

Strategies to facilitate the long-term expression of transgenes for treatment of chronic liver disease

**Thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy**



Joost van Haasteren

**Gene Medicine Research Group
Radcliffe Department of Medicine**

and

**Lady Margaret Hall
University of Oxford**

Supervisors: Prof. Deborah R. Gill and Prof. Stephen C. Hyde

Trinity Term 2019

Strategies to facilitate the long-term expression of transgenes for treatment of chronic liver disease

Joost van Haasteren

Radcliffe Department of Medicine and Lady Margaret Hall, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2019

Abstract

The development of new gene therapy products offers ground-breaking opportunities for the treatment of human disease. To provide lasting benefit to patients, the expression of a correct copy of a non-functional gene needs to be life-long, and the maintenance of therapeutic levels of gene expression is a key challenge. Gene therapy for chronic liver disease using recombinant adeno-associated virus, is very encouraging, but as the vector exists as an extra-chromosomal episome, it can be lost during subsequent cell division, especially in paediatric liver disease.

In this study, maintenance of transgene expression was investigated through repeated administration of rAAV after inducing immunogenic tolerance to rAAV capsid proteins. Using a liver-specific rAAV2/8, induction of tolerance to the non-self transgene luciferase was successfully demonstrated in mice. However, rAAV2/8 delivery of the AAV8 capsid was unable to induce humoral tolerance to the AAV8 capsid itself. Similarly, a lentiviral directed approach in the liver did not induce tolerance to AAV8 capsid. An alternate method to achieve life-long transgene expression utilised a gene editing strategy called homology-independent targeted integration (HITI), that facilitates insertion of a transgene sequence in a precise location in the genome, allowing, for example, therapeutic transgene expression under the strict control the endogenous promoter of an otherwise defective target gene. A fluorescent reporter cell line was constructed to test the feasibility of HITI and was used to

confirm multiple permutations of the integration location and the sequence inserted, each of which offers a distinct therapeutic opportunity. Optimisation of the HITI donor sequence was required to prevent potential false positive expression in the analysis of *in vivo* HITI applications. One permutation was successful in addressing alpha-1 antitrypsin deficiency *in vitro* and subsequently in a disease mouse model. Two new quantitative analytical methods, Capture-Seq and CasEnrich, were used to determine the propensity for off-target integration in the models tested. While these next-generation sequencing approaches were optimised for analysis of HITI on- and off-target integration events, they can be readily applied to interrogate the genomic consequences other gene editing and gene therapy approaches.

In conclusion, in the absence of a successful protocol for repeat rAAV administration *in vivo*, the HITI approach provides a promising platform for targeted genomic integration leading to life-long expression of therapeutic transgenes from an endogenous promoter.

Acknowledgements

First of all, I would like to thank my supervisors and mentors, Prof Deborah Gill and Prof Steve Hyde for their wonderful support over the years, guiding me scientifically and professionally to where I am now.

I cannot overstate the impact my family has had during my DPhil as well as the path that led to it. My soulmate soothes my mind with love and *Thaikun* and makes all things well again. An unrelentingly supportive parental team 'ElTon' have given me everything I take for granted with exceeding generosity. Taking Afua in as your own, we make a great family of four. Maude and Nana, thank you for the *fufu* and raising such a wonderful daughter. My great friends Lynn and Esra for all the laughs and wine therapy.

The Gene Medicine lab is a truly delightful place to do science, not in the last place by all the members I have had the pleasure to share it with, past and present. Dr Ian Pringle, Dr Jean-François Gélinas, Dr Natasha Davies, Dr Agata Antepowicz, Dr Jack Tan, Dr Toby Gamlen Dr Kamran Miah, Dr Altar Munis, Dr Yue Du, the ones shortly becoming Dr's themselves: Rosie Munday, Barnabas Williams, Helena Meyer-Berg, Rachel Ashworth, Omar Habib, Eoin Mac Réamoinn and support staff Mariana Viegas, Rebecca Dean and Catriona Conway. Special mention must go out to Tina Garland, quite possibly the most important pillar of the lab, providing immeasurable support and kindness to all around her.

Technical support was provided by Dr James Davies, who I thank wholeheartedly for guiding my foray into NGS and bioinformatics, and Prof Meri Huch and Dr Gianmarco Mastrogiovanni for their training in liver organoids. Andrew Pinsent and the BMS JR did all mouse breeding. This work was primarily supported by an RDM DPhil scholarship, with additional funding for the organoid work from the Children's Liver Disease Foundation and mouse work by a joint grant from the Children's Liver Disease Foundation and Guts UK.

Declaration

This thesis has been composed by myself and has not been used in any previous application for a degree. The results presented were obtained by myself and contributions made by others have been acknowledged within the text and legends. All sources of published or unpublished information have been specifically acknowledged by means of a reference or personal communication notation, respectively.

List of abbreviations

AAP	assembly-activating protein
AAT	alpha-1 antitrypsin
AATD	AAT deficiency
AAV	adeno-associated virus
AAVR	AAV receptor
AAVS1	adeno-associated virus integration site 1
ABI1	Abl Interactor 1
ABE	adenine base editor
Ad	adenovirus
ADAR	adenosine deaminase acting on RNA
ADA-SCID	adenosine deaminase deficiency form of SCID
a-EJ	alternative end-joining
ALT	alanine aminotransferase
ANOVA	analysis of variance
APC	antigen-presenting cell
APoE/hAAT	apolipoprotein E enhancer/human AAT promoter
AST	aspartate aminotransferase
BLI	bioluminescence imaging
bp	base pairs
BSA	bovine serum albumin
CASI	CMV enhancer, chicken b-actin promoter and ubiquitin enhancer
Cap	capsid protein
Cas	CRISPR associated
CBE	cytosine base editor
cDNA	complementary DNA
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
CK8	creatine kinase 8
CMV	cytomegalovirus
CMVie	CMV immediate early promoter
cNHEJ	classic NHEJ
CNV	copy number variation
COPD	chronic obstructive pulmonary disease
CpG	5'-CG-3' dinucleotide
CRISPR	clustered regularly interspaced short palindromic repeats
DAPI	4',6-diamidino-2-phenylindole
DC	dendritic cell
dCas	'dead' or catalytically inactivated Cas
ddPCR	digital droplet PCR
DEPC	diethylpyrocarbonate
DMD	Duchenne's muscular dystrophy
DMEM	Dulbecco's modified eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNMT3A	DNA methyltransferase 3A

dNTPs	deoxynucleoside triphosphates
DSB	double-strand break
dsDNA	double stranded DNA
dsODN	double stranded oligodeoxynucleotide
EC₅₀	half maximal effective concentration
ECM	extracellular matrix
EF1α	human elongation factor-1 alpha
EGFP	enhanced fluorescent green protein
ELISA	enzyme-linked immunosorbent assay
ELISPOT	enzyme-linked immunosorbent spot
ERT	enzyme replacement therapy
FACS	fluorescence-activated cell sorting
FIX	factor IX
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GC	genome copies
GFP	green fluorescent protein
GLux	Gaussia luciferase
gRNA	(single) guide RNA
hAAT	human alpha-1-antitrypsin
HCC	hepatocellular carcinoma
HDR	homology-directed repair
HEK293T	human embryonic kidney cells containing the SV40 T-antigen
HITI	homology-independent targeted integration
HIV1	human immunodeficiency virus type 1
HLA	human leukocyte antigen
HFN1A	hepatocyte nuclear factor 1 homeobox A
HRP	horseradish peroxidase
IL10	interleukin 10
IM	intramuscular
IP	intraperitoneal
IRES	internal ribosome entry site
ITR	inverted terminal repeat
IV	intravenous
IVIG	intravenous immunoglobulin
KRAB	Krüppel-associated box
LB	lysogeny broth
LSEC	liver sinusoidal endothelial cell
LTR	long terminal repeat
Lux	firefly luciferase
mAAP	membrane-associated accessory protein
MHC	major histocompatibility complex
MMEJ	micro-homology mediated end-joining
miRNA	microRNA
miR-142-3pT	microRNA target site for miR-142-3p
MOI	multiplicity of infection
mRNA	messenger RNA
mWPRES	modified woodchuck hepatitis virus regulatory element
nAbs	neutralising antibodies

NHEJ	non-homologous end-joining
NHP	non-human primate
<i>ObLiGaRe</i>	<i>obligate ligation-gated recombination</i>
ORF	open reading frame
OVA	ovalbumin
p300	human acetyltransferase p300
PAM	protospacer-adjacent motif
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEI	polyethylenimine
PFA	paraformaldehyde
PI*S	glu264val substitution mutation of AAT
PI*Z	glu342lys substitution mutation of AAT
PITCH	precise integration into target chromosomes
polyA	polyadenylation
PYL1	abscisic acid receptor PYL1
qPCR	quantitative PCR
rAAV	recombinant adeno-associated virus
rAd	recombinant adenovirus
RLU	relative light unit
rLV	recombinant lentiviral vector
RNA	ribonucleic acid
RNP	ribonucleoprotein
SCID	severe combined immunodeficiency
scRNA	CRISPR RNAs
SD	standard deviation
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	standard error of the mean
shRNA	short hairpin RNA
siRNA	small interfering RNA
soGLux	CpG free, codon optimised version of GLux
SSA	single-strand annealing
ssDNA	single stranded DNA
SV40	simian virus 40
TALENs	transcription activator-like (TAL) effector nucleases
TBS	Tris-buffered saline
Tet-On	inducible tetracycline-regulated gene expression system
TGF-β	transforming growth factor beta
TMB	3,3',5,5'-tetramethylbenzidine
TNF-α	tumour necrosis factor α
tracrRNA	trans-activating crRNA
Tregs	regulatory T-cells
UK	The United Kingdom of Great Britain and Northern Ireland
USA	The United States of America
UTR	untranslated region
VPR	VP64-p65-Rta fusion protein
xg	times gravity
ZFN	zinc-finger nucleases

Table of contents

Abstract	I
Acknowledgements	III
Declaration	IV
List of abbreviations	V
Table of contents	VIII
List of figures	XIV
List of tables	XVII
Chapter 1 – Introduction	1
1.1 Gene therapy for the treatment of genetic diseases	1
1.1.1 The gene addition approach.....	1
1.1.2 Adeno-associated virus	3
1.1.2.1 Adeno-associated virus properties and genome structure	3
1.1.2.2 Recombinant adeno-associated virus as a delivery vector	6
1.1.2.3 AAV induced genotoxicity.....	11
1.1.3 Immunological pitfalls of rAAV therapeutics	13
1.1.3.1 Cellular response to rAAV	13
1.1.3.2 Humoral response to rAAV	14
1.1.3.3 rAAV induced immune toxicity.....	16
1.2 Genome editing	18
1.2.1 Genome editing tools.....	18
1.2.2 DNA repair	26
1.2.3 Genome editing approaches	32
1.2.4 Homology-independent targeted integration	34
1.2.5 Off-target effects of genome editing	41
1.2.5.1 Detecting off-target dsDNA cuts.....	41
1.2.5.2 Preventing off-target dsDNA cuts.....	42
1.3 Alpha-1 antitrypsin deficiency	45
1.3.1 Novel treatments for AATD.....	51
1.4 Aims	53
Chapter 2 – Materials and Methods	54
2.1 Common reagents and solutions	54

2.2	Cells.....	54
2.3	Liver organoids isolation and maintenance	54
2.4	Microscopy	55
2.4.1	Immunohistochemistry.....	55
2.4.1.1	Cells	55
2.4.1.2	Liver cryosections.....	56
2.4.2	Fluorescence microscopy imaging and cell counting	56
2.5	Flow cytometry	57
2.6	Plasmid cloning and transfection.....	57
2.6.1	Expression plasmids and HITI donor plasmids	57
2.6.2	CRISPR/Cas9 and HITI Donor plasmid gRNA insertion	58
2.6.3	Site-directed mutagenesis	58
2.6.4	Blunt TOPO cloning.....	59
2.6.5	Transformation and bacterial culture	59
2.6.6	Plasmid extraction and purification.....	60
2.6.7	Agarose gel electrophoresis	60
2.6.8	DNA fragment gel extraction	61
2.6.9	Plasmid transfection of cells	61
2.6.10	Plasmid transfection of liver organoids.....	61
2.6.11	Plasmids used in this study	63
2.7	Polymerase chain reaction.....	68
2.8	Sanger sequencing	68
2.9	Digital droplet PCR	68
2.10	T7EI mismatch detection assay.....	69
2.11	Virus production and transduction	73
2.11.1	Recombinant AAV virus production.....	73
2.11.2	Recombinant AAV virus titration.....	74
2.11.3	Lentiviral virus production.....	74
2.11.4	Viral transduction of organoids	75
2.11.5	Recombinant viruses used in this study	77
2.12	Animal studies.....	79
2.12.1	Mouse strains	79
2.12.2	PI*Z knock-in genotyping.....	79
2.12.3	In vivo virus delivery methods	81

2.12.3.1	Intramuscular injection	81
2.12.3.2	Intravenous injection	81
2.12.4	Blood sampling	81
2.12.5	In-vivo bioluminescence imaging.....	81
2.12.6	Organ harvesting	82
2.12.7	In-vivo viral copy number assay (qPCR)	82
2.12.8	Anti-AAV ELISA	83
2.12.9	Interferon- γ ELISPOT	83
2.12.10	 <i>Gaussia</i> luciferase assay	84
2.12.11	 Cryosectioning of mouse liver	84
2.13	Statistical analysis	85
Chapter 3	– Repeat rAAV2/8 administration through immune tolerance	87
3.1	Introduction	87
3.1.1	Anti-AAV humoral immunity and the repeat administration conundrum	87
3.1.2	Current approaches to achieve repeat administration	88
3.1.3	Immune tolerance as a repeat administration strategy	89
3.1.4	Aims.....	90
3.2	Results	94
3.2.1	Expression and biodistribution of rAAV2/8 vectors with ubiquitous promoters	94
3.2.2	Restricting transgene expression to the liver using a hepatocyte specific promoter	99
3.2.3	Liver-restricted expression induces cellular immune tolerance to luciferase.	102
3.2.4	Humoral immune response to AAV8 prevents repeat administration of rAAV2/8	106
3.2.5	rAAV2/8 mediated liver expression of AAV8 capsid does not induce tolerisation	107
3.2.6	Lentivirus mediated liver expression of AAV8 capsid does not induce tolerisation	111
3.3	Discussion	116
3.3.1	Restricting transgene expression to the liver and inducing immune tolerance	116
3.3.2	Use of rAAV2/8 to express AAV8 capsid protein in the liver does not induce humoral immune tolerance to rAAV2/8	116

3.3.3 AAV8 capsid protein expression in the liver with VSV-G pseudotyped lentivirus does not induce humoral immune tolerance to rAAV2/8	117
3.3.4 Potential adverse effects of capsid expression	119
3.3.5 Conclusion	119
Chapter 4 – Development of a reporter cell line to demonstrate HITI permutations	122
4.1 Introduction	122
4.1.1 Genome editing for large constructs	122
4.1.2 Aims	126
4.2 Results	126
4.2.1 Reporter cell line construct	126
4.2.2 5' and 3' integration of mNeonGreen with a transcription blocker or IRES... ..	127
4.2.3 Replacement of mCherry with mNeonGreen	137
4.2.4 Development of a HITI reporter mouse line	140
4.2.5 Strategies to prevent AAV2 ITR expression	142
4.2.5.1 Dual transcription blocker	144
4.2.5.2 HITI SpliceOut	146
4.2.5.3 pX601 ITR	150
4.3 Discussion	152
4.3.1 Feasibility of different permutations of HITI permit different therapeutic applications	152
4.3.2 HITI reporter mouse model combines confirmation of HITI permutations with <i>in vivo</i> targeting for translational purposes	153
4.3.3 The AAV2 ITR is a potent promoter	153
4.3.4 Conclusion	155
Chapter 5 – Investigation into HITI as a treatment for AAT deficiency	156
5.1 Introduction	156
5.1.1 Genome editing for alpha-1 antitrypsin deficiency	156
5.1.2 Liver organoids	159
5.1.3 Aims	160
5.2 Results	160
5.2.1 mNeonGreen expression in Huh-7 cells from the endogenous AAT locus prevents AAT expression	160
5.2.2 Characterising HITI integration in the AAT locus	169
5.2.3 Liver organoids can be transduced and transfected	172

5.3 Discussion	179
5.3.1 The AAT ‘double-hit’ genome editing strategy is viable for PI*Z AATD	179
5.3.2 Liver organoids provide a promising tool for gene therapy and genome editing research	180
5.3.3 Conclusion	182
Chapter 6 – <i>In vivo</i> HITI mediated integration for alpha-1 antitrypsin deficiency ...	183
6.1 Introduction	183
6.1.1 Animal models of alpha-1 antitrypsin deficiency	183
6.1.2 Aims	184
6.2 Results	185
6.2.1 Characterisation of PI*Z KI mouse line	185
6.2.1.1 Breeding patterns and ddPCR genotyping	185
6.2.1.2 Nanopore sequencing of the PI*Z locus	186
6.2.1.2.1 Background	186
6.2.1.2.2 Method	189
6.2.1.2.3 Results	190
6.2.2 <i>In vivo</i> HITI-mediated mNeonGreen integration in PI*Z	197
6.3 Discussion	203
6.3.1 PI*Z KI mice have eight copies of the AAT PI*Z gene per allele	203
6.3.2 The double-hit AAT approach using rAAV2/8 is achievable <i>in vivo</i>	203
6.3.3 Conclusion	205
Chapter 7 – Analysis of off target sites for the HITI approach	206
7.1 Introduction	206
7.1.1 Aims	206
7.2 Results	208
7.2.1 Off-target integration detection by Capture-seq	208
7.2.1.1 Background	208
7.2.1.2 Method	208
7.2.1.3 Results	209
7.2.2 Off-target integration detection by CasEnrich	217
7.2.2.1 Background	217
7.2.2.2 Methods	217
7.2.2.3 Results	219
7.3 Discussion	224

7.3.1 Capture-seq enriches for targeted integration and potential off-target sequences	224
7.3.2 CasEnrich is not effective in detecting off-target insertion sites	225
7.3.3 Conclusion.....	226
Chapter 8 – Discussion.....	227
8.1 Repeat administration of rAAV	227
8.2 Characteristics of HITI mediated integration.....	231
8.3 Expanding HITI applications	234
8.4 Translational potential.....	239
8.5 Conclusion and future directions	240
References	242
Appendix	261

List of figures

Figure 1.1 Diagram of the genomic structure of AAV	5
Figure 1.2 AAV virus production schematic.....	12
Figure 1.3 Targeted endonuclease structure and DNA recognition	25
Figure 1.4 Major DSB repair mechanism and associated genome editing strategies	29
Figure 1.5 DSB DNA repair is dependent on cell cycle	30
Figure 1.6 Mechanisms of homology-independent integration.....	38
Figure 1.7 Mechanism of common off-target DSB detection methods	44
Figure 1.8 Geographical distribution PI*S and PI*Z	47
Figure 2.1 Diagram of the T7EI mismatch detection assay.....	72
Figure 2.2 Anti-AAV8 ELISA mechanism	86
Figure 3.1 Proposed mechanism of tolerance induction using liver expression of transgenes.....	93
Figure 3.2 Schematic of vector constructs used in Chapter 3	96
Figure 3.3 In vivo bioluminescent imaging 14 days post-injection of IV or IM delivered rAAV2/8 or rAAV2/8-Y733F expressing Lux from CASI or CMVie promoter.....	97
Figure 3.4 Time course quantification of in vivo luciferase expression by bioluminescence imaging of mice injected IM or IV with rAAV2/8 or rAAV2/8-Y733F	98
Figure 3.5 In vivo bioluminescence imaging to compare the level and biodistribution of Lux expression from rAAV2/8 and rAAV2/8-Y733F under the control of the CASI or APoE/hAAT promoter	100
Figure 3.6 Vector copy number per cell in different mouse organs following IV injection of rAAV2/8	101
Figure 3.7 Interferon- γ ELISPOT performed on splenocytes harvested from mice injected IV with rAAV2/8 expressing Lux from CASI or APoE/hAAT promoter.....	104
Figure 3.8 Quantification of interferon- γ ELISPOT stimulated with AAV8 (A) or luciferase (B) peptide	105
Figure 3.9 Anti-AAV8 antibody titres after repeat injection of rAAV2/8	108
Figure 3.10 Gaussia luciferase activity measurements in serum after single or repeat doses of rAAV2/8.....	109
Figure 3.11 Anti-AAV8 antibody titres after IV injection of EGFP control or AAV8 capsid tolerising vectors.....	110
Figure 3.12 Biodistribution of rHIV1 VSV-G APoE/hAAT Lux after IV injection.....	113
Figure 3.13 Lentivirus based tolerisation experiment timeline.....	114
Figure 3.14 Anti-AAV8 antibodies after lentiviral tolerance induction to AAV8 capsid proteins prevent repeat administration	115

Figure 4.1 HITI reporter cell line construct and verification.	130
Figure 4.2 Diagram of possible HITI permutations in the reporter construct	132
Figure 4.3 Schematic of plasmid constructs used in HITI permutation experiments.....	133
Figure 4.4 Single cut HITI permutations allow selective mNeonGreen and/or mCherry expression	135
Figure 4.5 Sanger sequencing of HITI integration junctions shows high integration fidelity	136
Figure 4.6 Replacement of mCherry with mNeonGreen results in mNeonGreen expression without mCherry expression	139
Figure 4.7 Design of HITI reporter mouse for in vivo testing of genome editing permutations	141
Figure 4.8 Expression of the fluorescent reporter gene in HITI Donor plasmid.....	143
Figure 4.9 Dual transcription blocker in HITI donor constructs to reduce ITR transgene expression	145
Figure 4.10 Schematic showing the SpliceOut mechanism	147
Figure 4.11 Configuration and expression of SpliceOut plasmids	149
Figure 4.12 HNF1a transcription factor enhances ITR transcription activity	151
Figure 5.1 AAT HITI double-hit integration mechanism.....	163
Figure 5.2 AAT gRNA target site cleavage assessment and in-and-out PCR integration assay	165
Figure 5.3 Schematic of AAV2 based viral constructs used in Chapter 5	166
Figure 5.4 Fluorescent microscope images of AAT HITI integrated cells.....	168
Figure 5.5 Investigation into AAV2 plasmid backbone integration	170
Figure 5.6 Sequencing of integration junctions determines integration fidelity.....	171
Figure 5.7 Derivation of (human) liver organoids	173
Figure 5.8 Plasmid transfection of murine liver organoids.....	174
Figure 5.9 rAAV transduction of murine and human liver organoids	178
Figure 6.1 Mendelian inheritance of the PI*Z locus results in mice with 8 or 16 copies of PI*Z.....	188
Figure 6.2 Schematic of Cas9-mediated enrichment of PI*Z locus for Nanopore sequencing	193
Figure 6.3 PI*Z reference sequence and CasEnrich gRNA design.....	194
Figure 6.4 CasEnrich mediated Nanopore sequencing of the PI*Z locus	196
Figure 6.5 Schematic of PI*Z HITI vector genomes and ddPCR assay for 5' HITI integration junction	200
Figure 6.6 Fluorescence microscopy images of in vivo HITI genome editing of the AAT locus	202

Figure 7.1 Currently available methods of detection of dsDNA insert location.....	207
Figure 7.2 Schematic of Capture-seq protocol for dsDNA inserts	213
Figure 7.3 Ideogram of selected chromosomes of Capture-seq	215
Figure 7.4 Schematic of CasEnrich method for off-target detection of HITI insertions..	221
Figure 7.5 CasEnrich mediated Nanopore sequencing allows directional sequencing from a HITI donor.....	223
Figure 8.1 Proposed single-vector, liver-specific, AAV construct with inducible expression	230
Figure 8.2 HITI strategies for genes exceeding the rAAV packaging capacity.....	237
Figure 8.3 Diagram of integrase deficient Lentiviral vector (IDLV) ‘all-in-one’ HITI construct	238

List of tables

Table 1.1 rAAV serotype receptors and in vivo tropism	9
Table 1.2 Seroprevalence of neutralising antibodies to selected AAV serotypes	17
Table 1.3 Selected Cas proteins for DNA and RNA targeting	21
Table 1.4 Other uses of Cas proteins	23
Table 1.5 Genome editing approaches and their DNA repair pathways used after blunt double strand DNA break	31
Table 1.6 In vivo HITI gene insertion	39
Table 1.7 Selected AAT variants, cellular defects, and disease association	48
Table 1.8 Most prevalent AAT allele combinations and disease risk	49
Table 2.1 Flow cytometry parameters	57
Table 2.2 Expression plasmids	63
Table 2.3 CRISPR/Cas9 targeting plasmids	64
Table 2.4 HITI Donor plasmids – no gRNA targets	65
Table 2.5 HITI Donor targeting plasmids – containing gRNA target sites	66
Table 2.6 Other plasmids	67
Table 2.7 General ddPCR conditions for CNV assay	71
Table 2.8 ddPCR cycling conditions for Bio-Rad's C1000 Touch Thermal Cycler	71
Table 2.9 rAAV production plasmid transfection calculation	76
Table 2.10 Iodixanol gradient preparation for 6 gradients or 2 virus preps	76
Table 2.11 AAV qPCR titration mix	76
Table 2.12 List of recombinant lentivirus and AAV viruses used in this study	77
Table 2.13 PI*Z KI ddPCR primers	80
Table 2.14 Peptides sequences used in ELISPOT assay	80
Table 3.1 Strategies to circumvent anti-AAV antibody responses to aid vector repeat administration	91
Table 3.2 Comparison of rAAV and rLV delivery platforms for tolerisation requirements	121
Table 4.1 Current methods for genomic insertion of large sequences	124
Table 6.1 In vivo double hit HITI experimental groups	199
Table 7.1 Capture-Seq integration sites with greatest frequency	216

Chapter 1 – Introduction

1.1 Gene therapy for the treatment of genetic diseases

Gene therapy, in its original and simplest form, means the introduction of nucleic acids into a system (usually a cell) for therapeutic benefit. Gene therapy has been studied intensively for only three decades, but its potential can be called transformative for patients who previously had no chance of a cure. Genetic diseases such as cystic fibrosis (CF) or alpha-1 antitrypsin deficiency (AATD) are severely debilitating and sometimes even lethal in nature. They have a poor track record of curative treatment options which can lead them to having no treatment option at all (called orphan diseases), in large part due to their rarity [1]. Another contributing factor for the orphan disease designation is the fact that the cause of a genetic disease is fundamentally intrinsic. For a treatment option to be truly curative rather than symptomatic – as are many current treatments for genetic diseases, including enzyme replacement therapy – it should target the underlying genetic mutation.

1.1.1 The gene addition approach

Gene addition is the process of using nucleic acids to express a non-mutant version of a protein in a cell that can no longer make a (functional) version of that protein due to a mutation. This is a feasible approach for genetic diseases in which the mutation either renders a protein dysfunctional or completely abolishes its expression. Successful gene addition hinges upon the expression of the functional protein in the right cell, at the right time and at the right level.

Probably the most crucial aspect of successful gene addition is the spatial control of gene expression. Ideally the protein is expressed in the tissue in which the affected protein is normally made, since many proteins need co-factors, such as tissue specific transcription factors, to be expressed. Similarly, proteins may rely on post-translational modification or cleavage to be fully functional, the machinery for which may only be present in cells

programmed to make this protein. One way to make sure the nucleic acid is only delivered to the intended tissue is by choosing or modifying a delivery vector to only recognise and enter the chosen tissue. This can be achieved by choosing a virus that possesses a natural tropism to a specific tissue [2], modifying an existing virus so it recognises different receptor molecules on the cell surface [3, 4] or by attaching small molecules to the nucleic acids directly to persuade a cell to internalise it [5].

Another way to restrict expression of a gene to the target tissue is to utilise a cell type-specific promoter, such as APoE/hAAT for liver expression, or CK8 for expression in the muscle [6, 7]. These promoter sequences are based on known promoter and enhancer sequences that drive strong, tissue-specific expression in their native tissue type *in vivo*; the CK8 promoter for instance is based on the muscle creatine kinase control region. The transcription factors that drive expression of the muscle creatine kinase gene are only present in the muscle, and so when the vector is delivered to another organ by mistake the promoter will not be active. In an alternative approach, computational in-silico data-mining can be used to identify transcriptional enhancer elements that contain clusters of transcription factor binding sites that can help improve transcription levels in a tissue-specific manner [8]. Another safeguard, which can be used in combination with a tissue-specific promoter, is a microRNA target sequence. These are usually added to the 3' untranslated region (UTR) of a transgene cassette and are recognised by microRNAs present in a cell in which the gene should preferentially not be expressed. The microRNA will complex with the erroneously transcribed mRNA and degrade it and thus prevent translation. This approach has worked well to prevent expression in antigen-presenting cells (APCs, using a microRNA target which is highly expressed in APCs: miR-142-3p, used in this study – **Section 3.2.1**) and to de-target expression from the liver [9, 10].

Control of temporality or expression level may be necessary when the gene is detrimental when present at the wrong time or at supra-physiologic levels. This can be the case for anti-

inflammatory biologics such as the tumour necrosis factor- α (TNF- α) inhibiting biologic etanercept, targeting diseases such as rheumatoid arthritis, insulin expression in type I diabetes, or expression of potentially toxic genes in cancer gene therapy [11]. This control can be exerted at a transcriptional [12, 13], pre-translational [14] and post-translational [15] level.

1.1.2 Adeno-associated virus

1.1.2.1 Adeno-associated virus properties and genome structure

Gene addition approaches need an efficient and controlled gene delivery vector. This more often than not involves the use of modified viruses, which have taken advantage of millions of years of evolution to optimise delivery of nucleic acids (their genome) to target cells of their host. A widely used viral vector is the recombinant adeno-associated virus (rAAV). Adeno-associated virus (AAV) is a small (~25nm) non-enveloped virus of the dependoparvovirus genus with an icosahedral (T=1) capsid formed from 60 subunits of its three capsid protein constituents, VP1, VP2 and VP3, in a 1:1:10 stoichiometry respectively. Being a dependoparvovirus, AAV depends on superinfection of a helper virus for efficient replication, which can be either a herpes virus or adenovirus. This is highlighted by the fact that AAV was discovered during an investigation into adenovirus and was first ignored as degraded adenovirus particles [16] before being recognised as a 'defective' adenovirus associated virus [17]. The capsid of the AAV particle, and in particular the hypervariable regions that are found on the surface of the virus particle, determine the target cell that the particle will transduce by binding to different receptors. The capsid is also the main immunogenic protein of AAV and specific immunological responses to different capsids are reported even when capsid sequences are very similar (for instance in the case of rh32.33 in mice, **Section 1.1.3**). The wild-type AAV genome is a ssDNA expressing four rep proteins, involved in replication, from open reading frame 1 (ORF 1) and the aforementioned capsid proteins from ORF 2, sandwiched in between two inverted terminal repeats (ITR) (**Figure 1.1**). The two ITRs are short (145bp) sequences that form hairpin structures on

either end of the genome and are indispensable for virus replication and production. Relatively recently, three more proteins have been discovered to be expressed from the small AAV genome, revealing that AAV is more complex than originally thought (**Figure 1.1**). These proteins, while still under-characterised, are proposed to be important mainly in virus replication and assembly. They include X gene, a small (155 amino acids) computationally predicted protein, ostensibly expressed from a new promoter in the AAV genome, p81 [18], which was further characterised as supporting viral DNA replication [19]. Assembly-activating protein (AAP) is a small nucleolus-localised protein that is indispensable for the assembly of some, but not all [20, 21], rAAV capsids [22]. The AAP likely functions through the stabilisation of individual capsid proteins before oligomerisation and full capsid structures are formed [20]. The membrane-associated accessory protein (mAAP) was recently discovered through computational analysis of putative, frameshifted ORFs and was localised at the membrane of producer cells (presented at ASGCT 2019, [23]). Not much is known about mAAP, except that knock-out causes inefficient rAAV production.

