuclease treatment. A** – The WT structure of the CCR5 protein* as it is expressed in the human plasma membrane, showing the amino acid residues corresponding to sgRNA 7 and 8 gene target sites, the location of the CCR5-Δ32 mutation, two essential disulfide bridges and palmitoylation sites at the C terminus. **B** – A diagram of the CCR5 protein effects of two high frequency indels introduced by *Sa*Cas9 + sgRNA 7 treatment in TZM-bl cells as determined by targeted amplicon deep sequencing. The most common modification is the deletion of the GG dinucleotide (-2) at the expected cut site and the second most common is the single deletion (-1) of the G at position 3 from the PAM proximal end of sgRNA 7. The change in nucleotide sequence downstream of each modification is shown and the stop codon highlighted in red italics. The red amino acids within the sgRNA 7 target site are the

modified residues following the indel and prior to the stop codon. The exact resulting structure is not known but this illustration shows how the indels could result in the introduction of stop codons which would truncate the protein and prevent formation of one of the critical disulphide bridges as well as remove the C terminus. The inability to palmitoylate the protein would result in containment within intracellular compartments and inhibit trafficking to the cell surface. *This protein structure has been adapted from Parmentier *et al*. (2015).

As the result of both the CCR5-Δ32 mutation and *Sa*Cas9 sgRNA 7-mediated gene editing, the critical palmitoylation sites as well as one of the stabilising disulfide bridges are absent in the truncated proteins (Blanpain et al., 2001). It would be expected therefore that the defective proteins generated by the CRISPR/Cas9-mediated indels would behave in a similar manner to the CCR5-Δ32 variant if able to avoid the NMD pathway. The effect of the dinucleotide deletion resulting from *Sa*Cas9 and sgRNA 8 treatment (Figure 3.7) resulted in an altered primary protein sequence downstream of the modified site but no truncation owing to deletion of the 5′ CT pair in the CTCT di-repeat sequence. This phenomenon can be attributed to 'microhomology' sequences which are considered negatively for effective *in silico* sgRNA design by the Cas-Designer online tool (Bae et al., 2014, Park et al., 2015). The effect of an altered C terminal protein sequence on the cellular processing of the protein is unknown as it is unlike that of the CCR5-Δ32 or *Sa*Cas9 nuclease and sgRNA 7-treated variants. Other minor microhomology-independent mutations introduced into the *Sa*Cas9 sgRNA 8 target site would be expected to generate proteins that behave in a similar manner to the two aforementioned truncated proteins. The *Sa*Cas9 nickase and dual sgRNA 7/8 treatment mutations were revealed to be more extensive and variable than the nuclease-treated sample but each should exert a similar truncating effect on the resulting protein.

The qRT-PCR primers utilised in this study amplified mRNA sequences upstream of the CCR5-Δ32 and CRISPR/Cas9 target sites and thus would be able to detect all exon 3 ORF-containing transcripts present, regardless of downstream sequence alterations. At first glance, indel frequencies introduced into the target sites appear to correspond to the level of mRNA reduction, except that the *Sa*Cas9 and sgRNA 8 and *Sa*Cas9-D10A and dual sgRNA 7/8 treatments displayed higher mRNA reduction than expected. Four different mechanisms may have played a role in these resulting mRNA levels. The first is that CRISPR/Cas9-mediated indels ablate mRNA production by an NMD-like pathway, however, as this is not an effect of the CCR5-Δ32 mutation, it is unlikely that this was the cause and this phenomenon necessitates deeper investigation. It is important to note that CRISPR/Cas9 systems will only be active in the parent cell owing to plasmid loss upon replication. This means that 48 hours

post-transfection, some cells may still experience CRISPR/Cas9 activity although the exact proportion is unknown. The second mechanism that could potentially reduce mRNA levels is thus CRISPR/Cas9- or NHEJ DNA repair machinery-mediated interference of transcription at the target site. The former would, however, be expected to affect mRNA levels from negative control samples treated with the individual sgRNA and *Sa*Cas9 nickase and there was no significant reduction observed in this regard. An interesting alternative phenomenon has recently come to light through the discovery that the CRISPR/Cas9 system derived from *Francisella novicida* is able to target the positive sense single stranded Hepatitis C RNA genome (Price et al., 2015). It has since also been demonstrated that the catalytically inactive CRISPR/*Sp*Cas9 system is able to target host cellular mRNAs and inhibit translation in the absence of a PAM duplex (Liu et al., 2016b). These findings are potentially relevant as it has not been investigated whether the *Sa*Cas9 or *Sa*Cas9-D10A, in complex with a complementary sgRNA, can bind and cleave mRNA sequences thereby inducing nuclease-mediated degradation or an NMD-like effect. Critically, only reverse-orientated sgRNAs, such as sgRNA 7, would be complementary to the *CCR5* mRNA sequence and able to form an RNA homoduplex. Finally, it may be possible for pre-Cas9 bound sgRNAs to interact with mRNA transcripts and, beyond repressing translation, potentially engage in an RNAi-like pathway. It appears that no study has investigated whether a sgRNA of 21 – 22 nt in length, similar to that of siRNAs and miRNAs, enables interference and degradation of mRNA or if the secondary structure of the scaffold circumvents this.

A number of experiments may be necessary in order to elucidate the true effect of CRISPR/*Sa*Cas9 nuclease and nickase-mediated indels on mRNA levels. To determine the effect of CRISPR/Cas9-mediated indel formation, qRT-PCR should be carried out after an extended period of time following treatment to allow for turnover of the gene editing systems. To determine whether transcription is suppressed by CRISPR interference, a time course mRNA expression experiment could be carried out as transient plasmid-expressed CRISPR/Cas9 components are lost with every cell cycle. An *in vitro* assay could be carried out to determine whether mRNA is targeted and cleaved by the *Sa*Cas9 nuclease and nickase systems by visualisation of RNA fragments on a Southern blot. Finally, whether sgRNA:mRNA homoduplexes enter into an RNAi pathway could be determined by tracking of the sgRNA and staining RISC complexes in live cells (Nelles et al., 2016). A likely effect of PAM-independent CRISPR/Cas9 recognition of mRNA is suppression of translation, shown to be achievable in a recent study (Liu et al., 2016b).This also opens up the possibility

for suppression not only at the target site but also a vastly increased PAM-free potential off target landscape (Liu et al., 2016b). This, in combination with indel-mediated CCR5 protein structure alterations and CCR5-Δ32-like sequestration effects, could have resulted in the greater preclusion of R5-tropic infectivity exerted than was projected based on quantified indel frequencies.

In summary, it is imperative to elucidate the direct effect of CRISPR/Cas9 nuclease and nickase treatment on mRNA and protein levels. It is possible that not all treatments will affect downstream processes in the same way as the CCR5-Δ32 mutation or the CRISPR/Cas9 systems investigated in this study as the effect is expected to gene- and target locus-specific. Importantly, it remains to be ascertained whether the CRISPR/Cas9 systems employed in this study can directly cleave and induce degradation of mRNA transcripts in an NMD pathway-like manner. Additional levels of targeting by CRISPR/Cas9 could cause a paradigm shift in the current understanding of CRISPR/Cass9 gene editing specificity and the scope of downstream as well as other functional off target effects.

### 4.4. Future prospects of a CCR5-specific CRISPR-mediated anti-HIV-1 gene therapy

Precision genome engineering technology holds immense promise to uncovering single intervention treatments for a plethora of diseases for which current therapies are ineffective, extremely invasive or non-existent. Mathematical models have asserted that entry and early replication steps in the HIV-1 lifecycle must be the target of gene therapies in order to overcome the pathological effects of infection (von Laer et al., 2006). Mimicking the CCR5-Δ32 phenotype with no apparent natural immune consequence appears to hold the key for a novel approach to uncovering an HIV-1 functional cure. Access to donor-matched CCR5-Δ32 T cells for the cure of tens of millions of infected patients is practically infeasible and thus genome engineering-based gene therapies that are able to recapitulate the desired phenotypic effect have been an important novel focus (Reviewed in Cannon et al., 2014). Despite intense interest garnered since the discovery of CRISPR/Cas9 genome engineering, its novelty has resulted in employment of previous generation of genome engineering tools in the majority of gene therapies currently in clinical or preclinical trials (Reviewed in Song, 2017). Moving forward, CRISPR/Cas9 is expected to explode onto the clinical scene owing to simplicity of design and production. Recent advancements in the field of stem cell therapy (SCT) and improvements to gene therapy delivery modalities pave the way forward for broader, simpler and more cost-effective anti-HIV-1 therapeutics.

Beyond characterising the safety and efficacy of a CRISPR/Cas9-based gene therapy, understanding and optimising delivery to the target cells is critical. AAVs are valuable delivery vehicles for both *in vivo* and *ex vivo* gene therapy and have recently been engineered to suit desired clinical applications (Reviewed in Kotterman and Schaffer, 2014). The vector has recently received regulatory approval with the first *in vivo* gene therapy employing AAV serotype 8 for the delivery of the Factor IX transgene in haemophilia B patients (Nathwani et al., 2014). The tropism and relative expression levels of nine major serotypes was characterised in mice by Zincarelli and colleagues in 2008 and since has undergone extensive advancement (Zincarelli et al., 2008). Pertinent to anti-HIV-1 gene therapy development, AAV-DJ/8 is a novel recombinant variant that was effectively used to deliver HIV-1 LTR-specific CRISPR/*Sa*Cas9 systems systemically within a humanised mouse model (Yin et al., 2017). For a greater level of specificity, AAV serotype 6 has been shown to target T cells with high efficiency, useful for *ex vivo* modification of primary cells (Sather et al., 2015). Viral-based vectors appear to be the preferred method of delivery although advancements in the efficiency and safety of non-viral modalities are beginning to increase their practicality (Zuris et al., 2015). Small peptides have also recently been added to enhance AAV-mediated delivery by increased cell permeability (Liu et al., 2014). These are only a few examples among many advancements that will facilitate CRISPR/Cas9 gene therapy delivery both *ex vivo* and *in vivo* and the specifics of the selected system are dependent on the target cell model.

Stem cells are highly attractive cell models as they enable modification of every cell type along the defined lineage, for example, *CCR5*-specific targeting of $CD34^+$ haematopoietic stem/progenitor cell (HSPC) genomes will enable modification of each cell type clinically relevant to HIV-1 infection. The two major branches of SCT are being investigated to generate a permanent source of HIV-1 resistant cells in gene therapy development are adult and pluripotent stem cells. In a recent clinical trial led by Sangamo Therapeutics (CA, USA), adoptive cell transfer of ZFN-modified primary T cells to HIV-1 infected patients demonstrated the safety and efficacy of a *CCR5*-knock out strategy for HIV-1 functional cure, however, the resistant cells had a limited half-life of just under a year, necessitating the development of a longer-term approach (Tebas et al., 2014). Critically, not all of the HIV-1 clinically relevant cell types are modified with this approach which would not eliminate all cell-cell transmission events or preclude cell death. The follow-on clinical trial currently investigates the applicability of *ex vivo* modification and transplantation of autologous

(patient's own) HSPCs and has awakened the field of HIV-1 cure therapeutics to the array possibilities for *ex vivo* SCT-based genome engineering gene therapy (DiGiusto et al., 2016). The premise is that the most efficient gene editing systems can be delivered to the appropriate *ex vivo* stem cell model and cells that are successfully and safely modified selected, expanded and transplanted into the patient. This circumvents difficulties currently faced with *in vivo* therapy whereby a threshold number of modified cells for the therapeutic to be effective must be determined and reached and pre-screening circumvents the uncertainty surrounding potential undetected off target events. The issue faced now is whether autologous SCT is a viable option for the tens of millions of people infected with HIV-1 globally and whether alternative strategies need to be investigated to broaden the applicability of this type of therapy.

Induced pluripotent stem cells (iPSCs) offer an alternative approach and have recently been successfully modified by both CRISPR and TALENs for HIV-1 resistance by introducing the CCR5-Δ32 mutation (Ye et al., 2014, Smith et al., 2014). These cells were able to differentiate into monocytes and macrophages with resistance to HIV-1 infection. iPSCs can easily be derived from any adult tissue, such as skin, and reprogrammed to develop into a desired cell type. Scalability is still a drawback with this technology as, ideally, an iPSC-based therapeutic should be autologous-derived to circumvent graft-versus-host technicalities but the sheer number of HIV-1 infected patients to be treated prohibits this under these circumstances. An interesting additional benefit provided by the CCR5-Δ32 mutation is protection against graft-versus-host disease. Whether the absence of CCR5 is enough to confer a protective effect or if changes in gene expression levels of closely associated genes, such as CCRL2 and WD repeat domain 6, alters a patient's response to allogeneic stem cells remains to be fully elucidated (Hutter et al., 2011). A clinical trial (NCT00948753) is underway to investigate the role of CCR5 in this protective effect by blocking the receptor with Maraviroc in patients undergoing non-myeloablative allogeneic stem cell therapy (Reshef et al., 2012). Databases could play an integral role in developing effective stem cell-based therapeutics. One such example is HLA groupings of patients and donors so as to select stem cells for modification and allogeneic transplantation. The IciStem project is an example of a global effort to document and characterise HIV positive patient responses to allogeneic stem cell transplant for life-threatening haematological diseases (Deeks et al., 2016). The goal is to understand the mechanism of viral reduction and/or identify cases of remission but

this type of database can also be used to group HLA types according to patient responses and safely scale up stem cell gene therapies from autologous to allogeneic.

A final critical consideration for purely *CCR5*-targeted gene therapies is the role of alternative co-receptors, such as CXCR4, and tropism switching during the time course of infection in the long term efficacy of the therapeutic (Ribeiro et al., 2006, Jiang et al., 2011, Frange et al., 2013). Importantly, bystander cell death by apoptosis has been shown to be induced predominantly by X4-tropic isolates, highlighting the pathological role that other co-receptors may play in the progression to AIDS-related illness (Jekle et al., 2003). Although X4-tropic isolates typically emerge in chronic infection following rapid $CD4^+$ T cell decline, these variants must be critically evaluated as they will influence the efficacy of CCR5-targeted therapeutics in chronically ill patients. Combination gene therapies could offer an elegant solution to this as well as to the issue of allograft rejection in allogeneic stem cell transplantation. To enhance the HIV-1 entry barrier, *CCR5* disruption could be accompanied by the insertion of a fusion inhibitor, such as C46 (Sather et al., 2015), a modification of the approach recently shown to be effective in simian HIV infected macaques (Wolstein et al., 2014, Peterson et al., 2016). The multiplexed capacity of CRISPR/Cas9 enables modification of more than one gene within a single genome, offering the opportunity to knock out both *CCR5* and *B2M*, a gene that if non-functional confers hypoimmunogenicity to stem cells (Riolobos et al., 2013, Mandal et al., 2014). Furthermore, an HIV-1 tat-dependent suicide gene could be included in the therapeutic so as to initiate apoptosis in infected cells that are not successfully modified for resistance (Pandit and de Boer, 2015).

## 5. Conclusions

This study demonstrated that both the native and 'enhanced specificity' variants of the CRISPR/Cas9 system derived from *Staphylococcus aureus* were able to genetically modify the human *CCR5* gene and reduce mRNA as well as R5-tropic pseudovirus infectivity in cell culture. Target site cleavage activity was found to be variable, likely as a result of the sgRNA secondary structure and consequent recognition by the Cas9 protein. The *Sa*Cas9 nickase system functioned with reduced on target cleavage efficiency compared to the nuclease but also reduced off target activity within the selected *CCR2* locus making this a safer option. In an effort to maintain cleavage efficiency while enhancing specificity, the e*Sa*Cas9 nuclease was employed, displaying both equivalent and reduced activity at the two selected *CCR5* target sites. As no known e*Sa*Cas9 nickase system has been investigated, this was included in gene editing experimentation and found to function with minimal activity at the selected dual sgRNA target site. Indel frequency comparison between the WT and 'enhanced specificity' *Sa*Cas9 nuclease and nickase systems highlighted a likely trade-off effect between on target cleavage activity and specificity. Finally, *CCR5* mRNA production and R5-tropic HIV-1 pseudovirus infectivity was hampered by genetic modification of the *CCR5* gene. The complexities surrounding downstream functional effects of WT and 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase-mediated *CCR5*-specific gene editing, particularly on mRNA and protein production and function, remains to be fully elucidated.

In summary, it is imperative that future studies evaluate genome-wide unbiased off target cleavage activity of each of the *CCR5*-targeted CRISPR/*Sa*Cas9 nuclease and nickase systems. Thereafter, experimentation involving AAV-mediated delivery and cleavage activity in *ex vivo* HSCs or iPSCs will determine whether this is a promising avenue to explore for improved specificity and deliverability of gene therapies for the functional cure of HIV-1. The monumental scale of HIV-1 infection and highest burden within resource-limited settings necessitates the development of cost-effective and simply administered therapeutics. A reconstituted healthy and resistant immune system not only relieves patients of a lifelong ARV burden but also reduces transmission and could ultimately lead to eradication of this devastating disease.

## 6. References

### 6.1. Journals


Abbas, W. & Herbein, G. 2012. Molecular Understanding of HIV-1 Latency. *Adv Virol,* 2012**,** 574967.

Agosto, L. M., Uchil, P. D. & Mothes, W. 2015. HIV cell-to-cell transmission: effects on pathogenesis and antiretroviral therapy. *Trends Microbiol,* 23**,** 289-95.

Agrawal, L., Jin, Q., Altenburg, J., Meyer, L., Tubiana, R., Theodorou, I., et al. 2007. CCR5Delta32 protein expression and stability are critical for resistance to human immunodeficiency virus type 1 in vivo. *J Virol,* 81**,** 8041-9.

Bae, S., Kweon, J., Kim, H. S. & Kim, J. S. 2014. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods,* 11**,** 705-6.

Barmania, F. & Pepper, M. S. 2013. C-C chemokine receptor type five (CCR5): An emerging target for the control of HIV infection. *Applied & Translational Genomics,* 2**,** 3-16.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., et al. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science,* 220**,** 868-71.

Batra, J., Tripathi, S., Kumar, A., Katz, J. M., Cox, N. J., Lal, R. B., et al. 2016. Human Heat shock protein 40 (Hsp40/DnaJB1) promotes influenza A virus replication by assisting nuclear import of viral ribonucleoproteins. *Sci Rep,* 6**,** 19063.

Belew, A. T., Hepler, N. L., Jacobs, J. L. & Dinman, J. D. 2008. PRFdb: a database of computationally predicted eukaryotic programmed -1 ribosomal frameshift signals. *BMC Genomics,* 9**,** 339.

Belew, A. T., Meskauskas, A., Musalgaonkar, S., Advani, V. M., Sulima, S. O., Kasprzak, W. K., et al. 2014. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature,* 512**,** 265-9.

Berger, E. A., Murphy, P. M. & Farber, J. M. 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol,* 17**,** 657-700.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G., et al. 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol,* 21**,** 289-97.

Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161**,** 1169-75.

Blanpain, C., Wittamer, V., Vanderwinden, J. M., Boom, A., Renneboog, B., Lee, B., et al. 2001. Palmitoylation of CCR5 is critical for receptor trafficking and efficient activation of intracellular signaling pathways. *J Biol Chem,* 276**,** 23795-804.

Bothmer, A., Phadke, T., Barrera, L. A., Margulies, C. M., Lee, C. S., Buquicchio, F., et al. 2017. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun,* 8**,** 13905.

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

Bullen, C. K., Laird, G. M., Durand, C. M., Siliciano, J. D. & Siliciano, R. F. 2014. New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. *Nat Med,* 20**,** 425-9.

Cannon, P. M., Kohn, D. B. & Kiem, H. P. 2014. HIV eradication--from Berlin to Boston. *Nat Biotechnol,* 32**,** 315-6.

Cash-Padgett, T., Sawa, A. & Jaaro-Peled, H. 2016. Increased stereotypy in conditional Cxcr4 knockout mice. *Neurosci Res,* 105**,** 75-9.

Cavrois, M., Neidleman, J., Kreisberg, J. F., Fenard, D., Callebaut, C. & Greene, W. C. 2006. Human immunodeficiency virus fusion to dendritic cells declines as cells mature. *J Virol,* 80**,** 1992-9.

Chari, R., Yeo, N. C., Chavez, A. & Church, G. M. 2017. sgRNA Scorer 2.0: A Species-Independent Model To Predict CRISPR/Cas9 Activity. *ACS Synth Biol*.

Chen, P., Hubner, W., Spinelli, M. A. & Chen, B. K. 2007. Predominant mode of human immunodeficiency virus transfer between T cells is mediated by sustained Env-dependent neutralization-resistant virological synapses. *J Virol,* 81**,** 12582-95.

Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol,* 31**,** 230-2.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., et al. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res,* 24**,** 132-41.

Chun, T. W., Finzi, D., Margolick, J., Chadwick, K., Schwartz, D. & Siliciano, R. F. 1995. In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nat Med,* 1**,** 1284-90.

Chun, T. W., Stuyver, L., Mizell, S. B., Ehler, L. A., Mican, J. A., Baseler, M., et al. 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci U S A,* 94**,** 13193-7.

Combadiere, C., Ahuja, S. K., Tiffany, H. L. & Murphy, P. M. 1996. Cloning and functional expression of CC CKR5, a human monocyte CC chemokine receptor selective for MIP-1(alpha), MIP-1(beta), and RANTES. *J Leukoc Biol,* 60**,** 147-52.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

Cox, D. B., Platt, R. J. & Zhang, F. 2015. Therapeutic genome editing: prospects and challenges. *Nat Med,* 21**,** 121-31.

Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. 2013. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res,* 41**,** 9584-92.

Crosetto, N., Mitra, A., Silva, M. J., Bienko, M., Dojer, N., Wang, Q., et al. 2013. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods,* 10**,** 361-5.

Dale, B. M., Mcnerney, G. P., Thompson, D. L., Hubner, W., De Los Reyes, K., Chuang, F. Y., et al. 2011. Cell-to-cell transfer of HIV-1 via virological synapses leads to endosomal virion maturation that activates viral membrane fusion. *Cell Host Microbe,* 10**,** 551-62.

Daya, S. & Berns, K. I. 2008. Gene therapy using adeno-associated virus vectors. *Clin Microbiol Rev,* 21**,** 583-93.

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., et al. 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science,* 273**,** 1856-62.

Deeks, S. G. 2012. HIV: Shock and kill. *Nature,* 487**,** 439-40.

Deeks, S. G., Lewin, S. R., Ross, A. L., Ananworanich, J., Benkirane, M., Cannon, P., et al. 2016. International AIDS Society global scientific strategy: towards an HIV cure 2016. *Nat Med,* 22**,** 839-50.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., et al. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-7.

Deneka, M., Pelchen-Matthews, A., Byland, R., Ruiz-Mateos, E. & Marsh, M. 2007. In macrophages, HIV-1 assembles into an intracellular plasma membrane domain containing the tetraspanins CD81, CD9, and CD53. *J Cell Biol,* 177**,** 329-41.

Didigu, C. A., Wilen, C. B., Wang, J., Duong, J., Secreto, A. J., Danet-Desnoyers, G. A., et al. 2014. Simultaneous zinc-finger nuclease editing of the HIV coreceptors ccr5 and cxcr4 protects CD4+ T cells from HIV-1 infection. *Blood,* 123**,** 61-9.

Digiusto, D. L., Cannon, P. M., Holmes, M. C., Li, L., Rao, A., Wang, J., et al. 2016. Preclinical development and qualification of ZFN-mediated CCR5 disruption in human hematopoietic stem/progenitor cells. *Mol Ther Methods Clin Dev,* 3**,** 16067.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol,* 34**,** 184-91.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol,* 32**,** 1262-7.

Doitsh, G., Cavrois, M., Lassen, K. G., Zepeda, O., Yang, Z., Santiago, M. L., et al. 2010. Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. *Cell,* 143**,** 789-801.

Doitsh, G., Galloway, N. L., Geng, X., Yang, Z., Monroe, K. M., Zepeda, O., et al. 2014. Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature,* 505**,** 509-14.

Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., et al. 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature,* 381**,** 667-73.

Dragic, T., Trkola, A., Thompson, D. A., Cormier, E. G., Kajumo, F. A., Maxwell, E., et al. 2000. A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *Proc Natl Acad Sci U S A,* 97**,** 5639-44.

Dutartre, H., Claviere, M., Journo, C. & Mahieux, R. 2016. Cell-Free versus Cell-to-Cell Infection by Human Immunodeficiency Virus Type 1 and Human T-Lymphotropic Virus Type 1: Exploring the Link among Viral Source, Viral Trafficking, and Viral Replication. *J Virol,* 90**,** 7607-17.

Eugen-Olsen, J., Iversen, A. K., Garred, P., Koppelhus, U., Pedersen, C., Benfield, T. L., et al. 1997. Heterozygosity for a deletion in the CKR-5 gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. *AIDS,* 11**,** 305-10.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L. M., Buck, C., Chaisson, R. E., et al. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science,* 278**,** 1295-300.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K. S., Lecrivain, A. L., Bzdrenga, J., et al. 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.

Frange, P., Meyer, L., Ghosn, J., Deveau, C., Goujard, C., Duvivier, C., et al. 2013. Prevalence of CXCR4-tropic viruses in clustered transmission chains at the time of primary HIV-1 infection. *Clin Microbiol Infect,* 19**,** E252-5.

Friedland, A. E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., et al. 2015. Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol,* 16**,** 257.

Frock, R. L., Hu, J., Meyers, R. M., Ho, Y. J., Kii, E. & Alt, F. W. 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol,* 33**,** 179-86.

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.

Galloway, N. L., Doitsh, G., Monroe, K. M., Yang, Z., Munoz-Arias, I., Levy, D. N., et al. 2015. Cell-to-Cell Transmission of HIV-1 Is Required to Trigger Pyroptotic Death of Lymphoid-Tissue-Derived CD4 T Cells. *Cell Rep,* 12**,** 1555-63.

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.

Geijtenbeek, T. B., Kwon, D. S., Torensma, R., Van Vliet, S. J., Van Duijnhoven, G. C., Middel, J., et al. 2000. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell,* 100**,** 587-97.

Goldstone, D. C., Ennis-Adeniran, V., Hedden, J. J., Groom, H. C., Rice, G. I., Christodoulou, E., et al. 2011. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature,* 480**,** 379-82.

Gong, W., Howard, O. M., Turpin, J. A., Grimm, M. C., Ueda, H., Gray, P. W., et al. 1998. Monocyte chemotactic protein-2 activates CCR5 and blocks CD4/CCR5-mediated HIV-1 entry/replication. *J Biol Chem,* 273**,** 4289-92.

Gouws, E., Stanecki, K. A., Lyerla, R. & Ghys, P. D. 2008. The epidemiology of HIV infection among young people aged 15-24 years in southern Africa. *AIDS,* 22 Suppl 4**,** S5-16.

Guerreiro, C., Sousa, C., Santos, A., Gomes, R., Ferreira, J. & Fonseca, P. 2012. Prevalence of HIV 2: A retrospective study from a HIV/AIDS consult. *Journal of the International AIDS Society,* 15.

Gunthard, H. F., Saag, M. S., Benson, C. A., Del Rio, C., Eron, J. J., Gallant, J. E., et al. 2016. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2016 Recommendations of the International Antiviral Society-USA Panel. *JAMA,* 316**,** 191-210.

Hansasuta, P. & Rowland-Jones, S. L. 2001. HIV-1 transmission and acute HIV-1 infection. *Br Med Bull,* 58**,** 109-27.

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., et al. 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem,* 83**,** 8604-10.

Horuk, R. 1994. The interleukin-8-receptor family: from chemokines to malaria. *Immunol Today,* 15**,** 169-74.

Hou, P., Chen, S., Wang, S., Yu, X., Chen, Y., Jiang, M., et al. 2015. Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection. *Sci Rep,* 5**,** 15577.

Hrecka, K., Hao, C., Gierszewska, M., Swanson, S. K., Kesik-Brodacka, M., Srivastava, S., et al. 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature,* 474**,** 658-61.

Hrecka, K., Hao, C., Shun, M. C., Kaur, S., Swanson, S. K., Florens, L., et al. 2016. HIV-1 and HIV-2 exhibit divergent interactions with HLTF and UNG2 DNA repair proteins. *Proc Natl Acad Sci U S A,* 113**,** E3921-30.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol,* 31**,** 827-32.

Hu, W., Kaminski, R., Yang, F., Zhang, Y., Cosentino, L., Li, F., et al. 2014. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc Natl Acad Sci U S A,* 111**,** 11461-6.

Hubner, W., Mcnerney, G. P., Chen, P., Dale, B. M., Gordon, R. E., Chuang, F. Y., et al. 2009. Quantitative 3D video microscopy of HIV transfer across T cell virological synapses. *Science,* 323**,** 1743-7.

Hutter, G., Neumann, M., Nowak, D., Klein, S., Kluter, H. & Hofmann, W. K. 2011. The effect of the CCR5-delta32 deletion on global gene expression considering immune response and inflammation. *J Inflamm (Lond),* 8**,** 29.

Hutter, G., Nowak, D., Mossner, M., Ganepola, S., Mussig, A., Allers, K., et al. 2009. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med,* 360**,** 692-8.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol,* 169**,** 5429-33.

Iwami, S., Takeuchi, J. S., Nakaoka, S., Mammano, F., Clavel, F., Inaba, H., et al. 2015. Cell-to-cell infection by HIV contributes over half of virus infection. *Elife,* 4.

Izquierdo-Useros, N., Lorizate, M., Mclaren, P. J., Telenti, A., Krausslich, H. G. & Martinez-Picado, J. 2014. HIV-1 capture and transmission by dendritic cells: the role of viral glycolipids and the cellular receptor Siglec-1. *PLoS Pathog,* 10**,** e1004146.

Jekle, A., Keppler, O. T., De Clercq, E., Schols, D., Weinstein, M. & Goldsmith, M. A. 2003. In Vivo Evolution of Human Immunodeficiency Virus Type 1 toward Increased Pathogenicity through CXCR4-Mediated Killing of Uninfected CD4 T Cells. *Journal of Virology,* 77**,** 5846-54.

Jiang, C., Parrish, N. F., Wilen, C. B., Li, H., Chen, Y., Pavlicek, J. W., et al. 2011. Primary infection by a human immunodeficiency virus with atypical coreceptor tropism. *J Virol,* 85**,** 10669-81.

Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. 2015. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science,* 348**,** 1477-81.

Jin, Q., Agrawal, L., Meyer, L., Tubiana, R., Theodorou, I. & Alkhatib, G. 2008a. CCR5Delta32 59537-G/A promoter polymorphism is associated with low translational efficiency and the loss of CCR5Delta32 protective effects. *J Virol,* 82**,** 2418-26.

Jin, Q., Marsh, J., Cornetta, K. & Alkhatib, G. 2008b. Resistance to human immunodeficiency virus type 1 (HIV-1) generated by lentivirus vector-mediated delivery of the CCR5-Delta 32 gene despite detectable expression of the HIV-1 co-receptors. *J Gen Virol,* 89**,** 2611-21.

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., East, A., Cheng, A., Lin, S., Ma, E. & Doudna, J. 2013. RNA-programmed genome editing in human cells. *Elife,* 2**,** e00471.

Jolly, C., Kashefi, K., Hollinshead, M. & Sattentau, Q. J. 2004. HIV-1 cell to cell transfer across an Env-induced, actin-dependent synapse. *J Exp Med,* 199**,** 283-93.

Jolly, C., Mitar, I. & Sattentau, Q. J. 2007. Adhesion molecule interactions facilitate human immunodeficiency virus type 1-induced virological synapse formation between T cells. *J Virol,* 81**,** 13916-21.

Kaminski, R., Bella, R., Yin, C., Otte, J., Ferrante, P., Gendelman, H. E., et al. 2016a. Excision of HIV-1 DNA by gene editing: a proof-of-concept in vivo study. *Gene Ther,* 23**,** 696.

Kaminski, R., Chen, Y., Fischer, T., Tedaldi, E., Napoli, A., Zhang, Y., et al. 2016b. Elimination of HIV-1 Genomes from Human T-lymphoid Cells by CRISPR/Cas9 Gene Editing. *Sci Rep,* 6**,** 22555.

Kang, H., Minder, P., Park, M. A., Mesquitta, W. T., Torbett, B. E. & Slukvin, Ii 2015. CCR5 Disruption in Induced Pluripotent Stem Cells Using CRISPR/Cas9 Provides Selective Resistance of Immune Cells to CCR5-tropic HIV-1 Virus. *Mol Ther Nucleic Acids,* 4**,** e268.

Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., et al. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A,* 105**,** 7552-7.

Kibbe, W. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.*

Kim, E., Kim, S., Kim, D. H., Choi, B. S., Choi, I. Y. & Kim, J. S. 2012. Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res,* 22**,** 1327-33.

Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res,* 24**,** 1012-9.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et al. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature,* 529**,** 490-5.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., et al. 2015a. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol,* 33**,** 1293-8.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., et al. 2015b. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature,* 523**,** 481-5.

Kondo, N., Marin, M., Kim, J. H., Desai, T. M. & Melikyan, G. B. 2015. Distinct requirements for HIV-cell fusion and HIV-mediated cell-cell fusion. *J Biol Chem,* 290**,** 6558-73.

Kotterman, M. A. & Schaffer, D. V. 2014. Engineering adeno-associated viruses for clinical gene therapy. *Nat Rev Genet,* 15**,** 445-51.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol,* 32**,** 677-83.

Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E. C., Dragin, L., et al. 2012. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol,* 13**,** 223-8.

Lai, J. P., Yang, J. H., Douglas, S. D., Wang, X., Riedel, E. & Ho, W. Z. 2003. Quantification of CCR5 mRNA in Human Lymphocytes and Macrophages by Real-Time Reverse Transcriptase PCR Assay. *Clinical and Vaccine Immunology,* 10**,** 1123-8.

Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. 2000a. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J,* 19**,** 6860-9.

Le Hir, H., Moore, M. J. & Maquat, L. E. 2000b. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev,* 14**,** 1098-108.

Lee, B., Leslie, G., Soilleux, E., O'doherty, U., Baik, S., Levroney, E., et al. 2001. cis Expression of DC-SIGN allows for more efficient entry of human and simian immunodeficiency viruses via CD4 and a coreceptor. *J Virol,* 75**,** 12028-38.

Lee, B., Sharron, M., Montaner, L. J., Weissman, D. & Doms, R. W. 1999. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. *Proc Natl Acad Sci U S A,* 96**,** 5215-20.

Lee, J., Chung, J. H., Kim, H. M., Kim, D. W. & Kim, H. 2016. Designed nucleases for targeted genome editing. *Plant Biotechnol J,* 14**,** 448-62.

Li, M., Gao, F., Mascola, J. R., Stamatatos, L., Polonis, V. R., Koutsoukos, M., et al. 2005. Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol,* 79**,** 10108-25.

Libert, F., Cochaux, P., Beckman, G., Samson, M., Aksenova, M., Cao, A., et al. 1998. The deltaccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum Mol Genet,* 7**,** 399-406.

Lin, F. L., Sperle, K. & Sternberg, N. 1985. Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc Natl Acad Sci U S A,* 82**,** 1391-5.

Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., et al. 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell,* 86**,** 367-77.

Liu, R., Zhao, X., Gurney, T. A. & Landau, N. R. 1998. Functional analysis of the proximal CCR5 promoter. *AIDS Res Hum Retroviruses,* 14**,** 1509-19.

Liu, X., Homma, A., Sayadi, J., Yang, S., Ohashi, J. & Takumi, T. 2016a. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system. *Sci Rep,* 6**,** 19675.

Liu, Y., Chen, Z., He, A., Zhan, Y., Li, J., Liu, L., et al. 2016b. Targeting cellular mRNAs translation by CRISPR-Cas9. *Sci Rep,* 6**,** 29652.

Liu, Y., Kim, Y. J., Ji, M., Fang, J., Siriwon, N., Zhang, L. I., et al. 2014. Enhancing gene delivery of adeno-associated viruses by cell-permeable peptides. *Mol Ther Methods Clin Dev,* 1**,** 12.

Long, C., Mcanally, J. R., Shelton, J. M., Mireault, A. A., Bassel-Duby, R. & Olson, E. N. 2014. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science,* 345**,** 1184-8.

Longo, P. A., Kavran, J. M., Kim, M. S. & Leahy, D. J. 2013. Transient mammalian cell transfection with polyethylenimine (PEI). *Methods Enzymol,* 529**,** 227-40.

Ma, Q., Jones, D., Borghesani, P. R., Segal, R. A., Nagasawa, T., Kishimoto, T., et al. 1998. Impaired B-lymphopoiesis, myelopoiesis, and derailed cerebellar neuron migration in CXCR4- and SDF-1-deficient mice. *Proc Natl Acad Sci U S A,* 95**,** 9448-53.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol,* 13**,** 722-36.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol,* 31**,** 833-8.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., Dicarlo, J. E., et al. 2013b. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

Mandal, P. K., Ferreira, L. M., Collins, R., Meissner, T. B., Boutwell, C. L., Friesen, M., et al. 2014. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell,* 15**,** 643-52.

Marmor, M., Sheppard, H. W., Donnell, D., Bozeman, S., Celum, C., Buchbinder, S., et al. 2001. Homozygous and heterozygous CCR5-Delta32 genotypes are associated with resistance to HIV infection. *J Acquir Immune Defic Syndr,* 27**,** 472-81.

Mashal, R. D., Koontz, J. & Sklar, J. 1995. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat Genet,* 9**,** 177-83.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., et al. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol,* 29**,** 143-8.

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.

Mojica, F. J., Diez-Villasenor, C., Soria, E. & Juez, G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol,* 36**,** 244-6.

Monel, B., Beaumont, E., Vendrame, D., Schwartz, O., Brand, D. & Mammano, F. 2012. HIV cell-to-cell transmission requires the production of infectious virus particles and does not proceed through env-mediated fusion pores. *J Virol,* 86**,** 3924-33.

Montefiori, D. C. 2009. Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol Biol,* 485**,** 395-405.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., et al. 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods,* 12**,** 982-8.

Mummidi, S., Adams, L. M., Vancompernolle, S. E., Kalkonde, M., Camargo, J. F., Kulkarni, H., et al. 2007. Production of specific mRNA transcripts, usage of an alternate promoter, and octamer-binding transcription factors influence the surface expression levels of the HIV coreceptor CCR5 on primary T cells. *J Immunol,* 178**,** 5668-81.

Mummidi, S., Ahuja, S. S., Mcdaniel, B. L. & Ahuja, S. K. 1997. The human CC chemokine receptor 5 (CCR5) gene. Multiple transcripts with 5′-end heterogeneity, dual promoter usage, and evidence for polymorphisms within the regulatory regions and noncoding exons. *J Biol Chem,* 272**,** 30662-71.

Nathwani, A. C., Reiss, U. M., Tuddenham, E. G., Rosales, C., Chowdary, P., Mcintosh, J., et al. 2014. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med,* 371**,** 1994-2004.

Nelles, D. A., Fang, M. Y., O'connell, M. R., Xu, J. L., Markmiller, S. J., Doudna, J. A., et al. 2016. Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell,* 165**,** 488-96.

Nishimasu, H., Cong, L., Yan, W. X., Ran, F. A., Zetsche, B., Li, Y., et al. 2015. Crystal Structure of Staphylococcus aureus Cas9. *Cell,* 162**,** 1113-26.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell,* 156**,** 935-49.

Pandit, A. & De Boer, R. J. 2015. HIV-1 CCR5 gene therapy will fail unless it is combined with a suicide gene. *Sci Rep,* 5**,** 18088.

Park, J., Bae, S. & Kim, J. S. 2015. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics,* 31**,** 4014-6.

Parmentier, M. 2015. CCR5 and HIV Infection, a View from Brussels. *Front Immunol,* 6**,** 295.

Passaes, C. P., Bruel, T., Decalf, J., David, A., Angin, M., Monceaux, V., et al. 2017. Ultrasensitive HIV-1 p24 Assay Detects Single Infected Cells and Differences in Reservoir Induction by Latency Reversal Agents. *J Virol,* 91.

Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., et al. 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol,* 26**,** 808-16.

Peterson, C. W., Haworth, K. G., Burke, B. P., Polacino, P., Norman, K. K., Adair, J. E., et al. 2016. Multilineage polyclonal engraftment of Cal-1 gene-modified cells and in vivo selection after SHIV infection in a nonhuman primate model of AIDS. *Mol Ther Methods Clin Dev,* 3**,** 16007.

Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B. & Kabat, D. 1998. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J Virol,* 72**,** 2855-64.

Pliatsika, V., Rigoutsos, I. 2015. Off-Spotter: very fast and exhaustive enumeration of genomic lookalikes for designing CRISPR/Cas guide RNAs. *Biology Direct.*

Price, A. A., Sampson, T. R., Ratner, H. K., Grakoui, A. & Weiss, D. S. 2015. Cas9-mediated targeting of viral RNA in eukaryotic cells. *Proc Natl Acad Sci U S A,* 112**,** 6164-9.