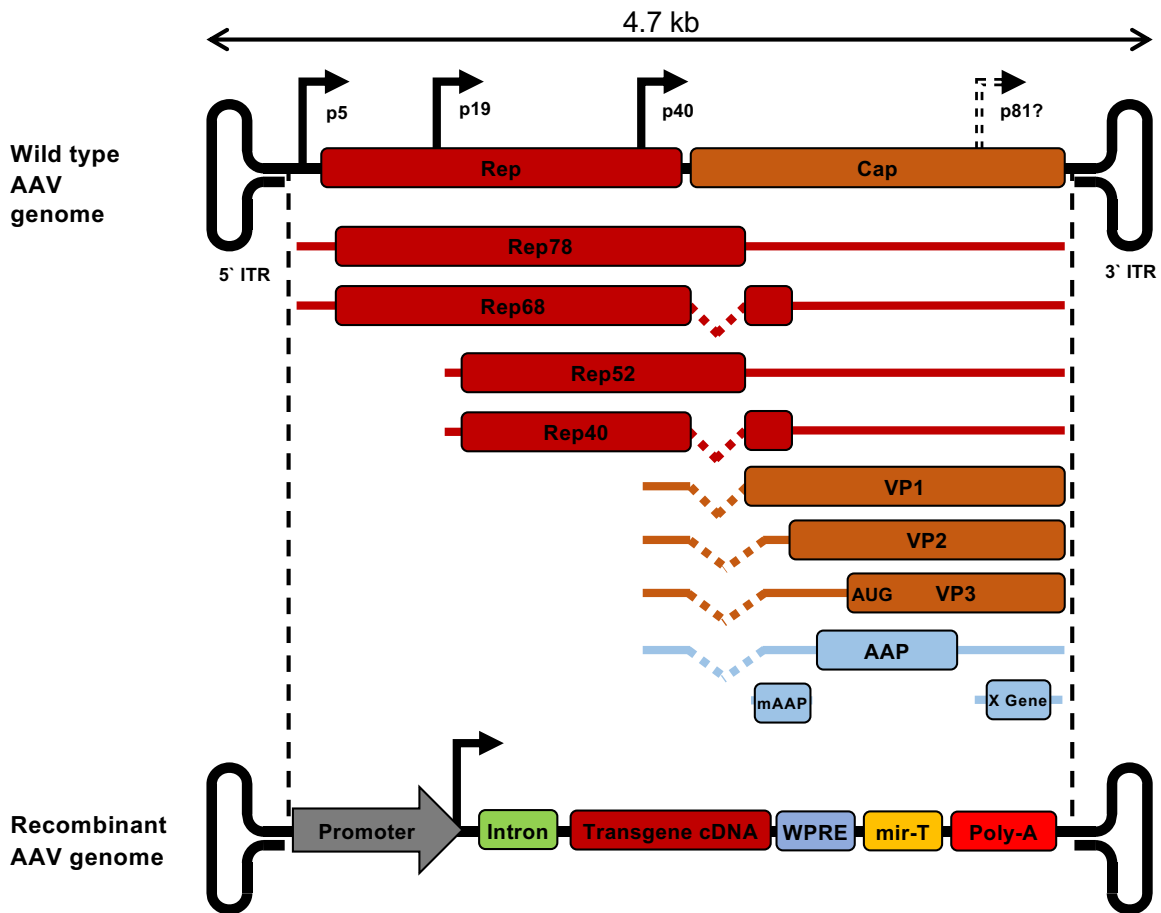


Figure 1.1 | Diagram of the genomic structure of AAV. AAV uses multiple ORFs to fit multiple proteins in a small DNA footprint (~4.7kb in total). The Rep proteins, required for genome replication and packaging, are expressed from transcripts transcribed from the p5 and p19 promoters. All four Rep proteins are translated from the same ORF, ORF1. The p40 promoter drives the expression of all three capsid proteins; VP1, VP2 and VP3. Differential splicing accounts for the differential translation of VP1 and VP2, whereas VP3 utilises a non-canonical translation start codon (AUG). AAP is expressed from the same ORF (ORF2) as the VP proteins but mAAP and the X Gene are reported to have putative alternative ORFs. Removing all the viral sequence between the two Inverted Terminal Repeats (ITR) allows the construction of a recombinant AAV genome for gene therapy application. AAP, assembly-activating protein; mAAP membrane-associated accessory protein; mir-T, microRNA target; WPRE, woodchuck hepatitis virus regulatory element.

1.1.2.2 Recombinant adeno-associated virus as a delivery vector

Recognising the potential of AAV as a gene therapy vector, Samulski and colleagues started deleting all of the viral genes, leaving only the ITRs on the left and right of the single-stranded DNA virus genome, which are required for packaging of the genome into the capsid sequence [24]. The removal of viral genes provided space for an expression cassette that can express a therapeutic protein after transduction (**Figure 1.1**). Since AAV cannot replicate solely using its own or cellular factors, in order to make rAAV for use in gene transfer, the necessary factors from a helper virus in plasmid form (in most cases adenovirus), along with the rep and cap genes, have to be provided *in trans* (**Figure 1.2**).

Since its first use as a gene delivery platform, rAAV has grown to be the most widely adopted gene therapy tool. Many naturally occurring AAV serotypes have high capsid sequence homology but vary widely in tissue tropism governed by receptor-mediated tissue specificity. The archetypal AAV variant, AAV2, was widely believed to be non-specific in its transduction ability in the decades after its discovery [17], in part due to its ability to infect multiple tissues across many species. Later, a receptor was found to facilitate rAAV2 binding to [25], and infection of [26] cells. This receptor for rAAV2, heparan sulphate proteoglycan (HSPG), is found on many cell types and species, providing an explanation of its wide infectivity. It may also explain, however, why the first clinical trial using rAAV as a gene transfer vehicle by Flotte and colleagues was not as successful as hoped [27]. This gene therapy clinical trial was aiming to treat the lungs in CF patients and, since HSPG is expressed on the basolateral membrane of the targeted lung epithelial cells, the receptor is not easily accessed by inhaled rAAV2 [28]. Recently, a genome wide gene knock-out screen identified KIAA0319L (later named as the AAV receptor or AAVR) as a receptor required for efficient cell transduction for nearly all AAV serotypes [29] with the exception of those that are derived from AAV4 [30] (see **Table 1.1**). Interestingly, there is some preliminary evidence that AAVR is also expressed primarily on the basolateral side of lung

epithelial cells [31], something for researchers to consider in their choice of lung-targeting AAV serotypes.

To date there are 13 naturally occurring capsid serotypes and more than a hundred natural variants of the main serotypes (called serovars) of AAV described in the literature. These capsid variants, while sharing a fair amount of sequence homology, have distinct transduction profiles and tissue tropism [2], largely determined by their receptors. The primary and co-receptors and tissue tropism of the main serotypes are summarised in **Table 1.1**. After binding to its (co-)receptor(s), rAAV enters the cell within 30 minutes (measured *in vitro* in [32]). The method in which rAAV gets into a target cell has not been determined definitively, but clathrin-mediated endocytosis was identified as the primary pathway [32]. However, knock-down of dynamin (a protein indispensable for clathrin-mediated endocytosis) did not completely abolish rAAV transduction [33], suggesting that alternative entry routes exist. Micropinocytosis [34] and the Clathrin-Independent Carriers/GPI-Enriched Endocytic Compartment (CLIC/GEEC) endocytic pathway [35] are contenders for these alternative entry routes of rAAV but both these pathways and clathrin-mediated endocytosis have only been described *in vitro* [36]. Contrary to rAAV cell entry, intracellular trafficking of AAV particles is a slower process – it takes 16-20h to reach the nucleus – that is inefficient as most viral particles remain perinuclear in *in vitro* experiments [32]. As with particle uptake, trafficking of rAAV is not well understood and the majority of experimental data has come from an *in vitro* setting. The fact that the principal uptake mechanism utilised by a cell differs between cell types and the suggestion that the same could hold true for tissues *in vivo* [37] warrants further investigation in appropriate *in vivo* models to allow any predictions to be made for clinical settings.

If a desired AAV cannot be selected from the natural repertoire there are approaches to improve existing capsids through a process of directed evolution (reviewed in [38]). Briefly, a library of novel capsid sequences is prepared, modified by DNA shuffling, peptide display,

or error prone PCR, and then used to transduce a desired target tissue (positive selection), or provide inhibitory factors such as neutralising antibodies (negative selection), and the capsid variants that overcome these barriers are isolated. These approaches have resulted in an extensive variety of AAV capsids and has led to rAAV being used in many therapeutic applications (reviewed in [39]).

One of the main disadvantages of rAAV as a gene therapy delivery platform is the limited packaging capacity, with a maximum of around 5kb, including ITRs, before production yields begin to be affected [40]. Considering the necessity of transcriptional elements such as a tissue-specific promoter and polyadenylation sequence this means that there are diseases where the therapeutic cDNA would be too big to fit in the rAAV (e.g. CFTR for cystic fibrosis). Producing recombinant viruses with oversized genomes results in a heterogeneous mixture of AAV particles with genomes that are truncated at the 5' or 3' ends [40]. These viruses are able to express the transgene cassette, most likely due to recombination of homologous sequences, but at a greatly reduced efficacy that hampers reliable clinical translation [41]. One solution to overcome this limitation includes splitting the transgene cassette into two halves that can recombine in the cell after co-transduction of two viruses, either via homologous recombination, ITR head-to-tail concatemerisation or trans-splicing (reviewed in [42]). Another interesting approach attempts to decrease the negative luminal charge of the capsid, to increase capacity for negatively charged DNA [43]. It is yet to be seen whether this approach will reach pre-clinical stage.

Table 1.1 | rAAV serotype receptors and *in vivo* tropism

AAV Serotype	Origin	Primary receptor(s) [44]	Co-receptor(s) [45]	AAVR dependent [29, 30]	Reported tropism [46-48]		
					Mouse	Human	Other (indicated)
AAV1	Human	$\alpha 2,3/\alpha 2,6$ N-linked sialic acid	-	Yes	skeletal muscle, CNS, airway, retina, pancreas, cardiac muscle, liver	skeletal muscle	skeletal muscle (NHP, canine), CNS (NHP), airway (NHP), cardiac (pig)
AAV2	Human	heparan sulphate proteoglycan (HSPG)	$\alpha V\beta 5$ integrin, fibroblast growth factor receptor-1 (FGFR1), hepatocyte growth factor receptor (c-MET), A5 β 1 integrin, CD9 tetraspanin, Laminin receptor (LamR)	Yes	skeletal muscle, CNS, liver, kidney, retina	skeletal muscle, CNS, retina, airway, liver, joint	CNS (NHP), retina (murine, canine, NHP)
AAV3	Human	heparan sulphate proteoglycan (HSPG)	Hepatocyte growth factor receptor (c-MET), Laminin receptor (LamR)	Yes	skeletal muscle, liver	-	-
AAV4	NHP	$\alpha 2-3$ O-linked sialic acid	-	No	CNS, retina, lung, kidney	-	retina (canine), lung (NHP)
AAV5	Human	$\alpha 2-3$ and $\alpha 2-6$ N-linked sialic acid	Platelet-derived growth factor receptor (PDGFR)	Yes	skeletal muscle, CNS, airway, retina	liver	CNS (canine)
AAV6	Human	$\alpha 2-3$ and $\alpha 2-6$ N-linked sialic acid, HSPG	Epidermal growth factor receptor (EGFR)	Yes	skeletal muscle, heart muscle, airway	-	skeletal muscle (canine), heart muscle (canine, sheep, pig),

									airway (canine, NHP)
AAV7	Rhesus Macaque	-	-	-	-	-	-	skeletal muscle, retina, CNS, liver	-
AAV8	Rhesus Macaque	-	Laminin receptor (LamR)	Yes	-	-	-	skeletal muscle, liver, CNS, retina, pancreas, heart, kidney	liver (NHP, canine, human), CNS (canine, NHP), retina (canine)
AAV9	Human	N-linked galactose	Laminin receptor (LamR)	Yes	-	-	-	skeletal muscle, liver, heart, pancreas, CNS, retina, testes, kidney	heart (NHP, pig, NHP) skeletal muscle (canine), CNS (NHP), retina (NHP)
AAV10	Cynomolgus monkey	-	-	-	-	-	-	liver, heart, skeletal muscle, lung, kidney, and uterus	-
AAV11	Cynomolgus monkey	-	-	-	-	-	-	skeletal muscle, kidney, spleen, lung, heart, and stomach	-
AAV12	Vervet monkey	-	-	-	-	-	-	salivary gland, skeletal muscle	-
AAV13	NHP	heparan sulphate proteoglycan (HSPG)	-	-	-	-	-	-	-

‘-’ indicates no information is available for the indicated serotype. NHP, non-human primate; CNS, central nervous system.

1.1.2.3 AAV induced genotoxicity

The potentially troubling finding that AAV genome integration can contribute to the development of hepatocellular carcinoma (HCC) could argue against the perceived non-pathogenicity of AAV. There are reports that AAV integrates in higher numbers in a particular region of the murine genome encoding for several microRNAs and small nucleolar RNAs [49, 50]. The AAV integration led to the overexpression of these RNAs in HCC tumours, leading to the hypothesis that AAV integration is directly causal in HCC development [50]. Integration of AAV was found in several cancer driver genes in HCC tumours in humans [51]. Considering that only a portion of the AAV ITR sequence was purported to be necessary for a transcriptional effect of genes near the integration site, gutted rAAV genomes (in which the only viral sequences are the ITRs) could have a similar effect [51]. This could act through the bidirectional hepatocyte-specific transcription – hepatocyte nuclear factor 1 homeobox A (HFN1A) – which was recently found in the AAV2 ITR [52]. In the light of these reports, however, it must also be noted that there is little evidence directly linking AAV to an increased incidence of HCC, for either natural AAV infection (which is widespread see **Section 1.1.3 and Table 1.2**), or for rAAV vectors in the context of liver-targeting clinical trials [53, 54].

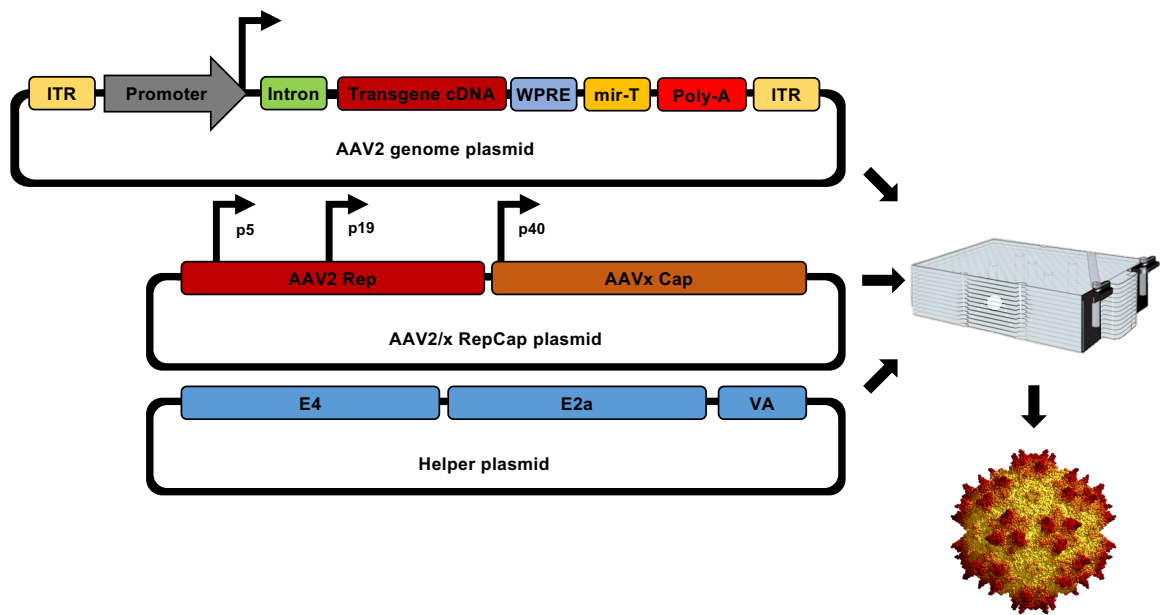


Figure 1.2 | AAV virus production schematic. Vector is produced using a triple transfection method ([55] and **Section 2.11.1**). The genome plasmid (top) is co-transfected with the pRepCap (middle) and pAd Delta F6 helper plasmid (bottom). The pRepCap plasmid typically expresses the AAV2 rep proteins (corresponding to the AAV2 ITR sequences in the AAV genome plasmid) and the capsid proteins of the serotype used. The helper plasmid expresses adenoviral proteins E4, E2a and VA. The adenoviral protein E1 is also necessary for AAV production and is typically present in the production cell line (e.g. HEK293T cells). Following transfection, the virus is harvested, purified and concentrated.

1.1.3 Immunological pitfalls of rAAV therapeutics

1.1.3.1 Cellular response to rAAV

While rAAV has many advantages as a gene therapy agent and is notable for its lack of pathogenicity in humans and other mammals, it still elicits a potent immune response in both human and commonly used animal models. This immune response is both humoral and cellular, each of which has created limitations to the use of rAAV in gene therapy, most notable in human clinical trials. The cellular immune response is exemplified by a pivotal clinical trial of haemophilia B [56], whereby rAAV2/2 was chosen to express factor IX from a liver-specific promoter to restrict expression to the liver, the natural tissue for factor IX expression. Initially the results were impressive, with levels of factor IX within the therapeutic range, but this effect waned after week 4 with a concomitant increase of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) in the serum, indicative of liver toxicity. After investigation, it was determined that liver cells were killed by a cellular CD8⁺ T-cell response against the AAV2 capsid proteins [56]. Importantly, this anti-AAV2 capsid response was never observed in pre-clinical models [57]. Notably, AAV capsid rh32.33 has been reported to uniquely elicit a T-cell response in mice when used as a gene delivery vehicle [58]. Significant sequence homology to other capsids that do not provoke such a response, indicate that small changes in both peptide sequence and protein structure could potentially have a dramatic impact on the efficacy of gene transfer tools.

It must be noted, however, that the observed 7-9 week time interval between administration of the viral vector and the first observation of an increase of AST/ALT and concomitant decrease in transgene expression [56, 59] seems incongruent with the theory that the CD8⁺ T-cell response is directed at the capsid proteins of the injected viral vector. A typical cellular antiviral response peaks between one and two weeks post-inoculation [60]. It was hypothesised that the anti-capsid response could have arisen from capsid proteins synthesised in the hepatocytes after transduction (Amit Nathwani, personal communication and [61]). While the rAAV vector genomes are gutted and thus contain no viral sequences

other than the ITRs (see **Figure 1.1**), it has long been known that DNA sequences other than the viral genome can be packaged into rAAV particles during production [62]. This DNA can consist of genomic DNA from the cells in which the virus is made, bacterial sequences such as antibiotic resistance genes from the plasmid backbones of the viral genome or helper plasmids (**Figure 1.2**) or wild-type AAV rep/cap sequences from the pRep/Cap plasmid used to create a virus with the desired serotype. Virus preps can contain up to 0.01%–0.11% of rep/cap sequences [63] and capsid expressing cells have been identified *in vivo* [61]. Even though the fraction of viruses that contains capsid sequences seems low, the high viral titres required for successful gene therapy could make the number of rAAV particles that contains capsid DNA functionally significant. For instance, it was estimated that in a clinical trial for FIX in which a strong and deleterious anti-capsid response was measured [56], the total number of viral particles capable of expressing capsid could be around 10,000 per dose [61]. Besides eliciting a CD8⁺ T-cell response, the expressed capsid could also be deleterious to transgene expression due to its reported cellular toxicity (Amit Nathwani, personal communication). Recently, strategies have shown to be able to largely, but not completely, prevent the packaging of capsid sequences by introducing a large intron in to the capsid sequence, thereby making it too large to completely fit into the space-limited rAAV particles [61, 64].

1.1.3.2 Humoral response to rAAV

The other main adaptive immune response to AAV is also aimed at its capsid, with binding and neutralising antibodies directed to the capsid proteins upon exposure. This has two main implications. Patients previously exposed to naturally occurring AAV will have neutralising antibodies (nAbs) that could prevent successful transduction of rAAV by sequestering the virus before it can reach and transduce the target tissue. The prevalence of so-called ‘pre-existing’ anti-AAV neutralising antibodies for some serotypes is summarised in **Table 1.2**. Another implication is that even after a successful first administration of therapeutic rAAV, the humoral immune response against the administered

serotype will prevent an effective repeat administration ([65, 66] and **Chapter 3**). This is important since rAAV genomes persist in cells in episomal form [67], meaning that expression of a therapeutic transgene will wane over time when the targeted cells divide and dilute out the virus genome. Having a practical strategy to permit successful repeat administration of rAAV could potentially alleviate this problem.

Pre-existing neutralising antibodies against some AAV capsids were also detected in widely used animal models, including mice, dogs and non-human primates [68, 69]. This implies that a lack of observed transduction in the target tissue, when moving from rodents to larger animal models, could have been falsely attributed to the model tested whereas previous exposure to AAV could actually have been the problem. As a notable example, Wang and colleagues observed that pre-existing immunity to AAV8 in macaques prevented transduction of rAAV2/8 to the liver, instead sequestering the virus in the spleen [70]. This highlights the necessity of screening in both clinical and pre-clinical experimental procedures.

It is important to note that there is currently no gold standard for the measurement of nAbs, with measurement methods differing in the cells used for transduction and transgene employed to measure transduction efficacy. Similarly, there is no consensus on a cut-off value of the titre of nAbs that can be considered seropositive, i.e. a level that would severely hamper the efficacy of an rAAV based treatment [71]. While perhaps impractical to scale for clinical application, it was reported that using human intravenous immunoglobulin (IVIG) in severe combined immunodeficiency (SCID) mice is a sensitive method in determining neutralising antibody titres [72]. Lower nAbs titres have been shown to prevent transduction *in vivo* as compared to *in vitro*, suggesting *in vitro* methods can underestimate the titre of AAV-clearing nAbs. Correlation between the detection of anti-AAV8 binding antibodies (antibodies that bind but may not neutralise) and neutralisation is reported to be strong [73], but another report suggests that binding antibodies can actually increase transduction in

the livers of mice [74]. Given that a naturally occurring response to AAV capsids will likely contain a polyclonal mixture of both binding and neutralising antibodies, the safest way forward is likely testing for both [75].

Another complicating factor is the fact that many people have been exposed to multiple naturally occurring AAV serotypes and thus make antibodies to all of them. Furthermore, antibodies can, because of the close homology in capsid sequence between serotypes, often bind and neutralise serotypes other than the one the antibody was raised against. These factors make predictive interpretation of anti-AAV binding and neutralising antibody titres difficult [76].

1.1.3.3 rAAV induced immune toxicity

There is also evidence that there is a potent innate immune response that can be deleterious to gene therapy outcomes. High-dose (2×10^{14} genome copies (GC) per kilogram) of rAAV administered intravenously to non-human primates (NHPs) and piglets resulted in severe acute toxicity [77, 78]. The vectors used in these studies are closely related to AAV9, which is reported to be both liver- and neurotropic as it can cross the blood-brain barrier. The NHPs in the studies suffered from strong increases in liver transaminases, independent of CD8⁺ T-cell response, and both NHPs and piglets showed lesions in the dorsal root ganglia sensory neurons severe enough to necessitate euthanasia. These reports should be assessed as the backdrop of recent successes in rAAV9 gene therapy trials for spinal muscular atrophy, and recently approved for clinical use as Zolgensma [79]. In this treatment, a similar amount of rAAV2/9 was injected with a vector genome identical to the pre-clinical trials of above. While deaths in the clinical trial were reported, patients did not present the same symptoms as in the above reports [80], although post-mortem results are required before firm conclusions can be drawn. However, it seems clear that an appropriate response to this is necessary to protect the wellbeing of current and future patients [81].

Table 1.2 | Seroprevalence of neutralising antibodies to selected AAV serotypes

AAV serotype	nAbs %	Average nAbs %	References
1	50 - 59 (France) 30 (Australia) 27 (Europe) 43 (Africa) 20 (United States)	38	[73, 82, 83]
2	59 (France) 35 (Australia) 35 (Europe) 56 (Africa) 28 - 32 (United States) 82 (Netherlands)	41	[73, 82, 84, 85]
5	3 (France) 27 (Netherlands)	15	[73, 85]
6	37 (France) 64 (Netherlands)	50	[73, 85]
8	19 (France) 27 (Australia) 22 (Europe) 31 (Africa) 14 (United States) 59 (Netherlands)	23	[73, 82, 85]
9	33 (France)	-	[73]

1.2 Genome editing

Genome editing aims to correct the underlying mutation in a genetic disease rather than providing a wild-type copy as in gene addition (**Section 1.1.1**). This correction is achieved by reverting the mutation back to the normal sequence, preventing expression of a deleterious gene, or inserting a therapeutic DNA sequence into a precise spot in the genome. It was, arguably, pioneering work on DNA repair pathways that spurred the genome editing field to the behemoth it is now. In these studies, it was found that the introduction of a double strand break (DSB), in this case through the homing endonuclease *I-SceI*, significantly increases the frequency of integration of a DNA fragment flanked by homology arms [86]. This observation allowed researchers to manipulate the genome within the cells with appreciably higher frequency. The reason that introducing a DSB increases recombination so much is that the DSB induces a DNA repair response and it is the DNA repair machinery that helps achieve the desired genomic manipulation. The discovery and development of enzymes that can introduce a DSB into any location in the genome, called programmable endonucleases, has driven the field to near unlimited opportunities.

1.2.1 Genome editing tools

Programmable endonucleases allow cuts at any location in the genome and thus permit genetic manipulation of these genomic sequences. The first innovation that made this possible in an efficient manner were zinc-finger nucleases. Zinc-fingers are conserved structural motifs commonly found in a variety of proteins, some of which are capable of binding DNA, often as part of transcription factor complexes. Engineered zinc-fingers recognise a 3bp stretch of DNA and placing multiple zinc-fingers in a row in a modular fashion allows specific binding of a stretch of DNA, usually 15 or 18bp long. As shown in **Figure 1.3A**, a *FokI* endonuclease domain on a polypeptide spacer attached to a zinc-finger (together making the zinc-finger nuclease) will allow cutting of DNA when two zinc-finger nucleases are placed within 6bp of each other on the opposite DNA strand on which the *FokI* domains can dimerise and become catalytically active [87, 88]. This dimerization

increases the recognised sequence and thus specificity two-fold. While this can be a very efficient system, creating a library of 3bp recognising zinc-fingers and pairing them to efficiently recognise a given target requires a lot of engineering. Shortly after ZFNs another system was developed which tackles some of these drawbacks. This system is based on transcription activator-like (TAL) effectors, derived from bacteria, which, like ZFNs, recognise DNA but on a single base pair level in this case [89]. Like ZFNs, TAL effector nucleases (TALENs) use *FokI* domain dimerization for its endonuclease activity [90] (**Figure 1.3B**). TALENs are easier than ZFNs to design and produce sequence specific molecules against a large number of targets, but synthesis and validation remain an obstacle for TALENs to become a widespread genome editing tool.

The real democratisation of genome editing came only after the advent of the most recent programmable endonuclease, Cas9. This system is derived from bacteria, in which it functions as an adaptive immune system against bacteriophages [91]. Bacteria ‘record’ small pieces of the phage’s genome in their own genome followed by a repetitive sequence, together they are termed clustered regularly interspaced palindromic repeats (CRISPRs). This creates a ‘memory’ of the phage genome and upon secondary infection with the same phage, an RNA copy of the ‘memory’ phage and repeat sequence together are used to guide an endonuclease to the phage genome and cut it up, rendering it incapacitated. The endonucleases in this case are CRISPR-associated (Cas) proteins, the most studied and widely used is Cas9. By combining the repeat sequence (tracrRNA) and complementing it with a crRNA containing the sequence that is targeted (combined called a single guide RNA or gRNA), the Cas9 can be guided to a specific part in the genome to exert its endonuclease activity [92] (**Figure 1.3C**). The targeting sequence can be virtually anywhere in the genome, bar some design restrictions that include GC content, repeats and most importantly the proximity to a protospacer adjacent motif (PAM). The PAM sequence is specific to the choice of Cas orthologue used, the PAM for *Streptococcus pyogenes* Cas9 (SpCas9) for instance is NGG (**Table 1.3**). The fact that targeting a new genomic sequence

is as straightforward as designing a new RNA oligo allows researchers all over the world to employ Cas-based genome engineering with ease in a wide variety of models and diseases [93]. Spurred on by the relative simplicity of gRNA design, oligo synthesis, and plasmid and viral library preparation, multidisciplinary approaches including machine learning have led to advances in our understanding of the rules that govern the efficiency (how well the Cas endonuclease can cut) and specificity (whether it cuts in the right location in the genome, see **Section 1.2.3**) of the gRNA/Cas combination [94-97].

The first Cas protein used as an endonuclease was SpCas9, derived from the bacteria *Streptococcus pyogenes* [92, 98, 99]. SpCas9 has been studied intensively and remains the prototypical Cas protein, due to the relatively high cutting efficiency (routinely up to 90% in multiple settings), stability as a ribonucleoprotein (RNP) and relatively frequent PAM distribution (approximately every 8bp) [98, 100]. New and improved Cas variants have been discovered: *Staphylococcus aureus* (SaCas9) and *Campylobacter jejuni* (CjCas9), at only 1053 and 986 amino acids respectively, have allowed the incorporation of Cas9 into rAAV genomes for effective *in vivo* genome editing [101, 102]. More recently, Cas proteins of the Cas13 family have been derived that recognise and cut RNA rather than dsDNA and, using a catalytically inactive mutant and attached accessory deaminase, it is possible to edit the RNA directly, changing adenines to inosines [103, 104]. Further Cas proteins have been summarised in **Table 1.3**. Cas enzymes including Cas12 and Cas13 have been used for nucleic acid detection, to allow for high sensitivity detection of DNA and RNA based viruses respectively [105-107]. Other uses of Cas proteins are summarised in **Table 1.4**.