Prykhozhij, S. V., Rajan, V., Gaston, D. & Berman, J. N. 2015. CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. *PLoS One,* 10**,** e0119372.

Qi, C., Jia, X., Lu, L., Ma, P. & Wei, M. 2016. HEK293T Cells Are Heterozygous for CCR5 Delta 32 Mutation. *PLoS One,* 11**,** e0152975.

Ramirez, C. L., Certo, M. T., Mussolino, C., Goodwin, M. J., Cradick, T. J., Mccaffrey, A. P., et al. 2012. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res,* 40**,** 5560-8.

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., et al. 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature,* 520**,** 186-91.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-9.

Reshef, R., Luger, S. M., Hexner, E. O., Loren, A. W., Frey, N. V., Nasta, S. D., et al. 2012. Blockade of lymphocyte chemotaxis in visceral graft-versus-host disease. *N Engl J Med,* 367**,** 135-45.

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

Reynes, J., Portales, P., Segondy, M., Baillat, V., Andre, P., Reant, B., et al. 2000. CD4+ T cell surface CCR5 density as a determining factor of virus load in persons infected with human immunodeficiency virus type 1. *J Infect Dis,* 181**,** 927-32.

Ribeiro, R. M., Hazenberg, M. D., Perelson, A. S. & Davenport, M. P. 2006. Naive and memory cell turnover as drivers of CCR5-to-CXCR4 tropism switch in human immunodeficiency virus type 1: implications for therapy. *J Virol,* 80**,** 802-9.

Richardson, M. W., Jadlowsky, J., Didigu, C. A., Doms, R. W. & Riley, J. L. 2012. Kruppel-like factor 2 modulates CCR5 expression and susceptibility to HIV-1 infection. *J Immunol,* 189**,** 3815-21.

Richon, V. M. 2006. Cancer biology: mechanism of antitumour action of vorinostat (suberoylanilide hydroxamic acid), a novel histone deacetylase inhibitor. *British Journal of Cancer,* 95**,** S2-S6.

Riolobos, L., Hirata, R. K., Turtle, C. J., Wang, P. R., Gornalusse, G. G., Zavajlevski, M., et al. 2013. HLA engineering of human pluripotent stem cells. *Mol Ther,* 21**,** 1232-41.

Rosen, L. E., Morrison, H. A., Masri, S., Brown, M. J., Springstubb, B., Sussman, D., et al. 2006. Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res,* 34**,** 4791-800.

Rudin, N. & Haber, J. E. 1988. Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences. *Mol Cell Biol,* 8**,** 3918-28.

Saayman, S. M., Lazar, D. C., Scott, T. A., Hart, J. R., Takahashi, M., Burnett, J. C., et al. 2016. Potent and Targeted Activation of Latent HIV-1 Using the CRISPR/dCas9 Activator Complex. *Mol Ther,* 24**,** 488-98.

Salkowitz, J. R., Bruse, S. E., Meyerson, H., Valdez, H., Mosier, D. E., Harding, C. V., et al. 2003. CCR5 promoter polymorphism determines macrophage CCR5 density and magnitude of HIV-1 propagation in vitro. *Clin Immunol,* 108**,** 234-40.

Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. 1996a. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry,* 35**,** 3362-7.

Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C. M., et al. 1996b. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature,* 382**,** 722-5.

Sarzotti-Kelsoe, M., Bailer, R. T., Turk, E., Lin, C. L., Bilska, M., Greene, K. M., et al. 2014. Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J Immunol Methods,* 409**,** 131-46.

Sather, B. D., Romano Ibarra, G. S., Sommer, K., Curinga, G., Hale, M., Khan, I. F., et al. 2015. Efficient modification of CCR5 in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci Transl Med,* 7**,** 307ra156.

Satomura, A., Nishioka, R., Mori, H., Sato, K., Kuroda, K. & Ueda, M. 2017. Precise genome-wide base editing by the CRISPR Nickase system in yeast. *Sci Rep,* 7**,** 2095.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 9**,** 671-5.

Schroder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R. & Bushman, F. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell,* 110**,** 521-9.

Schumann, K., Lin, S., Boyer, E., Simeonov, D. R., Subramaniam, M., Gate, R. E., et al. 2015. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc Natl Acad Sci U S A,* 112**,** 10437-42.

Schweighardt, B., Roy, A. M., Meiklejohn, D. A., Grace, E. J., 2nd, Moretto, W. J., Heymann, J. J., et al. 2004. R5 human immunodeficiency virus type 1 (HIV-1) replicates more efficiently in primary CD4+ T-cell cultures than X4 HIV-1. *J Virol,* 78**,** 9164-73.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., et al. 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods,* 11**,** 399-402.

Shen, R., Richter, H. E. & Smith, P. D. 2011. Early HIV-1 target cells in human vaginal and ectocervical mucosa. *Am J Reprod Immunol,* 65**,** 261-7.

Shimizu, S., Kamata, M., Kittipongdaja, P., Chen, K. N., Kim, S., Pang, S., et al. 2009. Characterization of a potent non-cytotoxic shRNA directed to the HIV-1 co-receptor CCR5. *Genet Vaccines Ther,* 7**,** 8.

Shimizu, S., Ringpis, G. E., Marsden, M. D., Cortado, R. V., Wilhalme, H. M., Elashoff, D., et al. 2015. RNAi-Mediated CCR5 Knockdown Provides HIV-1 Resistance to Memory T Cells in Humanized BLT Mice. *Mol Ther Nucleic Acids,* 4**,** e227.

Siliciano, R. F. & Greene, W. C. 2011. HIV latency. *Cold Spring Harb Perspect Med,* 1**,** a007096.

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X. & Zhang, F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science,* 351**,** 84-8.

Sloan, R. D., Kuhl, B. D., Mesplede, T., Munch, J., Donahue, D. A. & Wainberg, M. A. 2013. Productive entry of HIV-1 during cell-to-cell transmission via dynamin-dependent endocytosis. *J Virol,* 87**,** 8110-23.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., et al. 2014. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell,* 15**,** 12-3.

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.

Song, M. 2017. The CRISPR/Cas9 system: Their delivery, in vivo and ex vivo applications and clinical development by startups. *Biotechnol Prog*.

Sourisseau, M., Sol-Foulon, N., Porrot, F., Blanchet, F. & Schwartz, O. 2007. Inefficient human immunodeficiency virus replication in mobile lymphocytes. *J Virol,* 81**,** 1000-12.

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.

Sundquist, W. I. & Krausslich, H. G. 2012. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med,* 2**,** a006924.

Sussman, D., Chadsey, M., Fauce, S., Engel, A., Bruett, A., Monnat, R., Jr., et al. 2004. Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J Mol Biol,* 342**,** 31-41.

Tabebordbar, M., Zhu, K., Cheng, J. K., Chew, W. L., Widrick, J. J., Yan, W. X., et al. 2016. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science,* 351**,** 407-11.

Tebas, P., Stein, D., Tang, W. W., Frank, I., Wang, S. Q., Lee, G., et al. 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med,* 370**,** 901-10.

Thomas, K. R., Folger, K. R. & Capecchi, M. R. 1986. High frequency targeting of genes to specific sites in the mammalian genome. *Cell,* 44**,** 419-28.

Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. 2016. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun,* 7**,** 11750.

Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., et al. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol,* 33**,** 187-97.

Tycko, J., Myer, V. E. & Hsu, P. D. 2016. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol Cell,* 63**,** 355-70.

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., et al. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature,* 435**,** 646-51.

Von Laer, D., Hasselmann, S. & Hasselmann, K. 2006. Gene therapy for HIV infection: what does it need to make it work? *J Gene Med,* 8**,** 658-67.

Vriend, L. E., Prakash, R., Chen, C. C., Vanoli, F., Cavallo, F., Zhang, Y., et al. 2016. Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. *Nucleic Acids Res,* 44**,** 5204-17.

Wang, G., Zhao, N., Berkhout, B. & Das, A. T. 2016a. CRISPR-Cas9 Can Inhibit HIV-1 Replication but NHEJ Repair Facilitates Virus Escape. *Mol Ther,* 24**,** 522-6.

Wang, J., Friedman, G., Doyon, Y., Wang, N. S., Li, C. J., Miller, J. C., et al. 2012. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Research,* 22**,** 1316-26.

Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science,* 343**,** 80-4.

Wang, Z., Pan, Q., Gendron, P., Zhu, W., Guo, F., Cen, S., et al. 2016b. CRISPR/Cas9-Derived Mutations Both Inhibit HIV-1 Replication and Accelerate Viral Escape. *Cell Rep,* 15**,** 481-9.

Welsch, S., Keppler, O. T., Habermann, A., Allespach, I., Krijnse-Locker, J. & Krausslich, H. G. 2007. HIV-1 buds predominantly at the plasma membrane of primary human macrophages. *PLoS Pathog,* 3**,** e36.

Wolstein, O., Boyd, M., Millington, M., Impey, H., Boyer, J., Howe, A., et al. 2014. Preclinical safety and efficacy of an anti-HIV-1 lentiviral vector containing a short hairpin RNA to CCR5 and the C46 fusion inhibitor. *Mol Ther Methods Clin Dev,* 1**,** 11.

Wu, L., Paxton, W. A., Kassam, N., Ruffing, N., Rottman, J. B., Sullivan, N., et al. 1997. CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, in vitro. *J Exp Med,* 185**,** 1681-91.

Wu, Z., Yang, H. & Colosi, P. 2010. Effect of genome size on AAV vector packaging. *Mol Ther,* 18**,** 80-6.

Xu, H., Xiao, T., Chen, C. H., Li, W., Meyer, C. A., Wu, Q., et al. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Res,* 25**,** 1147-57.

Xu, X., Duan, D. & Chen, S. J. 2017. CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Sci Rep,* 7**,** 143.

Ye, L., Wang, J., Beyer, A. I., Teque, F., Cradick, T. J., Qi, Z., et al. 2014. Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. *Proc Natl Acad Sci U S A,* 111**,** 9591-6.

Yin, C., Zhang, T., Qu, X., Zhang, Y., Putatunda, R., Xiao, X., et al. 2017. In Vivo Excision of HIV-1 Provirus by saCas9 and Multiplex Single-Guide RNAs in Animal Models. *Mol Ther*.

Zagordi, O., Klein, R., Daumer, M. & Beerenwinkel, N. 2010. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res,* 38**,** 7400-9.

Zhu, L. J., Holmes, B. R., Aronin, N. & Brodsky, M. H. 2014. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One,* 9**,** e108424.

Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. 2008. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol Ther,* 16**,** 1073-80.

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., et al. 2015. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol,* 33**,** 73-80.

### 6.2. Websites

Mathews lab, University of Rochester. 2016. RNAstructure, Version 5.8.1. Available from: https://rna.urmc.rochester.edu/RNAstructure.html. New York, USA. [Accessed 15 March 2017].

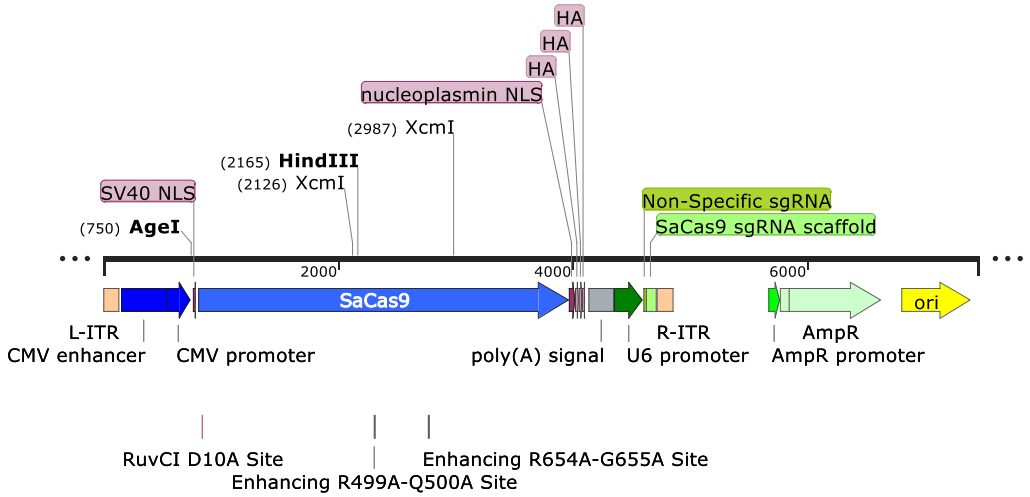
Sangamo Therapeutics. 2014. A phase 1/2, open-label study to assess the safety and tolerability of repeat doses of autologous T-cells genetically modified at the *CCR5* gene by zinc finger nucleases in HIV-infected subjects following cyclophosphamide conditioning [Online]. Available from: https://clinicaltrials.gov/ct2/show/NCT02225665 [Accessed 10 May 2017].

UNAIDS. 2015. AIDSinfo [Online]. Available from: http://www.unaids.org/sites/default/files/media_asset/20150714_FS_MDG6_Report_en.pdf. [Accessed 28 April 2017].

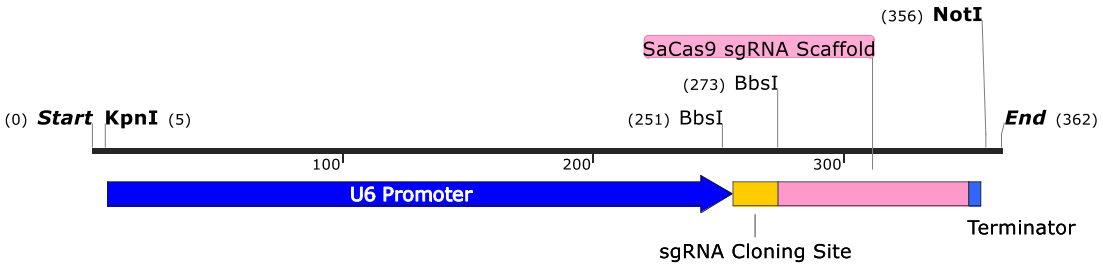
# 7. Appendix

## 7.1. Plasmid construct maps

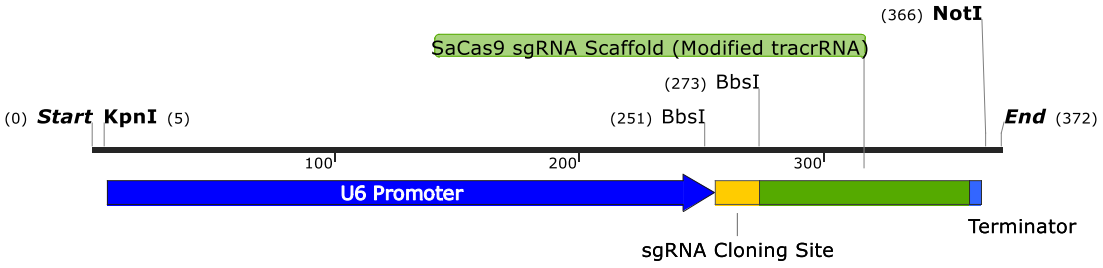
### 7.1.1. pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA vector (pAAV-CMV_SaCas9)




### 7.1.2. pTZ57R/T_U6-BbsI-sgRNA-SaCas9_tracrRNA vector (U6-sgRNA)




### 7.1.3. pTZ57R/T_U6-BbsI-sgRNA-SaCas9_modtracrRNA (U6-sgRNA_modtracr)



## 7.2. Supplementary Sequences

### 7.2.1. *Sa*Cas9

AAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTA
CGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAAC
GAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAG
AGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAA
CCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCC
TGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAA
CGAGCTGTCCACCAgAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCG
AACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGAC
CAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACC
AGAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGC
GAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCAC
CTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGA
ACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTC
CAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAA
TCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACGCCGA
GCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAAC
TGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTAT
ACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAA
CGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC
AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTC
ATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTAT
CGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGG
AACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGT
ACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGC
GTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGG
GCAACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAG
AAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCT
GGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATA
CCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGAC
GTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAA
AGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCA
TCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGA
AAgGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCC
CCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAG

CCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACACCCT
GATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACA
AGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATT
ATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCT
GACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAAC
TGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCC
CTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAA
TCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGA
AGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAG
ATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAA
CATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTG
TATGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGC

### 7.2.2. *Sa*Cas9-D10A

AAGCGGAACTACATCCTGGGCCTGGCCATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTA
CGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAAC
GAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAG
AGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAA
CCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCC
TGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAA
CGAGCTGTCCACCAgAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCG
AACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGAC
CAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACC
AGAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGC
GAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCAC
CTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGA
ACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTC
CAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAA
TCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACGCCGA
GCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAAC
TGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTAT
ACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAA
CGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC
AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTC
ATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTAT
CGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGG
AACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGT
ACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC

ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGC
GTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGG
GCAACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAG
AAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCT
GGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATA
CCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGAC
GTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAA
AGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCA
TCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGA
AAgGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCC
CCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAG
CCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACACCCT
GATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACA
AGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATT
ATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCT
GACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAAC
TGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCC
CTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAA
TCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGA
AGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAG
ATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAA
CATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTG
TATGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGC

### 7.2.3. e*Sa*Cas9

AAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTA
CGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAAC
GAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAG
AGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAA
CCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCC
TGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAA
CGAGCTGTCCACCAgAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCG
AACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGAC
CAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACC
AGAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGC
GAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCAC
CTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGA
ACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTC
CAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAA

TCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACC
AACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACGCCGA
GCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAAC
TGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTAT
ACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAA
CGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGC
AGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTC
ATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTAT
CGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGG
AACGCCGCCACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGT
ACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGCC
ATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGC
GTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGG
GCAACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAG
AAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCT
GGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATA
CCAGATACGCCACCGCCGCCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGAC
GTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAA
AGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCA
TCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGA
AAgGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCC
CCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAG
CCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACACCCT
GATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACA
AGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATT
ATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCT
GACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAAC
TGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCC
CTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAA
TCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGA
AGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAG
ATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAA
CATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTG
TATGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGC

### 7.2.4. e*Sa*Cas9-D10A

AAGCGGAACTACATCCTGGGCCTGGCCATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTA
CGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAAC
GAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAG

AGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAA
CCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCC
TGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAA
CGAGCTGTCCACCAGAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCC
GAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGA
CCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGAC
CAGAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGG
CGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCA
CCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTG
AACGACCTGAACAATCTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTT
CCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAA
ATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCAC
CAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACGCCG
AGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAA
CTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTA
TACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCA
ACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAG
CAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTT
CATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTA
TCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCG
GAACGCCGCCACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAG
TACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCTGGAAGC
CATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAG
CGTGTCCTTCGACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGG
GCAACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAG
AAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCT
GGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATA
CCAGATACGCCACCGCCGCCCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGAC
GTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAA
AGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCA
TCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGA
AAgGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCC
CCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAG
CCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACACCCT
GATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACA
AGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAGAAACTGAAGCTGATT
ATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCT
GACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAAC
TGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCC
CTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAA

TCTGGATGTGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGA
AGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAG
ATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAA
CATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCA
TTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTG
TATGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGC

### 7.2.5. *Sa*Cas9 U6-sgRNA cloning vector

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATT
GGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT
TCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGA
AAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGTCTTCGAGAAGA
CCTGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAA
CTTGTTGGCGAGATTTTTGCGGCCGC

U6 promoter, *BbsI* sgRNA cloning site and *Sa*Cas9-specific sgRNA tracrRNA (scaffold)

### 7.2.6. *Sa*Cas9 U6-sgRNA_modtracr cloning vector

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATT
GGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT
TCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGA
AAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGTCTTCGAGAAGA
CCTGTTTAAGTACTCTGTGCTGGAAACAGCACAGAATCTACTTAAACAAGGCAAAATGCCGTGTTT
ATCTCGTCAACTTGTTGGCGAGATTTTTTTTTTGCGGCCGC

U6 promoter, *BbsI* sgRNA cloning site and A-U Flip as well as 5 bp extended tetraloop sequences within the *Sa*Cas9-specific sgRNA tracrRNA (scaffold)

### 7.2.7. pTZ57R/T-H1_sh1005 construct

CTCGAGTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCATAAACGTGAAATGTCTTTG
GATTTGGGAATCTTATAAGTTCTGTATGAGACCACTCAGATCCGAGCAAGCTCAGTTTACACCTTG
TCCGACGGTGTAAACTGAGCTTGCTCTTTTTTGGCGCGCC

*XhoI* and *AscI* restriction sites, H1 promoter and sense-loop-antisense sequences

### 7.2.8. Amplicon Sequences for Deep Sequencing

(sgRNA 7 and sgRNA 8 target sites shown, including PAM recognition sites)

### *7.2.8.1. CCR5*

CTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAATCCTAAAAACTCTGCTTCGGTGTCGAAAT
GAGAAGAAGAGGCACAGGGCTGTGAGGCTTATCTTCACCATCATGATTGTTTATTTTCTCTTCTGG
GCTCCCTACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAATAATTGCAGTA
GCTCTAACAGGTTGGACCAAGCTATGCAGGTGACAGAGACTCTTGGGATGACGCACTGCTGCATC
AACCCCATCATCTATGCCTTTGTCGGGGAGAAGTTCA

### *7.2.8.2. LCP1*

CCAGCCACTGCAGGAGTGAGTTCTGTCCTGACTATGGAACGACTGTGACTGTGACCCTCCAGGGTC
CCATTCGTGTTGTGTTTCTCAGGCTCTCCTTATTTTTCCTCTTTTCTCCCTTTTAGATATGCCATCTCT
ATGGCCCGAAAAATTGGAGCAAGAGTGTATGCCCTGCCAGAAGACCTGGTTGAAGTGAACCCCAA
AATGGTCATGACCGTGTTTGCCTGCCTCATGGGGAAAGGAATGAAGAGGGTGTGAGGCCCAATGG
GGCTGGGTGGGAGGCGGTGCACTCACTCCTGACT

### *7.2.8.3. CNTN3*

CACCCATATGTATGAATCATGTGCCAATGATGAAATTATTTTAAATATTAAAAAAACATTGCATTA
TCTTTTAGATAAATATTACTAAATGCAAGTACCATTTCATCTCTGTTTCTCTGACCTGTATAGTTCC
AGAAGTGCCTCCTTCTGAAGTCAATGGAGGAGGCGGAAGCCGGTCTGAACTTGTGATAACCTGGG
ATGTAAGTGTTTGGGCAGCATCTTCCCCATCAGGGTCTTGCTGTTCCCATGATCATTATGGCTTCCC
AGGTGGTGGTGCCAATGGTGCAGGTGGTGGTGGTAGTGC

### *7.2.8.4. SLITRK2*

CAACAGTTAGCCTGCTCCAGCCCCCCCAGTATCGAATCTATCAGCTTTTTCTCAATGGAAACCTCTT
GACAAGACTGTATCCAAACGAATTTGTCAATTACTCCAACGCGGTGACTCTTCACCTAGGTAACAA
CGGGTTACAGGAGATCCGAACGGGGGCATTCAGTGGCCTGAAAACTCTCAAAAGACTGCATCTCA
ACAACAACAAGCTTGAGATATTGAGGGAGGACACCTTCCTAGGCCTGGAGAGCCTGGAGTATCTC
CAGGCCGACTACAATTACATCAGTGCCATCGAGGCTG

### *7.2.8.5. BMPR1B*

GCCTTAGAAAACCCAGACACATAGCAGTGGAAATGAAAGAAATGGCATCAGAAGTGACTTAATTT
AGCAATTGTGATTCCTCTTGTAAAACAAAACAAAAAAACAATGCCATATTTTTTGGAGAAAAGTTG
GCAATATAGGGGTTTCGTTGTCTGTTTCACAAGAAGACTCATTTGTTCTTTTGGGGGAACCAGTGC
CTTACAGATTTTGTATATACTGTAATTATTCAGGACTAGGGAACAAACAATTGTATTGTATTTGTTA
CAGATTGTATATGGCTTTGTTTTAACATTCCCCTAAATAAAATGGCTTCATTCTCCCCTTGGAAA

### *7.2.8.6. FLNB*

CTCAAGTACCTGTCCCCTCGCTCTGGTGATTATTTCTTGCAGAATCACCACACGAGACCATCCCGG
CAGTCATGGTTTTGCTTTAGTTTTCCAAGTCCGTTTCAGTCCCTTCCTTGGTCTGAAGAAATTCTGC

AGTGGCGAGCAGTTTCCCACTTGCCAAAGATCCCTTTTAACCAACACTAGCCCTTGTTTTTAACAC
ACGCTCCAGCCCTTCATCAGCCTGGGCAGTCTTACCAAAATGTTTAAAGTGATCTCAGAGGGGCCC
ATGGATTAACGCCCTCATCCCAAGGTCCGTCCCATG

### *7.2.8.7. CCR2*

GGCAGTGAGAGTCATCTTCACCATCATGATTGTTTACTTTCTCTTCTGGACTCCCTATAATATTGTC
ATTCTCCTGAACACCTTCCAGGAATTCTTCGGCCTGAGTAACTGTGAAAGCACCAGTCAACTGGAC
CAAGCCACGCAGGTGACAGAGACTCTTGGGATGACTCACTGCTGCATCAATCCCATCATCTATGCC
TTCGTTGGGGAGAAGTTCAGAAGGTATCTCTCGGTGTTCTTCCGAAAGCACATCACCAAGCGCTTC
TGCAAACAATGTCCAGTTTTCTACAGGGAGACAGTG

### *7.2.9.8. MCM3AP*

CTGTCCTGGAACTCTCATCTGTGGGCTGGTCTCCATCCAGGCAACATGTGCCTTAGTGCCATCATTG
TGAGCATATGTGTCCCTGCTCCTCCACTTGGAAGCTCTGTGACCTTGGGCAAGTTCCCTAGCGTCTC
TGCCACCTGTGTATGGGTACTGTTACTTATTATCCCTATTTTACTCAGAGACGGTAAAATAAACTGC
CCAAGCCTCTGCAGGTGCTGTCTGGTGGAGATGCTCATCTGGCTCTTGAGGATAAAATGAGTGCAA
ACACAGACTGCACAGGACAGTGCCTGGCACAG

### *7.2.8.9. SRL*

GCCACCAAGGCTTAACATTGACCTTCCCGCCTGACCTTGATGCAGATGTCCACTGAACACACCGCA
GGAAAGCCAGGGCCTTCAATACCAATAAGTGTGAATATGTGTGTATGTTGTCCAAGAGAGATTAG
GGAGATCACATAGACTCTAGGGAGTAGAGAACTTGTAACAGTCTTGCAAGGCTAGCATGCACGGC
TCCACAGCAGGTGGTGGGGAGCAGAGGGGCAGGACCTGCAGGGAAGAAGCAGCCTTTGGATGGT
GAAATGTGCATGGTGCACAGTCTGTGCATGCCCAGGAGA

### *7.2.8.10. TOMM5*

ACAGGACATCACATATGAATGCACGATATGAAGAGCCTGGTTACAGTTTCGACTCCTCTCTGCAAG
TGAATAGGCCCAGAAAGGTGTAAGAGACTCTTTGAATGGACATAAAATTCTGCTTGTTAAGAACA
AGTTTGGCTCTGGTAACTGACCTTCAAAGCTAAAATATAAAACTATTTGGGAAGTATGAAACGATG
TCTCGTGATCTGGTGTACCCTTATCCCTGTGACGTTTGGCCTCTGACAATACTGGTATAATTGTAAA
TAATGTCAAACTCCGTTTTCTAGCAAGTATTAAGGGAGCTGTGTCTG

### *7.2.8.11. POLL*

GCTGACTCGGAAGCTATTCTGGCCATTTGCCCTCCTTCCCCCCTTCGTCCGCTCTCATTGGCTCTGC
TGGTAAGTGGTCTATTCCTGCCCACCCCCGGGTGACTAGCTTGGCCAGTAGTCGACCCCACCCGGG
GACCGACTCTGGGGGTTGGAGAGACTCTTGGGGCCGGGGTCGGGCACTCCAGCTTTCTTCTAGCCC
CGAGCTGGGATTCCCTGGCCTGCGCCAGCTGCGTACACGGCGAGTACACCGCACCTGCCCGGGAC
TTCACCCGCAGCTGCGAGACTCCTCCATTCCCGGA

### 7.3. Supplementary reagents and recipes

#### 7.3.1. 1 X Phosphate Buffered Saline solution

1 PBS Tablet (Sigma-Aldrich, USA)
Double distilled $H_20$ ($ddH_20$)

One PBS tablet was dissolved in 500 ml of $ddH_20$ and autoclaved at 121 °C and 2 $kg/cm^2$ for 20 minutes to sterilise.

#### 7.3.2. Freeze medium

High Glucose DMEM (Thermo Fisher Scientific, USA)
FBS (Thermo Fisher Scientific, USA)
Dimethyl sulfoxide (DMSO; Merck, Germany)

A solution containing 40 % DMEM, 50 % FBS and 10 % dimethyl sulfoxide (DMSO) was prepared, filter sterilised with a 0.8 µm syringe filter and stored at 4 °C.

#### 7.3.3. 1 X TAE Buffer

50X Tris-acetate EDTA (TAE) Buffer containing 40 mM Tris, 20 mM acetic acid, 1 mM Ethylenediaminetetraacetic acid (EDTA; Thermo Scientific™, Lithuania)

50X TAE Buffer was diluted to 1X using $ddH_20$

#### 7.3.4. Restriction enzyme digestion reactions and buffer components

10 U *AgeI/* 10 U *HindIII* (Thermo Fisher Scientific, USA)
1 X R Buffer composition is proprietary information (Thermo Fisher Scientific, USA). Restriction enzyme double digestion was carried out at 37 °C for one hour and enzymes inactivated at 80 °C for twenty minutes.

10 U *BbsI* (New England Biolabs®, USA)
1 X NEB Buffer 2.1 contains 50 mM of NaCℓ, 10 mM Tris-HCℓ, 10 mM $MgC\ell_2$ and 100 µg/ml BSA at pH 7.9 (New England Biolabs®, USA). Restriction enzyme digestion was carried out at 37 °C for one hour and enzymes inactivated at 65 °C for twenty minutes.

10 U *XcmI* (New England Biolabs®, USA)

1 X NEB Buffer 2.1 contains 50 mM of NaCℓ, 10 mM Tris-HCℓ, 10 mM $MgC\ell_2$ and 100 µg/ml BSA at pH 7.9 (New England Biolabs®, USA). Restriction enzyme digestion was carried out at 37 °C for one hour and enzymes inactivated at 65 °C for twenty minutes.

10 U *XhoI*/ 20 U *AscI* (Thermo Fisher Scientific, USA)
1 X R Buffer composition is proprietary information (Thermo Fisher Scientific, USA).
Restriction enzyme double digestion was carried out at 37 °C for one hour and enzymes inactivated at 80 °C for twenty minutes.

### 7.3.5. PIPES transformation buffer

Glycerol (Merck, Germany)
Calcium Chloride ($CaCl_2.2H_2O$; Associated Chemical Enterprises, South Africa)
PIPES (Sigma-Aldrich, USA)

Piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES) transformation buffer was made by mixing 15 ml of glycerol, 1.47 g of calcium chloride and 302.4 mg of PIPES.HCℓ, made up to 80 ml with $ddH_2O$. The solution pH was set to 7.0 by HCℓ titration using the Orion Star A211 pH meter (Thermo Fisher Scientific, USA). The PIPES solution was autoclaved at 121 °C and 2 $kg/cm^2$ for 20 minutes to sterilise and kept at -20 °C.

### 7.3.6. SOC outgrowth medium

1 X Super Optimal broth with Catabolite repression (SOC) Outgrowth Medium (New England Biolabs®, USA) is sterile and contains:

2% Vegetable Peptone
0.5% Yeast Extract
10 mM NaCl
2.5 mM KCl
10 mM MgCl2
10 mM MgSO4
20 mM Glucose

### 7.3.7. LB/Amp broth and plates

10 g of Tryptone (Oxoid™, Thermo Fisher Scientific, USA)
5 g Yeast Extract (Oxoid™, Thermo Fisher Scientific, USA)

5 g Sodium Chloride (NaCℓ, Merck, Germany)
dd$H_2O$
12 g Agar Bacteriological (Oxoid™, Thermo Fisher Scientific, USA)

Lysogeny Broth (LB, Lennox) was made by dissolving 10 g of Tryptone, 5 g of Yeast Extract and 5 g of NaCℓ in 1 L of dd$H_2O$. LB Agar was made in the same way with the addition of 12 g of Agar Bacteriological. Solutions were autoclaved at 121 °C and 2 kg/$cm^2$ for 20 minutes to sterilise. An Ampicillin stock solution of 100 mg/ml was made by dissolving 1 g of Ampicillin (Sodium Salt, Irradiated; Thermo Fisher Scientific, USA) in 10 ml of dd$H_20$ and this was stored at -20 °C. A final concentration of 100µg/ml was used as required. LB/Amp plates were made by pouring the cooled liquid solution into petri dishes and these were then allowed to set. LB plates and broth were stored at 4°C for no longer than two months.

### 7.3.8. Polyethylenimine "Max" Mw 40 000 solution

Polyethylenimine "Max" Mw 40 000 (PEI Max; Polysciences, USA)
dd $H_20$
10 M sodium hydroxide (NaOH)

To make a 5 mg/ml stock solution, 75 mg of PEI Max was weighed and suspended in 13.5 ml dd $H_20$ in a beaker. The mixture was stirred until the powder completely dissolved and 10 M NaOH added dropwise until it reached pH 7. The solution was transferred to a graduated cylinder and the final volume adjusted to 15 ml before filter sterilising using a 0.22 µm syringe filter. The stock solution was stored at 4 °C and diluted to 1 mg/ml using dd$H_20$ to create a working stock solution.

## 7.4. Supplementary protocols

### 7.4.1. Agarose gel electrophoresis

SeaKem® LE Agarose (Lonza Group, USA ) was used in the preparation of agarose gels for electrophoretic separation of DNA fragments based on molecular weight. Either a 1% or 2% gel was made for improved clarity in separation of larger or smaller fragments respectively. The required mass (grams) of SeaKem® LE Agarose was dissolved in 1X TAE electrophoresis buffer (Thermo Scientific™, USA; Appendix 7.3.3.) in the microwave. A total of 2 µl of 10mg/ml Ethidium Bromide Solution (Molecular Grade, Promega, USA) per 100 µl of agarose solution was added and the flask swirled before pouring the slightly cooled liquid into the gel tray. The comb was inserted and the gel allowed to set. The set gel and tray were placed into the electrophoresis chamber (Cleaver Scientific, England) and the comb removed carefully. The chamber was filled with 1X TAE buffer making sure to cover the entire gel and wells. Either 2µl of 6X BlueJuice™ (Invitrogen™, Thermo Scientific™, USA) or 6X Orange DNA Loading Dye (Thermo Scientific™, USA) was added to each sample depending on whether larger or smaller fragments were to be run on the gel. Between 1 and 4 µl of O'GeneRuler DNA Ladder Mix (Thermo Scientific™, USA) was loaded alongside the sample lanes for band size determination. The 1 % gels were resolved at 60V for 15 minutes and 100V thereafter and 2 % gels resolved at 50V for 15 minutes and 80V thereafter. The initial low voltage run allows for the DNA to exit the wells without band distortion and thereafter the voltage is increased to speed up the resolution process. All gels were imaged using UV light in the ChemiDoc™ XRS+ imager and analysed with Quantity One® 1-D Analysis Software, v4.6.9 (Bio-Rad Laboratories, Inc., California, USA). Exposure times were determined based on EtBr staining intensity.

### 7.4.2. GeneJET PCR Purification

All centrifugation steps were carried out at 13 000 rpm at room temperature (Centrifuge 5 415, Eppendorf, Germany). A 1:1 ratio of Binding Buffer (Thermo Fisher Scientific, USA) to PCR product was made up, with 10 µl of 3M sodium acetate (pH 7.0) if the solution was not orange and not yellow. This was vortexed and transferred to the GeneJET PCR Purification column and centrifuged for one minute. The flow-through was discarded and 700 µl of Wash Buffer added to the column. The column was centrifuged for one minute and the flow-through discarded before another one minute spin to ensure that residual ethanol from the wash buffer was removed as this can affect the purified DNA quality. To elute the DNA, the

column was transferred to a 1.5 ml microcentrifuge tube and 20 µl of Ambion® nuclease-free water (Thermo Fisher Scientific, USA) was applied directly to the column filter. This was allowed to stand for one minute before a one minute centrifugation. The sample concentrations and purities were determined using the NanoDrop® Spectrophotometer ND-1000 v3.8.1 (Thermo Fisher Scientific, USA).

### 7.4.3. QIAquick® Gel Extraction

All centrifugation steps were carried out at 13 000 rpm at room temperature (Centrifuge 5 415, Eppendorf, Germany). The AccuBlock™ Digital Dry Bath heating block (Labnet International, Taiwan) was set to just below 50 °C before starting. A gel with the product/s to be excised was allowed to run and the appropriate bands visualised with an ultraviolet (UV) light box (UVP Inc., USA) and excised using a clean sharp scalpel. A plastic face shield was worn when using the UV light box to protect the facial skin and corneas from the harsh DNA-mutagenic light and care was taken to reduce the exposure time of the DNA to UV so as not to introduce mutations. Gel fragments were weighed and placed in a 1.5 ml microcentrifuge tube. An equivalent volume of Buffer QG (Qiagen, Germany) to mass of gel was added to each tube and placed at 50 °C in the heating block for no longer than ten minutes to dissolve. One gel volume of room temperature isopropanol was added and the tube inverted to mix. The sample was added to a QIAquick spin column in a 2 ml flow-through tube and centrifuged for one minute. The flow-through was discarded and 500 µl of Buffer QG added and the column centrifuged for one minute. The flow-through was discarded, 750 µl of Buffer PE (containing ethanol) was added and the tube centrifuged for one minute. The flow-through was discarded and the column centrifuged for another minute to remove residual wash buffer. The column was placed in a sterile 1.5 ml microcentrifuge tube and 20 µl of Ambion® nuclease-free water (Thermo Fisher Scientific, USA) applied to the filter. The column was left to stand for one minute at room temperature and a final minute centrifuge carried out to elute the DNA. The sample concentration and purity was determined using the NanoDrop® Spectrophotometer ND-1000 v3.8.1 (Thermo Fisher Scientific, USA).

### 7.4.4. Transformation of *E. coli* DH5α cells

Transformation of NEB Stable Competent *E. coli* (High Efficiency) cells (New England Biolabs, USA) or chemically competent *E. coli* DH5α cells (Appendix 7.4.5.) was carried out by the addition of 8 µl of ligation mixture to between 25 and 100 µl of cells which were thawed on ice. The tubes were left on ice for a further 30 minutes and the reaction tubes were

heat-shocked at 42 °C GLS Aqua 12 Plus water bath (Grant, UK) for 90 seconds, or 45 seconds for Gibson Assembly® (New England Biolabs® Inc., Massachusettes, USA). The cells were recovered on ice for 2 minutes before being plated on sterile LB/Amp agar and grown overnight at 37 °C in an incubator (IncoTherm, Labotec, South Africa). Plates used for blue-white screening of constructs ligated into the pTZ57R/T backbone containing a beta-galactosidase (β-Gal) reporter were treated with 40 µl of 20 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal, Roche, Mannheim, Germany) and 8 µl of 100 mg/ml Isopropyl β-D-1-thiogalactopyranoside (IPTG, Roche, Mannheim, Germany) spread onto them at least fifteen minutes before plating and left with the plate lid slightly ajar to allow for evaporation of the toxic dissolving agent of Xgal, N,N-dimethylformamide (DMF, Sigma-Aldrich, USA). White colonies were an indicator on clone harbouring inserts.

### 7.4.5. Generation of chemically competent *E. coli* DH5α cells

For the generation of chemically competent *E. coli* DH5α cells, a starter culture from was grown in 4 ml of LB/Amp overnight at 37 °C in the orbital shaker incubator (Labotec, South Africa) and 2 ml of this was added to 120 ml of LB/Amp$^-$ (bacterial cells were not resistant to Amp at this point) in a 200 ml conical flask. The cell count was measured using the Biowave S2100 UV/Vis Diode Array Spectrophotometer (Labotec, South Africa) at various points until cells reached an optical density (OD) of between 0.4 and 0.6 (600 nm light wavelength) as this was indicative of log phase growth. Cells were transferred to 50 ml Falcon™ Conical Centrifuge tubes (Thermo Fisher Scientific, USA) and spun at 4 °C and 1 500 rpm for 10 minutes in the Eppendorf Centrifuge 5804R (Eppendorf, Germany). The supernatant was poured off gently and pellets suspended gently in 1 ml of PIPES transformation buffer (Appendix 7.3.5.). A further 4 ml of PIPES buffer was added and cells incubated on ice for 30min. Cells were centrifuged at 4 °C and 1 000 rpm for 10 min, with acceleration and deceleration set on low (at speed setting 4 out of 9). The supernatant was removed and the pellet suspended gently in 2 ml of PIPES buffer. Previously cooled 1.5 ml microcentrifuge tubes on ice were used for 100 µl aliquots of competent cells and these were kept at -80 °C until needed, at which stage they would be thawed on ice.