Table 1.3 | Selected Cas proteins for DNA and RNA targeting

Name	First identified in	PAM or PFS *	Cut site	Cleaves DNA/RNA	Size (aa) **	Selected references
SpCas9	<i>S. pyogenes</i>	5'-NGG-3'	3bp 5' of PAM	DNA, blunt	1368	[92, 98]
SaCas9	<i>S. aureus</i>	5'-NNGRRT-3'	3bp 5' of PAM	DNA, blunt	1053	[102]
NmeCas9	<i>N. meningitidis</i>	5'-NNNNGATT-3' (DNA) 5'-NNNNCUAA-3' (RNA)	3bp 5' of PAM	DNA, blunt and RNA	1082	[108, 109]
CjCas9	<i>C. jejuni</i>	5'-NNNRYAC-3'	3bp 5' of PAM	DNA, blunt	986	[101]
StCas9	<i>S. thermophilus</i>	5'-NNAGAAW-3'	3bp 5' of PAM	DNA, blunt	1121	[108, 110]
Cas12a (formerly known as cpf1)	<i>Francisella cf. novicida</i>	5'-(T)TTN-3'	17-18bp 3' of PAM	DNA, 5bp 5' overhang	1307	[111, 112]
Cas12b (formerly known as c2c1)	<i>Alicyclobacillus acidoterrestris</i>	5'-TTN-3'	17-18bp 3' of PAM	DNA, 7bp 5' overhang	1129	[112]
Cas12c (formerly known as c2c3)	<i>Oleiphilus sp.</i>	Unknown	Unknown	Unknown	1252	[112]
Cas12d (also known as CasY)	<i>Katanobacteria</i>	5'-TA-3'	Unknown	Unknown	1200	[113]
Cas12e (also known as CasX)	<i>Deltaproteobacteria bacterium</i>	5'-TTCN-3'	12-14bp 3' of PAM	DNA, 10bp 5' overhang	986	[113, 114]
Cas13a (formerly known as c2c2)	<i>Leptotrichia shahii</i>	5'-H-3'	• Not dictated by gRNA sequence • Cuts outside of the gRNA • Preference for uracil	RNA	1389	[103, 115]
Cas13b	<i>Prevotella buccae</i>	None	• Not dictated by gRNA sequence • Cuts outside of the gRNA • Preference for uracil	RNA	1224	[116]
Cas13c	<i>Fusobacterium perfoetens</i>	Unknown	Unknown	RNA (predicted)	1120	[117]
Cas13d (also known as CasRx)	<i>Ruminococcus flavefaciens</i>	None	• Variable • Outside of gRNA sequence • Preference for uracil	RNA	930	[118]

**DNA targeting and RNA targeting use PAM and PFS nomenclature respectively to describe the sequence adjacent to the protospacer to allow discrimination between both. ** If multiple orthologues exist in the literature, the size of the first identified variant is given. PAM, protospacer adjacent motif; PFS, protospacer flanking site*

Table 1.4 | Other uses of Cas proteins

Use	Accessory protein	Summary	Selected references
DNA/RNA base editing	ABE, CBE, ADAR	dCas recruited to target site used base editing accessory enzymes to modify bases in a 5-6 bp window	[104, 119, 120]
Transcription activation/repression	VPR, KRAB	dCas recruited to vicinity of transcription start site to influence expression by recruiting transcription factors (TF). dCas9 itself can repress expression by steric hindrance of transcription factors	[121, 122]
Epigenome editing	DNMT3A, p300	dCas conjugated epigenome editing accessory proteins can exert stable regulation even after cessation of expression	[123, 124]
DNA/RNA detection (research)	Sun Tag-GFP	dCas binds DNA/RNA and presence and resistance can be detected by fluorescent probe attached to dCas. SunTag recruits multiple GFP proteins (up to 24) for increased signal. Cas13a is used for RNA detection	[103, 125]
DNA/RNA detection (diagnostic)	ssDNA cleavage reporter	Uses the collateral ssDNA cleavage activity of Cas12a and Cas13a to detect DNA and RNA respectively	[105-107]
DNA/RNA isolation	His, Flag	dCas bound to target (DNA or RNA) can be pulled down and DNA/RNA can be sequenced, or bound proteins can be subjected to mass spectrometry	[103, 126]
DNA looping	ABI1/PYL1 dimer	Using 2 dCas9 enzymes tagged to each half of the dimer targeting two regions of the genome brings those together (looping) to modulate gene expression	[127]

ABE, adenine base editor; CBE, cytosine base editor; dCas, catalytically inactivated Cas; ADAR, adenosine deaminase acting on RNA; VPR, VP64-p65-Rta fusion protein; KRAB, Krüppel-associated box; DNMT3A, DNA methyltransferase 3A; p300, human acetyltransferase p300; ABI1, Abl Interactor 1; PYL1, Abscissic acid receptor PYL1.

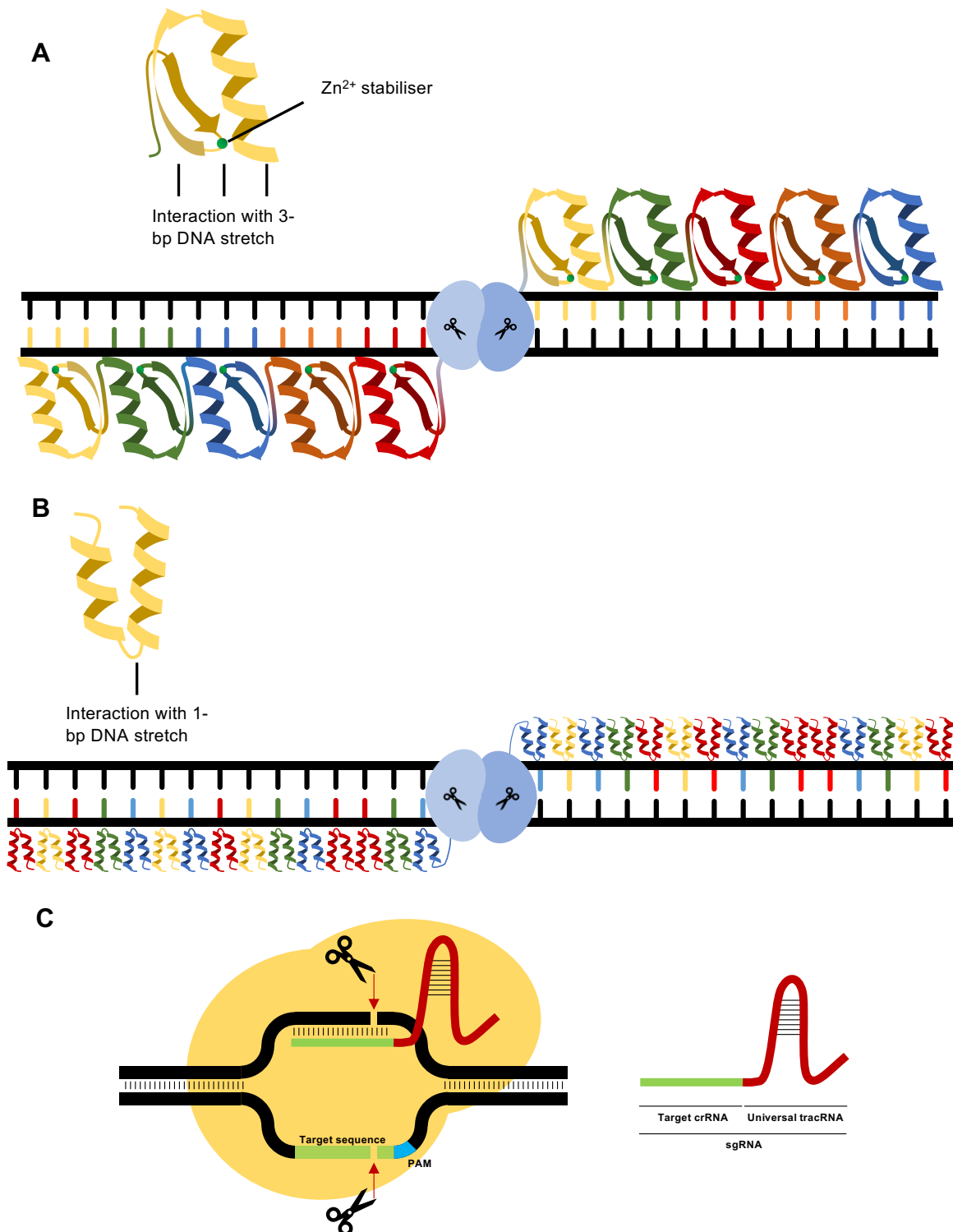


Figure 1.3 | Targeted endonuclease structure and DNA recognition. Zinc-finger nucleases (**A**) are comprised of one alpha-helix and two beta barrels, totalling around 30 amino acids, complexed with a Zn^{2+} ion. Each ZFN subunit interacts with 3 bp in the major groove of DNA. Upon dimerization the FokI nucleases will create a DSB. The FokI endonuclease domains of two 5-subunit ZFN molecules will only dimerise when both recognise the correct DNA sequence. In a similar fashion to ZFNs, TALENs (**B**) are multimeric units of a highly repetitive nature. In contrast to ZFNs, TALEN units only interact with a single base pair, simplifying the design process. TALENs also require FokI dimerization to create a DSB. Many Cas proteins however, such as Cas9, have intrinsic endonuclease activity (**C**). Rather than protein-DNA interaction to find its DSB target site, Cas9 uses a guide RNA that, upon Watson-Crick base pairing, induces a conformational change in the Cas protein to induce a blunt double strand break 3 bp from the PAM sequence. All Cas proteins so far use RNA to target their endonuclease activity, but differences in PAM selectivity, DSB placement and break pattern (blunt or overhang) and endonuclease substrate (DNA or RNA) exist (**Table 1.3**).

1.2.2 DNA repair

Genome editing outcomes are dependent on the repair machinery that attempts to fix the DSB that has been created (**Figure 1.4**). In broad terms, the main mechanisms for repair can be classified as homology-dependent or homology-independent. The homology-independent repair is known as classic non-homologous end-joining (cNHEJ), specifically termed classic to differentiate it from alternative end-joining, and is the predominant route for repair, being more active than any of the homology-dependent repair pathways in most cell types and stages of cell cycle (**Figure 1.5**) [128, 129]. Reportedly, even in S/G2 phase, where homology-directed repair (HDR) is more active, cNHEJ outcompetes HDR by a ratio of approximately 4:1 due to the overabundance of Ku70 and Ku80 proteins [128, 130]. This is consistent with the fact that humans with mutations in major NHEJ enzymes are very rare, implying that the other pathways are unlikely to be able to offer redundancy, highlighting cNHEJ as the major repair pathway [131]. Depending on Ku70 and Ku80 end-protection of primarily blunt DSBs, DNA ligase IV will join the chromosomal ends together without any need for end-resection, and with high fidelity [132]. The homology-dependent pathways all need some degree of end-resection and, as the name implies, a nucleic acid sequence that is homologous to the DNA around the DSB. For alternative end-joining (a-EJ), also known as micro-homology mediated end-joining (MMEJ), this homology is minimal with reports of 2-20 bp being sufficient [133]. This homology is often derived from micro-homologies just 5' - and 3' of the induced DSB. The a-EJ process is very error-prone as it recesses a longer stretch of DNA than it uses to repair, leading to deletions. This is likely a main reason for increased occurrence of deletions after endonuclease induced DSBs at sites with microhomology, where the preferred repair pathway might shift from cNHEJ to a-EJ [134]. The a-EJ process is active mainly during G1 and early S-phases, where blunt DSBs are normally repaired by cNHEJ and more complex breaks are favoured by a-EJ (**Figure 1.5**, [135]). Conversely, the other homology-dependent repair pathways excel in their accuracy. HDR uses relatively long (>100bp) homology, normally provided by the sister chromatid, to be used as a template for repair even when end-resection has resulted

in nucleotide deletions. This dependence on a sister chromosome is likely linked to the observation that HDR activity is negligible in cell cycle phases other than S/G2 as that is the time where the chromosomes are co-located (**Figure 1.5**). The remaining main repair pathway is single strand template repair or single strand annealing (SSA), which uses a ssDNA template rather than a double stranded one. The different repair pathways are summarised in **Table 1.5**.

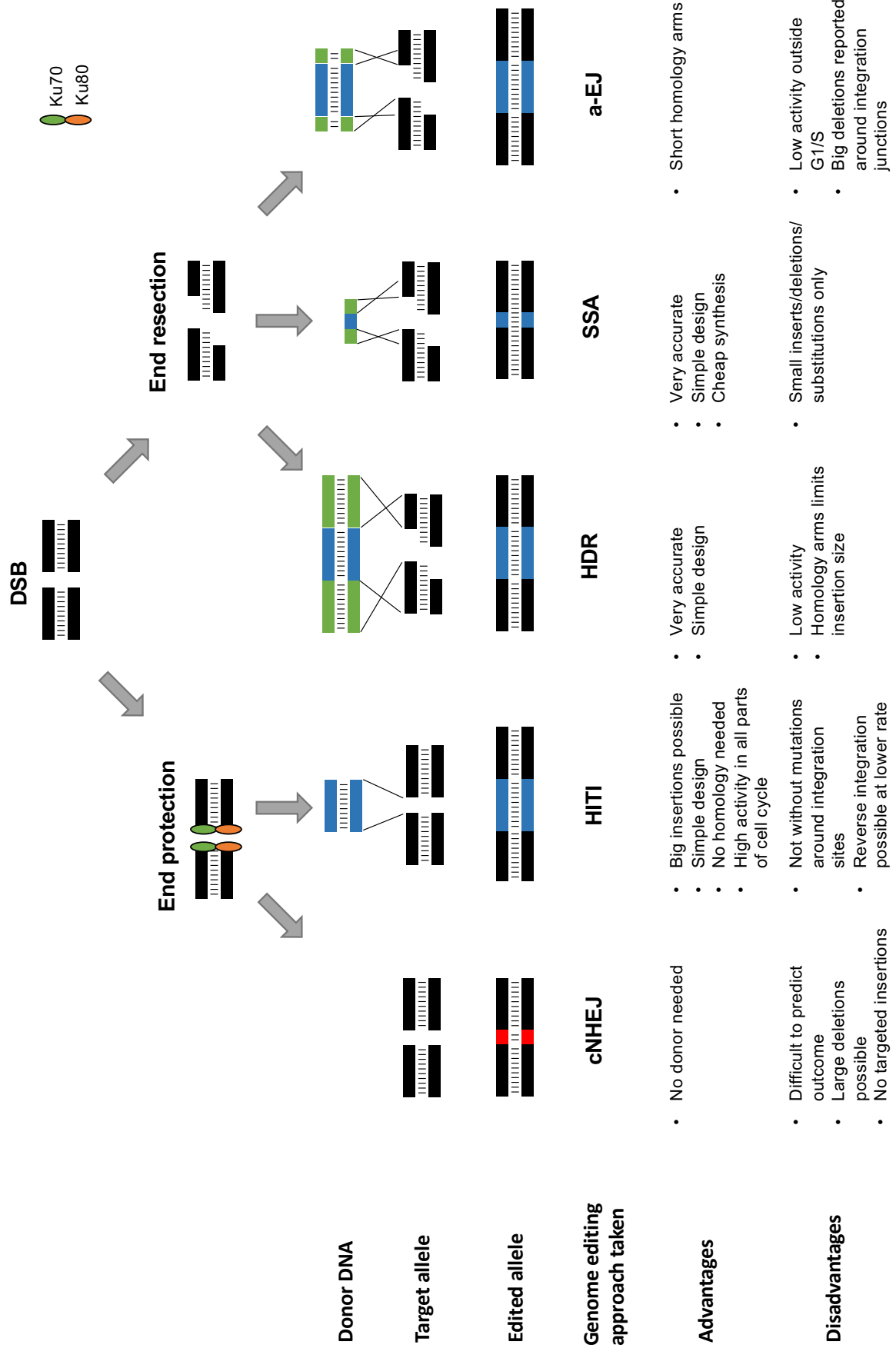


Figure 1.4 | Major DSB repair mechanism and associated genome editing strategies.

The repair pathway through which a DSB is repaired can be manipulated by the addition of appropriate repair donors to achieve a particular genome editing outcome. When no donor is provided, the default repair pathway, classical non-homologous end joining (cNHEJ) repairs the DSB with reasonable fidelity [132]. When the DSB is induced by an endonuclease that is expressed for an extended period of time, the target site is available for re-cutting after repair until an error occurs. When a dsDNA donor sequence with blunt ends is provided, this can be integrated into the DSB through cNHEJ, in a process termed homology-independent targeted integration (HITI) [136]. When a repair template is provided that contains homology to the sequence around the DSB, the repair mechanism shifts towards repair that commences with end-resection. Short (2-20bp) homology arms, reminiscent of microhomologies, induce alternative end-joining (a-EJ) used in the PITCH genome editing method [137]. A single strand repair template with short homology arms induces single strand annealing (SSA), useful for small repairs. Longer homology arms (>100bp) induce the high-fidelity repair mechanism homology-directed repair widely used by the genome editing community. Adapted from [138].

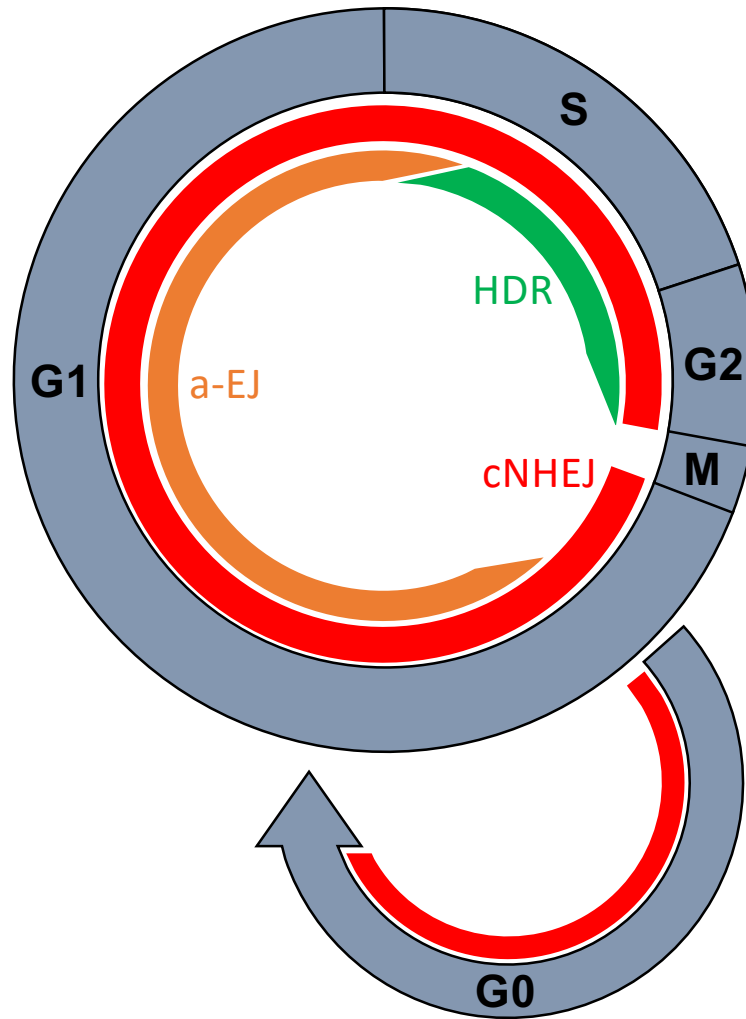


Figure 1.5 | DSB DNA repair is dependent on cell cycle. DSB repair mediated by cNHEJ (red) is active throughout the cell cycle, with the exception of the M-phase in which no DNA repair occurs at all [139] and it is therefore the most utilised pathway in naturally induced DSB events and a versatile pathway to be employed for genome editing. Repair via a-EJ (orange) is active mainly in the G1 phase and is therefore feasible in non-dividing cells. The HDR pathway (green) is active only in the S and G2 phases meaning that only actively cycling cells would be able to use this pathway. This has implications for timing of genome editing. For instance, improved HDR mediated genome editing was observed in cell cycle synchronised cells, when editing was forced to occur during the G2 phase [140].

Table 1.5 | Genome editing approaches and their DNA repair pathways used after blunt double strand DNA break

Repair pathway	Genome editing approach	Main factors involved	Active in cell cycle phase	Activity *	Homologous sequence necessary for repair?	DNA end resection?	ssDNA strand invasion?	Accuracy
cNHEJ	Targeted indel formation, HITI	Ku70-Ku80, Ligase IV	All but M	+++	No, but can use <4bp	No, end protection by Ku70-Ku80	No	+
a-EJ	MMEJ/PITCh	PARP1, Pol θ	G1/early S	+	Yes (2-20bp)	Yes	Yes	--
HDR	HDR mediated repair or integration	RAD51, RAD54, BRCA1, BRCA2	S/G2	-	Yes (sister chromatid or provided >100bp)	Yes	Yes, >100bp	+++
SSA	SST-R	RAD52, (XPF)–ERCC1, FANCA, FANCD2	S/G2	-	Yes (single stranded oligo)	Yes	Yes, >20 bp	+++

* Information derived from proteomic analysis of U2OS cells [141] summarised in [128]. Information for this table was derived from [128, 142].

cNHEJ (classic non-homologous end joining), a-EJ (alternative end joining), HDR (homology directed repair), SSA (single strand annealing),

HITI (homology-independent targeted integration), MMEJ (microhomology mediated end joining), SST-R (single-strand template repair)

1.2.3 Genome editing approaches

Classic NHEJ has been exploited in two major genome editing approaches: knocking out a gene through insertion or deletion (indel) formation and via homology-independent targeted integration (HITI, **Section 1.2.4**). The former is a rather crude way of genome editing that relies solely on the endonuclease to cut, and continue to keep cutting, until the cNHEJ pathway has created an indel sufficiently large that the target site can no longer be recognised. If an exon is targeted, and if the indel size is not a multiple of three, in two out of three occasions this will lead to a frameshift mutation, that potentially leads to a stop codon. As discussed in **Section 1.2.2**, targeting a site with some degree of microhomology can increase the introduction of deletions possibly due to a shift to a-EJ repair, and some gRNA design tools have incorporated this into their design criteria [134]. It could also help explain the observation of large deletions (>5kb in some cases) and translocations after Cas9 cutting [143, 144]. More recently, machine learning algorithms have been used to analyse the indel formation after Cas9 cutting which have shown that some degree of prediction can be made with regards to the outcome of indel formation [145, 146]. Classic NHEJ can also be used to specifically cut out a region of the genome, by cutting twice in the same chromosome. This has been used in an attempt to turn Duchenne's muscular dystrophy (DMD) into the less severe Becker's syndrome [147-149]. This excision approach will be tested in an *in vivo* clinical trial by the genome editing company Editas (NCT03872479) where a deleterious splice mutant in the eyes of patients with Leber's congenital amaurosis type 10 will be attempted to be excised using SaCas9 delivered by rAAV2/5 [150]. The HITI approach relies on cNHEJ ligase activity to insert large dsDNA sequences into the targeted DSB. It does so by simultaneously cutting the genome and a provided dsDNA donor sequence, creating two sets of blunt ended DSBs that can be recognised by Ku70-Ku80 and ligated together by DNA Ligase IV (**Figures 1.5 and 1.6**; for HITI see **Section 1.2.4**).

Genome editing mediated by a-EJ/MMEJ conversely uses relatively small homology arms to guide DSB repair to insert the sequence between the homologous sequences (**Figure 1.4**). The PITCH system uses TALENs or Cas9 to create small (8-9bp) microhomology between the genomic sequence and provided donor to enable integration of large (>1kb) cassettes in cells and frog embryos [137]. In this report, the integration junction showed evidence of low-fidelity repair with deletions reported of >1600bp. The a-EJ can also be used advantageously without a donor sequence when aiming to delete microduplications; when a DSB is introduced close to the duplication junction, it will be repaired by a-EJ resulting in a deletion of one of the duplications [151].

Genome editing through the SSA pathway (**Section 1.2.2**) primarily uses single stranded oligodeoxynucleotides (ssODNs) as the donor DNA. These can be synthesised inexpensively at high concentrations and co-delivered with an endonuclease of choice [152]. Clever design of the size of homology-arms left and right of the cut site can improve genome editing efficacy [153]. The physical tethering of the ssODN to the nuclease also increases the chance that the ssODN is used for repair, rather than the sister chromosome or relying on cNHEJ [154]. Given the limited size of the oligos, and the fact that with larger sized homology the repair mechanism will shift towards HDR, this method is mostly limited to small insertions, deletions or substitutions.

The current genome editing method of choice is probably HDR due to the unrivalled repair accuracy and simplicity of HDR arm design, combined with the versatility of Cas genome editing; this means that development of genome edited cell and mouse lines are within reach for all. In these systems, the fact that HDR is much less active than NHEJ can be overcome by antibiotic selection or by simply screening a large number of clones. Experiments in cells show that HDR activity can be modulated with small molecule inhibitors of NHEJ, such as Scr7 that inhibits DNA Ligase IV [155]. This approach is likely to be too risky to be used *in vivo*, as it is likely that whole body inhibition of NHEJ could ostensibly

lead to an irresponsibly high chance of mutations. *Ex vivo* genome editing of hematopoietic stem cells is very effective when electroporating Cas9 RNPs followed by transduction of an rAAV2/6 vector with homology arms, with efficacies of around 50% [156]. There are currently two HDR-based *in vivo* genome editing clinical trials underway that use ZFNs and an HDR donor delivered by rAAV to the liver in patients with Hunter syndrome and haemophilia B [157]. In these approaches, the wild-type cDNA is attempted to be integrated in the albumin locus which, due to its high expression, should be able to achieve therapeutic levels of the cDNA even when only a small fraction of liver cells is successfully edited.

Due to the relatively long homology arms needed to induce HDR, taking 800bp or more from a packaging restricted, donor delivery vector such as rAAV, can limit the use of HDR for some diseases. Pan-mutation approaches in which cDNAs are integrated, preventing the need to create new genome editing reagents for each single mutation, will be limited to cDNAs of <4kb. This can be overcome in *ex vivo* settings by sequential genome editing, where the second editing event is targeted to the first integrated sequence, concatenating two integration cassettes [158]. How well this would work *in vivo*, given the difficulties in repeat administration using rAAV vectors (**Section 1.1.3.2**), remains to be seen.

1.2.4 Homology-independent targeted integration

Homology-independent targeted integration (HITI) is a genome editing technique that hinges on cNHEJ to integrate a dsDNA sequence into a DSB made by a programmable endonuclease. Probably the first report proving the potential of this method was published by the genome editing company Sangamo in 2010, using ZFNs as the nuclease and a dsDNA oligo provided in trans to be ligated into the DSB [159]. While this was reasonably successful, the group struggled to insert larger sequences into the genome of cells, which they attributed to nuclease digestion of the donor DNA before it could get integrated into the genome. In an attempt to remedy this, the donor DNA was provided in a plasmid containing a ZFN target site [160] (**Figure 1.6A**). This way the plasmid is linearized and

ready for insertion at the same time as the ZFN cuts the genome, preventing premature degradation of the donor DNA. This change made it possible to integrate large (up to 14kb) cassettes into a precise location in the genome. Of note is that the ZFNs used in these reports were programmed to make staggered cuts, thus creating an overhang that was used to induce direction specific integration, analogous to restriction cloning.

These staggered cuts resulted in many deletions around the integration site, with an average of 30bp [160]. A ‘double nicking’ approach with Cas9-nickases, where a dsDNA oligo with overhangs that are complementary to the nick overhangs was provided, resulted in integration but only at low efficiency (one out of 37 clones) [161]. The high mutation rates and low efficacy of HITI using overhangs is most likely due to the fact that any overhangs, whether compatible or not, will be resected, and only the resulting blunt ends will be ligated by Ligase IV in NHEJ repair [128]. The above is in line with experience using Cas12a (formerly known as Cpf1) that induces a staggered cut; this was proposed to be the reason for increased HDR repair (>4x) compared with Cas9, which produces a blunt end DSB [162, 163]. This promotion of HDR is likely due to a suppression of NHEJ, as Ku70-Ku80 will preferentially bind to a blunt DSB and will contribute to the low efficacy of overhang guided HITI. Consequently, HITI using Cas12a has been attempted but was not deemed to be effective (Patrick Harrison, 3rd Annual Genome Editing Congress, London 2016).

Another peculiarity of these early HITI approaches is the fact that after the desired seamless integration in the correct orientation has been achieved, the integration site becomes a target site for the endonuclease, which is possibly still expressed at that time. This likely also contributed to the large mutations around the integration junction, as the ZFNs will probably keep cutting until mutations render the target site unrecognisable (this is the approach taken in knock-out genome editing using cNHEJ, **Figure 1.4**). If two recognisable sites remain on both sides of the integration, it is possible that the integrated sequence will be excised completely. To overcome this, a novel approach relies on the endonuclease

target sites being in reverse complement in the donor DNA relative to the genomic target (ObLiGaRe, **Figure 1.6B**) [164]. The result of this configuration is that after ligation of the insert in the desired direction, the ZFN recognition sites are palindromic repeats, and the *FokI* nuclease domains can no longer dimerise to catalyse the nuclease reaction. A similar methodology was developed using Cas9 as the targeted endonuclease which, probably due to the publication in a high-impact journal, is now generally known as the 'HITI method' (**Figure 1.6C**) [136]. In this report, several novel permutations were investigated including: using a minicircle as donor (which lacks the bacterial backbone and would not be integrated in the genome), cutting on both sides of the donor sequence, releasing the insert completely (again an attempt to reduce insertion of the backbone) and using rAAV as the delivery method of the Cas9 endonuclease and HITI donor DNA.

As to date, there are 11 reports highlighting the use of HITI in a variety of *in vivo* models, from zebrafish to ferrets, using TALEN, ZFN and Cas9 for either the development of animal models or for therapeutic application in disease models (summarised in **Table 1.6**). This highlights the potential of HITI in systems outside of cell culture and potentially in the clinic in the future.

Figure 1.6 | Mechanisms of homology-independent integration. Different methods have been employed to utilise cNHEJ to integrate sequences in endonuclease induced DSBs. In the first **(A)** a single cut is made in both the genome and donor sequence and the donor is integrated whole, i.e. in the case of a plasmid the backbone will be inserted. In case of a correct integration orientation the ZFN target sites align to again form targetable sites. This, in combination with the staggered cuts of early ZFN DSB approaches, which induces end-resections during repair, are probably the reason for the observed mutations around the integration junctions [160]. Improving on this method, ObLiGaRe was developed in which the ZFN target sites in the donor are in the reverse complement orientation with respect to the genome target **(B)** [164]. When the target is integrated in the intended direction the ZFN target site can no longer be cut. Conversely, when the donor is integrated in the wrong direction it can be cut out again and re-ligated in the right direction. An improvement on the ObLiGaRe method was published in 2016 [136] where ZFNs had been replaced with Cas9 and the donor sequence was cut twice around the insert sequence rather than just once, preventing insertion of the backbone sequence **(C)**. N.B. the genome in this figure is visualised as a simplified expression cassette with promoter and mutated target gene but the integration location and insert sequence is of course up to the researcher and driven by the targeted mutation.

Table 1.6 | In vivo HITI gene insertion.