### 7.4.6. QIAprep® Spin Miniprep

Colonies from the appropriate LB/Amp plates were picked and then grown in 4 ml of LB/Amp (Appendix 7.3.7.) overnight at 37 °C in an Orbital Shaker incubator (Labotec, South Africa). Cells were spun down in 50 ml Falcon™ Conical Centrifuge tubes (Thermo Fisher

Scientific, USA) at 4 000 rpm for 15 minutes at 4 °C and the supernatant discarded. From here on, all centrifugation steps were carried out at 13 000 rpm at room temperature (Centrifuge 5 415, Eppendorf, Germany).The pellet was suspended in 250 µl of suspension Buffer P1 (Qiagen, Germany) containing RNase A to remove any bacterial RNA and transferred to a 1.5 ml microcentrifuge tube. Cell lysis was performed by the addition of 250 µl of Buffer P2 containing LyseBlue reagent and the tube inverted four to six times to mix. This step was not allowed to occur for more than five minutes before 350 µl of Buffer N3 was added for neutralisation. The tubes were immediately inverted four to six times for mixing, noted to be complete when the blue colour turned to milky white. The samples were centrifuged for ten minutes and the supernatant transferred to a 2 ml QIAprep® Spin column. This was centrifuged for one minute and the flow-through discarded. To wash the column, 750 µl of Buffer PE containing ethanol was added and the columns were centrifuged for one minute. The flow-through was discarded and the columns centrifuged for another minute to remove residual Buffer PE as the ethanol could reduce the eluted DNA quality. The columns were placed into sterile 1.5 ml microcentrifuge tubes and 30 – 50 µl of Ambion® nuclease-free water (Thermo Fisher Scientific, USA) was added to the column filter which was left to stand for one minute at room temperature. DNA was eluted by one minute centrifugation and sample concentrations and purities were determined using the NanoDrop® Spectrophotometer ND-1000 v3.8.1 (Thermo Fisher Scientific, USA).

### 7.4.7. QIAGEN® Plasmid Midi and Maxiprep

The method is described for Midi and Maxiprep simultaneously with volumes mentioned for these respectively. Overnight cultures were grown in LB/Amp (Appendix 7.3.7.) at 37 °C or, for prepping of pAAV-CMV-*Sa*Cas9, cultures were grown at 30 °C reduce the risk of AAV inverted terminal repeat (ITR) region loss through recombination. Glycerol stocks for all plasmid Midi or Maxipreps were made by mixing 500 µl of bacterial cell culture with 200 µl of sterile glycerol in a 1.5 ml microcentrifuge tube and these were stored at -80 °C. In 50 ml Falcon™ Conical Centrifuge tubes (Thermo Fisher Scientific, USA), 100/200 ml of culture was pelleted at 4 000 rpm for 15 minutes (Centrifuge 5804R, Eppendorf, Germany). The supernatant was discarded and cells suspended in 4/10 ml of Buffer P1 (Qiagen, Germany) containing RNase A to degrade bacterial RNA. An equivalent volume of Buffer P2 was added and the tubes inverted thoroughly to mix before incubation at room temperature for five minutes. Thereafter, 4/10 ml of pre-chilled Buffer P3 was added and again tubes mixed thoroughly by inversion. These were incubated on ice for fifteen/twenty minutes and then

centrifuged at 5 000 rpm for 90 minutes. A QIAGEN-tip 100/500 was equilibrated with 4/10 ml of Buffer QBT and from here on, all solutions were allowed to pass through the column by gravity flow. The clear supernatant was passed through the column which was then washed with two times 10/30 ml of Buffer QC. The column was transferred to a 15/50 ml Falcon™ Conical Centrifuge tube (Thermo Fisher Scientific, USA) and DNA eluted by the addition of 5/15 ml of Buffer QF. DNA was precipitated by the addition of 3.5/5 ml room temperature isopropanol which was mixed and placed at -80 °C for one hour to improve DNA precipitation. The sample was centrifuged at 5 000 rpm (Centrifuge 5804R, Eppendorf, Germany) for 60 minutes and the supernatant carefully removed and discarded. The DNA pellet was washed with 2/5 ml room temperature 70 % ethanol and centrifuged at 5 000 rpm for fifteen minutes. The supernatant was carefully removed and the DNA pellet air-dried before suspending in 100-200 µl of Ambion® nuclease-free water (Thermo Fisher Scientific, USA). The concentration and purity of the DNA was determined using the NanoDrop® Spectrophotometer ND-1000 v3.8.1 (Thermo Fisher Scientific, USA).

### 7.4.8. KAPA Express Extract DNA Extraction

Cells were lysed by adding 1 X KAPA Express Extract Buffer (KAPA Biosystems, South Africa) and 1 U of Express Extract Enzyme (proprietary information; KAPA Biosystems, South Africa) made up to a final volume of 50 µl with Ambion® nuclease-free water (Thermo Fisher Scientific, USA). The reaction was carried out in a thermal cycler at 75 °C for ten minutes followed by an inactivation step at 95 °C for five minutes. Samples were cooled to 4 °C before centrifuging at 13 000 rpm for one minute to pellet any cellular debris.

### 7.4.9. QIAamp DNA Mini gDNA Extraction

The DNA Purification from Blood or Body Fluids Spin Protocol was followed for genomic DNA (gDNA) extraction using the QIAamp DNA Mini Kit (Qiagen, Germany). All centrifugation steps were carried out at at room temperature (Centrifuge 5 415, Eppendorf, Germany). Pelleted cells were suspended in 200 µl of 1X PBS (Appendix 7.3.1) in 1.5 ml microcentrifuge tubes and 20 µl of QIAGEN Protease was added. The samples were vortexed thoroughly and 200 µl of Buffer AL added. Samples were mixed by pulse vortexing for fifteen seconds and incubated in a heating block at 56 °C for ten minutes. Tubes were spun down briefly and 200 µl of room temperature 96-100 % ethanol added. Samples were mixed again by pulse vortexing for fifteen seconds and spun down to remove droplets from the lid. The mixture was transferred to the QIAamp Mini spin column in a 2 ml collection tube and

centrifuged at 8 000 rpm for one minute. The column was placed into a new 2 ml collection tube and 500 µl of Buffer AW1 added. Samples were centrifuged at 8 000 rpm for one minute, the column placed into a new 2 ml collection tube and 500 µl of Buffer AW2 added. Samples were centrifuged at 13 000 rpm for three minutes, the flow-through discarded, and centrifuged for a further minute. The column was transferred to a clean sterile 1.5 ml microcentrifuge tube and 30 µl of Ambion® nuclease-free water (Thermo Fisher Scientific, USA) added directly to the filter. The column was left to stand for one minute and then centrifuged at 8 000 rpm for one minute to elute the DNA. The sample concentrations and purities were determined using the NanoDrop® Spectrophotometer ND-1000 v3.8.1 (Thermo Fisher Scientific, USA).

### 7.4.10. Qubit® dsDNA High Sensitivity (HS) Quantification

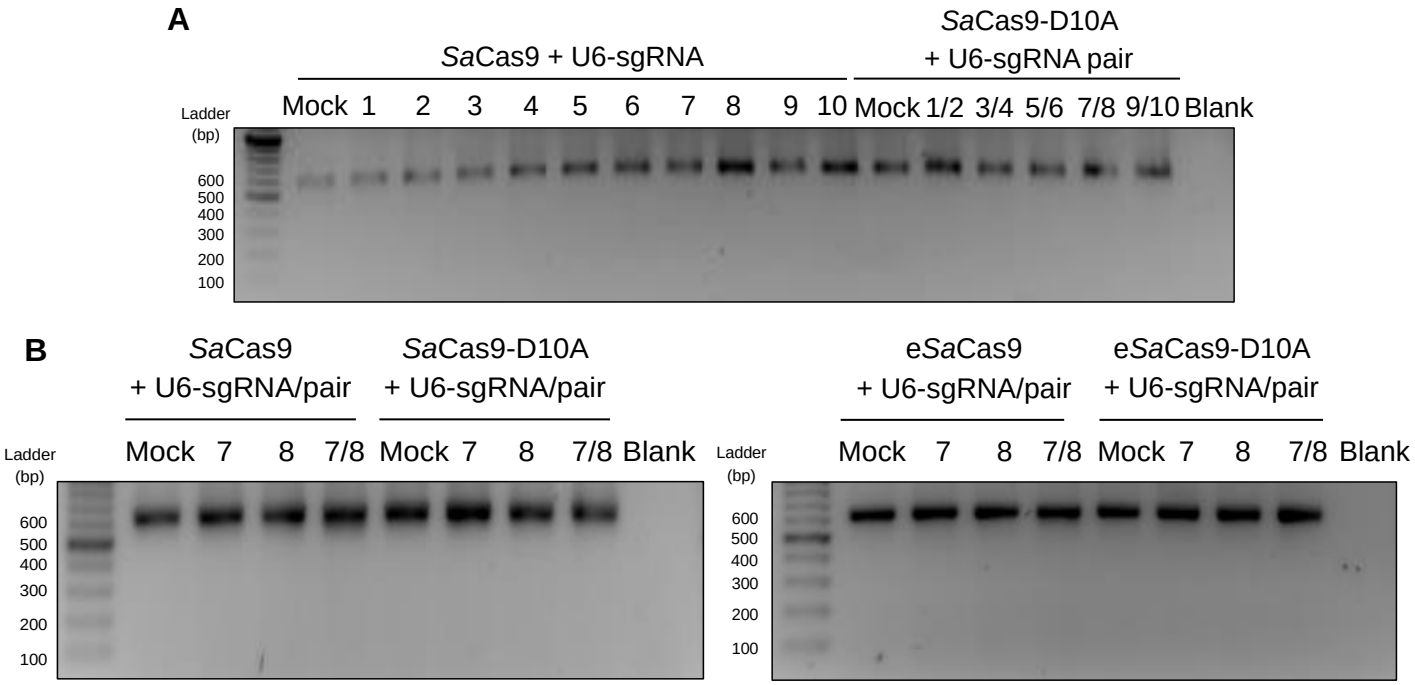
A 1/200 working solution of the Qubit® dsDNA HS Reagent (Thermo Fisher Scientific, USA) in Buffer was prepared and protected from light. Two standards (1 and 2) were used to calibrate the Qubit® Fluorometer 3.0. In a thin-wall clear 0.5 ml Qubit® assay tube, 10 µl of Standard 1 was added to 190 µl of the probe dilution, vortexed for two to three seconds and incubated at room temperature for two minutes before inserting into the Fluorometer. On the home screen, 'DNA' was selected followed by 'dsDNA High Sensitivity' and then 'read standards'. This was repeated for Standard 2 before samples were quantified. For each sample, 198 µl of the working solution and 2 µl of purified dsDNA were added to the tube. The sample was vortexed for two to three seconds and incubated at room temperature for two minutes before being placed into the Fluorometer. The lid was closed and 'read sample' selected on the screen. The calculated concentration was converted automatically to ng/µl by the Fluorometer.

### 7.4.11. TRIzol® Reagent RNA Extraction

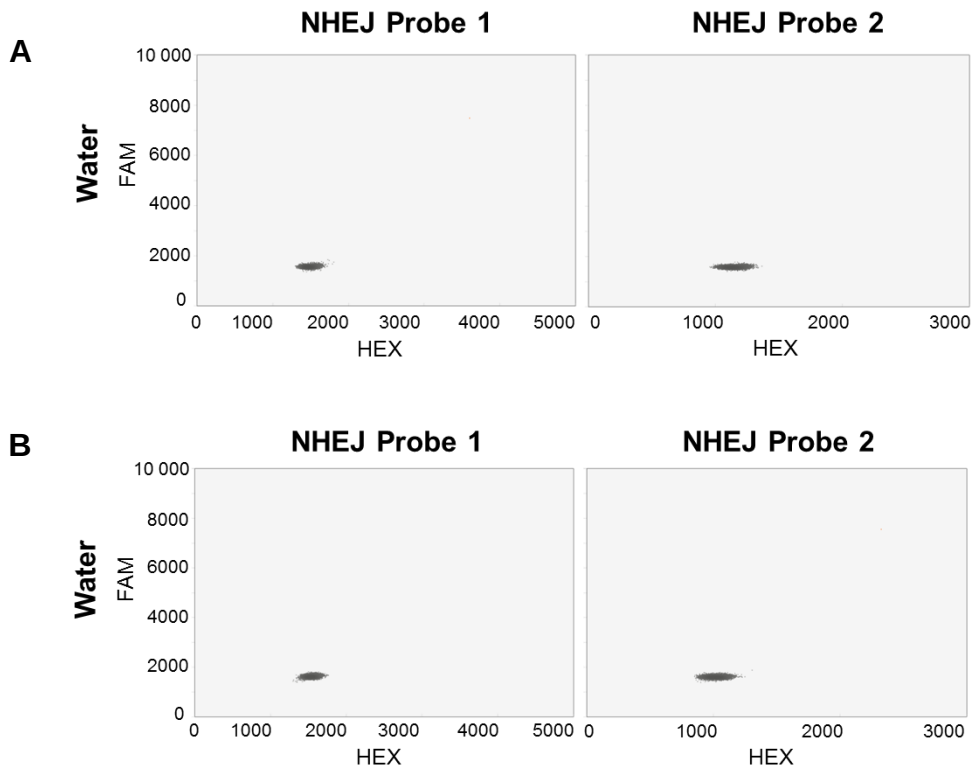
In order to extract RNA from TZM-bl cells transfected in a 24-well plate, 0.3 ml of TRIzol® Reagent (Thermo Fisher Scientific, USA) was added to the washed cells, the mixture pipetted up and down and the sample left to homogenise at room temperature for five minutes. The homogenised sample was then transferred to a 1.5 ml microcentrifuge tube, 60 µl of chloroform added and the tube mixed vigorously for fifteen seconds by hand. After a two minute incubation at room temperature, the tubes were centrifuged at 13 000 rpm for fifteen minutes at 4°C (Centrifuge 5 415, Eppendorf, Germany). The top layer (aqueous phase containing the RNA) was removed carefully and transferred to a sterile 1.5 ml

microcentrifuge tube. To precipitate the RNA, 150 µl of 100 % isopropanol was added and the sample incubated at room temperature for ten minutes. The tubes were centrifuged at 13 000 rpm for ten minutes at 4°C to pellet the RNA and the supernatant was removed. The pellet was washed with 300 µl of 75% ethanol before centrifugation at 8 000 rpm for five minutes at 4°C. The supernatant was again removed, the pellet air dried for ten minutes and the RNA suspended in 50 µl of Ambion® nuclease-free water (Thermo Fisher Scientific, USA). The sample was incubated at 55 °C in a heating block for ten minutes and the RNA samples stored at -80 °C.

## 7.5. Supplementary Figures and Tables

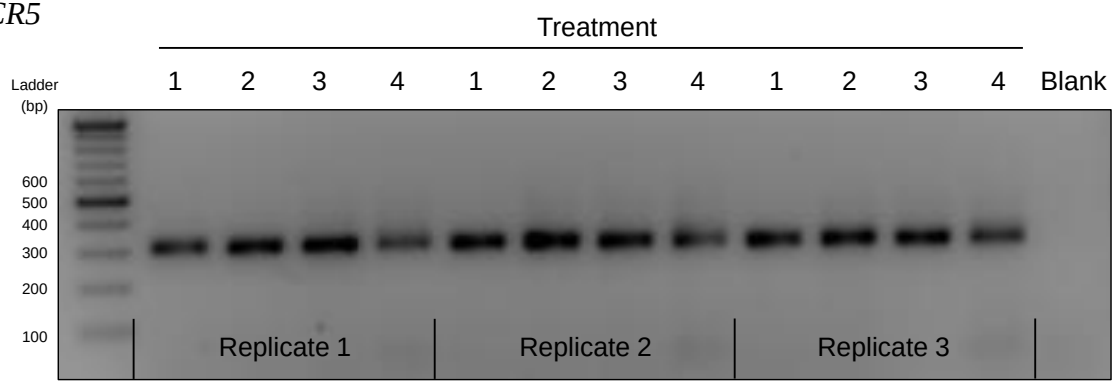



**Supplementary Figure 7.5.1: Agarose gels showing PCR products from amplification of the *CCR5* target site.** PCR gels from the T7EI assays were run for confirmation of a single 620 bp *CCR5* target site amplicon. gDNA was extracted from HEK293T cells 48 hours following transfection and PCR amplification carried out using *CCR5* F and R primers. Products, including those from the Mock controls and a blank containing no DNA template, were resolved on 1 % agarose gels alongside a molecular weight marker from the following treatments: *Sa*Cas9 + U6-sgRNAs 1 – 10 and *Sa*Cas9-D10A + U6-sgRNAs 1/2 – 9/10 (**A**) and either *Sa*Cas9 or e*Sa*Cas9 + U6-sgRNAs 7, 8 or 7/8 (**B**).

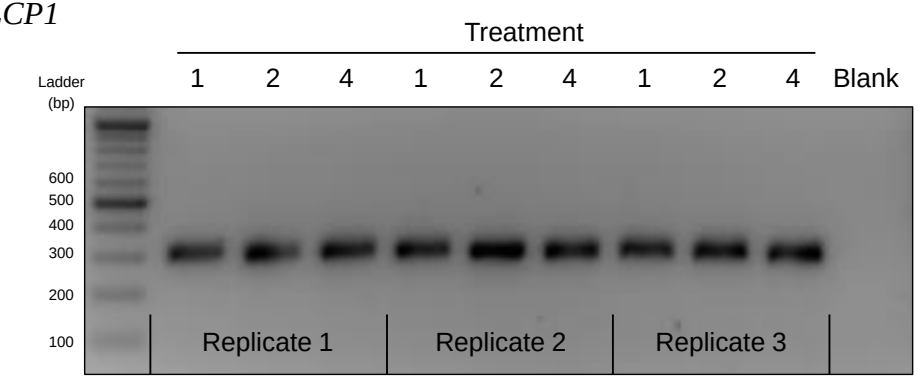



**Supplementary Figure 7.5.2: Quantification of *CCR5* target site indel formation in the no gDNA negative control by the ddPCR™ drop-off assay. A** and **B** – 2D-Plots of HEX (NHEJ probe 1 or 2) and FAM (ref probe) fluorescence measured by the ddPCR™ drop-off assay with no gDNA input as a negative control for *CCR5*-specific target site indel quantification. **A** – No gDNA ddPCR™ drop-off assay control to set the FAM fluorescence threshold at 2 250 for TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNA 7 or 8 and *Sa*Cas9-D10A + U6-sgRNA 7/8 for three independent transfections (Figure 3.6). **B** – No gDNA ddPCR™ drop-off assay control to set the FAM fluorescence threshold at 2 250 for TZM-bl cell treatment with *Sa*Cas9/-D10A or e*Sa*Cas9/-D10A + U6-sgRNA 7, 8 or 7/8 (Figure 3.10).