Organism	Nuclease Used	Maximum Integration size	Delivery Method	Efficiency (% cells with targeted integration)	Notes	Reference
Zebrafish	Cas9	5.7kb	Cas9/gRNA mRNA and donor plasmid DNA injection in single cell embryos	31%	ObLiGaRe	[165]
Zebrafish	TALEN	~ 1kb	TALEN mRNA and donor plasmid DNA injection in single cell embryos	30%	ObLiGaRe	[165]
Zebrafish	Cas9	NR	Cas9/gRNA mRNA and donor plasmid DNA injection in single cell embryos	27-54%	Used different gRNA for genome and donor, preventing re-cutting after integration	[166]
Zebrafish	Cas9	NR	Cas9/gRNA mRNA and donor plasmid DNA injection in single cell embryos	NR	Based on ObLiGaRe	[167]
Zebrafish	Cas9	NR	Cas9/gRNA mRNA and donor plasmid DNA injection in single cell embryos	2-20%		[168]
Fungus fly (<i>Ciara Coprophila</i>) embryos	ZFN	6.5kb	ZFN mRNA and donor plasmid DNA injection in male single cell embryos	27%	Based on ObLiGaRe	[169]
Xenopus Tropicalis	Cas9	NR	Cas9/gRNA mRNA and donor plasmid DNA injection in single cell embryos	2.2%		[170]
Mouse	Cas9	NR	In utero (E15.5) electroporation of Cas9 plasmid and donor minicircle	0.1%		[136]

Mouse	Cas9	~740bp	Direct dual rAAV8 injection into visual cortex	3.5%		[136]
Mouse	Cas9	~740bp	Dual rAAV8 or rAAV9 injection into tail vein of p1 mice	4.2% (liver), 3.4% (heart), 10.0% (quadriceps muscle)	Integration also observed in brain, choroid plexus, kidney, adrenal gland, spleen, and lung	[136]
Rat	Cas9	NR	Direct sub-retinal rAAV8 injection	NR, mRNA expression 4.5% of normal		[136]
Bread wheat	TALEN	NR	Particle bombardment of wheat protoplasts with TALEN and donor plasmids	6.5%		[171]
Axolotl (<i>Ambystoma mexicanum</i>)	Cas9	~5kb	Injection of RNP and donor plasmid in single cell stage fertilised eggs	5-15%		[172]
Water Flea (<i>daphnia magna</i>)	Cas9	5kb	Injection of RNP and donor plasmid in eggs	1%	Three donor plasmids were integrated in tandem	[173]
Ferret	Cas9	8kb	Injection of RNP and donor plasmid in fertilised single cell embryos	22%		[174]

NR = Not reported

1.2.5 Off-target effects of genome editing

One of the major shortcomings of genome editing is the potential for the genome editing machinery to induce off-target effects. A distinction must be made, specifically in the context of *in vivo* genome editing, where ‘off-target’ can mean erroneous genome editing in a wrong part of the genome, or in the wrong tissue. The latter is governed by the location of expression of genome editing components (ZFN, TALENs and Cas + gRNA), the control of which is determined by the delivery vector and transgene cassette configuration. This potential off-target effect has been a risk of gene addition since its conception (see **Section 1.1.1**), from which the genome editing field can borrow at will. The former off-target effect, editing in a wrong part of the genome, can potentially have disastrous consequences and is the subject of intense research in this fast-paced field [175]. When genome editing is used in a clinical setting, making genomic alterations in a wrong (and often random) position could potentially disrupt normal gene function, the results of which can be unpredictable. On the other hand, when using genome editing to create new research tools such as cell lines or transgenic mice, avoiding off-target genome edits is imperative for correct interpretation of the model’s characteristics/phenotype caused by the edit, and to prevent confounding future experiments using these tools.

1.2.5.1 Detecting off-target dsDNA cuts

Off-target DSBs using the CRISPR/Cas9 system were first identified in *E.Coli* [176], followed by eukaryotic cells [177]. Since then, multiple, more sensitive methods have been developed to detect off-target DSB formation caused by programmable nucleases. The most widely-used (Guide-seq and Circle-seq), and one recent method (Discover-seq), are shown in **Figure 1.7**. Guide-seq uses a double-stranded oligodeoxynucleotide (dsODN) that is co-transfected into the cells with the endonuclease [178]. When a DSB occurs, the oligo will be ligated into the break (in a manner analogous to HITI), providing a known sequence used for targeted amplification and subsequent deep sequencing, with the genomic sequence directly adjacent to the dsODN oligo sequence corresponding to the site

of the DSB. Circle-seq was developed by the same lab as Guide-seq but is completely cell-free [179, 180]. In short, genomic DNA of a relevant model is fragmented and circularised, then incubated with an RNP which will cut circles that contain the desired site in the genome and possible off targets. The linearized DNA is then available for adapter ligation, amplification and deep sequencing. Finally, Discover-seq is the newest method that promises unbiased off-target detection [181]. It uses chromatin immunoprecipitation to capture a marker of DSBs called MRE11, and the DNA which co-precipitates with MRE11 is ostensibly DNA that was cut and is identified through deep sequencing.

1.2.5.2 Preventing off-target dsDNA cuts

Off-target effects can be addressed in part by titrating the expression of the endonuclease, for instance by using mRNA or RNPs rather than plasmids or integrating viruses, so that less endonuclease is present in the cell for a shorter amount of time [100]. This in turn reduces the chance that the endonuclease cuts in a location other than intended. At the same time, protein engineering approaches have evolved endonucleases such as Cas9 to increase the fidelity with which it cuts. Catalytically inactive Cas9 can be attached to *FokI* nucleases in a manner similar to ZFNs and TALENs (**Figure 1.3**) to double the amount of complementary DNA to which both Cas9 fusions need to bind [182, 183]. Mutations to the DNA binding domains to make off-target binding less energetically favourable, or improving proofreading capacity, significantly reduce off-target effects of Cas9 [184-187]. Designing gRNAs to truncate the gRNA sequence by 2bp (to decrease the stability of off-target RNA-DNA Watson-Crick base pairing) can also have a major impact on off-target effects [97, 188]. Taken together, these improvements have significantly reduced the chances of deleterious off-target effects. Given the many current and planned genome editing clinical trials, off-targets have not held back progress by much. However, care should still be taken to ensure that for a chosen target, delivery method and genome editing approach, the results are as close as possible to that anticipated, before making the jump to humans.

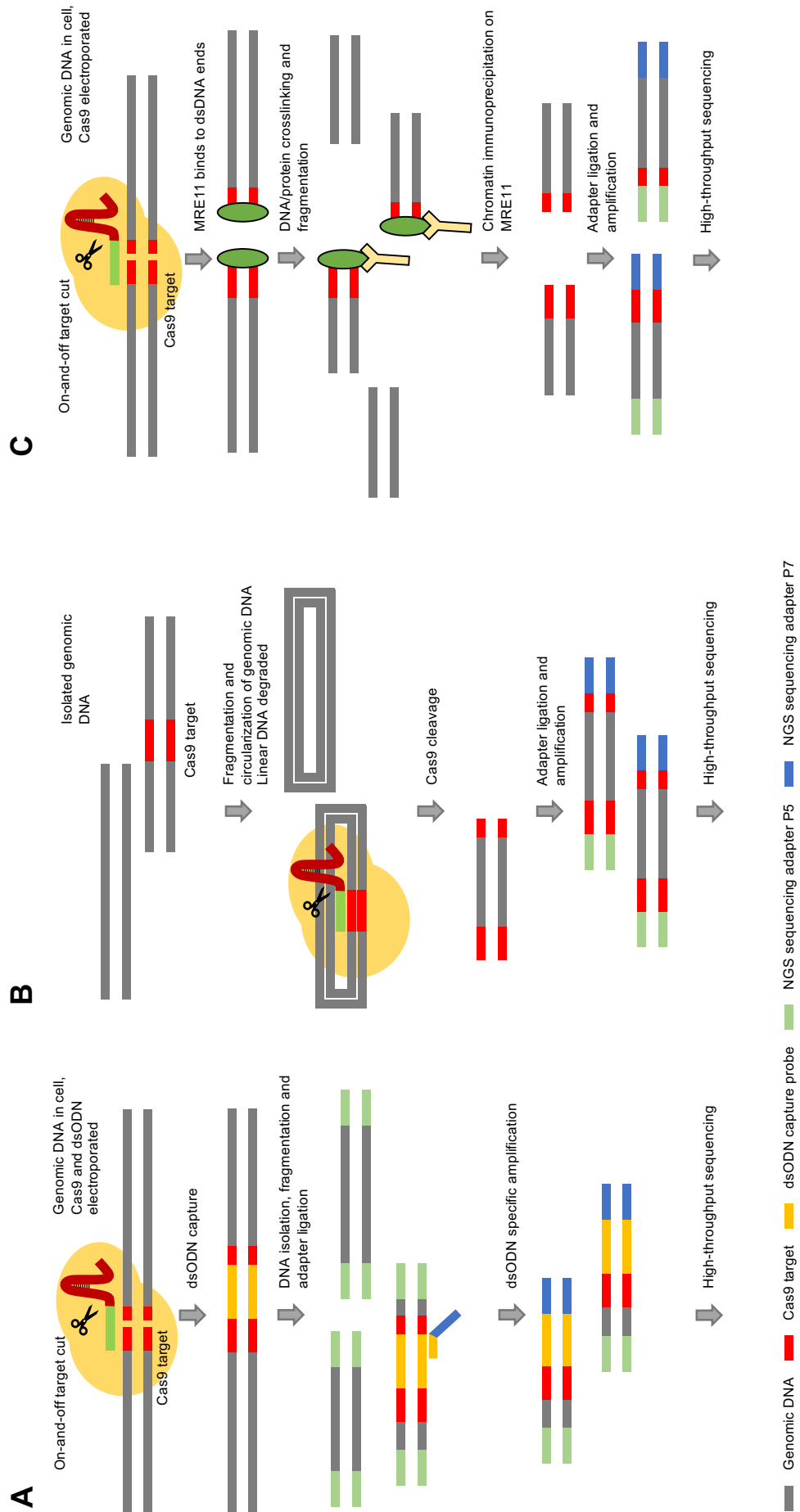


Figure 1.7 | Mechanism of common off-target DSB detection methods. In Guide-seq ([178], **A**) Cas9 with target gRNA and a double stranded oligodeoxynucleotide (dsODN) are introduced to a target cell through a method of choice (e.g. electroporation). After DSB formation by Cas9, whether on- or off-target, the dsODN will be ligated in the DSB through cNHEJ (analogous to HITI). The DNA can then be extracted and the DNA around the inserted dsODN can be selectively amplified and sequenced. The sequence directly adjacent to the dsODN corresponds to the Cas9 target site for that DSB (red). In Circle-seq ([179, 180], **B**) the genome of a target organism is first extracted, fragmented and linearized. Upon incubation of an RNP, circular genomic DNA will be linearized only when the RNP can cut (i.e. at on-target and off-target sites). Only these linearized fragments are amenable to adapter ligation and subsequent amplification and sequencing. The sequences directly adjacent to the adapters correspond to the DSB site. Enrichment of DSB sites in Discover-seq [181] conversely, relies on chromatin immunoprecipitation (**C**). Here, Cas9 is introduced into target cells and DSBs are induced. In a reaction to the DNA breaks, the cell sends MRE11 to the blunt ends of the DSBs. At this point the cells are cross-linked with formaldehyde and the MRE11 proteins, still bound to the DSB ends, are pulled down using chromatin immunoprecipitation methods. This enriches for DNA targeted by Cas9 and can then be amplified and sequenced. Note that while these methods were originally developed to detect off-target Cas9 cuts, Guide-seq and Discover-seq can readily be applied to ZFN or TALEN induced (blunt) DSB cuts. If a researcher is in possession of recombinant ZFN or TALEN proteins, Circle-seq can ostensibly be applied too.

1.3 Alpha-1 antitrypsin deficiency

Alpha-1 antitrypsin deficiency (AATD) is an inherited disease that can lead to predisposition to chronic obstructive pulmonary disease (COPD), with the main symptoms being lung emphysema, chronic bronchitis and bronchiectasis, and is, depending on the underlying mutation, associated with liver fibrosis. It affects between 1 in 3500-5000 people worldwide, with increased incidence in northern Europe (**Figure 1.8**). The gene affected is alpha-1 antitrypsin (AAT) or serpin peptidase inhibitor, clade A, member 1 (SERPINA1). This small (52 kDa) secreted glycoprotein is primarily produced in liver hepatocytes, with a small amount originating from resident alveolar macrophages, neutrophils and airway epithelial cells [189-191]. When secreted into the bloodstream, AAT will travel to the lung, where it inhibits serine proteases in the lung lumen. Null mutations of AAT (see **Table 1.7**), which almost invariably cause emphysema, have led to the protease/anti-protease balance hypothesis, where a disruption in anti-protease activity causes overactive protease activity with destruction of alveolar structure and mucus homeostasis, leading to inflammation and reduced mucociliary clearance [192]. While the name implies that the main function of AAT is the inhibition of trypsin, the main target is actually neutrophil elastase, followed by proteinase 3 and cathepsin G, all of which can induce some of the hallmark features of COPD when their activity is unchecked [193]. Normal levels of AAT vary between 20-53 μM (1-3 g/L), but serum AAT concentrations below 11 μM (~0.6 g/L) are associated with disease. People homozygous for the most common disease causing mutation of AAT, termed PI*Z (homozygotes are indicated by PI*ZZ), have concentrations between 1.2-7.7 μM (0.06-0.41 g/L) [194].

Mutations relating to AATD are summarised in **Table 1.7**, with their associated lung and liver phenotype. The most prevalent mutations in AATD are the S and Z mutations, also termed PI*S and PI*Z (the normal version of AAT is termed PI*M). The relatively common PI*S allele comprises a substitution of valine for glutamic acid at position 264, leading to intracellular protein degradation before it can be secreted [195]. While homozygosity of PI*S

does not lead to AATD, being compound heterozygous for PI*S with PI*Z, or a null variant, increases the risk of emphysema (**Table 1.8**). Considering the relatively high frequency of PI*S and PI*Z, heterozygosity is fairly common, and the geographical distributions of these alleles are displayed in **Figure 1.8**. The PI*Z mutation has a single bp mutation that causes a substitution of glutamic acid to lysine at position 342 which leads to loop-sheet polymerisation between PI*Z AAT molecules [196]. This polymerisation leads to a reduction of AAT secretion to roughly 15% of normal and can lead to ER stress in the hepatocytes [197], most likely the reason for increased incidence in liver cirrhosis in patients homozygous for PI*Z.

Diverse and varying phenotypes (see **Table 1.8**), even in individuals with the same mutation, and symptom similarities with COPD (induced by smoking), has led to under diagnosis of AATD [198]. This has also led to a delay in diagnosis (time between symptoms and formal diagnosis) of 5.6 years in 2003 [199] with many of the diagnoses reached only after consultation with a second or third physician [200]. Interestingly, liver injury in AATD is more frequent in males than females, indicating a sex-dependent phenotypic variation [201]. This is potentially rooted in hormonal differences, as female PI*Z mice, which normally have fewer polymer globules compared to males, increase this toxic liver pathology after testosterone injection [202].

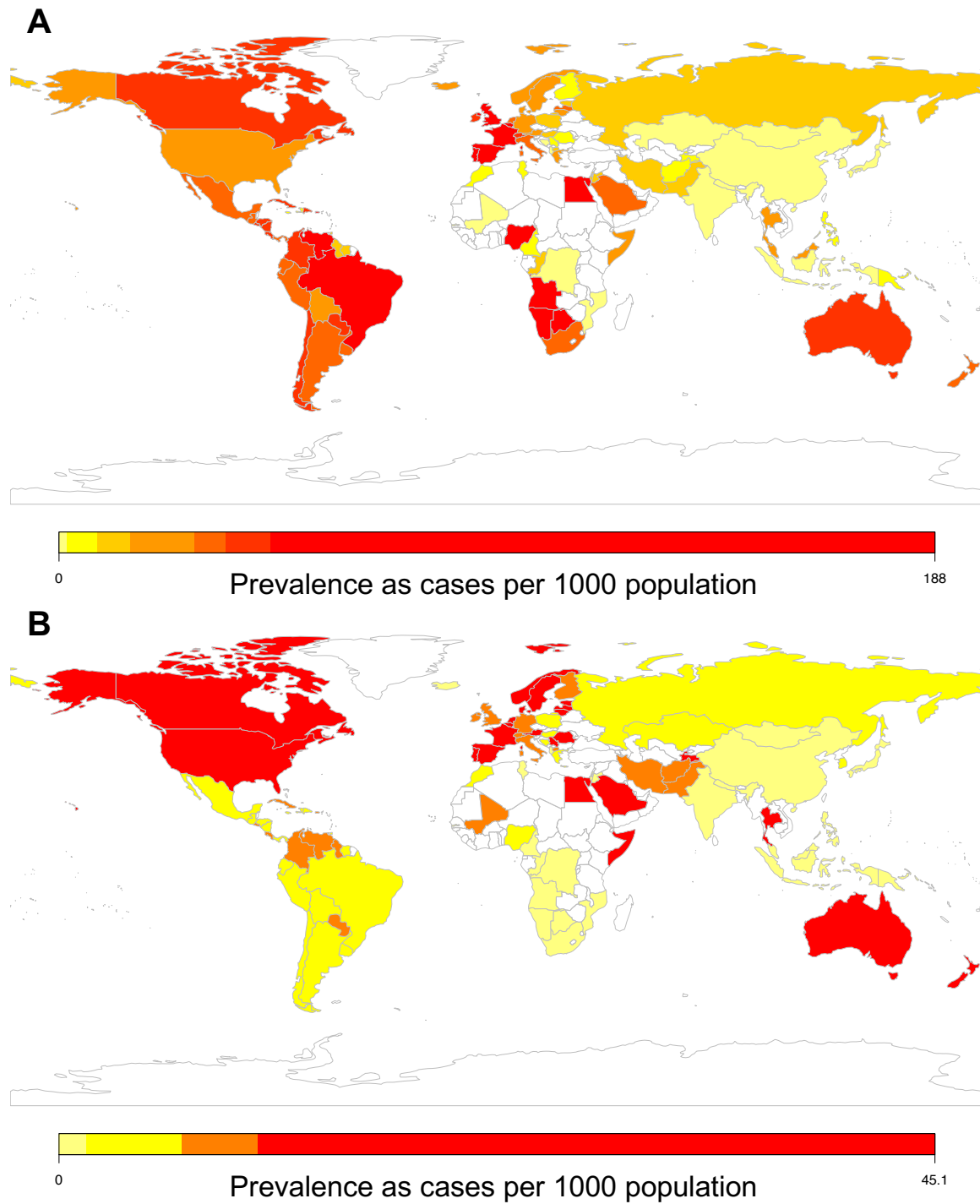


Figure 1.8 | Geographical distribution PI^*S and PI^*Z . Heatmap of PI^*S (A) and PI^*Z (B) allele distribution. Data derived from [203, 204]. Map created with the R package *rworldmap* [205].

Table 1.7 | Selected AAT variants, cellular defects, and disease association

PI allele	Type of mutation	Cellular defect	Disease association
Normal alleles			
PI*M (various subtypes)	1bp substitution polymorphisms	None	Normal
PI*X _{Christchurch}	Substitution (1 bp, lys363glu)	None	Normal
Deficiency alleles			
PI*S	Substitution (1 bp, glu264val)	Intracellular degradation	Lung
PI*Z*	Substitution (1 bp, glu342lys)	Intracellular accumulation	Lung, liver
PI*M _{Malton}	Deletion (3 bp, phe52del)	Intracellular accumulation	Lung, liver
PI*S _{Iiyama}	Substitution (1 bp, ser53phe)	Intracellular accumulation	Lung
PI*M _{Heerlen}	Substitution (1 bp, pro369leu)	Intracellular degradation	Lung
PI*M _{Procida}	Substitution (1 bp, leu41pro)	Intracellular degradation	Lung
PI*M _{Mineral springs} *	Substitution (1 bp, gly67glu)	Intracellular degradation	Lung
Null alleles			
PI*Q _{Q₀Bellingham}	Substitution (1bp, lys217stop)		Lung
PI*Q _{Q₀Bolton}	Deletion (1bp, frame-shift 373stop)	Stop codon; truncated protein	Lung
PI*Q _{Q₀Granite falls}	Deletion (1 bp, tyr160ter)	Stop codon; no mRNA	Lung
PI*Q _{Q₀Nulmattawa}	Insertion (1bp, leu353phe)	Stop codon; truncated protein	Lung
PI*Q _{Q₀Ludwigshafen}	Substitution (1 bp, ile92asn)	No protein	Lung
PI*Q _{Q₀Hongkong-1}	Deletion (2 bp, frame-shift 334stop)	Truncated; Intracellular accumulation	Lung
PI*Q _{Q₀Isola di Procida}	Deletion (17 kbp)	Deletion of coding regions; no mRNA	Lung
Dysfunctional alleles			
PI*M _{Mineral springs} *	Substitution (1 bp, gly67glu)	Defective inhibition of neutrophil elastase	Lung
PI*Z*	Substitution (1 bp, glu342lys)	Defective inhibition of neutrophil elastase	Lung, liver
PI*Pittsburgh**	Substitution (1 bp, met358arg)	Increased anti-thrombin 3 and factor XI activity	Bleeding diathesis

* Note that *M_{Mineral springs}* and *Z* mutations result in dysfunctional characteristics through a lack of functional inhibition of serine proteases but also display deficiency in the serum. ** The Pittsburgh mutation is a mutation at the active site of AAT, causing a change in activity from inhibition of serine proteases to an inhibitor of thrombin 3 and factor XI, leading to bleeding disorders [206]. IC=intracellular, bp=base pair(s). Data derived from [207] and OMIM entry 107400.

Table 1.8 | Most prevalent AAT allele combinations and disease risk

Alleles	Serum level	Increased risk of lung disease	Increased risk of liver disease
M + M	Normal	No	No
M + S	Normal or slightly decreased	No	No
S + S	Normal or slightly decreased	No	No
M + Z	Normal to intermediate	Yes	Yes
M + null	Normal to intermediate	No	No
S + Z	Low	Yes	Yes
Z + Z	Very low (<15%)	Yes	Yes
Z + null	Very low (<15%)	Yes	Yes
Null + null	Absent	Yes	No

Data derived from [207].

The current most effective treatment regimen for AATD is enzyme replacement therapy (ERT), given intravenously, usually at weekly doses of 60 mg/kg. This was recently supported by a large systematic review and was found to slow the progression of emphysema, while also being associated with a slight increase in COPD exacerbations [208]. However, ERT is not readily available in many countries, likely due to a high treatment costs (cited at >\$80,000 per year [209]) and lack of clinical trials showing significant benefit (critical Cochrane reviews emphasise lack of evidence [210, 211]). Airway delivery of AAT has been proposed to alleviate the need for weekly intravenous access, allowing self-delivery, and potentially lowering the required dosage. Evidence of efficacy of this method has so far been scant and the fact that the trials that have been conducted and completed have been reluctant to share data does not bode well (reviewed in [212]). Access to alveoli is impeded in regions of the lung that already have been compromised, lowering efficacy in regions that need it most. Importantly, while ERT can relieve the symptoms of lung disease caused by the lack of (functional) AAT in the lung, it will not be able to address AAT mutations with a liver phenotype such as PI*Z.

Currently, the only curative treatment for liver disease induced by AATD is a liver transplant [213], with most of these occurring in adults [201]. As the donor livers will express a normal version of AAT, both the liver disease and lung disease will be treated. Outcomes of liver transplantation in AATD are generally good in both children and adults with around 90% 5-year patient survival, similar to other diseases [214, 215]. The constant shortage of organ donors, however, makes liver transplant an unlikely first line treatment option for all AATD patients.

Potential future treatments for AATD ideally aim to alleviate both lung and liver symptoms. One approach is to address the pathways involved in misfolding and polymerisation of PI*Z. For instance, chemical chaperones were shown to be effective in increasing the level of functional AAT in a mouse model of PI*Z, although this effect was not replicated in humans

[216, 217]. Disposing of the polymeric deposits by upregulating autophagy pathways with rapamycin can decrease intrahepatic accumulation of PI*Z [218], but so far this has not yet progressed beyond mouse models.

1.3.1 Novel treatments for AATD

Gene therapy has several theoretical advantages over ERT, namely the possibility of a curative single dose, with stable AAT expression rather than peaks and troughs following injection, as well as a potential long-term cost reduction. However, like ERT, gene addition of normal AAT to any organ will not have any influence on the PI*Z liver phenotype. Many different delivery vectors (non-viral, retrovirus, adenovirus and AAV) targeting various organs (liver, lung, muscle and pleural mesothelium) have been investigated in pre-clinical studies (excellently reviewed in [219]). Early *in vivo* gene addition approaches for AATD therapy have been attempted using non-viral vectors (naked DNA and liposome coated DNA) to the liver and lungs of mice and rabbits [220-222]. Retroviruses directed to the liver did not reach therapeutically relevant levels of serum AAT expression [223, 224]. Lung-targeted gene delivery using a lentiviral vector pseudotyped with a lung-tropic pseudotype likewise failed to reach therapeutic serum levels of AAT but did reach therapeutic concentrations of AAT in the epithelial lining fluid of mouse lungs [225]. Recombinant adenovirus (rAd) was evaluated for lung delivery with some success in terms of expression level [226, 227] and subsequent development of helper-dependent rAd (devoid of most viral genes and thus less immunogenic) helped increase longevity of expression [228]. Using the less immunogenic rAAV (which has capacity for the relatively small – 1257bp – hAAT cDNA), successful delivery was reported to the muscle, lung, and liver of mice with high AAT expression as a result [229-231]. Other delivery routes to target these organs have also been investigated, including intrapleural delivery in an attempt to target the pleural mesothelium [232, 233] and limb perfusion to the limb muscle to prevent the need for intramuscular injections [234]. Interestingly, preliminary data show that lung delivery of rAAV2/9 transduces both the lung but also the liver, providing the opportunity for a non-

invasive delivery method that can target both affected organs in AATD (unpublished report by [235]).

A very interesting approach, that overcomes the fact that simple gene addition does not address the PI*Z polymerisation, is interference with PI*Z mRNA pre-translationally. Studies show that siRNAs [236, 237], antisense oligonucleotides [238] and miRNA [239] are capable of significantly reducing both intracellular and serum PI*Z levels. The latter report showed that complementing the anti-PI*Z miRNA with a PI*M AAT expression cassette can reduce PI*Z toxicity and simultaneously increase the level of normal AAT to therapeutically relevant levels [239]. Genome editing for AATD can potentially achieve the same goal, treating both lung and liver phenotype of AATD. A first approach used cNHEJ to knock out the PI*Z mutant in a mouse line using IV delivery of Ad5 expressing SpCas9, which was effective but does not address the lung disease [240]. Correction of the PI*Z mutation was the strategy of two back-to-back reports [241, 242], where the number of modified cells was around 1-2% and serum levels of ~8% of the normal threshold level. These studies are discussed in more detail in **Chapter 5**.

Limited clinical studies for AATD gene therapy have been performed so far. One of the studies followed on from aerosolised liposome delivery to the lungs of rabbits [221], with intranasal delivery to a single nostril, the other functioning as a control [243]. Significant increases in AAT of nasal lavage fluid expression only lasted for 5 days, with concomitant increases in local immunological markers suggesting an anti-AAT immune response to be the culprit of the loss of expression. Other clinical trials have revolved around rAAV as the delivery vector, with rAAV1 and rAAV2 as the serotypes used to deliver AAT to the muscle [244-246]. So far, the maximum achieved stable AAT serum level was ~240nM, far below the therapeutic range [246, 247]. This was achieved through 100 injections injecting a total of 135ml of virus, leading to questions about the scalability of the intramuscular method of viral delivery. A recent trial using intrapleural delivery of rAAVrh10 [248] was discontinued

as the sponsor determined that “the data did not demonstrate the potential to reach M-protein threshold levels of 11 μ M” [249].

Two clinical trials applying siRNA to prevent toxic aggregation of PI*Z protein in the liver have been instigated. The first, built on experience with antisense oligonucleotides [238] showed clear reduction of serum PI*Z in both healthy volunteers and PI*ZZ patients, but was stopped prematurely after adverse findings in an NHP study using a similar delivery vector (NCT02363946) [250]. The second is currently underway (NCT03767829) and uses GalNac-conjugated siRNA, to ensure efficient uptake in the liver by the sialoglycoprotein receptor (ASGPR) receptor.

1.4 Aims

Even with multiple gene therapy treatment modalities approved for clinical use, the question of the longevity of treatment effect remains unanswered. Time will tell whether a gene addition protocol using a single administration of rAAV to deliver nucleic acids to the liver is effective in life-long transgene expression. This study aims to investigate and overcome the limitations of single administration gene addition protocols as described above. First, a novel approach to attempt readministration of rAAV2/8 is investigated, relying on immune tolerisation of AAV8 capsid proteins through liver specific expression via delivery with rAAV or a lentiviral vector. Second, genome editing using the homology-independent targeted-integration (HITI) method was investigated to potentially allow life-long transgene expression controlled by the endogenous promoter following a single rAAV delivery through integration into the cellular genome.

Chapter 2 – Materials and Methods

2.1 Common reagents and solutions

All nucleic acid preparations were dissolved in nuclease free water (Promega, Southampton, UK) and all buffers were made in Milli-Q reverse osmosis ultrapure water (Merck Millipore, Feltham, UK). Solutions or materials were sterilised by autoclaving for a minimum of 20 minutes at 121 °C whereas solutions could also be sterilised by filtration with a 0.22 µm surfactant-free cellulose acetate filter (Thermo Fisher Scientific, Loughborough, UK). Lysogeny broth (LB) was prepared using 1% NaCl (Sigma Aldrich, St Louis, USA), 0.5% granulated yeast (Merck Millipore) and 1% tryptone (Sigma Aldrich) in Milli-Q water. To make LB-Agar an additional 1.5% of Select Agar (Life Technologies, Paisley, UK) was added. Both LB and LB-Agar were sterilised by autoclaving. Where solution concentrations are indicated by percentage, they refer to weight:volume (w/v) ratios for solids at RT unless otherwise specified.

2.2 Cells

HEK293T (ATCC, Manassas, USA) and Hepa1-6 (ATCC) cell lines were maintained in high-glucose DMEM (Life Technologies) containing 10% FBS (Sigma Aldrich), 2 mM L-Glutamine (Life Technologies), 100 U/mL penicillin and 100 µg/mL streptomycin (Life Technologies). Huh-7 cells (gift from Yanchun Peng, WIMM, University of Oxford, UK) were maintained in the same medium as above but in low-glucose DMEM. Cells were cultured at 5% CO₂ at 37 °C and passaged once confluency reached 70-90%. Cells were passaged according to need or 1:10 for HEK293T, 1:15 for Hepa1-6 and 1:5 for Huh-7 cells for maintenance, to accommodate duplication dynamics.

2.3 Liver organoids isolation and maintenance

Human liver tissue was procured from the Oxford Radcliffe Biobank (ORB, biobank reference 18/A106) under the ORB research tissue bank ethics reference 09/H0606/5+5.

Tissue was collected fresh in 25 ml of Advanced DMEM/F-12 (Life Technologies) and kept at 4 °C for a maximum of 48h until processing. Mouse liver tissue was resected from mice and kept in Advanced DMEM/F-12 until processing at 4 °C the same day. Human and mouse liver organoids were isolated according to the protocol of Broutier et al. [251]. Human and mouse liver organoids were resuspended in upright domes of the appropriate extracellular matrix (ECM) using BME2 RGF PathClear (Amsbio, Milton, UK) for human and Corning Matrigel Growth Factor Reduced (GFR) Basement Membrane Matrix (Corning, New York, USA) for mouse organoids. They were maintained in human and mouse specific liver organoid growth medium [251] and kept in a 5% CO₂ incubator at 37 °C. Organoids were passaged at 1:2 to 1:4 when the lumen of the organoids showed significant amounts of cell debris.

2.4 Microscopy

2.4.1 Immunohistochemistry

2.4.1.1 Cells

Cells were grown on autoclaved glass coverslips in 6-well plates. Cells on coverslips were kept in the plate for all steps until mounting. The coverslips were washed 2x with 1X PBS and blocked with 4% donkey serum (Sigma Aldrich) in PBS-T (0.1% Triton-X100 (Sigma Aldrich) in 1X PBS (Life Technologies)) for 30 minutes at room temperature. The cells were then incubated with primary antibody (rabbit anti-human AAT (ab9373, Abcam, Cambridge, UK) at 1:1000 dilution in 1% serum PBS-T for 2h at room temperature and then washed 2x with in PBS-T. Secondary antibody (Alexa Fluor 555 Goat anti-rabbit IgG, Life technologies) was incubated at 1:1000 dilution in 1% serum PBS-T for 1h at room temperature. Cells were washed 2x with in PBS-T and the coverslips were mounted on a SuperFrost Plus microscope slide (Thermo Fischer Scientific) using ProLong Gold Antifade Mountant with DAPI (Thermo Fischer Scientific). Slides were dried in the dark overnight before imaging.

2.4.1.2 Liver cryosections

Liver sections cut at 5 μm (see **Section 2.12.11**) were allowed to warm to room temperature and sections were encircled with a hydrophobic pen. Slides were washed twice for 10 minutes in 1% serum TBS-T (1% donkey serum in 0.4% Triton-X100 in 1X TBS (Alfa Aesar, Ward Hill, USA)). Then the sections were blocked in 5% donkey serum TBS-T for 30 minutes at room temperature in the dark. Following that the sections were incubated with the primary antibody (rabbit anti-human AAT, ab9373, Abcam) at 1:1000 dilution in 1% serum TBS-T for two hours at room temperature and then overnight at 4 °C in a humidified chamber. The following morning the sections were washed twice for 10 minutes in TBS-T without serum and subsequently incubated with the secondary antibody (Alexa Fluor 555 Goat anti-rabbit IgG, Life technologies) at a 1:500 dilution in 1% serum TBS-T for 2h at room temperature in the dark. Sections were washed twice for 10 minutes in TBS-T without serum and a coverslip was mounted using ProLong Diamond Antifade Mountant with DAPI (Thermo Fischer Scientific). Slides were dried in the dark overnight before imaging.

2.4.2 Fluorescence microscopy imaging and cell counting

Fluorescence microscopy images of 6- and 12-well plates were imaged on a Spectramax i3x with MiniMax 300 Imaging Cytometer expansion at a single 4x magnification (Molecular Devices, San Jose, USA). Percentage of GFP positive cells were determined by discrete object analysis using the Softmax Pro 7 software provided with the Spectramax i3x. GFP positive cells were counted using the settings: minimal size: 5nm, maximum size: 30nm, intensity above background: 100. Total cell number was counted by creating a custom (user-defined) analysis setting by drawing the cell shapes and determining background. Fraction GFP was determined by dividing the number of GFP positive cells by the total number of cells. Cells and tissues on microscope slides were visualised on an Evos FL Auto 2 Cell Imaging System (Thermo Fisher Scientific) at the indicated magnification using the Celleste Image Analysis Software provided with the microscope.

2.5 Flow cytometry

Cells were harvested, centrifuged for 5 minutes at 200 xg and washed once in 1X PBS and centrifuged again. Supernatant was removed and the cells were resuspended PBS at 1e6 cells per ml in a FACS tube (Starlab, Milton Keynes, UK). Cells were examined for EGFP/mNeonGreen and/or mCherry fluorescence using a LSRII (BD Biosciences, San Jose, USA) with the parameters in **Table 2.1** for all experiments. Results were analysed with FlowJo X (BD Biosciences).

Table 2.1 | Flow cytometry parameters

	Dichroic mirror – Bandpass filter	Voltage (V)
EGFP/mNeonGreen	505LP – 515/20	300
mCherry	600LP – 610/20	320
Forward scatter (FSC)	-	360
Side scatter (SSC)	-	240

2.6 Plasmid cloning and transfection

2.6.1 Expression plasmids and HITI donor plasmids

Typically, plasmid backbones and inserts were constructed by restriction digest of existing plasmids and gel extraction from an agarose gel based on size (**Section 2.6.7**). In other cases, a Polymerase Chain Reaction (PCR) mixture (see **Section 2.7**) was used to amplify a target sequence whilst simultaneously appending suitable restriction sites on the ends of the amplicon. These were then gel extracted from an agarose gel based on correct size. Amplicons were digested with appropriate restriction enzymes and purified with a Qiaquick Gel Extraction Kit (Qiagen, Hilden, Germany) by mixing the digestion product with 30 µl of the provided GC buffer and 30 µl isopropanol before application to the filter and a wash with 750 µl wash buffer. Approximately 25 ng of plasmid backbone was ligated with the insert at a 1:10 backbone:insert molar ratio in 20 µl total volume with 400 units of T4 DNA ligase (NEB, Ipswich, USA). Ligation products were processed as in **Section 2.6.5**. Expression plasmids and HITI donor plasmids used in this study are summarised in **Table 2.2** and **Table 2.4** respectively.

2.6.2 CRISPR/Cas9 and HITI Donor plasmid gRNA insertion

A CRISPR/Cas9 plasmid such as pX601 or HITI donor plasmid (**Tables 2.3 and 2.4**) were digested with 20 units of *BsaI* (NEB) for 2h at 37 °C. The digestion product was purified using Qiaquick Gel Extraction Kit (Qiagen) by mixing the digestion product with 30 µl of the provided GC buffer and 30 µl isopropanol before application to the filter and a wash with 750 µl wash buffer. Digested plasmid backbone was eluted in 20 µl of nuclease-free water. All gRNA target sequences were designed by picking the top scoring gRNA sites found in both Benchling (<http://www.benchling.com>) and CRISPOR [252] prediction tools. SaCas9 gRNAs were designed to be 21 bp long and adjacent to the NNGRRT PAM sequence. The gRNA sequences were purchased as single strand oligonucleotides (Sigma Aldrich) with 5'-CACC-3' and 5'-AAAC-3' appendages for cloning purposes. Oligonucleotides were resuspended in 20 µl of 1x T4 ligase buffer (NEB) and then denatured and reannealed in a thermocycler by heating to 95 °C for 5 min and then ramping down to 25 °C at 5 °C/min. The reannealed oligonucleotides (4 µl) were added to 25 ng of digested backbone and ligated at RT for 2h using 400 units of T4 ligase (NEB). The ligation product was further processed according to **Section 2.6.5**. To insert a second gRNA target site in a HITI Donor plasmid, the process was repeated but with *BbsI* as the restriction enzyme to target the 3'-gRNA target site. CRISPR/Cas9 targeting plasmids and HITI Donor targeting plasmids used in this study are summarised in **Table 2.3** and **Table 2.5** respectively, with their respective gRNA targets.

2.6.3 Site-directed mutagenesis

To produce the APoE/hAAT_Mut promoter plasmids used in **Chapter 6**, the APoE/hAAT promoter was subjected to site directed mutagenesis using the Q5 Site-Directed Mutagenesis Kit (NEB). Primers used are shown below, with the mutation sequence in lowercase.

APoEhAAT_MUT_Q5_F: 5`- gtaAATCCTCTAAGGTAAATATAAAATTTT
TAAGTGTATAATGTG-3`

APoEhAAT_MUT_Q5_R: 5`-acacCCAGGTCAGTGGTGGTGC-3`

PCR was performed with 10 ng of template (pGM621) with a T_a of 66 °C and extension time of 3 minutes. The KLD enzyme mix reaction (NEB) was executed as per the manufacturer's protocol and ligation products were processed as in **Section 2.6.5**.

2.6.4 Blunt TOPO cloning

PCR products (**Section 2.7**) or digestion reactions with blunt ends were ligated into an acceptor plasmid (pCR-Blunt II-TOPO Vector) using the Zero Blunt TOPO PCR Cloning Kit (Thermo Fischer Scientific) according to the manufacturers protocol. 10 to 50 ng of input product was ligated for 10 minutes at room temperature. The ligation product was further processed according to **Section 2.6.5**.

2.6.5 Transformation and bacterial culture

All plasmid transformations in this study used One Shot Stbl3 *E. coli* chemically competent cells (Thermo Fisher Scientific) unless indicated otherwise. Approximately 5 µl of ligation product was added to 25 µl of competent cells, flicked to mix, and incubated on ice for 30 minutes. Subsequently the cells were heat-shocked in a water bath at 42 °C for 45 seconds and placed back on ice for 5 minutes. *E. coli* S.O.C. medium (125 µl; Thermo Fisher Scientific) was added and the cells incubated at 225 rpm for 1h at 37 °C. Next, dilutions (12.5 and 125 µl) of cells were spread on 10 cm diameter petri dishes containing LB agar with 50 µg/ml of Kanamycin or 100 µg/ml of Ampicillin and incubated at 37 °C overnight. Single colonies were picked with an inoculating loop and deposited in LB medium containing 50 µg/ml of Kanamycin or 100 µg/ml of Ampicillin and placed in a shaking incubator (225 rpm overnight at 37 °C). Plasmid DNA was extracted from cells using the Wizard Plus SV MiniPrep DNA Purification System according to the manufacturer's instructions and

subjected to diagnostic restriction digest and agarose gel size evaluation to confirm correct ligation (**Section 2.6.7**) Cells (1 ml) were deposited in 200 ml (for subsequent HiSpeed Plasmid Maxi kit extraction) or 400 ml (for subsequent EndoFree Plasmid Mega kit extraction) in LB containing appropriate antibiotics and grown overnight as before.

2.6.6 Plasmid extraction and purification

Small-scale preparation of plasmid (Miniprep) from of cultured cells was performed using the Wizard Plus SV MiniPrep DNA Purification System (Promega) according to the manufacturer's protocol and eluted in of nuclease-free water. Plasmids for mammalian cell culture transfection experiments were extracted from 150-200 ml of cells using the HiSpeed Plasmid Maxi kit (Qiagen) following the manufacturer's protocol and elution in 1 ml of nuclease-free water. Plasmids used in rAAV and lentiviral vector production were extracted from 400-500 ml of cells using the EndoFree Plasmid Mega kit (Qiagen) following the manufacturer's protocol and elution in 1 ml of nuclease-free water. Plasmids were assessed for quality and concentration using a NanoDrop Spectrophotometer 2000 (Thermo Fisher Scientific) with the ratio of optical density (OD) at 260 nm (OD_{260}) divided by the OD_{280} indicative of protein contamination and a value of <1.8 resulting in a plasmid preparation being rejected.

2.6.7 Agarose gel electrophoresis

DNA was separated based on size by agarose gel electrophoresis. A gel in 1X Tris-acetate-ethylenediaminetetraacetic acid buffer (Sigma Aldrich) containing 1% w/v agarose (Life Technologies) and 1X SYBR Safe gel stain (Thermo Fischer Scientific) was made by dissolving the agarose by heating and cooling it down in a mould. DNA samples (PCR products (10 μ l per reaction) or 1 μ g of plasmid) were mixed with 1x purple loading dye (NEB) and ran at 12 V/cm for one hour. A DNA ladder was added to one of the wells to allow readout of the approximate sizes of input DNA. Gels were imaged at 302nm using an iBright FL1500 imager (Thermo Fisher Scientific).

2.6.8 DNA fragment gel extraction

Agarose gels of **Section 2.6.7** were visualised by eye using the iBright FL1500 imager's band excision setting. The desired band was excised using a scalpel and DNA was extracted using the Qiaquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

2.6.9 Plasmid transfection of cells

Adherent mammalian cells in culture were grown to 70-80% confluency prior to transfection. Unless otherwise indicated, cells were transfected with Lipofectamine 3000 (Thermo Fischer Scientific). Cells in 6-well plates (approximately 500,000 for HEK293T and Hepa1-6 cells and 400,000 for Huh-7 cells) were transfected per well with a transfection mix containing 1.25 µg per plasmid and 2 µg p3000 and 1.5 µg Lipofectamine 3000 per µg of plasmid DNA, made up in 50 µl of Optimem (Life Technologies), added directly to the cells.

2.6.10 Plasmid transfection of liver organoids

Organoids were harvested by resuspension of extracellular matrix domes containing the organoids in basal organoid medium (Advanced DMEM/F-12 with 1% Pen/Strep, 2mM Glutamax and 10 mM HEPES, all Life Technologies). The ECM was removed by vigorous pipetting and washing with basal organoid medium. After removal of supernatant, the organoids (three wells per 15ml Falcon tube) were incubated with 5 ml of TrypLE (Life Technologies) until ~90% of the organoids were digested into single cells. The cells were washed with 15 ml of basal medium to stop the digestion reaction and centrifuged for 5 minutes at 200 xg. Supernatant was removed and cells were resuspended at 2e6 cells per ml of liver isolation medium [251]. The cell suspension (2500 µl; approximately 500,000 cells) was deposited in a 48-well plate (Greiner) and 50 µl of transfection mix in OptiMem was added. Plasmid/transfection mix for organoids consists of 800 ng of plasmid in 50 µl of Optimem combined with either 1.28 µg P3000 and 1 µg of Lipofectamine for Lipofectamine 3000 transfection (Thermo Fischer Scientific), 4 µg of Lipofectamine 2000 for Lipofectamine

2000 transfection (Thermo Fischer Scientific) or 0.4 µg of PEIpro for PEIpro transfection (Polyplus-transfection, Illkirch, France). The plates were centrifuged at RT for 1h at 600 xg and subsequently incubated in a 5% CO₂ incubator at 37 °C for 5-6h. After incubation the cells were resuspended and centrifuged in a 1.5ml Eppendorf tube at 600 xg for 5 minutes and supernatant was removed. Cell pellets were resuspended in the appropriate ECM and after 10 minutes of hardening covered with the appropriate liver isolation medium [251]. After 3-4 days of incubation at 37 °C in a 5% CO₂ incubator the medium was replaced with growth medium.

2.6.11 Plasmids used in this study

Table 2.2 | *Expression plasmids*

Name	Promoter	Transgene	3' UTR + pA	Backbone	Source	Chapter used
pGM497	CASI	Luciferase	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome *	This study. Original plasmid a gift from Tiong Kit Tan	3
pGM498	CMVie	Luciferase	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	This study	3
pGM513	CASI	EGFP	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	This study	5
pGM514	APoE/haAT	Ovalbumin	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	This study	3
pGM515	APoE/haAT	Luciferase	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	This study	3
pGM516	APoE/haAT	EGFP	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	This study	3
pGM518	APoE/haAT	AAV8 Capsid	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	This study	3
pGM518	APoE/haAT	soGLux	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	This study	3
pGM545	APoE/haAT (no intron)	AAV8 Capsid	mWPRES – miR-142-3pT – SV40 pA	HIV1 Kan lentiviral genome	This study. Original plasmid a gift from lan Pringle.	3
pGM546	CMVie	EGFP	mWPRES – miR-142-3pT – SV40 pA	pZac2.1 AAV2 Amp Genome	Gift from Barney Williams	5
pGM616	APoE/haAT (no intron)	Luciferase	mWPRES – miR-142-3pT – SV40 pA	HIV1 Kan lentiviral genome	This study	3
pX601	CMV	SaCas9	bGH pA – U6 gRNA cassette	pX601 AAV2 Amp Genome	Addgene #61591 and [102]	3
pGM621	APoE/haAT	SaCas9	bGH pA – U6 gRNA cassette	pX601 AAV2 Amp Genome	This study	4,5

pGM689	APoE/hAAT_Mut	SaCas9	bGH pA – U6 gRNA cassette	pX601 AAV2 Amp Genome	This study	4,5
pGM715	APoE/hAAT_Mut	SaCas9	bGH pA	pX601 AAV2 Amp Genome	This study	6

*pZac2.1 was received from the Penn Vector core (Philadelphia, USA). Amp, ampicillin resistance gene; APoE/hAAT, promoter made from apolipoprotein E locus control region and human AAT promoter; CASI, promoter made out of CMV enhancer, chicken β -actin promoter and ubiquitin C intron; CMVie, CMV immediate early promoter; EGFP, enhanced green fluorescent protein; Kan, kanamycin resistance gene; miR-142-3pT, microRNA target for miR-142-3pT; mWPPE, mutant woodchuck hepatitis virus posttranscriptional regulatory element; soGLux, CpG free, codon optimised version of Gaussia luciferase; SV40 pA, simian virus 40 poly A; U6, U6 promoter.

Table 2.3 | CRISPR/Cas9 targeting plasmids

Name	Based on	gRNA target	gRNA target sequence	Source	Chapter used
pGM551	pX601	SERPINA1 Exon 1 – gRNA 2	CACCACCACCTGACCTGGGACA	This study	5,7
pGM559	pX601	AAVS1 Locus – gRNA 4	GTGTCCAGGCAAGAAAGCAA	This study	4
pGM575	pX601	Rosa26 – gRNA 2	GGGTGCAAGCACGTTTCCGAC	This study	4
pGM577	pX601	Rosa26 – gRNA 4	ATTGGGACATGGATTCTCCG	This study	4
pGM711	pGM689	SERPINA1 Exon 1 – gRNA 2	CACCACCACCTGACCTGGGACA	This study	6

Table 2.4 | HITI Donor plasmids – no gRNA targets

Name	5' of gRNA target	Insert	3' of gRNA target	Backbone	Source	Chapter used
pGM562		SA – EGFP - TB		pZac2.1 AAV2 Amp Genome	This study	4
pGM581	SpliceOut Donor	mNeonGreen – TB	SpliceOut acceptor	pZac2.1 AAV2 Amp Genome	This study	4
pGM585		mNeonGreen – TB		pZac2.1 AAV2 Amp Genome	This study	4
pGM586	TB – SpliceOut Donor	mNeonGreen – TB	SpliceOut acceptor	pZac2.1 AAV2 Amp Genome	This study	4
pGM587	TB	mNeonGreen – TB		pZac2.1 AAV2 Amp Genome	This study	4
pGM588	U6 gRNA cassette – SpliceOut Donor	mNeonGreen – TB	SpliceOut acceptor	pZac2.1 AAV2 Amp Genome	This study	4
pGM589	U6 gRNA cassette	mNeonGreen – TB		pZac2.1 AAV2 Amp Genome	This study	4
pGM635		mNeonGreen – TB		pX601 AAV2 Amp Genome	This study	4
pGM636	TB	mNeonGreen – TB		pX601 AAV2 Amp Genome	This study	4
pGM651	TB	SA – mNeonGreen – TB		pX601 AAV2 Amp Genome	This study	4

Amp, ampicillin resistance gene; EGFP, enhanced green fluorescent protein; SA, splice acceptor; TB, transcription blocker; U6, U6 promoter.

Table 2.5 | HITI Donor targeting plasmids – containing gRNA target sites

Name	Based on	5'-gRNA target	5'-gRNA target sequence (PAM)	Insert	3'-gRNA target	3'-gRNA target sequence (PAM)	Source	Chapter used
pGM565	pGM562	SERPINA1 Exon 1 – gRNA 2	CACCACCACTGACCT GGGACA(GTGAAT)	SA – EGFP – TB	SERPINA1 Exon 1 – gRNA 2	CACCACCACTGACC TGGGACA(GTGAAT)	This study	5,7
pGM653	pGM651	SERPINA1 Exon 1 – gRNA 2	CACCACCACTGACCT GGGACA(GTGAAT)	SA – mNeonGreen – TB	SERPINA1 Exon 1 – gRNA 2	CACCACCACTGACC TGGGACA(GTGAAT)	This study	5
pGM656	pGM636	Rosa26 – gRNA 2	GGGTGCAAGCACGT TTCCGAC(TTGAGT)	mNeonGreen – TB	Rosa26 – gRNA 2	GGGTGCAAGCACG TTCCGAC(TTGAGT)	This study	4
pGM658	pGM636	Rosa26 – gRNA 2	GGGTGCAAGCACGT TTCCGAC(TTGAGT)	mNeonGreen – IRES	Rosa26 – gRNA 2	GGGTGCAAGCACG TTCCGAC(TTGAGT)	This study	4
pGM659	pGM636	Rosa26 – gRNA 4	ATTGGGACATGGATT TCTCCG(GTGAAT)	IRES – mNeonGreen – TB	Rosa26 – gRNA 4	ATTGGGACATGGAT TTCTCCG(GTGAAT)	This study	4
pGM684	pGM636	Rosa26 – gRNA 2	GGGTGCAAGCACGT TTCCGAC(TTGAGT)	mNeonGreen – TB	Rosa26 – gRNA 4	ATTGGGACATGGAT TTCTCCG(GTGAAT)	This study	4
pGM702*	pGM651	SERPINA1 Exon 1 – gRNA 2	CACCACCACTGACCT GGGACA(GTGAAT)	SA – mNeonGreen – TB	SERPINA1 Exon 1 – gRNA 2	CACCACCACTGACC TGGGACA(GTGAAT)	This study	6

* plasmids contain a U6 gRNA expression cassette for their respective gRNA targets sites between the 5' TB and the 5' gRNA target site.

EGFP, enhanced green fluorescent protein; IRES, internal ribosomal entry site; SA, splice acceptor; TB, transcription blocker; U6, U6 promoter.

Table 2.6 | Other plasmids

Name	Description	Source	Chapter used
pAd Delta F6	Helper plasmid for AAV production	Penn Vector core (Philadelphia, USA)	2
pRepCap2/x	Rep and Cap protein expression plasmid for AAV production. Expressed AAV2 Rep and AAVx Capsid where x is any serotype	Penn Vector core (Philadelphia, USA)	2
pGM281	Integration competent Gag-Pol expressing plasmid for HIV1 lentivirus production	Adapted from Addgene #12251 by Genemedicine Research Group	2
pRSV-Rev	Rev expressing plasmid for HIV1 lentivirus production	Addgene #12253 and [253]	2
pMD2-G	VSV-G pseudotyping plasmid for HIV1 lentivirus production	Addgene #12259	2
pGM584	Targeting plasmid for HITI Reporter cell line	This study	4
DG001	Targeting plasmid for HITI Reporter mouse line	This study	4

Rep, AAV rep proteins, Cap, AAV capsid proteins, VSV-G, vesicular stomatitis virus G protein; HIV1, human immunodeficiency virus 1; HITI, homology-independent targeted integration;

2.7 Polymerase chain reaction

Polymerase chain reaction (PCR) amplification was performed using the Phusion HotStart II DNA Polymerase kit (Thermo Fisher Scientific). PCR primers were designed using Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) and verified for off-target amplification using Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome). Amplification was performed on an Applied Biosystems 2720 Thermal Cycler (Thermo Fisher Scientific) according to the manufacturer's instructions at T_m calculated using their online tool (<https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html>). For amplification of a genomic target, 50ng of input was used for all reactions. Nested PCR was done by taking 0.1 μ l or 1 ng of the original PCR product as the template for a new PCR reaction using the nested primers. TOPO clone inserts (**Section 2.6.4**) were amplified directly from the colonies on the agar plates by depositing a bacterial colony in 50 μ l of nuclease free water using an inoculating loop. 5 μ l of the bacterial suspension was taken forward for amplification with M13F (CAGGAAACAGCTATGAC) and M13R (CAGGAAACAGCTATGAC) primers. All PCR products were verified using agarose gel electrophoresis.

2.8 Sanger sequencing

PCR amplicons were diluted to 10 ng/ μ l and plasmids were diluted to 100 ng/ μ l. 7 μ l of either was used as Sanger sequencing template. Primers (if not using stock primers) were provided at 3.2 pmol/ μ l. All Sanger sequencing was performed as a service by Source BioScience (Nottingham, UK).

2.9 Digital droplet PCR

Digital droplet PCR (ddPCR) was performed on a QX200 ddPCR system (Bio-rad) according to the manufacturer's recommendation for copy number variation (CNV) assays

on genomic DNA. All reagents and consumables were purchased from Bio-rad. All genomic DNA was digested with 10 units of the appropriate restriction enzyme(s) (all NEB) per μg of DNA. The PCR conditions are shown in **Table 2.7**, which include a 10% overage such that 20 μl of the reaction mix will be used for droplet generation. The reactions were performed as in **Table 2.8**, with any deviations indicated if necessary.

2.10 T7EI mismatch detection assay

Mammalian cells of the appropriate species (Huh-7 for human and Hepa1-6 for murine) were transfected as in **Section 2.6.9** with a Cas9 plasmid containing the appropriate gRNA target sequence and DNA was extracted 48 hours post-transfection using a QIAamp kit (Qiagen) according to the manufacturer's recommendations. Nested pairs of PCR primers were designed around the gRNA target site, and a nested PCR was performed (see **Section 2.7**) where 200 ng of nested PCR product of ~500 bp around the on-target Cas9 cut site was resuspended in 1x NEBuffer 2 (NEB) in 19 μl and denatured and reannealed in a thermocycler by heating to 95 °C for 5 min and then ramping down to 25 °C at 5C/min. A 10% TBE gel with 1.5mm spacing was prepared using the following recipe:

- 1.2 ml of 10x Tris/Borate/EDTA buffer:
 - o 108 g Tris Base (Sigma Aldrich)
 - o 55 g Boric acid (Sigma Aldrich)
 - o 7.44 g Na_2EDTA (Sigma Aldrich)
 - o Made up to 1L distilled water
- 6.8 ml distilled water
- 4 ml of 30% acrylamide (Bio-rad)
- 100 μl of 20% ammonium persulfate (Bio-rad)
- 10 μl Tetramethylethylenediamine (Bio-rad)

The T7EI enzyme (1 µl; NEB) was added to the reannealed PCR product and incubated at 37 °C for 15 minutes after which the reaction was stopped by the addition of 1.5 µl of 0.25M EDTA (Sigma Aldrich). Purple gel loading dye (5 µl, NEB) was added to the reaction products and electrophoresed in the 10% TBE gel at 150V until the pink band reached $\frac{3}{4}$ of the gel. Purple, as opposed to blue dye, was used because purple dye uses ficoll (rather than glycerol), which prevents streaking and ‘smiling’ as it does not stick to the glass plate. The gel was incubated in 1x SYBR green dye (Thermo Fisher Scientific) in 1x TBE buffer for 15 minutes and imaged on an iBright FL1500 imager (Thermo Fisher Scientific). Cleavage percentage was calculated according to Formula 2.1 (see **Figure 2.1** for the corresponding bands in the formula). Only gRNA targets that showed >15% of cleavage were chosen for further experiments. This would correspond with ~50% cleavage as Huh-7 and Hepa1-6 transfection efficiency is around 30% of total cells.

$$\textbf{Formula 2. 1: } Cleavage \% = 100 * \left(1 - \sqrt{1 - \left(\frac{b + c}{a + b + c} \right)} \right)$$

Table 2.7 | General ddPCR conditions for CNV assay

Component	Volume per reaction	Final Concentration
2x ddPCR Supermix for Probes (No dUTP)	11	1x
20x target primers/probe (FAM or HEX)	1.1	1x
20x reference primers/probe (HEX or FAM)	1.1	1x
DNA sample	1.1	22 ng
Water	Up to 22 µl	/
Total Volume	22	/

Table 2.8 | ddPCR cycling conditions for Bio-Rad's C1000 Touch Thermal Cycler.

Use a heated lid set to 105°C and set the sample volume to 40 µl

Cycling Step	Temperature in Celsius	Time	Ramp Rate	Number of cycles
Enzyme Activation	95	10 min	2C/sec	1
Denaturation	94	30 sec		40
Annealing/Extension	60 or primer dependent	1 min		
Extension step (in case Annealing is not extension temp)	60	(1 min)		
Enzyme Deactivation	98	10 min		1
Hold	4	Infinite	1C/sec	1

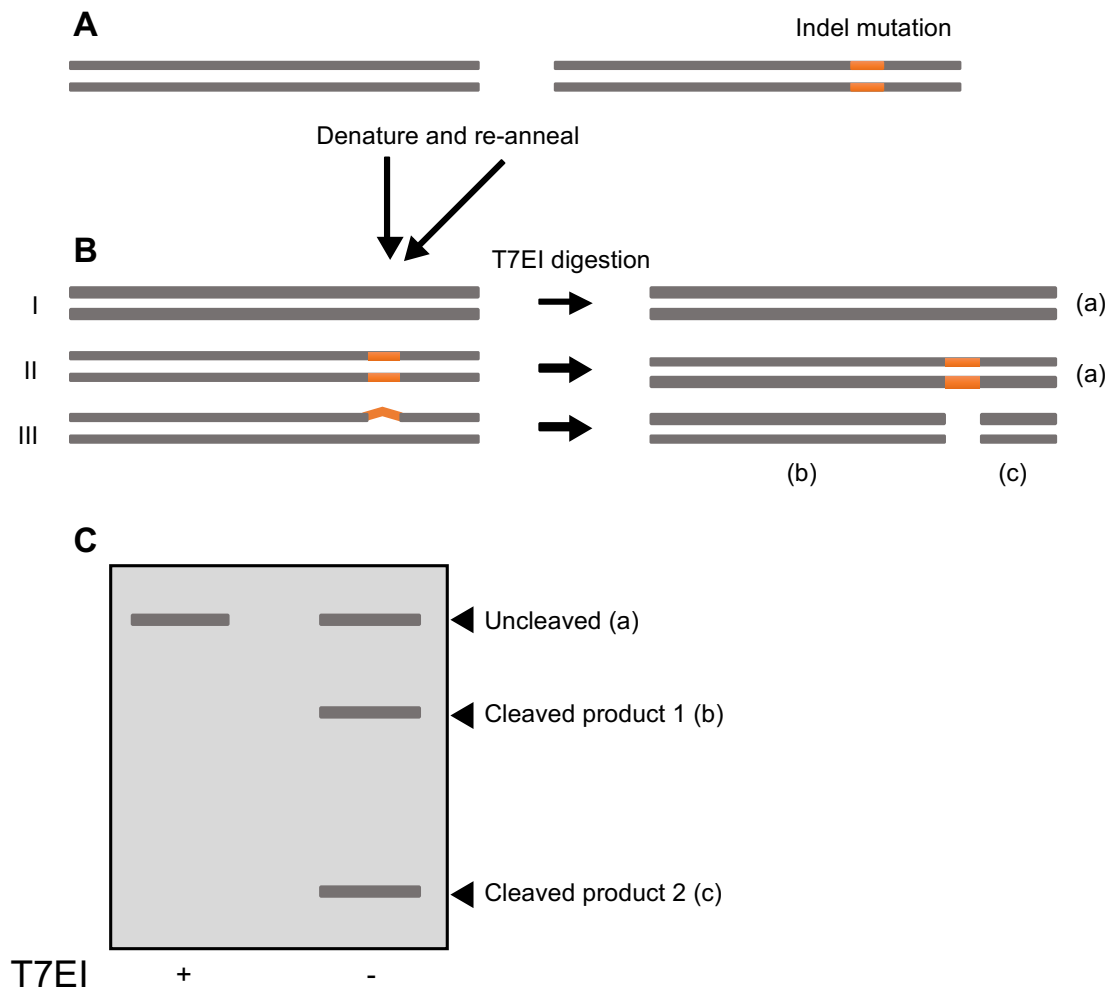


Figure 2.1 | Diagram of the T7EI mismatch detection assay. Input PCR product around the site of suspected indel formation caused by DSB cleavage by Cas9 (**A**, mutation in orange) is denatured and re-annealed by heating and subsequent cooling (**B**). The T7EI enzyme digests the re-annealed products that contain a bulge because of mismatches caused by the indels. The resulting cleavage products are visualised on a TBE gel (**C**) and cleavage percentage can be calculated using **Formula 2.1**, providing an indication of Cas9 cutting efficiency on that site.