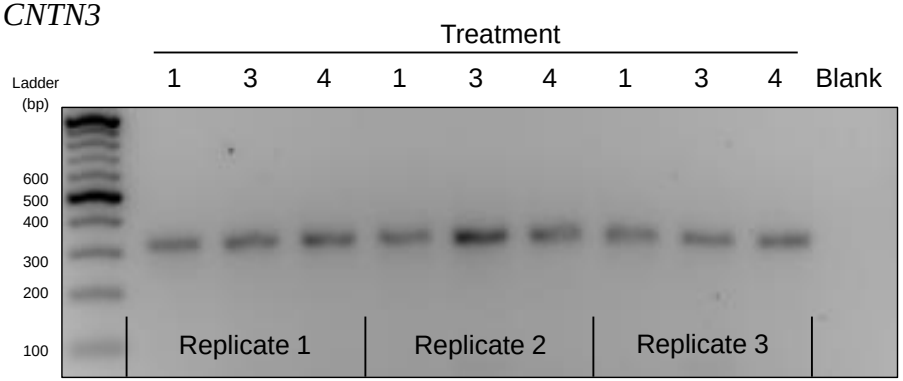
*CCR5*



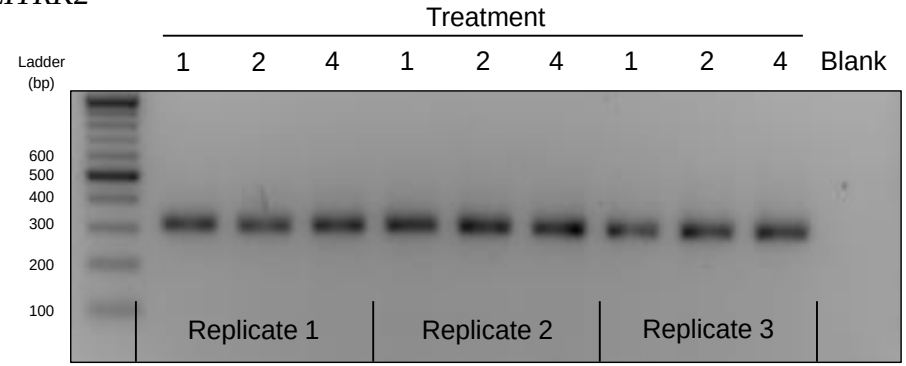

*LCP1*



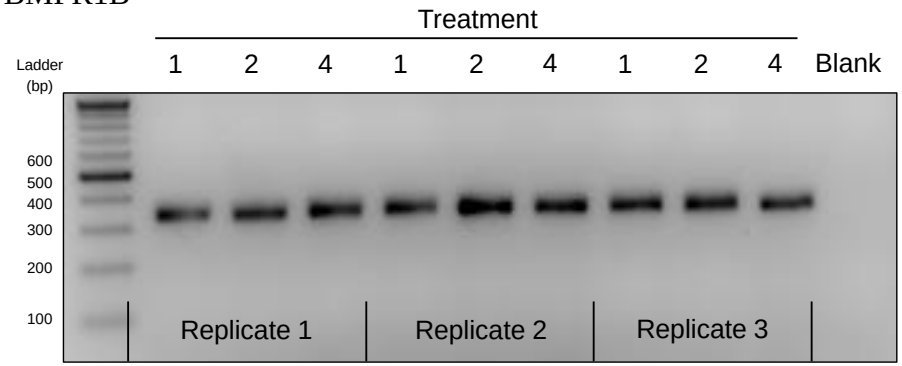

*CNTN3*



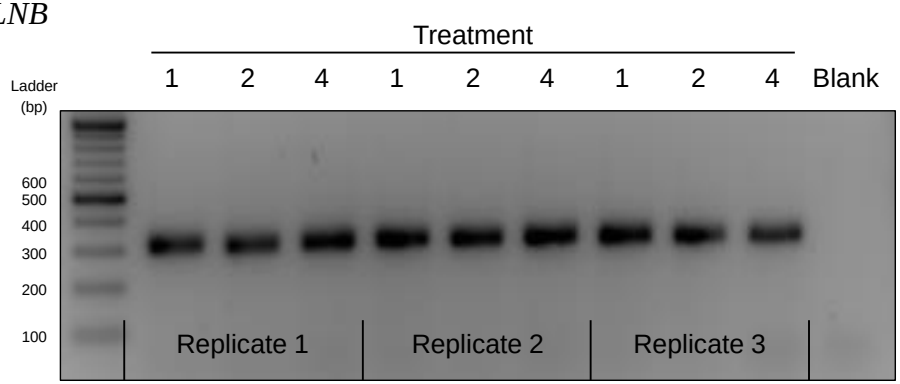

*SLITRK2*



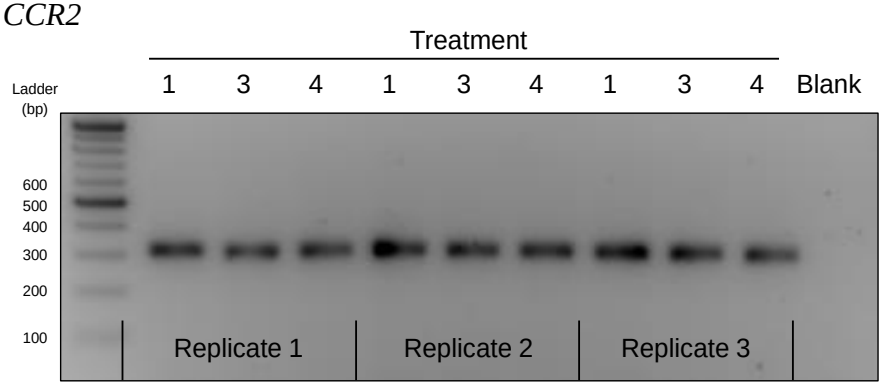

BMPR1B



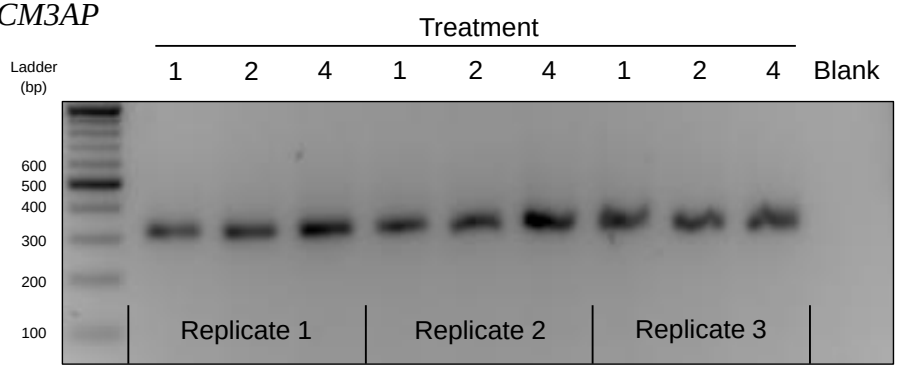
*FLNB*



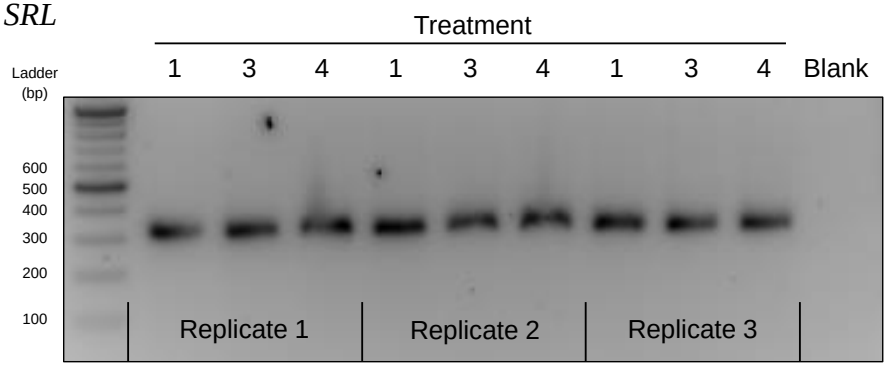

*CCR2*



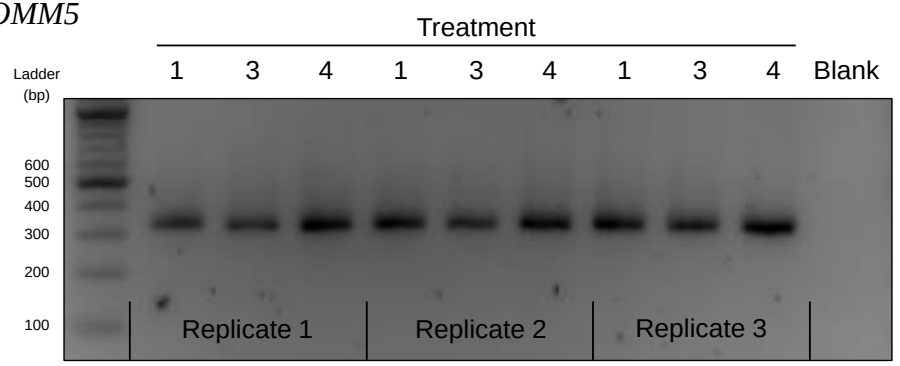

*MCM3AP*



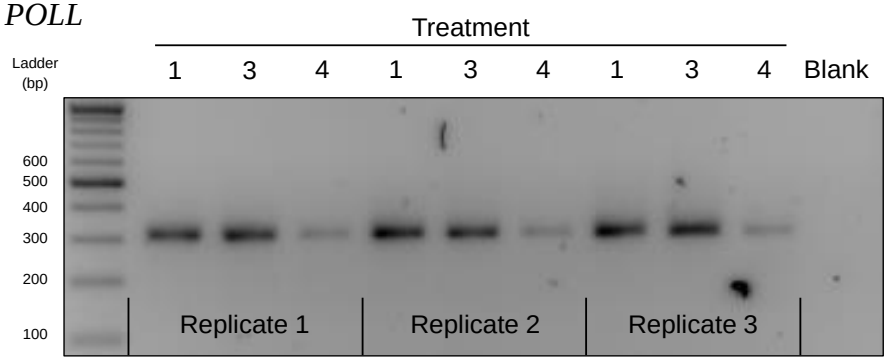

*SRL*




*TOMM5*




*POLL*



**Supplementary Figure 7.5.3: Agarose gels showing PCR products from amplification of the *CCR5* target site and ten off target sites for targeted amplicon deep sequencing.** PCR gels of ~ 300 bp amplicons sequenced by the Illumina MiSeq platform to assess indel formation in the *CCR5* target site and ten computationally predicted exonic off target sites, five with sequence similarity to the sgRNA target site (*LCP1*, *CNTN3*, *SLITRK2*, *BMPR1B* and *FLNB*) and five to sgRNA 8 (*CCR2*, *MCM3AP*, *SRL*, *TOMM5* and *POLL*). TZM-bl cells were treated with *Sa*Cas9 + empty U6-sgRNA as a mock control (1), *Sa*Cas9 + U6-sgRNA 7 (2), *Sa*Cas9 + U6-sgRNA 8 (3) and *Sa*Cas9-D10A + U6-sgRNA 7/8 (4) in three independent transfections (replicates 1 – 3). gDNA was extracted 48 hours following transfection and PCR amplification carried out using F and R primers listed in Table 2.14. Products were resolved on 1 % agarose gels alongside a molecular weight marker and a PCR blank sample (no template).

**Supplementary Table 7.5.1: Results obtained from targeted amplicon deep sequencing of the *CCR5* target region following TZM-bl cell treatment with *Sa*Cas9 + U6-sgRNAs 7 and 8 as well as *Sa*Cas9-D10A + U6-sgRNA 7/8**