2.11 Virus production and transduction

2.11.1 Recombinant AAV virus production

Production of rAAV used the triple-transfection method of Xiao et al. [55]. In brief, HEK293T cells were grown in HYPERFlasks (Corning) in 550 ml of HEK293T media (**Section 2.2**) until 70-80% confluent. To prepare the transfection mix, 10 ml of 150 mM NaCl containing 1.125 ml of 1 mg/ml polyethylenimine (in PBS, pH 7.4) was pipetted drop-wise into 10 ml of 150 mM NaCl containing 500 µg of total plasmid DNA, including genome and helper plasmids (**Table 2.6**) according to the calculation in **Table 2.9**. The 20 ml transfection mix was added to 500 ml of transfection medium (2% FBS, 1% Pen/Strep and 1 mM L-Glutamine in high-Glucose DMEM, all Life Technologies) and carefully added to the cells in the HYPERFlask. After 72h cells were dislodged and transferred, including medium, to a 500 ml centrifuge tube (Corning) and centrifuged for 10 minutes at 1000 xg. Most of the supernatant was removed, leaving approximately 30 ml, and left over medium and cells were transferred to a 50 ml Falcon tube and centrifuged again for 10 minutes at 1000 xg. Supernatant was removed and cells were resuspended in 15 ml of lysis buffer (1 M Tris(hydroxymethyl)aminomethane and 150 mM NaCl in diethylpyrocarbonate (DEPC) treated water, both Sigma Aldrich) and 0.5 ml of protease inhibitor cocktail (1 cOmplete mini EDTA-free protease inhibitor cocktail tablet (Roche) dissolved in 1ml of lysis buffer). Lysed cells were freeze-thawed three times by incubation for 1h at -80 °C followed by 15 minutes at 37 °C. On the final thaw, 50 U/ml of benzonase (Merck Millipore) was added and incubated at 37 °C for 45 minutes. In the meantime, an iodixanol gradient was prepared in 32.5 ml ultracentrifuge tubes (Beckman Coulter) according to **Table 2.10**. The cell lysate was centrifuged for 20 minutes at 3700 xg to pellet cell debris and the supernatant was carefully pipetted onto the top layer of the iodixanol gradient. The tubes were sealed and centrifuged for 1.5h at 20 °C at 59.000 rpm (~360.000 xg) in an Optima XPN-80 Ultracentrifuge with a 70Ti rotor (both Beckman Coulter). After centrifugation the 40% fraction was extracted with an 18-gauge syringe. The virus was concentrated using an Amicon Ultra-15 filter with a 100.000 kDa cut-off, which included 3 washes with 1x PBS,

according to the manufacturer's instructions. Virus was concentrated to 250-500 μ l and stored in 50 μ l aliquots at -80 °C.

2.11.2 Recombinant AAV virus titration

Recombinant AAV vector was titrated for DNase-resistant genome copies per ml (GC/ml) using 1 μ l of rAAV incubated with 2 units of DNase I in 20 μ l of 1x DNase I buffer (both Life Technologies) for 20 minutes at 37 °C to remove unpackaged viral DNA and contaminating genome plasmid. The virus was then diluted 1:10 in nuclease-free water and incubated for 10 minutes at 99.9 °C to denature the capsid and release the viral genome. The free viral DNA was diluted 1:10, 1:100 and 1:1000 in nuclease-free water and 4 μ l of each deposited in triplicate in 6 μ l of the qPCR mix shown (**Table 2.11**). Primers were chosen according to the viral genome, where mWPRES was used for viruses containing the mWPRES sequence and AAV2 ITR primers were used in all other cases (see **Table 2.12**). The qPCR was performed using the following conditions: 95 °C for 120 seconds followed by 40 cycles of 95 °C for 1 second and 60 °C for 20 seconds. The viral titre in GC/mL was calculated against a standard curve of a known concentration containing both the mWPRES and AAV2 ITR sequence.

2.11.3 Lentiviral virus production

Lentiviral vector was produced and titrated by the Virus Production team in the Gene Medicine Research Group (NDCLS, University of Oxford, UK). In brief, the HIV1-based expression plasmids pGM545 and pGM616 (**Table 2.2**) were used to create HIV1-based lentivirus, pseudotyped with VSV-G and using the APoE/hAAT promoter to express AAV8 capsid and Luciferase respectively. The SV40 intron normally present in the APoE/hAAT promoter was removed since lentiviruses have positive strand RNA genomes where introns can negatively impact virus titres. Plasmids used for lentivirus production were pGM281 (gag-pol), pRSV-Rev (rev protein) and pMD2-G (VSV-G pseudotype) (**Table 2.6**). Virus titres were based on transducing units (TU) calculated using a qPCR assay to detect

integrated lentiviral genomes after transduction of HEK293F cells. The qPCR was targeted against the mWPRES region (primers shown in **Table 2.11**) and an endogenous control and quantified using a standard curve.

2.11.4 Viral transduction of organoids

Organoids were transfected with plasmids in a similar manner as the plasmid transfection protocol as in section **2.6.10**, except that the 50 µl of transfection mix is replaced with 50 µl of virus diluted in 1x PBS at the indicated multiplicity of infection (MOI).

Table 2.9 | rAAV production plasmid transfection calculation

Plasmid	Size(kb)	Amount per transfection (µg)
pAd Delta F6	15.4	15.4 x R
pRepCap 2/x	7.3	7.3 x R
AAV2 genome plasmid	T	T x R
Total	K	500 µg
Ratio (R) mass/kb	500/K	

Table 2.10 | Iodixanol gradient preparation for 6 gradients or 2 virus preps

Fraction	Iodixanol	5M NaCl	5X PBS-MK	H ₂ O	Phenol Red
15%	12ml	9.6ml	9.6ml	16.8ml	-
25%	15ml	-	7.2ml	13.8ml	60µl
40%	20.1ml	-	6ml	3.9ml	-
60%	30ml	-	-	-	60µl

PBS-MK: 5XPBS with 5mM magnesium chloride and 12.5 mM potassium chloride in DEPC treated water (all Sigma Aldrich)

Table 2.11 | AAV qPCR titration mix

	mWPRE primers (final concentration)	AAV2 ITR primers (final concentration) [254]
Forward primer	TGGCGTGGTGTGCACTGT (300 nM)	GGAACCCCTAGTGATGGAGTT (100 nM)
Reverse Primer	CCCGGAAAGGAGCTGACA (300 nM)	CGGCCTCAGTGAGCGA (340 nM)
Probe	FAM-TTGCTGACGCAACCCCCACTGG-TAMRA (100 nM)	FAM-CACTCCCTCTCTGCGCGCTCG-TAMRA (100 nM)
Taqman Fast Universal PCR Master Mix	1X	1X

2.11.5 Recombinant viruses used in this study

Table 2.12 | List of recombinant lentivirus and AAV viruses used in this study

Name	Virus type	Serotype/pseudotype	Viral genome	Genome contents	Laboratory reference	Source	Chapter used
Lentiviral vectors							
HIV1 VSV-G ApoE/hAAT AAV8Capsid	rLV	VSV-G	pGM545	APoE/hAAT (no intron) - AAV8 Capsid - mWPRE – miR-142-3pT - SV40 pA	vGM199	Gene Medicine Research Group	3
HIV1 VSV-G ApoE/hAAT Lux	rLV	VSV-G	pGM616	APoE/hAAT (no intron) - Lux - mWPRE - miR-142-3pT - SV40 pA	vGM200	Gene Medicine Research Group	3
rAAV vectors							
rAAV2/8 CASI Lux	rAAV	2/8	pGM497	CASI Lux - mWPRE - miR-142-3pT - SV40 pA	-	This study	3
rAAV2/8 CMVie Lux	rAAV	2/8	pGM498	CMVie - Lux - mWPRE - miR-142-3pT - SV40 pA	-	This study	3
rAAV2/8 APoE/hAAT OVA	rAAV	2/8	pGM514	APoE/hAAT – Ovalbumin - mWPRE - miR-142-3pT - SV40 pA	-	This study	3
rAAV2/8 APoE/hAAT Lux	rAAV	2/8	pGM515	APoE/hAAT – Lux - mWPRE - miR-142-3pT - SV40 pA	-	This study	3
rAAV2/8 APoE/hAAT EGFP	rAAV	2/8	pGM516	APoE/hAAT - EGFP - mWPRE – miR-142-3pT – SV40 pA	-	This study	3
rAAV2/8 APoE/hAAT AAV8 Capsid	rAAV	2/8	pGM518	ApoE/hAAT – AAV8 Capsid - mWPRE - miR-142-3pT - SV40 pA		This study	3
rAAV2/8 APoE/hAAT soGLux	rAAV	2/8	pGM533	APoE/hAAT - soGLux- mWPRE - miR-142-3pT - SV40 pA		This study	3
rAAV2/1 CMVie EGFP	rAAV	2/1	pGM546	CMVie – EGFP - mWPRE - miR-142-3pT - SV40 pA	-	Gift from Omar Habib	5

rAAV2/2 CMVie EGFP	rAAV	2/2	pGM546	CMVie – EGFP - mWPPE - miR-142-3pT - SV40 pA	-	Gift from Omar Habib	5
rAAV2/5 CMVie EGFP	rAAV	2/5	pGM546	CMVie – EGFP - mWPPE - miR-142-3pT - SV40 pA	-	Gift from Omar Habib	5
rAAV2/6.2 CMVie EGFP	rAAV	2/6.2	pGM546	CMVie – EGFP - mWPPE - miR-142-3pT - SV40 pA	-	Gift from Omar Habib	5
rAAV2/8 CMVie EGFP	rAAV	2/8	pGM546	CMVie – EGFP - mWPPE - miR-142-3pT - SV40 pA	-	Gift from Omar Habib	5
rAAV2/9 CMVie EGFP	rAAV	2/9	pGM546	CMVie – EGFP - mWPPE - miR-142-3pT - SV40 pA	-	Gift from Omar Habib	5
rAAV2/8 AAT Exon1 HITI donor mNeonGreen	rAAV	2/8	pGM702	SERPINA 1 gRNA 2 target - SA – mNeonGreen – TB - SERPINA 1 gRNA 2 target	-	This study	6,7
rAAV2/8 ApoE/hAAT SaCas9	rAAV	2/8	pGM715	ApoE/hAAT_MUT - SaCas9 – bGH pA	-	This study	6,7

APoE/hAAT, promoter made from apolipoprotein E locus control region and human AAT promoter; *bGH*, bovine growth hormone poly A; *CAS1*, promoter made out of CMV enhancer, chicken β -actin promoter and ubiquitin C intron; *CMVie*, CMV immediate early promoter; *EGFP*, enhanced green fluorescent protein; *Kan*, kanamycin resistance gene; *miR-142-3pT*, microRNA target for 142-3pT; *mWPPE*, mutant woodchuck hepatitis virus posttranscriptional regulatory element; *rAAV*, recombinant AAV; *rLV*, recombinant lentivirus; *soGLux*, CpG free, codon optimised version of *Gaussia luciferase*; *SV40 pA*, simian virus 40 poly A; *U6*, U6 promoter; *VSV-G* vesicular stomatitis virus G protein.

2.12 Animal studies

2.12.1 Mouse strains

All mice were held in a specific pathogen free environment supplied with normal chow diet and water ad-libitum. Mice used in **Chapter 3** were 6-8-week-old BALB/c mice that were purchased from Envigo (Bicester, UK) and allowed to acclimatise for 1 week before any experimental procedure. In **Chapters 6 and 7**, the PI*Z knock-in mouse line was used (exact nomenclature: Tg(SERPINA1*E342K)#Slcw). This was acquired through a collaboration with Professor Christian Mueller from University of Massachusetts Medical School (Worcester, USA). Founder mice (three of each gender) were received and bred in pairs by local Biomedical Services (BMS) staff at the John Radcliffe Hospital. All animal procedures, experiments and breeding were performed under PPL 30/3259.

2.12.2 PI*Z knock-in genotyping

The PI*Z knock-in mice have the entire human locus of the PI*Z mutant of SERPINA1 inserted into their genome [255]. The insertion has multiple copies of the gene in a discrete (but random) region of the genome, most likely on chromosome 2 [240]. The precise genomic make-up of this mouse line is discussed in more detail in **Chapter 6**. Genotyping relied on detecting the presence and number of copies of the PI*Z knock-in sequence by ddPCR (**Section 2.9**). DNA from mouse ear clips was extracted as in using a QIAamp Blood and Tissue kit (Qiagen) following the manufacturer's instructions, after which 500 ng of DNA in 25 µl was digested with 5 units of *BsaWI* and *ScaI* (both NEB) for 1h at 37 °C and 1h at 65 °C. The ddPCR reaction was set up (**Table 2.7**) with 1.1 µl (22ng) of the digestion product, using the primers in **Table 2.13**, and 20 µl of the reaction used for the PCR at T_a of 59 °C for 1 minute and extension at 60 °C for 1 minute for 40 cycles in total. The ddPCR was assayed using Bio-rad's CNV=2 protocol, which assumes that the copy number of the endogenous control (murine *Rpp30*, NCBI Gene ID: 54364) is two and calculates the number of copies of SERPINA1 PI*Z.

Table 2.13 | PI*Z KI ddPCR primers. All at 3.6:1 ratio of primers:probe

Name	Sequence
<i>Rpp30</i> mouse Forward	CCAGCTCCGTTTGTGATAGT
<i>Rpp30</i> mouse Reverse	CAAGGCAGAGATGCCCATAA
<i>Rpp30</i> mouse Probe	HEX-CTGTGCACACATGCATTTGAGAGGT-Iowa Black FW
SERPINA1 human Forward	GTCCTGCAAGACAGAGATGG
SERPINA1 human Reverse	GTCAATCCCTGATCACTGGG
SERPINA1 human Probe	FAM-CCCGAGTCAAGGCACATGATGACT- Iowa Black FW

Table 2.14 | Peptides sequences used in ELISPOT assay

Target protein	Immuno-dominant peptide sequence for BALB/c haplotype (H2kd)	Sequence obtained from
AAV8 Capsid	IPQYGYLTL	[256]
Luciferase	GFQSMYTFV	[257]
EGFP	HYLSTQSAL	[258]

2.12.3 In vivo virus delivery methods

2.12.3.1 Intramuscular injection

The hind limb of a mouse was shaved with Veet (Reckitt Benckiser, Slough, UK) 1 day prior to injection. The limb was numbed with EMLA cream (AstraZeneca, Cambridge, UK) 15 minutes before injection. Mice were subsequently anaesthetised with 4% isoflurane (Piramal Critical Care, West Drayton, UK) and injected with 40 µl of virus diluted in PBS into the right gastrocnemius muscle.

2.12.3.2 Intravenous injection

The tails of mice were numbed with EMLA cream 30 minutes before injection and mice placed in a hot-box (42 °C) for 20 minutes. Mice were anaesthetised with 4% isoflurane and injected in the lateral tail vein with 40 or 100 µl of virus diluted in PBS.

2.12.4 Blood sampling

The tails of mice were numbed with EMLA cream 30 minutes before blood sampling and mice were placed in a hot-box (42 °C) for 20 minutes. Mice were immobilised and a scalpel was used to create a small nick in the lateral tail vein. Blood was collected in a CB300 microvette capillary tube lined with clotting activator (Sarstedt, Leicester, UK) either directly using capillary force or using a pipette. Blood was stored at 4 °C overnight and centrifuged at 10.000 xg for 5 minutes. The supernatant containing serum was collected in an Eppendorf tube and kept at -80 °C for further use.

2.12.5 In-vivo bioluminescence imaging

Approximately 10 minutes prior to imaging, mice were injected intraperitoneally with 150 µl of 15 mg/ml D-luciferin (Perkin Elmer, Waltham, United States) diluted in PBS, into both sides of the mouse (300 µl total). The mice were then anaesthetised with 4% isoflurane and maintained at 2% when placed in an *in vivo* bioluminescent imaging system (IVIS Lumina LT, Perkin Elmer, Seer Green, UK). Five mice were simultaneously imaged using the 'auto'

protocol using Living Image software version 4.5. Region-of-interest (ROI) analysis was performed on captured images using the same software.

2.12.6 Organ harvesting

Mice were culled using cervical dislocation and death was verified by cessation of circulation. Organs were harvested immediately afterwards. Blood was pipetted into CB300 microvettes and processed as in **Section 2.12.4**. Organs destined for DNA were cut up using a scalpel, deposited into a 1.5 ml Eppendorf tube and frozen at -80 °C until extraction using the QIAamp DNA Blood and Tissue kit (Qiagen) according to the manufacturer's instructions unless otherwise specified. For RNA extraction, tissues were cut up and placed in RNALater buffer (Ambion, Austin, USA) in a Lysing Matrix D tube (MP Biomedicals, Santa Ana, USA) before mechanical homogenisation using the FastPrep (MP Biomedicals) six times at 6.5 m/s² for 45 seconds. RNA was then extracted using the RNeasy mini kit (Qiagen) according to the manufacturer's instruction and stored at -80 °C. Organs for immunohistochemistry (**Section 2.4.1**) were excised whole and placed in 4% paraformaldehyde in PBS (Sigma Aldrich) at 4 °C overnight before further processing (**Section 2.12.11**). Spleens for ELISPOT (**Section 2.12.9**) were excised whole and placed in ice-cold complete medium (RPMI 1640, 10% FBS, 1% pen/strep and 25mM HEPES, 2mM L-Glutamine, all Life Technologies) before immediate processing (see **Section 2.12.9**).

2.12.7 In-vivo viral copy number assay (qPCR)

QIAamp extracted total DNA from mouse tissues (**Section 2.12.6**) was subjected to standard TaqMan qPCR assay on the mWPRES region of the rAAV genome. Primers are the same as in **Table 2.11**. The qPCR was performed using the following conditions: 95 °C for 120 seconds followed by 40 cycles of 95 °C for 1 second and 60 °C for 20 seconds. The number of viral copies per cell was calculated by normalisation according to DNA input

content (as measured by NanoDrop) and multiplied by 2.82404×10^{-6} to retrieve the number of viral copies per murine genome/cell.

2.12.8 Anti-AAV ELISA

Anti-AAV8 antibodies in the serum of mice were detected using a protocol adapted from Martino et al. [259] (a schematic of the mechanism of this assay is shown in **Figure 2.2**). Wells of a 96-well plate (Greiner) were coated with 5×10^9 GC of rAAV2/8 in coating buffer (Cambridge Bioscience, Cambridge, UK) for 2h at 37 °C in a shaking incubator at 100 rpm. The same virus batch (rAAV2/8 CMVie EGFP) was used throughout the study. Wells were washed three times with 100 μ l of 0.05% triton in PBS (PBS-T). Wells were then blocked with 1% BSA in PBS-T overnight at 4 °C. Wells were washed three times with 100 μ l of PBS-T. Serum was diluted in semi-log serial dilution starting at 1:10 (unless otherwise indicated) in 1% BSA in PBS-T and 50 μ l of the dilutions were added to the well and incubated for 2h at 37 °C in a shaking incubator at 100 rpm. Wells were washed three times with 100 μ l of PBS-T. Detection antibody (goat anti-mouse horseradish peroxidase (HRP), ab97265, Abcam) was added at 1:5000 in 1% BSA in PBST and incubated for 2h at 37 °C in a shaking incubator at 100 rpm. Wells were washed five times with 100 μ l of PBS-T. 100 μ l of 3, 3', 5, 5'-tetramethylbenzidine (TMB) buffer (Cambridge Bioscience) was added to activate the conjugated HRP and after 10 minutes at RT the reaction was stopped by the addition of 100 μ l of Stop Solution (Cambridge Bioscience). Absorption was measured at 410 nm using a Spectramax i3X (Molecular Devices).

2.12.9 Interferon- γ ELISPOT

An ELISPOT protocol was developed with modifications from [259, 260] using reagents from the Mabtech 3221-4HPW-2 ELISPOT kit. Spleens in complete medium (RPMI 1640, 10% FBS, 1% pen/strep and 25mM HEPES, 2mM L-Glutamine, all Life Technologies) were pulverised with the back of a syringe and connective tissue was removed. Cells were passed through a cell strainer (70 μ m, Falcon) which was rinsed with 40 ml of complete

medium. Cells were centrifuged (4 °C for 10 minutes at 350 xg), the supernatant removed, and then resuspended in 5 ml of red blood cell lysis buffer (Qiagen) for 5 minutes before washing with 45 ml of complete medium and centrifugation repeated as before. Cells were resuspended in complete medium and 500.000, 250.000 and 125.000 cells were put into an interferon- γ ELISPOT plate (Mabtech) and 50 μ l of each stimulating peptide (Sigma Aldrich, **Table 2.14**) at 2 μ g/ml or 1X cell activation cocktail (500X contains Monensin sodium salt (1.5 mM), Phorbol 12-myristate 13-acetate (0.0405 mM) and Ionomycin calcium salt (0.67 mM) in ethanol; R&D Systems, Minneapolis, USA) was added. Cells were incubated at 37 °C in a 5% CO₂ incubator. After 48h the cells were removed, and wells washed 5x with 200 μ l PBS. Detection antibody was diluted 1:100 in 0.5% FBS in PBS and 100 μ l was added to each well and incubated for 2h at RT. Wells were washed five times with 200 μ l PBS. Streptavidin-HRP conjugate was diluted at 1:1000 in 0.5% PBS and 100 μ l added to each well and incubated for 1h at RT. Wells were washed five times with 200 μ l PBS and 100 μ l of TMB buffer was added to each well and allowed to incubate for 30 minutes at RT in the dark. Wells were rinsed thoroughly from top and bottom and were allowed to air dry. Spots were counted on an ELISPOT reader (AID, Strassberg, Germany).

2.12.10 Gaussia luciferase assay

The activity of Gaussia luciferase (Glux) was measured in the serum of mice (Processed as in **Section 2.13.5**) with the BioLux Gaussia Luciferase Assay Kit (NEB) following the manufacturers protocol. 50 μ l of GLux assay solution was added to 10 μ l of serum and luminescence was measured on a GloMax 20/20 luminometer (Promega). Measurements were normalised to relative light units (RLU) per μ l of serum.

2.12.11 Cryosectioning of mouse liver

Mouse livers in 4% paraformaldehyde was washed twice with 1X PBS and placed in 30% sucrose (Sigma Aldrich) in 1X PBS overnight at 4 °C. Then the livers were gently dried with a paper towel, placed in a plastic embedding mould (Sigma Aldrich) and covered in optimal

cutting temperature embedding medium (Thermo Fischer Scientific). The moulds were snap frozen at -80 °C and kept at -80 °C until sectioning. The livers were sectioned at 5 µm using a Thermo Fischer Scientific CryoStar NX50 with the chuck at -15 °C. Sections were deposited on Superfrost Plus slides (thermo Fischer Scientific), air dried in a dark place overnight and kept at -80 °C until further processing.

2.13 Statistical analysis

All statistical analysis was performed using GraphPad Prism version 7 for Mac (Graphpad Software, San Diego, USA). All EC₅₀ values were determined using three-parameter agonist versus response non-linear regression. All performed tests were non-parametric. For comparison between two groups the two-tailed Mann-Whitney U test was performed. For comparison of multiple groups, a Kruskal-Wallis test with Dunn's multiple comparison test was applied. Values of $p < 0.05$ were considered significant.

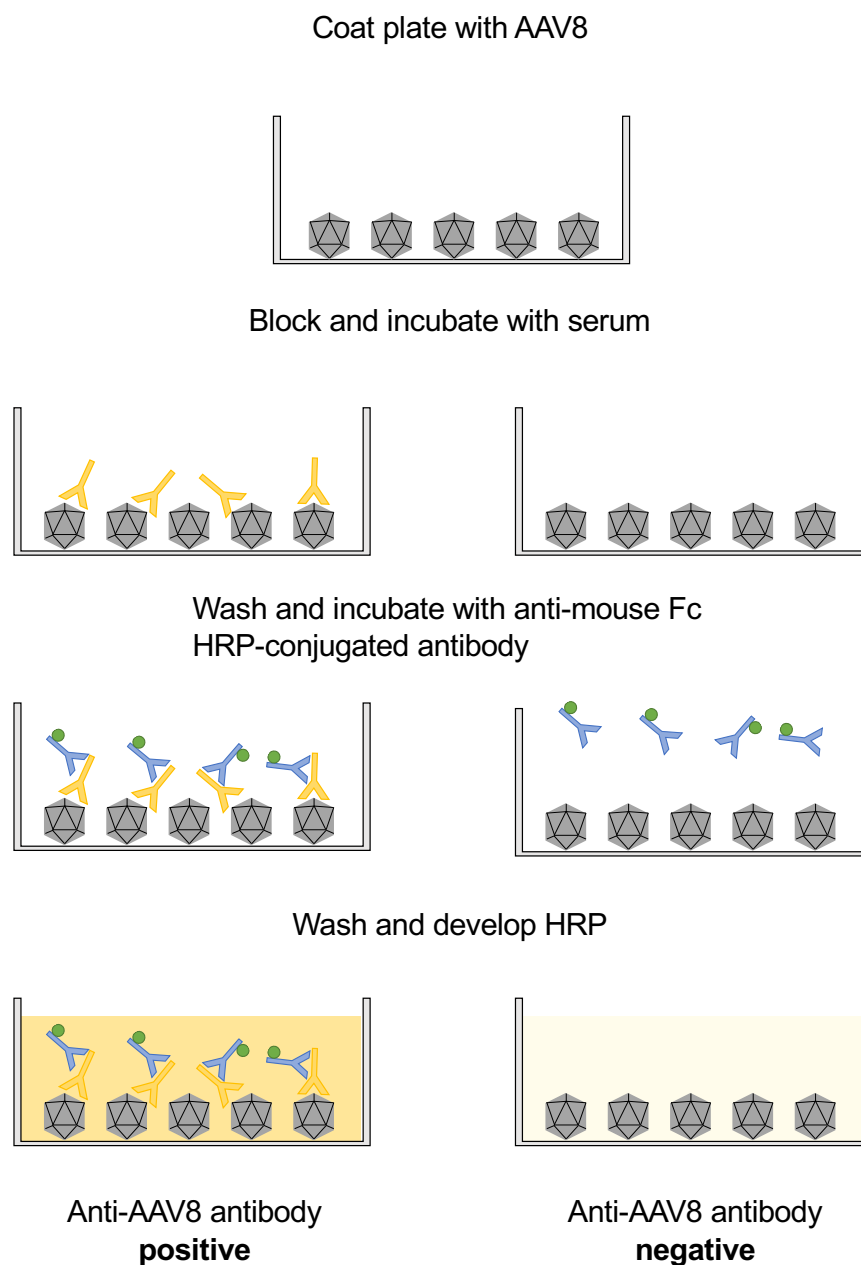


Figure 2.2 | Anti-AAV8 ELISA mechanism. Schematic of protocol in **Section 2.12.8**. ELISA plates were coated with 5×10^9 GC of AAV8 virus and blocked before incubation with serum that was possibly positive for anti-AAV8 antibodies. After washing, a pan anti-mouse Fc antibody conjugated with HRP was added and incubated. The wells were washed and HRP was developed and absorbance measured on a Spectramax i3x machine. Higher absorbance measurements indicated more HRP present and thus more anti-AAV8 antibodies in the serum that bind to AAV8 capsid.

Chapter 3 – Repeat rAAV2/8 administration through immune tolerance

3.1 Introduction

3.1.1 Anti-AAV humoral immunity and the repeat administration conundrum

Gene addition approaches mediated by rAAV have been very successful in clinical trials, with delivery of therapeutic transgenes for inherited genetic diseases to the liver [59, 261, 262], eye [263] and central nervous system [264]; and long-lasting expression at therapeutically relevant levels has spurred on regulatory approval for some of these drugs. While encouraging, it remains to be seen for how long expression will persist in a tissue such as the liver that, contrary to neurons or retinal epithelial cells, is mitotically active, albeit in very low levels (~0.01-1% of liver cells are reported to be proliferating at any given time [265, 266]). Should expression following gene addition strategies in adults prove to be life-long, the same is not expected for diseases where treatment is required to be administered during the early stages of life in order to prevent irreversible damage to the affected tissue [267]. As seen by experiments in neonatal mice, early treatment has the potential for the episomal AAV genome to be ‘diluted out’ during rounds of cell division, with concomitant loss of expression [268, 269]. If findings are similar in patients, repeat dosing of rAAV would be required. In addition, repeat AAV dosing would also allow more straightforward titration of vector doses, which is crucial to mitigate for toxicity seen in some high-dose experiments [77], and could provide opportunities to increase gene expression into a therapeutic range. Currently, many AAV gene therapy clinical trials, especially in Phase I, feature a dose (de)escalation protocol as a safety measure. When patients in the lowest dose cohort do not approach therapeutic levels of transgene expression [56, 270] they will subsequently be precluded from follow-up doses because of the humoral anti-AAV response, even if the gene therapy finds its way into clinical use. Furthermore, pre-existing antibodies are present in a large subset of people due to exposure to wild-type AAV (**Table 1.2**), which means that any administration of rAAV effectively constitutes a repeat administration in these people.

3.1.2 Current approaches to achieve repeat administration

Several different approaches have been proposed to circumvent the humoral anti-AAV response to allow for repeat dosing, all of which have their drawbacks (summarised in **Table 3.1**). These include the modification of the AAV surface by chemical modification [271-273], the incorporation of mutations into the capsid [274-276] or serotype switching between doses [277-280] with the aim to prevent the immune system from recognising the capsid and neutralising the virus. Other approaches address the immune system directly for instance by administration to neonates that have not yet developed a fully mature immune system [281], or by attempting to avoid the immune system spatially, by injecting the virus into an immune-privileged organs such as the eye [282]. Another way to tackle the immune system head on is by removing the anti-AAV antibodies altogether by plasmapheresis [283], the use of capsid decoys [284], or by immunosuppressing the immune system just before or at the time of virus administration [65, 85, 285-292].

Arguably the most successful approach so far is transient immunosuppression using rapamycin-containing nanoparticles [292], where nanoparticles are loaded with rapamycin (an immune suppressor that prevents B- and T-cell activation) and are co-administered with the AAV8 therapeutic vector. Consequently, anti-AAV8 humoral and cellular responses are abrogated in an antigen selective manner, i.e. only an immune response to the AAV8 capsid proteins was prevented. As a result, a successful repeat dose could be administered (again co-administered with the nanoparticles) as evidenced by expression of the transgene of the second dose [292]. Further investigation should elucidate whether or not the immunosuppression is long-lived and how many times a dose can be administered successfully. It is possible that anti-AAV antibody levels increase after a certain amount of time following injection, for which there is some evidence (Francesco Puzzo, personal communication).

3.1.3 Immune tolerance as a repeat administration strategy

One of the most prized features of rAAV as a gene delivery tool is its low immunogenicity in animal models. This is most likely a large contributing factor to the observation made by Mount and colleagues when they found factor IX (FIX) expression to be longer than expected in FIX-null dogs after AAV delivery of the transgene to the liver [293]. Normally when a protein is expressed in an organism that, due to a mutation, does not make that protein, a potent cellular and humoral immune response is launched targeted to that protein (then termed a non-self transgene protein), akin to a response to a foreign protein such as a viral capsid protein. Surprisingly, when FIX was expressed exclusively in the liver of dogs for which FIX is a non-self transgene, no anti-FIX neutralising antibody response was detected [293]. The tolerogenic nature of the liver is well known in the liver transplant field. For example, it has been observed that co-transplantation of part of the liver with another organ can prevent rejection of the second organ [294]. While a definitive mechanism for transgene tolerance after liver expression is not clear, a schematic of the current thinking is shown in **Figure 3.1**. Liver-mediated immune tolerance to transgenes is now actively used to prevent immune responses to therapeutic proteins. For instance, AAV mediated transgene tolerance has been investigated to prevent antibody formation in enzyme replacement therapy (ERT) by expressing the transgene at low levels in the liver before ERT [295]. Another interesting approach is co-administering two rAAV constructs, one of which targets the liver and induces tolerance, and the other directed at the therapeutic target tissue where expression would normally cause a therapy-threatening immune response [296]. Transgene tolerance has also been reported after muscle delivery, but this is more likely to be attributable to a lack of immune activation by AAV rather than active immune tolerance [297, 298]. Recombinant AAV has been the delivery vector of choice for inducing tolerance due to its low immunogenicity and the liver tropism of certain serotypes. Lentiviruses have also been used to induce tolerance even when inhibitory antibodies to the transgene were already present [9, 299-301].

In the light of these tolerising successes, the hypothesis emerged that if it were possible to induce active immune tolerance to AAV capsid proteins it could be possible to prevent anti-AAV antibody responses that inhibit successful re-administration of rAAV gene therapy vectors and potentially the cellular response that can diminish therapeutic transgene expression if not managed with immunosuppression [56].

3.1.4 Aims

This chapter describes studies to establish a gene delivery platform capable of selectively expressing transgenes in the liver, with the aim of inducing immunological tolerance to AAV8 capsid proteins, in order to allow for effective repeat administration of the gene therapy vector.

Table 3.1 | Strategies to circumvent anti-AAV antibody responses to aid vector repeat administration.

	Advantages	Disadvantages	Reference
Prevent detection of virus by coating (PEGylation, HPMA)	• Easy and cheap process • Proven to be effective for other viruses	• Loss of infectivity and tropism	[271-273]
Use different virus serotypes	• Many serotypes available	• Limited number of serotypes with similar tropism • Cross-reactive antibodies	[277-280]
Mutate virus capsid	• Capsid structure and nAb epitopes mostly known	• Not very effective in immune evasion	[274-276]
(Transient) Immunosuppression	• Inhibition of (part of) the immune system prevents or depletes anti-AAV antibodies • Some immunosuppressing drugs already approved for use in humans	• Stops working after cessation of suppression • Immunosuppression increases chance of opportunistic infection • Suppression only at the time of injection (rapamycin nanoparticles)	[65, 85, 285-292]
Exploit immature immune system	• Early injection (to a neonate) prevents immune response	• Only allows for two doses (one as a neonate and one as an adult) • First dose will dilute out when body/organ grows	[281]
Extend time between administrations	• Same serotype can be used	• Time of second injection could be long past time of need	[278, 302]

Capsid decoys	Sequesters any pre-existing nAbs	- Requires injection of high dose of particles of which only a fraction is effective as gene therapy. - Does not prevent immune reaction	[284]
Target immune-privileged organ	No alternate strategy needed	- Limits rAAV gene therapy to certain organs only - T-cell response to capsid still detected [303]	[282]
Plasmapheresis	Technique routinely in use clinically for several disorders	Does not remove 100% of antibodies and low antibody titre can still be neutralizing	[283, 304]

HPMA, poly-[N-(2-hydroxypropyl) methacrylamide]; nAbs, neutralising antibodies; rAAV, recombinant adeno-associated virus

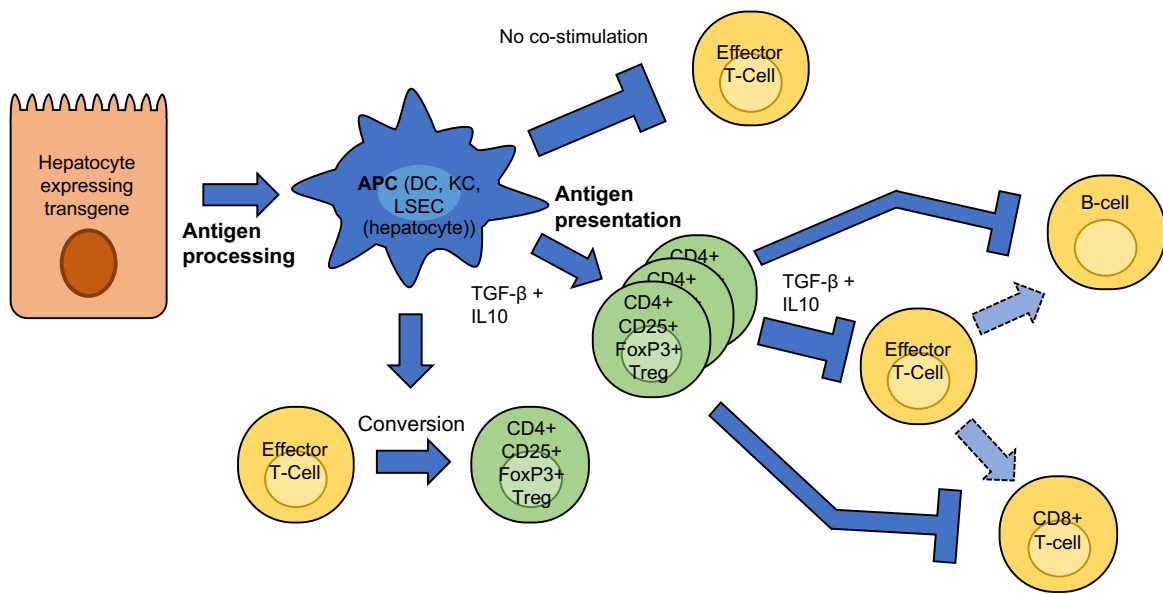


Figure 3.1 | Proposed mechanism of tolerance induction using liver expression of transgenes. After restricted expression of a transgene in hepatocytes, resident APCs will present the antigen in a tolerogenic milieu, due to the secretion of anti-inflammatory cytokines such as TGF-β and IL10. This induces the proliferation of CD4+CD25+FoxP3+ T-regulatory cells (Tregs) either directly or by trans-differentiation from effector T-cells. These Tregs actively inhibit antigen specific B- and CD8+ T-cell formation. The effect of APCs is likely local to the liver but circulating Tregs can assert their effect throughout the body. Information for the preparation of this figure is derived from [305-307]. APC, antigen-presenting cell; DC, dendritic cell; KC, Kupffer cell; LSEC, liver sinusoidal endothelial cell; TGF-β, transforming growth factor beta; IL10, interleukin 10.

3.2 Results

3.2.1 Expression and biodistribution of rAAV2/8 vectors with ubiquitous promoters

Gene therapy vectors based on rAAV2/8 have been used successfully in liver directed pre-clinical trials in mice, dogs and NHPs [308, 309] and have also shown efficacy in humans [262]. In mice in particular, rAAV2/8 has the ability to transduce a high percentage of hepatocytes without obvious adverse effects, given a sufficiently high dose [310]. This makes rAAV2/8 a prime candidate to use for liver specific expression with the aim of inducing transgene tolerance. To confirm the reported liver tropism, two rAAV2/8 based vectors were designed and manufactured to express the luciferase (Lux) reporter gene from a ubiquitous promoter, either CASI or CMVie (diagrams of each of the genome plasmids used for vector production of all vectors used in this study can be found in **Figure 3.2**). The pan-tissue expression characteristics of these promoters [2, 311, 312] combined with the reporter gene should allow the tracking of rAAV2/8 transduction *in vivo* over time. All vector genomes in this chapter contained a cassette comprising of 4 copies of the target site for miR-142-3p, which should prevent transgene expression in APCs [9]. The rAAV2/8 serotype was compared with rAAV2/8-Y733F, a single amino acid variant that has a tyrosine to phenylalanine substitution mutation, which is reported to enhance expression by inhibiting ubiquitination and degradation of virus particles following internalisation [313, 314].

A time-course study following the biodistribution of rAAV2/8 or rAAV2/8-Y733F expressing Lux from either a CASI or CMVie promoter was conducted by injecting 1e11 GC of each of the four vectors (viral vectors can be found in **Table 2.12**). Injection was performed on female 6-8-week-old BALB/c mice, either intravenously (IV) in the lateral tail vein or intramuscularly (IM) in the right gastrocnemius. The IV injection should readily deliver vector to the liver, while the IM route was investigated for liver delivery due to the simplicity in administration, as evidenced by its generic use for routine vaccination. Luciferase expression was followed over time via *in vivo* bioluminescence imaging (BLI) and

representative images are shown 14 days post-injection in **Figure 3.3**. Mice injected IM showed strong expression in the injected muscle, largely restricted to that area alone, although the high levels of expression (and therefore luminescence) in that location might obfuscate low-level expression elsewhere (**Figure 3.3A**). Tail vein injection shows a more widespread biodistribution pattern, with expression visible in the location of the liver, lower abdomen and tail (**Figure 3.3B**). The latter is likely due to damage-induced vector transduction at the injection site.

The time-course of luminescence, indicative of Lux expression, is shown in **Figure 3.4A and B** for IM and IV injections respectively. The ROI used for the measurements is shown in **Figure 3.4C and D**. Mice injected IM showed strong Lux expression across all capsid variants and promoters, indicating robust and long-lasting expression. Measurements of the liver region in mice injected intravenously likewise showed strong and long-lasting expression, with both the CASI and CMVie promoters. Taken together, these data show that rAAV2/8 is capable of transducing the livers of mice using an IV injection method but not IM. The biodistribution pattern of rAAV2/8 is more widespread than anticipated based on previous reports [315, 316] although other reports show a result more similar to what is shown here [2]. The capsid mutant rAAV2/8-Y733F did not offer any improvement over wild-type rAAV2/8, following IM or IV injection.

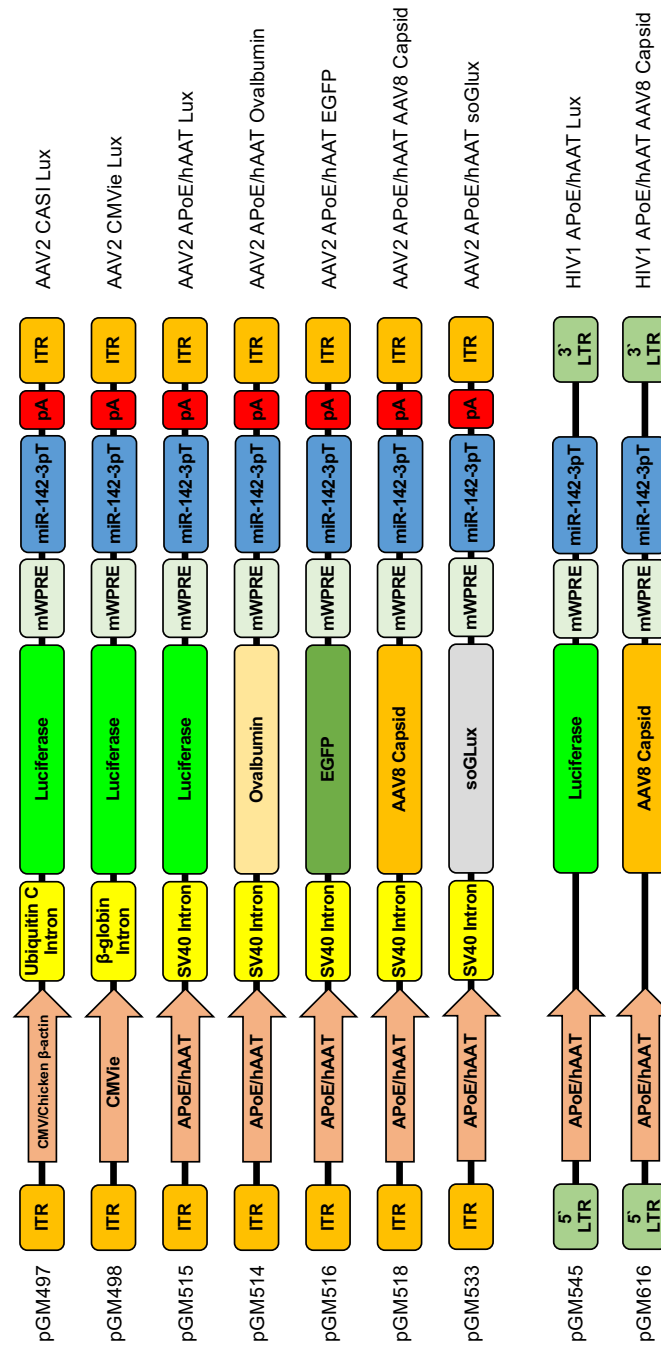


Figure 3.2 | Schematic of vector constructs used in Chapter 3. Vector constructs were used to make rAAV2/8 or rLV from AAV2 and HIV1 lentiviral backbones respectively (viral genomes can be found in **Table 2.2**, viruses made in **Table 2.12**). Internal plasmid names start with pGM and are shown on the left and naming convention used in this chapter are shown on the right. EGFP, enhanced green fluorescent protein; ITR, inverted terminal repeat; LTR; long terminal repeat; Lux, luciferase; miR-142-3pT, microRNA target sequence for APC expressed miR-142-3pT; mWPRES, mutant WPRES; pA, poly adenylation; soGLux, CpG free codon optimised Gaussia luciferase.

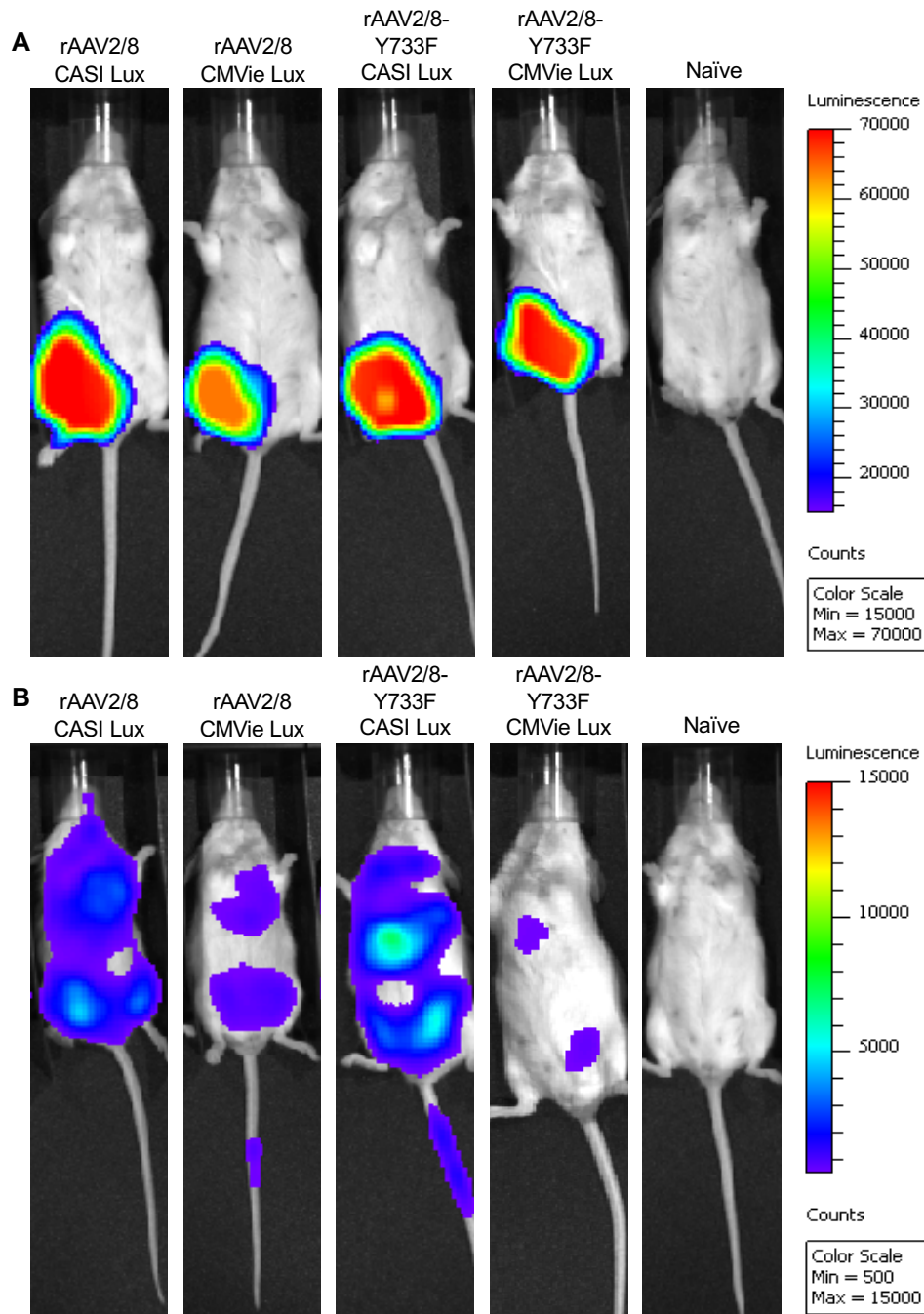


Figure 3.3 | In vivo bioluminescent imaging 14 days post-injection of IV or IM delivered rAAV2/8 or rAAV2/8-Y733F expressing Lux from CASI or CMVie promoter. Mice were injected in the right gastrocnemius muscle (**A**) or in the lateral tail vein (**B**) with $1e11$ GC of the indicated vector diluted to 40 μ l in PBS. After 14 days the mice were injected intraperitoneally with 300 μ l of D-luciferin and imaged 20 minutes later under general anaesthetic (protocol **Section 2.12.5**). Representative mouse shown from each group (n=5). Note the change in luminescence scale between **A** and **B**.

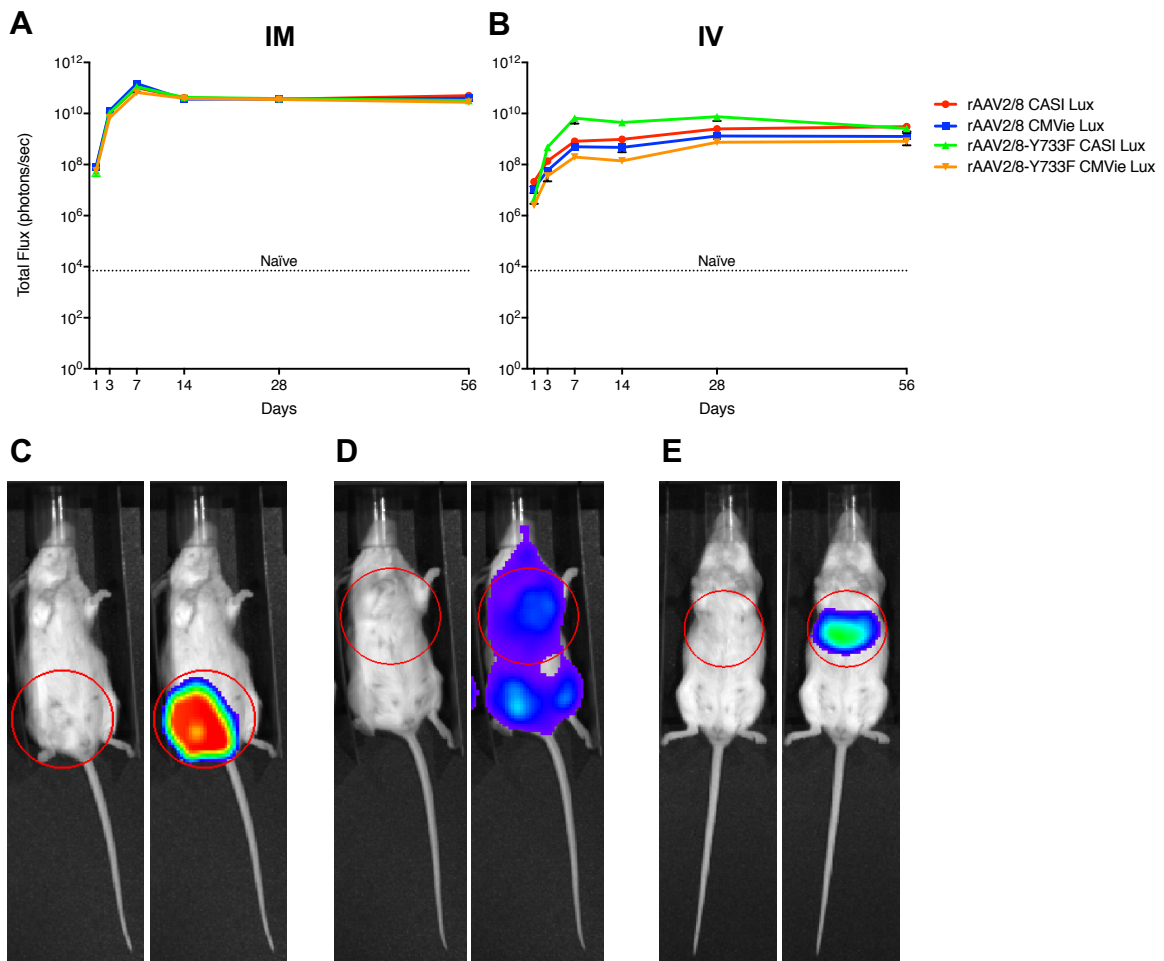


Figure 3.4 | Time course quantification of in vivo luciferase expression by bioluminescence imaging of mice injected IM or IV with rAAV2/8 or rAAV2/8-Y733F. Measurements of total Flux of a region of interest (ROI) taken around the right leg for IM injections (**A** and **C**, ROI is the red circle in which the measurement is quantified) and around the lower abdomen for IV (**B** and **D**) to capture muscle and liver expression respectively. Measurements are from mice of **Figure 3.3**. **E** shows the ROI for experiments with APoE/hAAT promoter (see **Figure 3.5**). Values shown are mean per group $\pm$ SEM ($n=5$). Representative mouse shown from each group in **C-E**, all at 14 days post-injection. Dotted line represents mean of naïve group measured at day 7.

3.2.2 Restricting transgene expression to the liver using a hepatocyte specific promoter

While rAAV2/8 is able to transduce the liver effectively, in order to induce tolerance the transgene must be expressed exclusively in the liver. Using vector tropism alone to target expression in the liver leads to unwanted expression in other tissues (**Figure 3.3B**). To address this issue, a liver-specific promoter was used to prevent ectopic expression; the APoE/hAAT promoter has been used successfully in clinical trials for haemophilia B [7, 59]. The ability of the APoE/hAAT promoter to modulate tissue-specific expression was tested *in vivo* by comparing it with the CASI promoter, whilst packaged in rAAV2/8 or rAAV2/8-Y733F injected IV as in the previous experiment. As seen in the BLI images of **Figure 3.5A**, APoE/hAAT is effective in restricting Lux expression to the liver of female BALB/c mice. The ROI measurements in the area of the liver (ROI according to **Figure 3.4E**, measurements in **Figure 3.5B**) show comparable levels between both promoter and capsid choice.

Vector DNA content was assessed in organs harvested 28 days post-injection using quantitative PCR (qPCR) against mWPPE which is present in the vector genome but absent from mouse genomic DNA (**Section 2.12.7**). The results shown in **Figure 3.6** indicate that the biodistribution of rAAV2/8 appears similar irrespective of promoter choice. Vector copy number (VCN) in the spleen appears to be higher in the group doses with rAAV2/8 APoE/hAAT Lux but this is not statistically significant (vs rAAV2/8 CASI Lux, $p=0.085$, Kruskal-Wallis with Dunn's multiple comparisons test). It is also evident that rAAV2/8 does not readily transduce organs other than the liver, although in this experiment the uteri and ovaries were not harvested. These latter organs are a reported target for rAAV2/8 [316] and may account for the Lux signal observed in the lower abdomen of IV injected mice expressing Lux from ubiquitous promoters (**Figure 3.4D and 3.5A**).

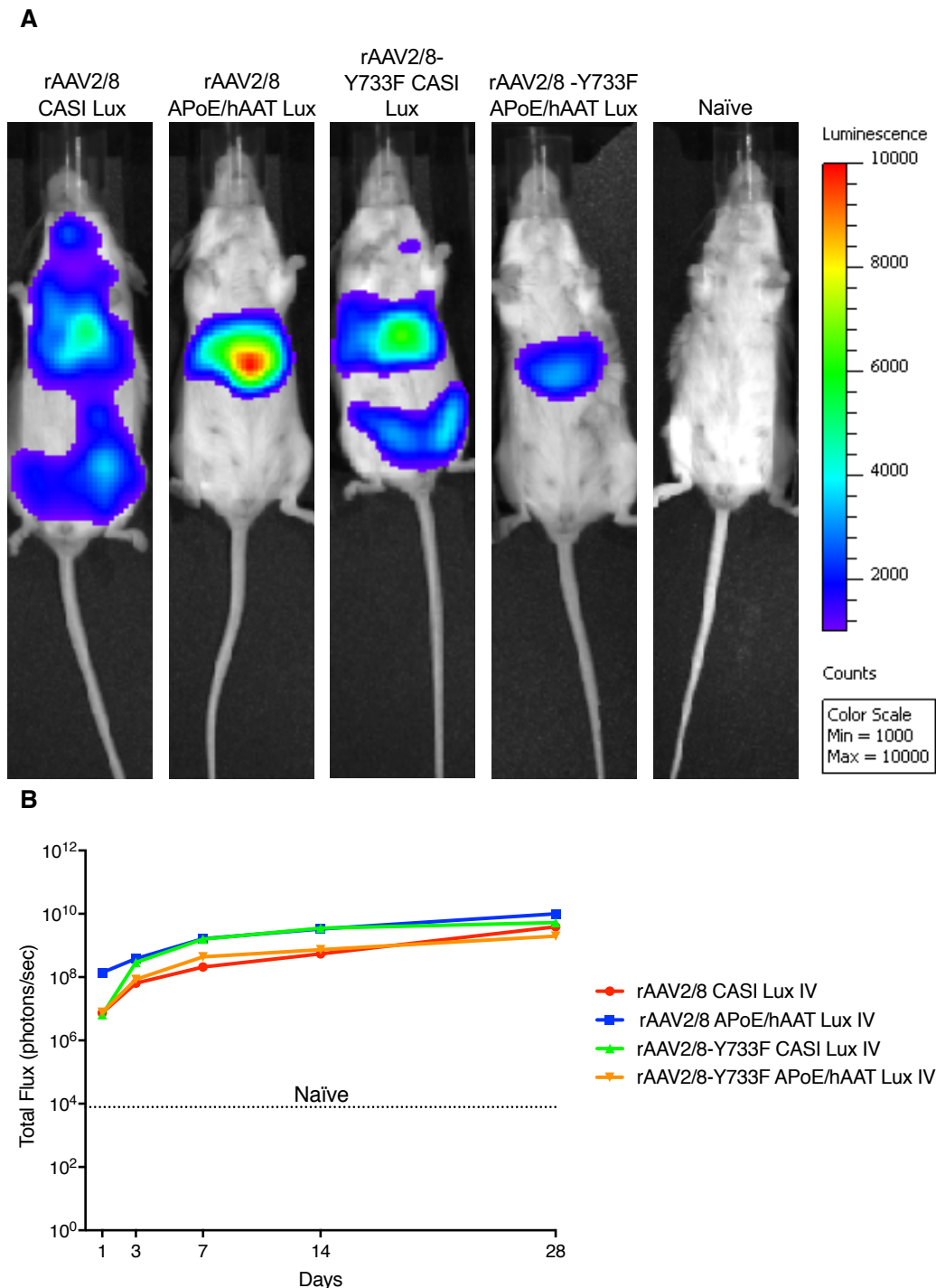


Figure 3.5 | In vivo bioluminescence imaging to compare the level and biodistribution of Lux expression from rAAV2/8 and rAAV2/8-Y733F under the control of the CASI or APoE/hAAT promoter. BLI images of mice 14 days after IV injection in the lateral tail vein with the indicated vectors (1×10^{11} GC dose per mouse) diluted in 100 μ l PBS (**A**). In (**B**), the time course of total Flux in an ROI in the lower abdomen is shown (see also **Figure 3.4E**). Representative images are shown; values are mean per group $\pm$ SEM ($n=5$). Dotted line represents mean of naïve group measured at day 7.

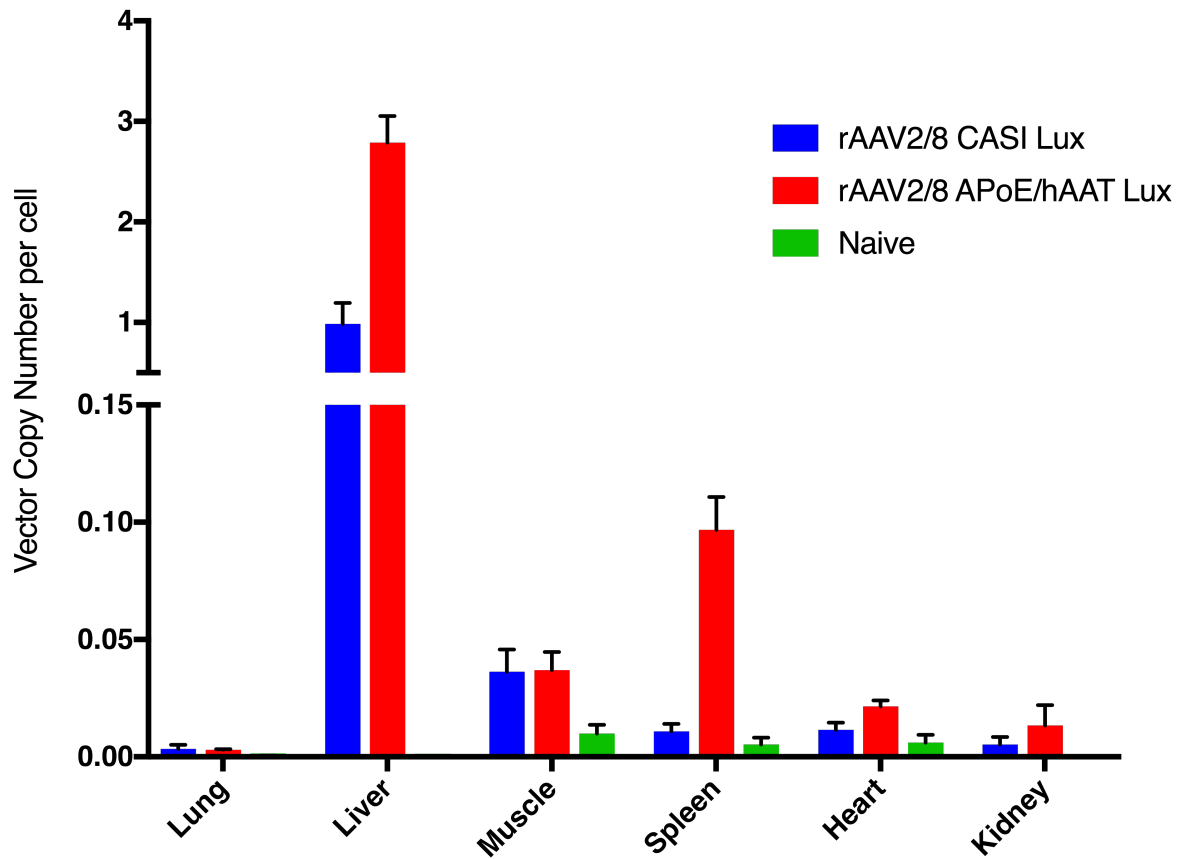
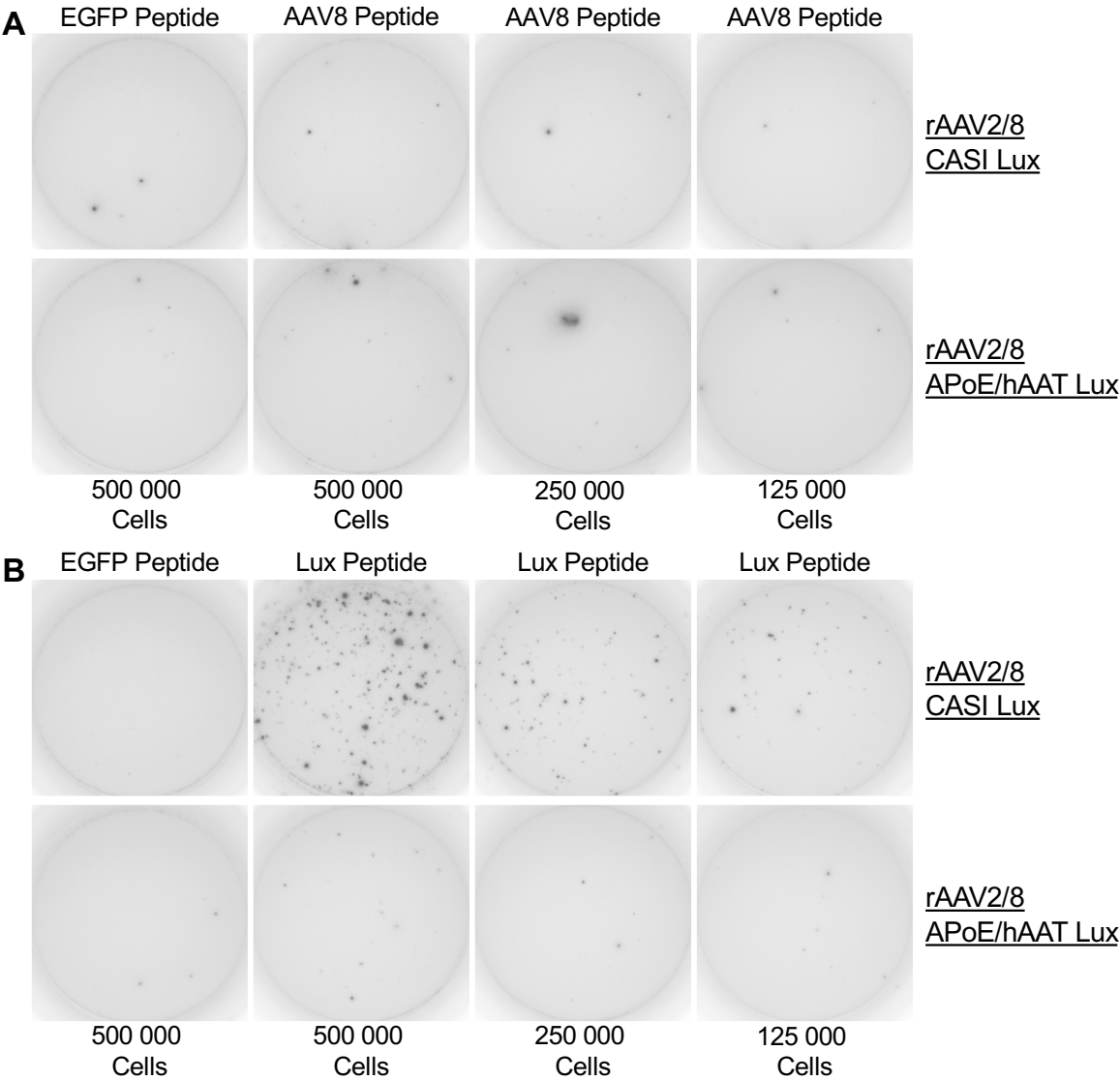


Figure 3.6 | Vector copy number per cell in different mouse organs following IV injection of rAAV2/8. Organs from mice in **Figure 3.5** were harvested on day 28 post-injection. The DNA was extracted, and virus genomes quantified by qPCR against the mWPRE sequence, normalised to total DNA content and converted to vector copy number per cell (**Section 2.12.7**). Values shown are mean per group $\pm$ SEM ($n=5$).