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| Mock Control | 46373686 | SNV | 1 | G | T | 8 | 7720 | 0.1 | 28.88 | - | - | - | - | 28 | 26506 | 0.11 | 26.14 |
| | 46373694 | SNV | 1 | T | C | 11 | 7720 | 0.14 | 35.7 | 51 | 13751 | 0.37 | 37.41 | 63 | 26757 | 0.24 | 37.27 |
| | 46373695 | Del | 2 | GG | - | 33 | 7720 | 0.43 | 36.81 | 40 | 13751 | 0.29 | 36.4 | 63 | 27051 | 0.23 | 36.75 |
| | 46373695 | Del | 1 | G | - | - | - | - | - | 14 | 13751 | 0.1 | 37.36 | 30 | 26637 | 0.11 | 36.47 |
| | 46373696 | Del | 4 | GCCT | - | - | - | - | - | 14 | 13751 | 0.1 | 35.83 | - | - | - | - |
| | 46373697 | Del | 1 | C | - | 22 | 7720 | 0.28 | 34.73 | 21 | 13751 | 0.15 | 36.29 | 33 | 26635 | 0.12 | 36.7 |
| | 46373708 | SNV | 1 | G | T | 10 | 7720 | 0.13 | 29.5 | - | - | - | - | - | - | - | - |
| | 46373709 | MNV | 4 | CAGT | TGAA | 8 | 7720 | 0.1 | 33.11 | - | - | - | - | - | - | - | - |
| | 46373716 | SNV | 1 | T | A | 8 | 7720 | 0.1 | 32.75 | - | - | - | - | - | - | - | - |
| | 46373720 | MNV | 6 | ACAGGT | GTCAAC | 8 | 7698 | 0.1 | 32.68 | - | - | - | - | - | - | - | - |
| | 46373736 | SNV | 1 | T | C | 10 | 7554 | 0.13 | 33.8 | - | - | - | - | - | - | - | - |
| | 46373738 | SNV | 1 | T | C | 8 | 7685 | 0.1 | 37 | - | - | - | - | - | - | - | - |
| | 46373744 | Del | 18 | TGACAGAGACTCTTGGGA | - | - | - | - | - | 14 | 13714 | 0.1 | 36.79 | - | - | - | - |
| | 46373747 | SNV | 1 | C | A | - | - | - | - | 16 | 13714 | 0.12 | 25.62 | - | - | - | - |
| | 46373752 | Del | 2 | AC | - | 8 | 7685 | 0.1 | 36.89 | - | - | - | - | - | - | - | - |
| | 46373753 | Del | 2 | CT | - | - | - | - | - | 21 | 13714 | 0.15 | 32.05 | 31 | 26986 | 0.11 | 36.23 |
| | 46373766 | SNV | 1 | G | T | 29 | 7685 | 0.38 | 34.82 | 31 | 13714 | 0.23 | 34.37 | 53 | 26170 | 0.2 | 35.64 |
| | 46373776 | SNV | 1 | A | C | 8 | 7429 | 0.11 | 29.88 | - | - | - | - | - | - | - | - |
| | **Total** | | | | | | | **2.2** | | | | **1.61** | | | | **1.12** | |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| Mock Control | 46373686 | SNV | 1 | G | T | 8 | 7720 | 0.1 | 28.88 | - | - | - | - | 28 | 26506 | 0.11 | 26.14 |
| | 46373694 | SNV | 1 | T | C | 11 | 7720 | 0.14 | 35.7 | 51 | 13751 | 0.37 | 37.41 | 63 | 26757 | 0.24 | 37.27 |
| | 46373695 | Del | 2 | GG | - | 33 | 7720 | 0.43 | 36.81 | 40 | 13751 | 0.29 | 36.4 | 63 | 27051 | 0.23 | 36.75 |
| | 46373695 | Del | 1 | G | - | - | - | - | - | 14 | 13751 | 0.1 | 37.36 | 30 | 26637 | 0.11 | 36.47 |
| | 46373696 | Del | 4 | GCCT | - | - | - | - | - | 14 | 13751 | 0.1 | 35.83 | - | - | - | - |
| | 46373697 | Del | 1 | C | - | 22 | 7720 | 0.28 | 34.73 | 21 | 13751 | 0.15 | 36.29 | 33 | 26635 | 0.12 | 36.7 |
| | 46373708 | SNV | 1 | G | T | 10 | 7720 | 0.13 | 29.5 | - | - | - | - | - | - | - | - |
| | 46373709 | MNV | 4 | CAGT | TGAA | 8 | 7720 | 0.1 | 33.11 | - | - | - | - | - | - | - | - |
| | 46373716 | SNV | 1 | T | A | 8 | 7720 | 0.1 | 32.75 | - | - | - | - | - | - | - | - |
| | 46373720 | MNV | 6 | ACAGGT | GTCAAC | 8 | 7698 | 0.1 | 32.68 | - | - | - | - | - | - | - | - |
| | 46373736 | SNV | 1 | T | C | 10 | 7554 | 0.13 | 33.8 | - | - | - | - | - | - | - | - |
| | 46373738 | SNV | 1 | T | C | 8 | 7685 | 0.1 | 37 | - | - | - | - | - | - | - | - |
| | 46373744 | Del | 18 | TGACAGAGACTCTTGGGA | - | - | - | - | - | 14 | 13714 | 0.1 | 36.79 | - | - | - | - |
| | 46373747 | SNV | 1 | C | A | - | - | - | - | 16 | 13714 | 0.12 | 25.62 | - | - | - | - |
| | 46373752 | Del | 2 | AC | - | 8 | 7685 | 0.1 | 36.89 | - | - | - | - | - | - | - | - |
| | 46373753 | Del | 2 | CT | - | - | - | - | - | 21 | 13714 | 0.15 | 32.05 | 31 | 26986 | 0.11 | 36.23 |
| | 46373766 | SNV | 1 | G | T | 29 | 7685 | 0.38 | 34.82 | 31 | 13714 | 0.23 | 34.37 | 53 | 26170 | 0.2 | 35.64 |
| | 46373776 | SNV | 1 | A | C | 8 | 7429 | 0.11 | 29.88 | - | - | - | - | - | - | - | - |
| | **Total** | | | | | | | **2.2** | | | | **1.61** | | | | **1.12** | |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 7** | 46373670 | Del | 27 | CCTGAACACCTTCCAGGAATTCTTTGG | - | - | - | - | - | - | - | - | - | 9 | 8991 | 0.1 | 37 |
| | 46373682 | Del | 15 | CCAGGAATTCTTTGG | - | 33 | 25188 | 0.13 | 36.65 | - | - | - | - | 26 | 8989 | 0.29 | 35.62 |
| | 46373683 | Del | 26 | CAGGAATTCTTTGGCCTGAATAATTG | - | 59 | 25188 | 0.23 | 34.93 | - | - | - | - | 16 | 8990 | 0.18 | 37.69 |
| | 46373683 | Del | 14 | CAGGAATTCTTTGG | - | - | - | - | - | 6 | 4721 | 0.13 | 37.83 | - | - | - | - |
| | 46373685 | Del | 10 | GGAATTCTTT | - | 24 | 25188 | 0.1 | 37 | - | - | - | - | 10 | 8988 | 0.11 | 34.7 |
| | 46373686 | Del | 14 | GAATTCTTTGGCCT | - | 63 | 25188 | 0.25 | 34.62 | 33 | 4721 | 0.7 | 34.88 | 33 | 8988 | 0.37 | 34.07 |
| | 46373686 | SNV | 1 | G | T | - | - | - | - | 6 | 4721 | 0.13 | 25.33 | - | - | - | - |
| | 46373687 | Del | 17 | AATTCTTTGGCCTGAAT | - | 101 | 25188 | 0.4 | 36.33 | 18 | 4721 | 0.38 | 36.72 | 16 | 8989 | 0.18 | 34.24 |
| | 46373688 | Del | 9 | ATTCTTTGG | - | - | - | - | - | 5 | 4721 | 0.11 | 38 | - | - | - | - |
| | 46373689 | Del | 8 | TTCTTTGG | - | 49 | 25188 | 0.19 | 35.22 | 7 | 4721 | 0.15 | 38 | 13 | 8987 | 0.14 | 35.23 |
| | 46373690 | Del | 7 | TCTTTGG | - | - | - | - | - | 5 | 4721 | 0.11 | 37.4 | 14 | 8987 | 0.16 | 37.86 |
| | 46373691 | Del | 6 | CTTTGG | - | 78 | 25188 | 0.31 | 37.22 | 22 | 4721 | 0.47 | 37.27 | 48 | 8987 | 0.53 | 36.78 |
| | 46373691 | Del | 7 | CTTTGGC | - | 85 | 25188 | 0.34 | 34.84 | 20 | 4721 | 0.42 | 36.65 | 13 | 8987 | 0.14 | 35.51 |
| | 46373692 | Del | 11 | TTTGGCCTGAA | - | 33 | 25188 | 0.13 | 36.97 | - | - | - | - | 9 | 8988 | 0.1 | 34.45 |
| | 46373692 | Del | 15 | TTTGGCCTGAATAAT | - | - | - | - | - | 6 | 4721 | 0.13 | 36.17 | 9 | 8988 | 0.1 | 34.89 |
| | 46373692 | Del | 5 | TTTGG | - | 51 | 25188 | 0.2 | 35.47 | 11 | 4721 | 0.23 | 35.18 | - | - | - | - |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 7** | 46373692 | Del | 14 | TTTGGCCTGAATAA | - | - | - | - | - | 12 | 4721 | 0.25 | 37.08 | - | - | - | - |
| | 46373693 | Del | 6 | TTGGCC | - | 35 | 25188 | 0.14 | 35.29 | 11 | 4721 | 0.23 | 37.45 | 23 | 8986 | 0.26 | 33.91 |
| | 46373693 | Del | 13 | TTGGCCTGAATAA | - | 60 | 25188 | 0.24 | 36.72 | 21 | 4721 | 0.44 | 37.19 | 21 | 8988 | 0.23 | 37.24 |
| | 46373693 | Del | 10 | TTGGCCTGAA | - | 73 | 25188 | 0.29 | 36.18 | 5 | 4721 | 0.11 | 34.2 | 18 | 8988 | 0.2 | 35.94 |
| | 46373693 | Del | 4 | TTGG | - | 136 | 25188 | 0.54 | 36.13 | 18 | 4721 | 0.38 | 36.5 | 64 | 8986 | 0.71 | 35.94 |
| | 46373693 | Del | 8 | TTGGCCTG | - | - | - | - | - | 5 | 4721 | 0.11 | 36.6 | - | - | - | - |
| | 46373694 | Del | 7 | TGGCCTG | - | 31 | 25188 | 0.12 | 36.51 | 6 | 4721 | 0.13 | 36.83 | 15 | 8986 | 0.17 | 36.93 |
| | 46373694 | Del | 9 | TGGCCTGAA | - | 52 | 25188 | 0.21 | 36.57 | 9 | 4721 | 0.19 | 36.22 | 22 | 8988 | 0.24 | 35.59 |
| | 46373694 | SNV | 1 | T | C | 113 | 25188 | 0.45 | 37.21 | 16 | 4721 | 0.34 | 37.81 | 22 | 8883 | 0.25 | 37.31 |
| | 46373694 | Del | 5 | TGGCC | - | 115 | 25188 | 0.46 | 36.4 | 19 | 4721 | 0.4 | 37.74 | 49 | 8986 | 0.55 | 36.96 |
| | 46373694 | Del | 3 | TGG | - | 200 | 25188 | 0.79 | 36.19 | 59 | 4721 | 1.25 | 35.59 | 54 | 8986 | 0.6 | 35.6 |
| | 46373694 | Del | 4 | TGGC | - | - | - | - | - | 7 | 4721 | 0.15 | 37.1 | - | - | - | - |
| | 46373695 | Del | 1 | G | - | 589 | 25188 | 2.34 | 36.79 | 108 | 4721 | 2.29 | 36.64 | 192 | 8841 | 2.17 | 36.39 |
| | 46373695 | Del | 2 | GG | - | 1666 | 25188 | 6.61 | 36.11 | 401 | 4721 | 8.49 | 35.54 | 478 | 8986 | 5.32 | 36.36 |
| | 46373695 | Del | 10 | GGCCTGAATA | - | - | - | - | - | 5 | 4721 | 0.11 | 37.2 | - | - | - | - |
| | 46373695 | Del | 7 | GGCCTGA | - | - | - | - | - | 6 | 4721 | 0.13 | 33.62 | - | - | - | - |
| | 46373696 | Del | 12 | GCCTGAATAATT | - | 28 | 25188 | 0.11 | 34.86 | 6 | 4721 | 0.13 | 37.83 | - | - | - | - |
| | 46373696 | Del | 4 | GCCT | - | 447 | 25188 | 1.77 | 36 | 117 | 4721 | 2.48 | 36.44 | 134 | 8986 | 1.49 | 34.99 |
| | 46373696 | Del | 7 | GCCTGAA | - | - | - | - | - | 5 | 4721 | 0.11 | 37.2 | - | - | - | - |
| | 46373696 | SNV | 1 | G | T | - | - | - | - | 10 | 4721 | 0.21 | 35.9 | - | - | - | - |
| | 46373697 | Del | 9 | CCTGAATAA | - | 27 | 25188 | 0.11 | 36.07 | 5 | 4721 | 0.11 | 37.6 | - | - | - | - |
| | 46373697 | Del | 8 | CCTGAATA | - | - | - | - | - | 11 | 4721 | 0.23 | 35.09 | - | - | - | - |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 7** | 46373697 | Del | 7 | CCTGAAT | - | 80 | 25188 | 0.32 | 35.8 | 17 | 4721 | 0.36 | 37.53 | 43 | 8988 | 0.48 | 34.57 |
| | 46373697 | Del | 5 | CCTGA | - | 103 | 25188 | 0.41 | 36.57 | 37 | 4721 | 0.78 | 36.68 | 42 | 8987 | 0.47 | 35.03 |
| | 46373697 | Del | 6 | CCTGAA | - | 130 | 25188 | 0.52 | 36.37 | 38 | 4721 | 0.8 | 36.69 | 58 | 8988 | 0.65 | 35.75 |
| | 46373697 | Del | 3 | CCT | - | 170 | 25188 | 0.67 | 35.39 | 35 | 4721 | 0.74 | 35.9 | 60 | 8986 | 0.67 | 36.37 |
| | 46373697 | Del | 2 | CC | - | 226 | 25188 | 0.9 | 35.64 | 48 | 4721 | 1.02 | 36.54 | 80 | 8986 | 0.89 | 35.27 |
| | 46373697 | Del | 1 | C | - | 1044 | 25188 | 4.14 | 36.25 | 159 | 4721 | 3.37 | 35.56 | 295 | 8866 | 3.33 | 36.1 |
| | 46373698 | MNV | 2 | CT | GG | - | - | - | - | 9 | 4721 | 0.19 | 37.04 | - | - | - | - |
| | 46373701 | SNV | 1 | A | G | - | - | - | - | 10 | 4721 | 0.21 | 36.8 | - | - | - | - |
| | 46373702 | SNV | 1 | A | G | 49 | 25188 | 0.19 | 37.41 | 6 | 4721 | 0.13 | 36.5 | 14 | 8831 | 0.16 | 36.79 |
| | 46373703 | SNV | 1 | T | G | - | - | - | - | 11 | 4721 | 0.23 | 37.55 | - | - | - | - |
| | 46373704 | SNV | 1 | A | C | - | - | - | - | 5 | 4721 | 0.11 | 31.4 | - | - | - | - |
| | 46373706 | SNV | 1 | T | C | 53 | 25188 | 0.21 | 36.85 | 6 | 4721 | 0.13 | 37.83 | 11 | 8916 | 0.12 | 37.18 |
| | 46373706 | SNV | 1 | T | G | - | - | - | - | 10 | 4721 | 0.21 | 38 | - | - | - | - |
| | 46373708 | SNV | 1 | G | T | - | - | - | - | 11 | 4721 | 0.23 | 36.36 | - | - | - | - |
| | 46373709 | MNV | 4 | CAGT | TGAA | 41 | 25188 | 0.16 | 35.82 | 5 | 4721 | 0.11 | 36.42 | 10 | 8983 | 0.11 | 35.88 |
| | 46373713 | SNV | 1 | A | C | - | - | - | - | 5 | 4721 | 0.11 | 34.2 | - | - | - | - |
| | 46373716 | SNV | 1 | T | A | 42 | 25188 | 0.17 | 37.21 | 5 | 4721 | 0.11 | 37.8 | 10 | 8793 | 0.11 | 37.2 |
| | 46373718 | SNV | 1 | T | C | 41 | 25188 | 0.16 | 36.8 | 7 | 4721 | 0.15 | 37.43 | 10 | 8902 | 0.11 | 37.6 |
| | 46373720 | MNV | 6 | ACAGGT | GTCAAC | 41 | 25188 | 0.16 | 35.89 | 5 | 4721 | 0.11 | 36.46 | 10 | 8975 | 0.11 | 37.09 |
| | 46373728 | SNV | 1 | G | A | - | - | - | - | - | - | - | - | 10 | 8689 | 0.12 | 36.4 |
| | 46373736 | SNV | 1 | T | C | 39 | 25113 | 0.16 | 37 | - | - | - | - | 12 | 8798 | 0.14 | 35.83 |
| | 46373738 | SNV | 1 | T | C | 39 | 25113 | 0.16 | 37.08 | - | - | - | - | 13 | 8861 | 0.15 | 36.92 |
| | 46373744 | Del | 18 | TGACAGAGACTCTTGGGA | - | - | - | - | - | 7 | 4695 | 0.15 | 36.71 | 15 | 8964 | 0.17 | 31.53 |
| | 46373752 | Del | 2 | AC | - | 24 | 25113 | 0.1 | 36.37 | 6 | 4695 | 0.13 | 35.5 | 12 | 8955 | 0.13 | 35.6 |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 7** | 46373753 | Del | 2 | CT | - | 35 | 25113 | 0.14 | 35.19 | 9 | 4695 | 0.19 | 33.94 | 15 | 8954 | 0.17 | 35.51 |
| | 46373760 | SNV | 1 | G | A | - | - | - | - | 6 | 4695 | 0.13 | 36.83 | - | - | - | - |
| | 46373766 | SNV | 1 | G | T | 73 | 25113 | 0.29 | 34.62 | 15 | 4695 | 0.32 | 34.47 | 22 | 8636 | 0.25 | 36 |
| | 46373767 | SNV | 1 | C | A | 32 | 25113 | 0.13 | 26.38 | 5 | 4695 | 0.11 | 26.4 | - | - | - | - |
| | **Total** | | | | | | | **25.45** | | | | **31.36** | | | | **22.93** | |
| ***Sa*Cas9 + U6-sgRNA 8** | 46373694 | SNV | 1 | T | C | 89 | 30568 | 0.29 | 37.2 | 47 | 18336 | 0.26 | 37.37 | 87 | 34801 | 0.25 | 37.39 |
| | 46373695 | Del | 1 | G | - | 41 | 30568 | 0.13 | 36.9 | - | - | - | - | 44 | 34578 | 0.13 | 35.73 |
| | 46373695 | Del | 2 | GG | - | 137 | 30568 | 0.45 | 36.77 | 54 | 18558 | 0.29 | 36.51 | 85 | 35217 | 0.24 | 35.74 |
| | 46373696 | SNV | 1 | G | T | 52 | 30568 | 0.17 | 36.78 | - | - | - | - | 54 | 34381 | 0.16 | 32.46 |
| | 46373697 | Del | 1 | C | - | 79 | 30568 | 0.26 | 35.41 | 35 | 18205 | 0.19 | 36.35 | 61 | 34562 | 0.18 | 36.49 |
| | 46373698 | MNV | 2 | CT | GG | 51 | 30568 | 0.17 | 36.26 | - | - | - | - | - | - | - | - |
| | 46373701 | SNV | 1 | A | G | 61 | 30568 | 0.2 | 37.62 | - | - | - | - | 37 | 34410 | 0.11 | 37.24 |
| | 46373703 | SNV | 1 | T | G | 60 | 30568 | 0.2 | 37.32 | 19 | 18197 | 0.1 | 35.84 | 36 | 34607 | 0.1 | 35.56 |
| | 46373706 | SNV | 1 | T | G | 60 | 30568 | 0.2 | 36.92 | - | - | - | - | 39 | 34887 | 0.11 | 35.41 |
| | 46373708 | SNV | 1 | G | T | 67 | 30568 | 0.22 | 36.54 | 21 | 18117 | 0.12 | 35.67 | 52 | 34340 | 0.15 | 34.12 |
| | 46373710 | Ins | 14 | - | AGCCTTTGTCGGGG | - | - | - | - | 21 | 18542 | 0.11 | 37.18 | - | - | - | - |
| | 46373712 | SNV | 1 | T | A | 61 | 30193 | 0.2 | 36.61 | 32 | 18341 | 0.17 | 36.59 | - | - | - | - |
| | 46373715 | SNV | 1 | C | T | 37 | 29773 | 0.12 | 36.27 | 22 | 18073 | 0.12 | 37.5 | - | - | - | - |
| | 46373723 | SNV | 1 | G | A | 47 | 29940 | 0.16 | 36.66 | - | - | - | - | - | - | - | - |
| | 46373725 | SNV | 1 | T | C | 61 | 30004 | 0.2 | 36.69 | - | - | - | - | - | - | - | - |
| | 46373727 | SNV | 1 | G | T | 31 | 30421 | 0.1 | 32.9 | - | - | - | - | - | - | - | - |
| | 46373728 | SNV | 1 | G | A | - | - | - | - | - | - | - | - | 58 | 33947 | 0.17 | 36.31 |
| | 46373736 | SNV | 1 | T | C | - | - | - | - | 20 | 18128 | 0.11 | 36.1 | - | - | - | - |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 8** | 46373737 | Del | 27 | ATGCAGGTGACAGAGACTCTTGGGATG | - | - | - | - | - | 36 | 18487 | 0.19 | 36.86 | - | - | - | - |
| | 46373738 | SNV | 1 | T | C | 31 | 30421 | 0.1 | 36.68 | - | - | - | - | - | - | - | - |
| | 46373738 | Del | 16 | TGCAGGTGACAGAGAC | - | 32 | 30421 | 0.11 | 36.04 | 21 | 18486 | 0.11 | 36.81 | - | - | - | - |
| | 46373738 | Del | 32 | TGCAGGTGACAGAGACTCTTGGGATGACGCAC | - | 41 | 30421 | 0.13 | 36.66 | - | - | - | - | - | - | - | - |
| | 46373738 | Del | 35 | TGCAGGTGACAGAGACTCTTGGGATGACGCACTGC | - | 57 | 30421 | 0.19 | 37.37 | - | - | - | - | - | - | - | - |
| | 46373739 | Del | 27 | GCAGGTGACAGAGACTCTTGGGATGAC | - | 46 | 30421 | 0.15 | 35.28 | 98 | 18485 | 0.53 | 36.58 | 57 | 35104 | 0.16 | 35.19 |
| | 46373739 | Del | 15 | GCAGGTGACAGAGAC | - | - | - | - | - | | | | | 39 | 35104 | 0.11 | 36.76 |
| | 46373742 | Del | 17 | GGTGACAGAGACTCTTG | - | 34 | 30421 | 0.11 | 36.1 | 44 | 18482 | 0.24 | 35.64 | - | - | - | - |
| | 46373742 | Del | 18 | GGTGACAGAGACTCTTGG | - | - | - | - | - | 27 | 18482 | 0.15 | 34.72 | - | - | - | - |
| | 46373743 | Del | 11 | GTGACAGAGAC | - | - | - | - | - | 27 | 18481 | 0.15 | 37 | - | - | - | - |
| | 46373743 | SNV | 1 | G | T | - | - | - | - | - | - | - | - | 37 | 33995 | 0.11 | 29.35 |
| | 46373744 | Del | 18 | TGACAGAGACTCTTGGGA | - | 220 | 30421 | 0.72 | 34.85 | 282 | 18481 | 1.53 | 35.15 | 428 | 35087 | 1.22 | 35.34 |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 8** | 46373744 | Del | 26 | TGACAGAGACTCTTGGGATGACGCAC | - | - | - | - | - | 36 | 18481 | 0.19 | 35.36 | - | - | - | - |
| | 46373744 | Del | 10 | TGACAGAGAC | - | - | - | - | - | - | - | - | - | 63 | 35087 | 0.18 | 35.02 |
| | 46373744 | Del | 29 | TGACAGAGACTCTTGGGATGACGCACTGC | - | - | - | - | - | - | - | - | - | 74 | 35087 | 0.21 | 35.51 |
| | 46373745 | Del | 6 | GACAGA | - | 73 | 30421 | 0.24 | 36.18 | 32 | 18480 | 0.17 | 34.53 | 84 | 35085 | 0.24 | 34.93 |
| | 46373745 | Del | 9 | GACAGAGAC | - | - | - | - | - | 22 | 18480 | 0.12 | 36.61 | - | - | - | - |
| | 46373746 | Del | 8 | ACAGAGAC | - | 54 | 30421 | 0.18 | 35.57 | 35 | 18474 | 0.19 | 36.34 | 75 | 35082 | 0.21 | 35.14 |
| | 46373746 | SNV | 1 | A | T | - | - | - | - | 34 | 18057 | 0.19 | 36.85 | - | - | - | - |
| | 46373746 | Del | 22 | ACAGAGACTCTTGGGATGACGC | - | - | - | - | - | - | - | - | - | 44 | 35082 | 0.13 | 35.75 |
| | 46373747 | Del | 20 | CAGAGACTCTTGGGATGACG | - | 31 | 30421 | 0.1 | 35.61 | - | - | - | - | 99 | 35078 | 0.28 | 35.61 |
| | 46373747 | Del | 8 | CAGAGACT | - | | | | | 32 | 18473 | 0.17 | 36.03 | - | - | - | - |
| | 46373749 | Del | 5 | GAGAC | - | 37 | 30421 | 0.12 | 35.03 | 33 | 18470 | 0.18 | 36.68 | 114 | 35050 | 0.33 | 36.37 |
| | 46373749 | Del | 17 | GAGACTCTTGGGATGAC | - | | | | | 28 | 18472 | 0.15 | 35.79 | - | - | - | - |
| | 46373750 | Del | 4 | AGAC | - | 61 | 30421 | 0.2 | 35.95 | 87 | 18465 | 0.47 | 35.9 | 105 | 35043 | 0.3 | 36.15 |
| | 46373751 | Del | 3 | GAC | - | 33 | 30421 | 0.11 | 36.1 | 55 | 18461 | 0.3 | 36.09 | 72 | 35038 | 0.21 | 34.82 |
| | 46373751 | Del | 9 | GACTCTTGG | - | - | - | - | - | 27 | 18465 | 0.15 | 36.78 | - | - | - | - |
| | 46373751 | Del | 7 | GACTCTT | - | - | - | - | - | - | - | - | - | 46 | 35038 | 0.13 | 36.97 |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 8** | 46373751 | Del | 12 | GACTCTTGGGAT | - | 48 | 30421 | 0.16 | 34.77 | 77 | 18465 | 0.42 | 35.21 | 97 | 35038 | 0.28 | 35.91 |
| | 46373752 | Del | 2 | AC | - | 493 | 30421 | 1.62 | 35.31 | 294 | 18461 | 1.59 | 35.93 | 401 | 35037 | 1.14 | 35.1 |
| | 46373752 | SNV | 1 | A | C | - | - | - | - | - | - | - | - | 35 | 34282 | 0.1 | 31.62 |
| | 46373752 | Del | 16 | ACTCTTGGGATGACGC | - | - | - | - | - | 21 | 18465 | 0.11 | 35.67 | - | - | - | - |
| | 46373753 | Ins | 1 | - | C | 31 | 30421 | 0.1 | 35.95 | 35 | 18460 | 0.19 | 37.54 | 82 | 35016 | 0.23 | 36.52 |
| | 46373753 | Del | 1 | C | - | 214 | 30421 | 0.7 | 36.63 | 162 | 17909 | 0.9 | 35.43 | 196 | 34012 | 0.58 | 36.48 |
| | 46373753 | Del | 2 | CT | - | 624 | 30421 | 2.05 | 35.36 | 375 | 18463 | 2.03 | 35.4 | 629 | 35035 | 1.8 | 34.97 |
| | 46373753 | SNV | 1 | C | A | - | - | - | - | 26 | 17909 | 0.15 | 30.42 | - | - | - | - |
| | 46373754 | Ins | 1 | - | A | 39 | 30421 | 0.13 | 36.52 | 79 | 18458 | 0.43 | 36.59 | 51 | 35003 | 0.15 | 37.29 |
| | 46373754 | Ins | 1 | - | G | - | - | - | - | 39 | 18458 | 0.21 | 36.02 | 50 | 35003 | 0.14 | 36.04 |
| | 46373754 | Ins | 1 | - | T | 42 | 30421 | 0.14 | 36.61 | 57 | 18458 | 0.31 | 37.12 | 44 | 35003 | 0.13 | 36.34 |
| | 46373754 | Del | 1 | T | - | 198 | 30421 | 0.65 | 34.85 | 138 | 18057 | 0.76 | 34.51 | 243 | 34208 | 0.71 | 34.77 |
| | 46373754 | Del | 3 | TCT | - | 98 | 30421 | 0.32 | 35.2 | 48 | 18460 | 0.26 | 34.54 | 36 | 35020 | 0.1 | 35.65 |
| | 46373754 | Del | 4 | TCTT | - | 89 | 30421 | 0.29 | 35.19 | 82 | 18461 | 0.44 | 35.58 | 65 | 35020 | 0.19 | 34.76 |
| | 46373754 | Del | 5 | TCTTG | - | 61 | 30421 | 0.2 | 34.87 | 33 | 18461 | 0.18 | 35.76 | - | - | - | - |
| | 46373754 | Del | 6 | TCTTGG | - | 82 | 30421 | 0.27 | 34.09 | 54 | 18462 | 0.29 | 36.29 | 64 | 35020 | 0.18 | 35.05 |
| | 46373754 | Del | 7 | TCTTGGG | - | 41 | 30421 | 0.13 | 34.98 | 46 | 18462 | 0.25 | 36.36 | - | - | - | - |
| | 46373754 | Del | 9 | TCTTGGGAT | - | 37 | 30421 | 0.12 | 36.09 | - | - | - | - | 99 | 35020 | 0.28 | 35.3 |
| | 46373754 | Del | 8 | TCTTGGGA | - | - | - | - | - | 29 | 18462 | 0.16 | 35.76 | - | - | - | - |
| | 46373755 | Ins | 4 | - | ATGC | - | - | - | - | 21 | 18432 | 0.11 | 36.6 | - | - | - | - |
| | 46373755 | SNV | 1 | C | G | - | - | - | - | 28 | 17922 | 0.16 | 35.14 | 39 | 33993 | 0.11 | 36.38 |
| | 46373757 | SNV | 1 | T | C | - | - | - | - | 29 | 18149 | 0.16 | 36.34 | 40 | 34364 | 0.12 | 35.6 |
| | 46373758 | Ins | 4 | - | TGTC | - | - | - | - | 25 | 18421 | 0.14 | 35.91 | - | - | - | - |
| | 46373761 | MNV | 2 | AT | GA | - | - | - | - | 51 | 18388 | 0.28 | 33.5 | 58 | 34876 | 0.17 | 35.03 |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9 + U6-sgRNA 8** | 46373765 | SNV | 1 | C | A | - | - | - | - | 72 | 17805 | 0.4 | 32.92 | 102 | 33904 | 0.3 | 32.3 |
| | 46373765 | SNV | 1 | C | T | - | - | - | - | 34 | 17805 | 0.19 | 36.97 | 37 | 33904 | 0.11 | 35.46 |
| | 46373767 | MNV | 2 | CA | TT | - | - | - | - | 49 | 18343 | 0.27 | 32.94 | 55 | 34844 | 0.16 | 36.37 |
| | 46373770 | SNV | 1 | T | A | - | - | - | - | 52 | 17803 | 0.29 | 35.52 | 58 | 33742 | 0.17 | 35.97 |
| | **Total** | | | | | | | **12.41** | | | | **17.33** | | | | **12.57** | |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | 46373694 | SNV | 1 | T | C | 65 | 22027 | 0.3 | 37.35 | - | - | - | - | - | - | - | - |
| | 46373695 | Del | 1 | G | - | 35 | 21889 | 0.16 | 36.1 | 19 | 17404 | 0.11 | 36.89 | 29 | 28880 | 0.1 | 36.5 |
| | 46373695 | Del | 2 | GG | - | 118 | 22247 | 0.53 | 37.19 | 63 | 17719 | 0.36 | 36.48 | 95 | 29373 | 0.32 | 36.96 |
| | 46373696 | Del | 4 | GCCT | - | - | - | - | - | 19 | 17721 | 0.11 | 35.76 | | | | |
| | 46373697 | Del | 1 | C | - | 91 | 21892 | 0.42 | 36.38 | 31 | 17440 | 0.18 | 35.98 | 64 | 28871 | 0.22 | 36.96 |
| | 46373698 | Del | 8 | CTGAATAA | - | - | - | - | - | - | - | - | - | 36 | 29376 | 0.12 | 36.94 |
| | 46373699 | Del | 8 | TGAATAAT | - | 31 | 22252 | 0.14 | 36.06 | - | - | - | - | - | - | - | - |
| | 46373699 | Del | 13 | TGAATAATTGCAG | - | 26 | 22253 | 0.12 | 35.54 | - | - | - | - | - | - | - | - |
| | 46373700 | Del | 11 | GAATAATTGCA | - | 24 | 22252 | 0.11 | 34.75 | - | - | - | - | - | - | - | - |
| | 46373706 | Ins | 31 | - | TTGCAGTAGCTCTAACAGGTTGGACCAAGCT | 32 | 22246 | 0.14 | 37.48 | 25 | 17712 | 0.14 | 37.1 | - | - | - | - |

| Treatment | Reference Position | Type | Length (bp) | Reference | Allele | Replicate 1 | | | | Replicate 2 | | | | Replicate 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality | Count | Coverage | % | Average quality |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | 46373707 | Ins | 31 | - | TGCAGTAGCTCTAACAGGTTGGACCAAGCTA | 47 | 22244 | 0.21 | 37.4 | 53 | 17709 | 0.3 | 37.01 | 40 | 29372 | 0.14 | 37.25 |
| | 46373728 | SNV | 1 | G | A | - | - | - | - | 23 | 17111 | 0.13 | 33.95 | - | - | - | - |
| | 46373744 | Del | 7 | TGACAGA | - | - | - | - | - | 20 | 17677 | 0.11 | 35.17 | 38 | 29329 | 0.13 | 35.79 |
| | 46373745 | Del | 6 | GACAGA | - | 71 | 22181 | 0.32 | 36.65 | 20 | 17677 | 0.11 | 34.62 | 37 | 29326 | 0.13 | 34.97 |
| | 46373750 | Ins | 4 | - | AGAC | - | - | - | - | 27 | 17667 | 0.15 | 36.83 | - | - | - | - |
| | 46373751 | Del | 2 | GA | - | 30 | 22163 | 0.14 | 36.92 | - | - | - | - | - | - | - | - |
| | 46373752 | Del | 1 | A | - | 32 | 21710 | 0.15 | 35.97 | - | - | - | - | - | - | - | - |
| | 46373754 | Del | 1 | T | - | 54 | 21741 | 0.25 | 36.61 | - | - | - | - | - | - | - | - |
| | 46373759 | Del | 3 | GGA | - | - | - | - | - | 19 | 17652 | 0.11 | 34.52 | - | - | - | - |
| | 46373759 | Del | 7 | GGATGAC | - | - | - | - | - | 20 | 17656 | 0.11 | 36.75 | - | - | - | - |
| | 46373760 | Del | 3 | GAT | - | - | - | - | - | - | - | - | - | 30 | 29286 | 0.1 | 36.04 |
| | **Total** | | | | | | | **2.45** | | | | **1.92** | | | | **1.26** | |

**Supplementary Table 7.5.2: Technical duplicate ddPCR™ drop-off assay results obtained from WT and 'enhanced specificity' CRISPR/*Sa*Cas9 nuclease and nickase treatment of TZM-bl cells**

| Treatment | NHEJ Probe | Fractional Abundance | Average Indel Frequency (%) | Standard Deviation | p-value |
|---|---|---|---|---|---|
| ***Sa*Cas9 + Empty U6-sgRNA** | 1 | 0.06<br>0.00 | 0.03 | 0.04 | - |
| | 2 | 0.05<br>0.00 | 0.03 | 0.04 | - |
| ***Sa*Cas9 + U6-sgRNA 7** | 1 | 33.04<br>32.61 | 32.82 | 0.31 | <0.0001 |
| ***Sa*Cas9 + U6-sgRNA 8** | 2 | 9.32<br>8.77 | 9.05 | 0.39 | 0.0009 |
| ***Sa*Cas9 + U6-sgRNA 7/8** | 1 | 34.22<br>33.48 | 33.85 | 0.52 | 0.0001 |
| | 2 | 28.76<br>30.54 | 29.65 | 1.26 | **0.0191** |
| **e*Sa*Cas9 + Empty U6-sgRNA** | 1 | 0.04<br>0.09 | 0.06 | 0.04 | - |
| | 2 | 0.00<br>0.03 | 0.01 | 0.02 | - |
| **e*Sa*Cas9 + U6-sgRNA 7** | 1 | 33.04<br>33.93 | 33.49 | 0.63 | 0.0002 |
| **e*Sa*Cas9 + U6-sgRNA 8** | 2 | 3.44<br>3.57 | 3.50 | 0.10 | 0.0004 |
| **e*Sa*Cas9 + U6-sgRNA 7/8** | 1 | 27.58<br>28.37 | 27.98 | 0.56 | 0.0002 |
| | 2 | 19.57<br>18.40 | 18.98 | 0.82 | **0.0196** |
| ***Sa*Cas9-D10A + Empty U6-sgRNA** | 1 | 0.07<br>0.09 | 0.08 | 0.02 | - |
| | 2 | 0.09<br>0.11 | 0.10 | 0.01 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 7** | 1 | 0.03<br>0.03 | 0.03 | 0.00 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 8** | 2 | 0.18<br>0.12 | 0.15 | 0.04 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | 1 | 5.03<br>5.40 | 5.21 | 0.26 | 0.0013 |
| | 2 | 3.28<br>3.19 | 3.24 | 0.06 | 0.0002 |
| **e*Sa*Cas9-D10A + Empty U6-sgRNA** | 1 | 0.04<br>0.00 | 0.02 | 0.03 | - |
| | 2 | 0.03<br>0.07 | 0.05 | 0.02 | - |
| **e*Sa*Cas9-D10A + U6-sgRNA 7** | 1 | 0.08<br>0.03 | 0.06 | 0.03 | - |

| Treatment | NHEJ Probe | Fractional Abundance | Average Indel Frequency (%) | Standard Deviation | p-value |
|---|---|---|---|---|---|
| **e*Sa*Cas9-D10A + U6-sgRNA 8** | 2 | 0.12<br>0.03 | 0.08 | 0.07 | - |
| **e*Sa*Cas9-D10A + U6-sgRNA 7/8** | 1 | 1.43<br>1.09 | 1.26 | 0.24 | 0.0185 |
| | 2 | 0.83<br>0.77 | 0.80 | 0.04 | 0.0023 |
| **Water** | 1 | 0.00<br>0.00 | 0.00 | 0.00 | - |
| | 2 | 0.00<br>0.00 | 0.00 | 0.00 | - |
| **NS shRNA** | 1 | 0.06<br>0.06 | 0.06 | 0.00 | - |
| | 2 | 0.03<br>0.11 | 0.07 | 0.06 | - |
| ***CCR5* shRNA** | 1 | 0.05<br>0.03 | 0.04 | 0.01 | - |
| | 2 | 0.06<br>0.03 | 0.04 | 0.02 | - |

Note: a Welch's correction was applied to obtain p-values in Bold.

**Supplementary Table 7.5.3: Biological triplicate qRT-PCR results for *CCR5* and the *β-actin* reference control obtained from Non-Specific (NS) shRNA and shRNA 1005 treatment of TZM-bl cells**

| Treatment | Gene Name | Threshold Value | Normalised Ratio | Relative Level of *CCR5* mRNA (%) | Relative Ratio Ave (%) | Relative Ratio StDev | p-value |
|---|---|---|---|---|---|---|---|
| **NS shRNA** | *CCR5* | 18.42 | 0.123 | 100.00 | 100 | - | - |
| | | 18.45 | 0.125 | 100.00 | | | |
| | | 17.19 | 0.289 | 100.00 | | | |
| | *β-actin* | 16.40 | | | - | | |
| | | 16.45 | | | | | |
| | | 16.40 | | | | | |
| ***CCR5* shRNA** | *CCR5* | 18.72 | 0.085 | 68.77 | 66.85 | 4.65 | 0.0002 |
| | | 18.74 | 0.088 | 70.22 | | | |
| | | 18.00 | 0.178 | 61.55 | | | |
| | *β-actin* | 16.16 | | | - | | |
| | | 16.23 | | | | | |
| | | 16.51 | | | | | |

**Supplementary Table 7.5.4: Technical duplicate qRT-PCR results for *CCR5* and the *β-actin* reference control obtained from CRISPR/*Sa*Cas9 as well as Non-Specific (NS) shRNA and shRNA 1005 treatment of TZM-bl cells**

| Treatment | Gene Name | Threshold Value | Normalised Ratio | Ave Ratio | StDev | Relative Ratio (%) | Relative Ratio StDev | p-value |
|---|---|---|---|---|---|---|---|---|
| ***Sa*Cas9 Mock Control** | *CCR5* | 17.36 | 0.376 | 0.372 | 0.005 | 100 | 1.462 | - |
| | | 17.47 | 0.369 | | | | | |
| | *β-actin* | 15.95 | - | | | - | | |
| | | 16.03 | - | | | | | |
| ***Sa*Cas9 + U6-sgRNA 7** | *CCR5* | 18.25 | 0.24 | 0.267 | 0.038 | 71.55 | 10.138 | 0.0581 |
| | | 17.81 | 0.293 | | | | | |
| | *β-actin* | 16.19 | - | | | - | | |
| | | 16.04 | - | | | | | |
| ***Sa*Cas9 + U6-sgRNA 8** | *CCR5* | 18.02 | 0.268 | 0.287 | 0.027 | 77 | 7.176 | 0.0475 |
| | | 17.66 | 0.306 | | | | | |
| | *β-actin* | 16.12 | - | | | - | | |
| | | 15.95 | - | | | | | |
| ***Sa*Cas9 + U6-sgRNA 7/8** | *CCR5* | 17.93 | 0.252 | 0.246 | 0.008 | 65.98 | 2.259 | 0.003 |
| | | 17.69 | 0.24 | | | | | |
| | *β-actin* | 15.94 | - | | | - | | |
| | | 15.63 | - | | | | | |
| **e*Sa*Cas9 Mock Control** | *CCR5* | 18 | 0.276 | 0.273 | 0.004 | 100 | 1.478 | - |
| | | 17.84 | 0.27 | | | | | |
| | *β-actin* | 16.14 | - | | | - | | |
| | | 15.95 | - | | | | | |
| **e*Sa*Cas9 + U6-sgRNA 7** | *CCR5* | 18.08 | 0.195 | 0.189 | 0.008 | 69.28 | 3.06 | 0.0063 |
| | | 18.11 | 0.183 | | | | | |
| | *β-actin* | 15.72 | - | | | - | | |
| | | 15.66 | - | | | | | |
| **e*Sa*Cas9 + U6-sgRNA 8** | *CCR5* | 17.67 | 0.344 | 0.31 | 0.048 | 113.59 | 17.739 | 0.3916 |
| | | 17.92 | 0.276 | | | | | |
| | *β-actin* | 16.13 | - | | | - | | |
| | | 16.06 | - | | | | | |
| **e*Sa*Cas9 + U6-sgRNA 7/8** | *CCR5* | 18.27 | 0.163 | 0.174 | 0.016 | 63.87 | 5.939 | 0.0142 |
| | | 17.85 | 0.186 | | | | | |
| | *β-actin* | 15.65 | - | | | - | | |
| | | 15.42 | - | | | | | |
| ***Sa*Cas9-D10A Mock Control** | *CCR5* | 17.76 | 0.319 | 0.31 | 0.012 | 100 | 3.9 | - |
| | | 17.8 | 0.302 | | | | | |
| | *β-actin* | 16.11 | - | | | - | | |
| | | 16.07 | - | | | | | |

| Treatment | Gene Name | Threshold Value | Normalised Ratio | Ave Ratio | StDev | Relative Ratio (%) | Relative Ratio StDev | p-value |
|---|---|---|---|---|---|---|---|---|
| ***Sa*Cas9-D10A + U6-sgRNA 7** | *CCR5* | 17.68<br>17.77 | 0.276<br>0.257 | 0.266 | 0.013 | 85.87 | 4.219 | 0.0747 |
| | *β-actin* | 15.82<br>15.81 | -<br>- | | | - | | |
| ***Sa*Cas9-D10A + U6-sgRNA 8** | *CCR5* | 17.89<br>17.84 | 0.366<br>0.297 | 0.332 | 0.049 | 106.97 | 15.668 | 0.6144 |
| | *β-actin* | 16.44<br>16.09 | -<br>- | | | - | | |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | *CCR5* | 18.65<br>18.09 | 0.209<br>0.222 | 0.216 | 0.009 | 69.51 | 3.056 | 0.0125 |
| | *β-actin* | 16.39<br>15.92 | -<br>- | | | - | | |
| **NS shRNA** | *CCR5* | 18.79<br>18.44 | 0.243<br>0.262 | 0.253 | 0.014 | 100 | 5.37 | - |
| | *β-actin* | 16.75<br>16.51 | -<br>- | | | - | | |
| ***CCR5* shRNA** | *CCR5* | 18.94<br>18.78 | 0.156<br>0.172 | 0.164 | 0.011 | 64.85 | 4.447 | 0.0191 |
| | *β-actin* | 16.26<br>16.24 | -<br>- | | | - | | |
| **NS shRNA No RT Control** | *CCR5* | -<br>30.95 | -<br>- | | | - | | |
| | *β-actin* | -<br>- | -<br>- | | | - | | |
| ***CCR5* shRNA No RT Control** | *CCR5* | -<br>- | -<br>- | | | - | | |
| | *β-actin* | -<br>- | -<br>- | | | - | | |
| **Negative Control** | *CCR5* | -<br>31.93 | -<br>- | | | - | | |
| | *β-actin* | -<br>- | -<br>- | | | - | | |

**Supplementary Table 7.5.5: Biological triplicate luciferase readings obtained from CRISPR/*Sa*Cas9; Non-Specific (NS) shRNA and shRNA 1005 treatment of TZM-bl cells**

| Treatment | Luciferase Reading | Average Reading | StDev | Relative Reading (%) | Relative StDev | p-value |
|---|---|---|---|---|---|---|
| ***Sa*Cas9 Mock Control** | 1.55E+06<br>2.02E+06<br>1.97E+06 | 1.85E+06 | 2.58E+05 | 100 | 13.99 | - |
| ***Sa*Cas9 + U6-sgRNA 7** | 1.30E+06<br>8.01E+05<br>1.29E+06 | 1.13E+06 | 2.85E+05 | 61.18 | 15.42 | 0.0321 |
| ***Sa*Cas9 + U6-sgRNA 8** | 1.29E+06<br>1.27E+06<br>1.04E+06 | 1.20E+06 | 1.40E+05 | 64.81 | 7.56 | 0.0188 |
| ***Sa*Cas9 + U6-sgRNA 7/8** | 5.57E+05<br>1.06E+06<br>1.02E+06 | 8.80E+05 | 2.81E+05 | 47.68 | 15.2 | 0.0116 |
| **e*Sa*Cas9 Mock Control** | 2.33E+06<br>2.68E+06<br>1.83E+06 | 2.28E+06 | 4.28E+05 | 100 | 18.75 | - |
| **e*Sa*Cas9 + U6-sgRNA 7** | 1.32E+06<br>1.46E+06<br>1.09E+06 | 1.29E+06 | 1.88E+05 | 56.48 | 8.23 | 0.0212 |
| **e*Sa*Cas9 + U6-sgRNA 8** | 1.52E+06<br>1.61E+06<br>1.19E+06 | 1.44E+06 | 2.22E+05 | 63.22 | 9.75 | 0.039 |
| **e*Sa*Cas9 + U6-sgRNA 7/8** | 1.17E+06<br>1.17E+06<br>9.55E+05 | 1.10E+06 | 1.23E+05 | 48.12 | 5.41 | 0.01 |
| ***Sa*Cas9-D10A Mock Control** | 2.05E+06<br>1.89E+06<br>1.38E+06 | 1.77E+06 | 3.48E+05 | 100 | 19.6 | - |
| ***Sa*Cas9-D10A + U6-sgRNA 7** | 1.53E+06<br>1.72E+06<br>1.49E+06 | 1.58E+06 | 1.24E+05 | 89.23 | 6.99 | 0.4176 |
| ***Sa*Cas9-D10A + U6-sgRNA 8** | 1.91E+06<br>1.74E+06<br>1.29E+06 | 1.64E+06 | 3.17E+05 | 92.69 | 17.9 | 0.6678 |
| ***Sa*Cas9-D10A + U6-sgRNA 7/8** | 1.35E+06<br>1.54E+06<br>1.10E+06 | 1.33E+06 | 2.22E+05 | 75 | 12.5 | 0.137 |
| **NS shRNA** | 8.65E+05<br>9.34E+05<br>9.27E+05 | 9.09E+05 | 3.79E+04 | 100 | 4.17 | - |
| ***CCR5* shRNA** | 5.32E+05<br>1.55E+06<br>2.02E+06 | 6.44E+05 | 1.02E+05 | 70.89 | 11.26 | 0.0137 |

## 7.6. Ethics Waiver Letter

**Human Research Ethics Committee (Medical)** 50 years 1966 – 2016

Research Office Secretariat: Faculty of Health Sciences, Phillip Tobias Building, 3rd Floor, Office 301,
29 Princess of Wales Terrace, Parktown, 2193 Tel +27 (0)11-717-1252 /1234/2658/2700
Private Bag 3, Wits 2050, email: zanele.ndlovu@wits.ac.za
Office email: hrec-medical.researchoffice@wits.ac.za
Website: www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/

Ref: W-CJ-170301-1 01/03/2017

***TO WHOM IT MAY CONCERN:***

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

**Investigator: Miss B Hanson (student no. 597766).**

**Project title: Gene editing of human *CCR5* by an enhanced CRISPR/*Sa*Cas9 Nickase System.**

**Reason:** this is a laboratory study using cell lines obtained from a repository. There are no human participants.

UNIVERSITY OF THE WITWATERSRAND
PROF PE CLEATON-JONES
HREC (MEDICAL)
2017 -03- 03
JOHANNESBURG

Professor Peter Cleaton-Jones

Chair: Human Research Ethics Committee (Medical)

Copy – HREC (Medical) Secretariat: Zanele Ndlovu, Rhulani Mkansi, Lebo Moeng.