3.2.3 Liver-restricted expression induces cellular immune tolerance to luciferase

To test the reported tolerogenic capacity of liver-restricted expression of non-self transgenes, splenocytes of mice from the study described in **Figure 3.5** (dosed with rAAV2/8 CASI Lux and rAAV2/8 APoE/hAAT Lux, but not rAAV2/8-Y733F) were harvested at day 28 post-injection and subjected to an ELISPOT assay (**Section 2.12.9**). The ELISPOT assay measures the number of cytotoxic T-cells that secrete interferon- γ (a mark of activity) following stimulation with immunogenic peptides from Lux or AAV8 capsid proteins. No AAV8 reactive T-cells were detected in either group of mice (**Figure 3.7A**). This might be due to the low immunogenicity of AAV8 in mice [317] or because splenocytes were harvested after the peak of CD8⁺ T-cell response (7-9 days [318]). On the other hand, a strong anti-Lux T-cell response was detected in the group dosed with rAAV2/8 CASI Lux, which was absent in the group expressing Lux from the liver-restricted APoE/hAAT promoter (**Figure 3.7B**). This statistically significant difference ($p = 0.03$, Mann Whitney U, **Figure 3.8B**) is likely due to the liver-restricted Lux expression by APoE/hAAT, inducing tolerance to the luciferase transgene product.



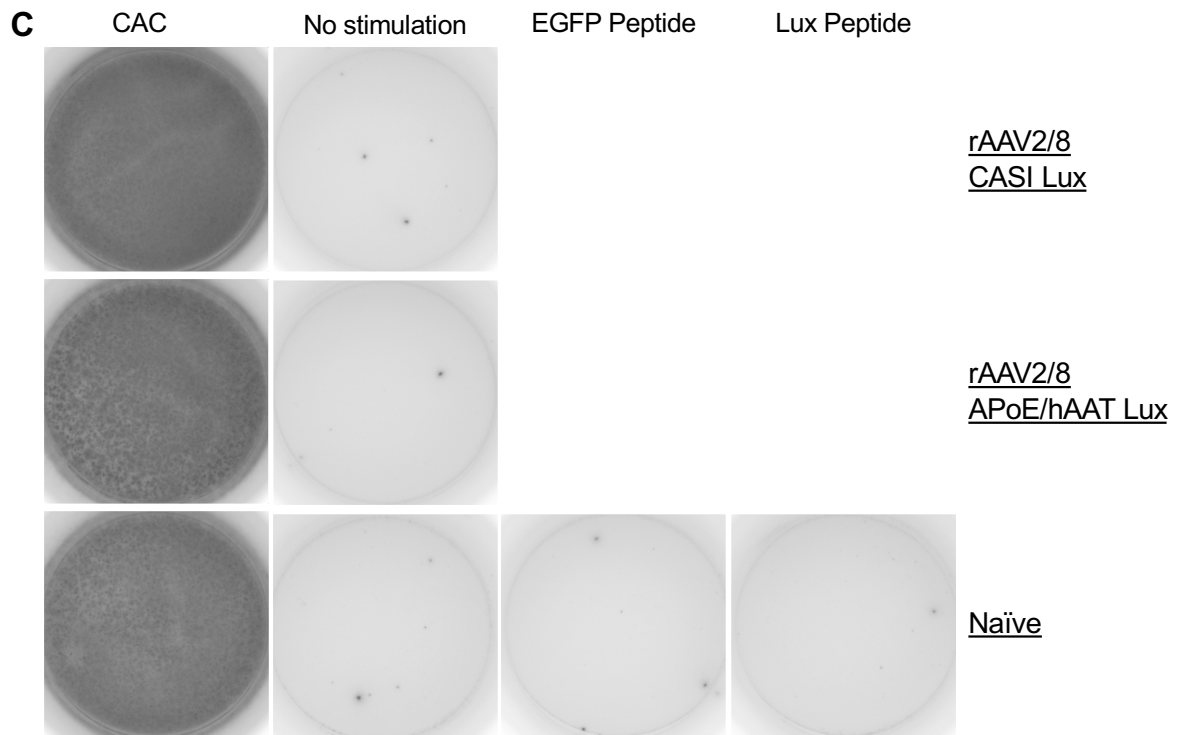


Figure 3.7 | Interferon- γ ELISPOT performed on splenocytes harvested from mice injected IV with rAAV2/8 expressing Lux from CASI or APoE/hAAT promoter. Female BALB/c mice were injected IV with rAAV2/8 CASI Lux or rAAV2/8 APoE/hAAT Lux and splenocytes were harvested 28 days post-injection. The cells were stimulated with 2 μ g/ml immunodominant peptides for AAV8 capsid (IPQYGYLTL, shown in **A**) or luciferase (GFQSMYTFV, shown in **B**) for the BALB/c haplotype (H2kd) in 10% RPMI for 48 hours. Panel **C** shows the positive control CAC (cell activation cocktail, an indiscriminate activator of T-cells) and unstimulated cells for all groups and confirms that cells from Naïve mice do not react to either the EGFP or Lux peptide. Cells from Naïve mice were not stimulated with AAV8 capsid peptide. The EGFP peptide (HYLSTQSAL) was used as a negative control and was not significantly different between both treatment groups (**B**) and Naïve mice (**C**). The mice in this figure are the same mice as shown in **Figure 3.5 and 3.6**. Representative well shown of each group (n=5). Protocol in **Section 2.12.9**. CAC, cell activation cocktail.

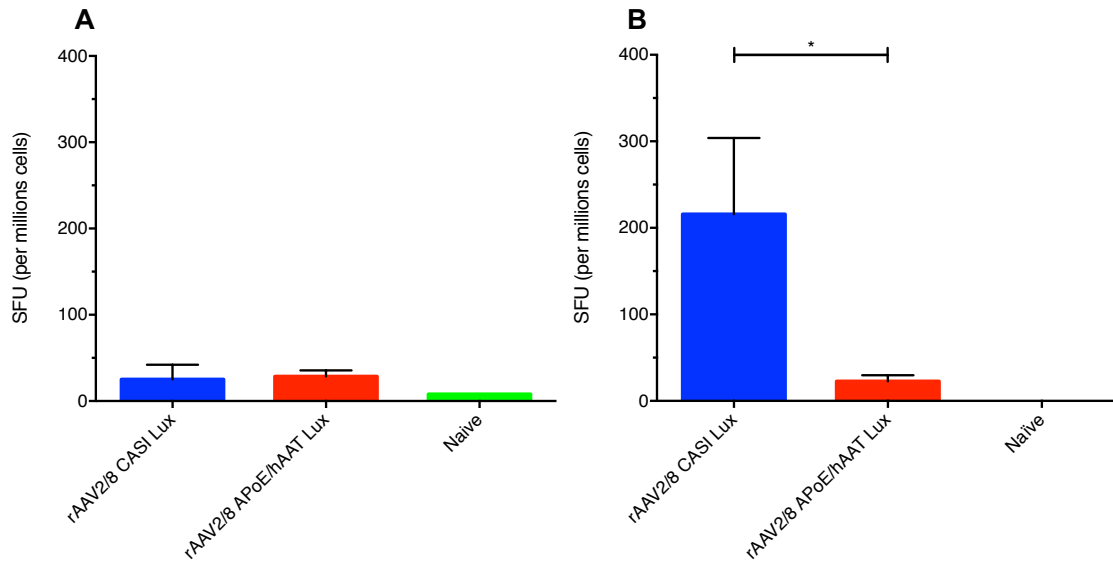


Figure 3.8 | Quantification of interferon-γ ELISPOT stimulated with AAV8 (A) or luciferase (B) peptide. Wells containing 500,000 cells from the study described in Figure 3.7 were counted with an ELISPOT reader and mean values $\pm$ SEM are shown here. The average of EGFP background was subtracted from all groups. SFU = Spot Forming Units. For cells stimulated with the luciferase peptide (B): rAAV2/8 CASI Lux versus rAAV2/8 APoE/hAAT Lux (*), $p=0.0317$ (Mann Whitney U test). $N=5$ for CASI Lux and APoE/hAAT Lux bars; Naïve is $n=1$.

3.2.4 Humoral immune response to AAV8 prevents repeat administration of rAAV2/8

To confirm that a humoral response to rAAV2/8 prevents effective re-administration (i.e. re-administration that results in measurable expression of the delivered transgene) an experiment was designed to specifically avoid the potential for a confounding anti-transgene response. To this end, three different viral constructs were created each expressing a different transgene (EGFP, Ovalbumin and soGLux) from the APoE/hAAT promoter (**Figure 3.2**), all packaged in AAV8 (**Table 2.12**). The soGLux transgene is a human, codon-optimized version of the Gaussia Luciferase reporter gene that can be secreted into the bloodstream allowing for a simple and sensitive measure of expression. The experimental protocol in **Figure 3.9A** highlights that mice in the repeat administration group received a total of three 1×10^{11} GC doses of rAAV2/8 expressing different transgenes, with soGLux being the final dose. The third injection is comparable to the first injection in the single administration group to allow read-out of soGLux expression after repeat or single administration.

To detect the anti-AAV8 humoral response, an anti-AAV8 antibody ELISA (protocol **Section 2.12.8** and a schematic of the procedure **Figure 2.2**) was performed on the serum 14 days post-injection of 1×10^{11} GC of rAAV2/8 APoE/hAAT EGFP. Four weeks following the first dose, a dose of 1×10^{11} GC rAAV2/8 APoE/hAAT Ovalbumin was delivered to the same mice and 14 days later another anti-AAV8 antibody ELISA was performed. **Figure 3.9B** shows the dilution curve of serum from mice receiving the single dose of rAAV2/8 expressing EGFP (red line) and the same mice having received both the EGFP and the Ovalbumin dose (blue line). There is a clear anti-AAV8 antibody response after a single dose of rAAV2/8 and the shift in the dilution curve and the increased (20x) associated EC_{50} value indicates that a second dose increases the response considerably, perhaps due to a memory response. Measuring soGLux activity in the serum of all groups of mice 14 days after the final dose of rAAV2/8 showed that soGLux is expressed in mice receiving the

soGLux vector alone, but not when the soGLux dose is preceded by two previous rAAV2/8 doses (**Figure 3.10**). This confirms that repeat administration of rAAV2/8 prevents successful expression of a transgene from a subsequent dose, most likely due to an anti-AAV8 antibody response that neutralises the vector before it can transduce the target.

3.2.5 rAAV2/8 mediated liver expression of AAV8 capsid does not induce tolerisation

In an effort to avoid the anti-AAV8 antibody response that prevented successful repeat administration of rAAV2/8 (**Figure 3.10**), tolerisation to AAV8 capsid was attempted. Using a similar construct utilised to successfully induce cellular tolerance to Lux protein, a batch of rAAV2/8 virus expressing AAV8 VP1 capsid from the APoE/hAAT promoter (**Figure 3.2**) was produced. While the entire capsid sequence is present, the original p40 promoter (See **Figure 1.1**) as well as the Rep proteins and Ad helper proteins are missing so it is unlikely that substantial quantities of VP2 and VP3 proteins are present because alternative splicing of the VP transcripts is dependent on these auxiliary factors. Western-blot analysis of cells transfected with viral genome plasmids expressing AAV8 capsid (pGM518, **Figure 3.2**) confirmed this hypothesis (data not shown). Nonetheless, since all VP proteins are in the same open reading frame (**Figure 1.1**), VP1 contains all of the peptides that can potentially be displayed on MHC complexes. Mice were injected with rAAV2/8 APoE/hAAT expressing AAV8 capsid or EGFP (as a control) and 14 days post-injection serum was collected for an anti-AAV8 antibody ELISA as before. As presented in **Figure 3.11**, expressing AAV8 capsid protein in the liver did not substantially reduce the antibody response to the vector from which it was expressed. While the difference in EC_{50} is statistically significant ($p=0.0317$, Mann Whitney U), the absolute reduction of less than three-fold was not sufficient to warrant an attempt at repeat dosing. Anti-AAV8 antibody levels did not drop significantly in the 6 months that the mice were monitored (data not shown).

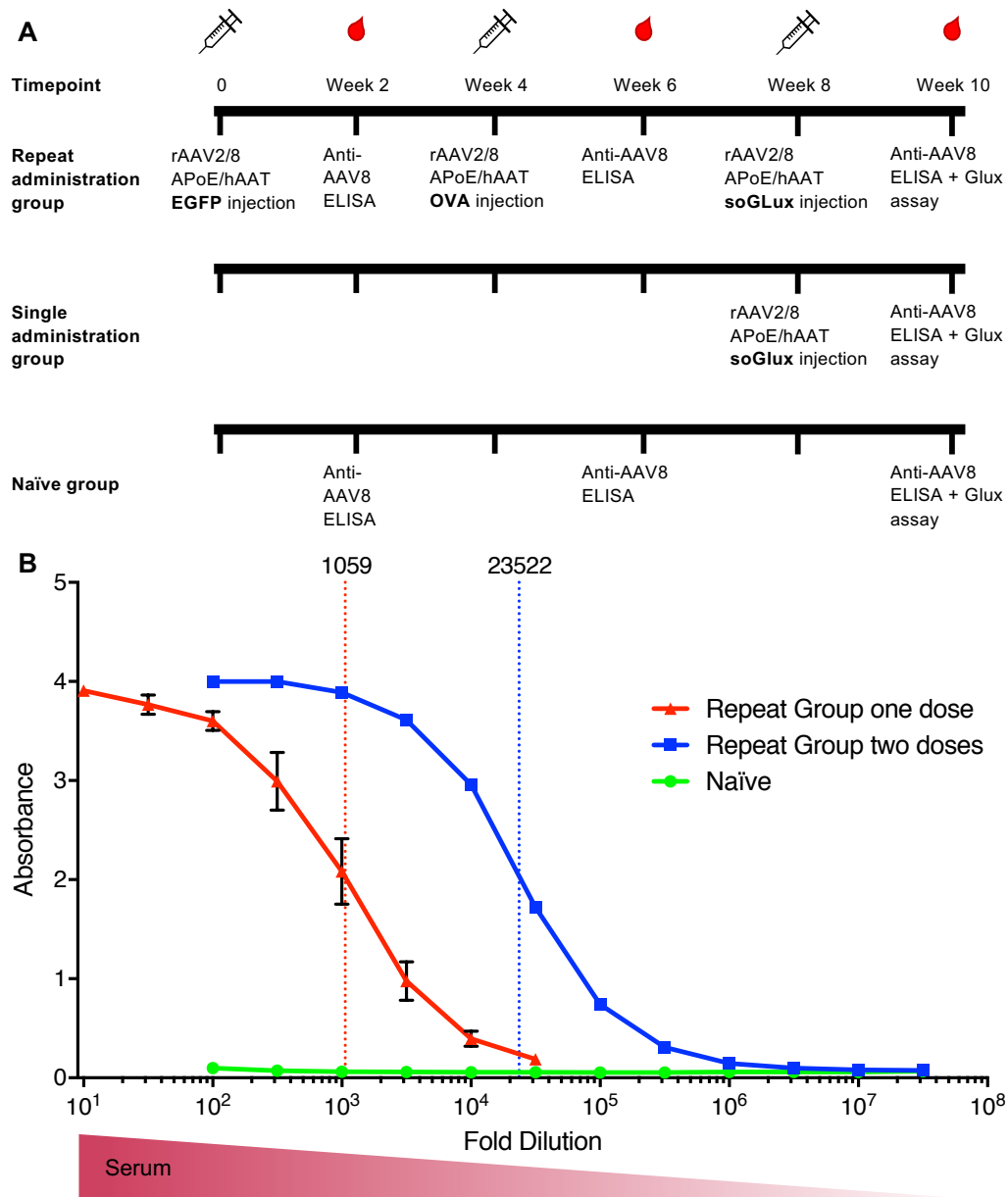


Figure 3.9 | Anti-AAV8 antibody titres after repeat injection of rAAV2/8. The experimental protocol is outlined in **A**. Mice were injected IV with $1e11$ GC of the indicated vectors. Serum was collected from the tail vein from at the indicated times and subjected to an anti-AAV8 antibody ELISA assay (**Figure 2.2** and **Section 2.12.8**). The results of the anti-AAV8 antibody ELISAs at week 2, when the mice in the Repeat administration group have received a single dose of rAAV2/8, and week 6 when mice in the Repeat administration group have received two doses or rAAV2/8 is shown in **B**. Dotted lines represent EC_{50} values calculated via three parameter linear regression. Values shown are mean per group $\pm$ SEM ($n=5$).

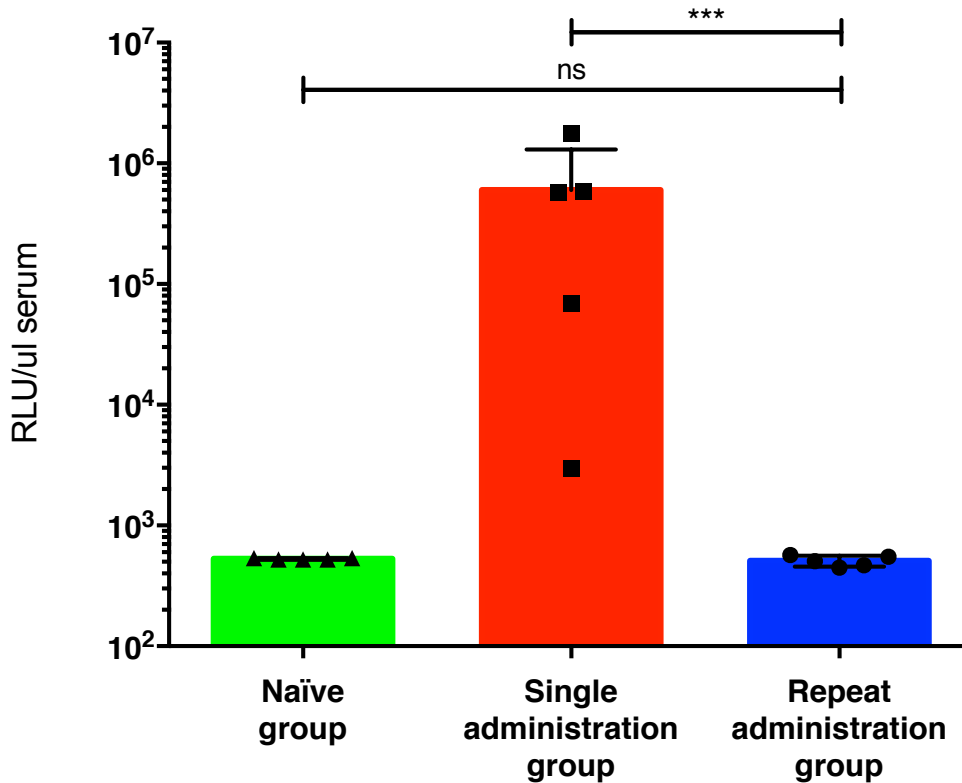


Figure 3.10 | Gaussia luciferase activity measurements in serum after single or repeat doses of rAAV2/8. Mice described in **Figure 3.9** received either a single dose of $1e11$ GC of rAAV2/8 APoE/hAAT soGLux (Single administration group) or they received a dose of $1e11$ GC rAAV2/8 APoE/hAAT soGLux after previously receiving two doses of $1e11$ GC of rAAV2/8 expressing a different transgene (rAAV2/8 APoE/hAAT EGFP and rAAV2/8 APoE/hAAT OVA, Repeat administration group). Serum was collected 14 days after the final dose and GLux measured (protocol **Section 2.12.10**). Mice were age matched at the time of rAAV2/8 APoE/hAAT soGLux dosing. Repeat versus Single (***) , $p = 0.0003$, Repeat versus Naïve (ns), $p > 0.999$ (Kruskall-Wallis with Dunn's multiple comparisons test). Values shown are mean per group $\pm$ SEM ($n=5$).

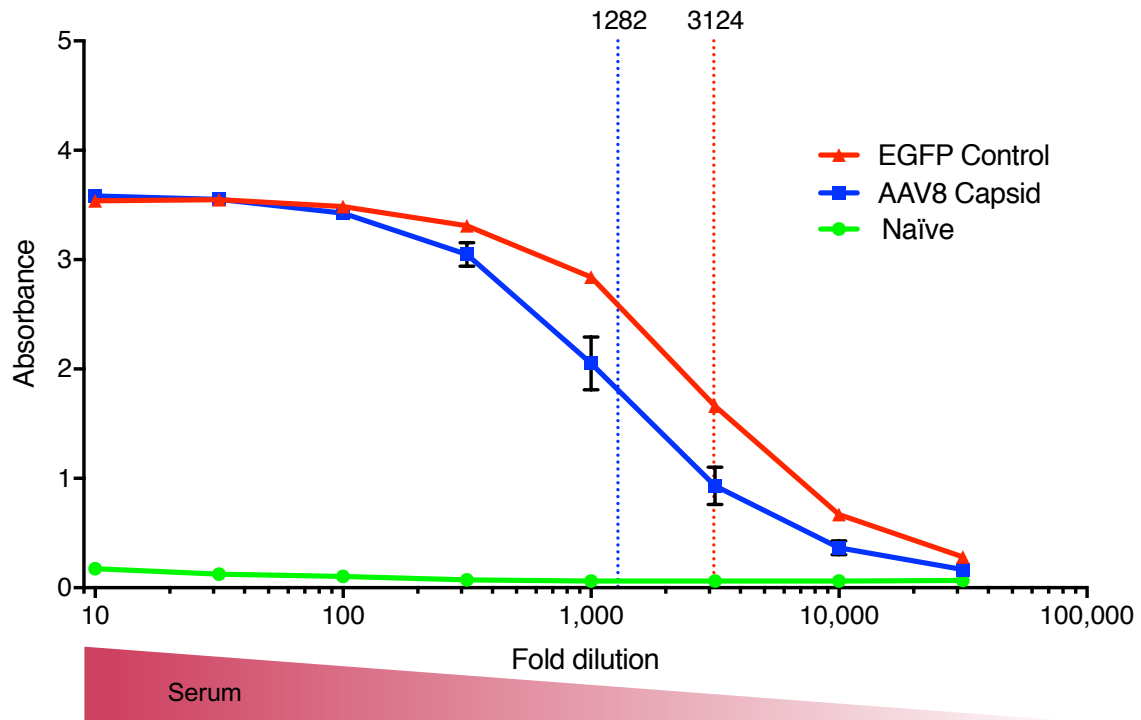


Figure 3.11 | Anti-AAV8 antibody titres after IV injection of EGFP control or AAV8 capsid tolerising vectors. Female BALB/c mice were injected with 1×10^{11} GC of rAAV2/8 APoE/hAAT EGFP (EGFP Control, red) or rAAV2/8 APoE/hAAT AAV8 Capsid (AAV8 Capsid, blue). Serum was collected 14 days post-injection and assayed for anti-AAV8 antibodies through an indirect ELISA. Dotted lines represent EC_{50} values calculated via three parameter linear regression. A mouse specific pan-IgG anti-Fc antibody was used to detect anti-AAV8 antibodies. Difference in EC_{50} is statistically significant ($p=0.0317$, Mann Whitney U). Values shown are mean per group $\pm$ SEM ($n=5$).

3.2.6 Lentivirus mediated liver expression of AAV8 capsid does not induce tolerisation

The results in **Section 3.2.5** indicated that using rAAV2/8 with the APoE/hAAT liver-specific promoter for AAV8 capsid expression to induce humoral immune tolerance to AAV8 capsid did not result in the hypothesised outcome. To the contrary, there was a strong anti-AAV8 antibody response measured after delivery of both the control and tolerising vector (**Figure 3.11**). The attempt to tolerise to the same capsid protein as was contained in the delivery vector was a potential confounding factor. Therefore, an alternative liver-specific delivery vector was identified. An HIV1-based lentiviral vector was previously shown to induce tolerance to FIX, even after pre-existing immunity [299], therefore, the APoE/hAAT AAV8 Capsid cassette (minus the SV40 intron sequence as introns can cause interference in lentivirus production) was cloned into a lentiviral genome plasmid (**Figure 3.2, Table 2.2**) and from that a VSV-G pseudotyped HIV1-based lentivirus was generated (**Table 2.12, Section 2.11.3**). This vector differed to that used by Annoni et al. [299] only in the liver-specific promoter used (mouse transthyretin in the indicated study versus APoE/hAAT in this study). Mice injected IV with 8e9 transducing units of lentivirus expressing Lux (instead of AAV8 capsid) from APoE/hAAT showed expression of Lux in the liver as expected (**Figure 3.12A**), for up to 2 months (**Figure 3.12B**) with a slight reduction in expression over time (2-fold, measured by total Flux in an ROI around the liver as in **Figure 3.4E**). Thus, the selected lentiviral genome showed suitable liver-restricted transgene expression – a pre-requisite for vector-induced tolerance. The experimental protocol of the lentiviral tolerance experiment is shown in **Figure 3.13**. At 14 days following the administration of the first dose of 8e9 transducing units of rHIV1 APoE/hAAT AAV8 Capsid, an ELISA for anti-AAV8 antibodies was performed to make sure capsid expression did not cause a humoral immune reaction (**Figure 3.14B**). A small increase in anti-AAV8 antibody titres can be observed which was not deemed strong enough to warrant changes in the experimental protocol.

Fourteen days after the dosing of the capsid expressing lentiviral vector, rAAV2/8 APoE/hAAT EGFP (1e10GC) was administered IV. Serum was collected and subjected to an anti-AAV8 antibody ELISA 14 days later (Week 4, **Figure 3.14B**). A strong antibody response was detected both in the control group and the capsid group of mice expressing AAV8 capsid in the liver, indicating that rLV mediated AAV8 capsid expression was not successful in preventing an anti-AAV8 antibody response. Interestingly, the anti-AAV8 antibody response waned slightly after 2 months (**Figure 3.14C**). To address whether this loss of antibody levels allowed for effective repeat administration of rAAV2/8, 11 weeks after the injection of rHIV1 VSV-G APoE/hAAT AAV8 Capsid, the mice in the Capsid and Control groups were injected with 1e10 GC of rAAV2/8 APoE/hAAT Lux. **Figure 3.14D** shows that no Lux expression was measurable in both the Control and Capsid group, indicating that the antibodies against AAV8 measured (**Figure 3.14C**) are sufficient to prevent successful repeat administration of rAAV2/8.

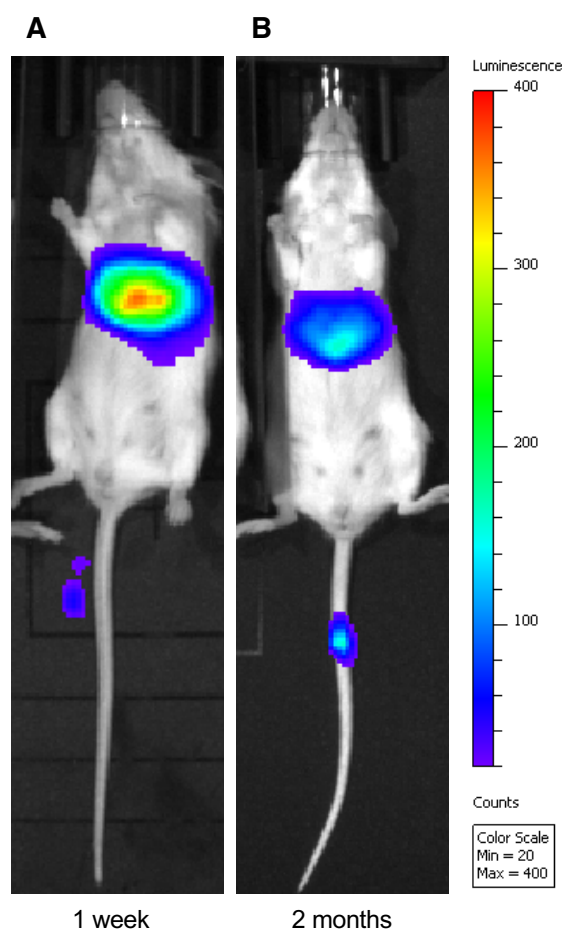


Figure 3.12 | Biodistribution of rHIV1 VSV-G APoE/hAAT Lux after IV injection. Female BALB/c mice were dosed with 8×10^9 transducing units of VGV-G pseudotyped HIV1-based lentivirus expressing luciferase from the APoE/hAAT promoter (rHIV1 VSV-G APoE/hAAT Lux, **Table 2.12**) and 7 days (A) and 2 months (B) post-injection Lux expression was assayed *in vivo* via bioluminescence imaging. Representative mouse shown ($n=5$).

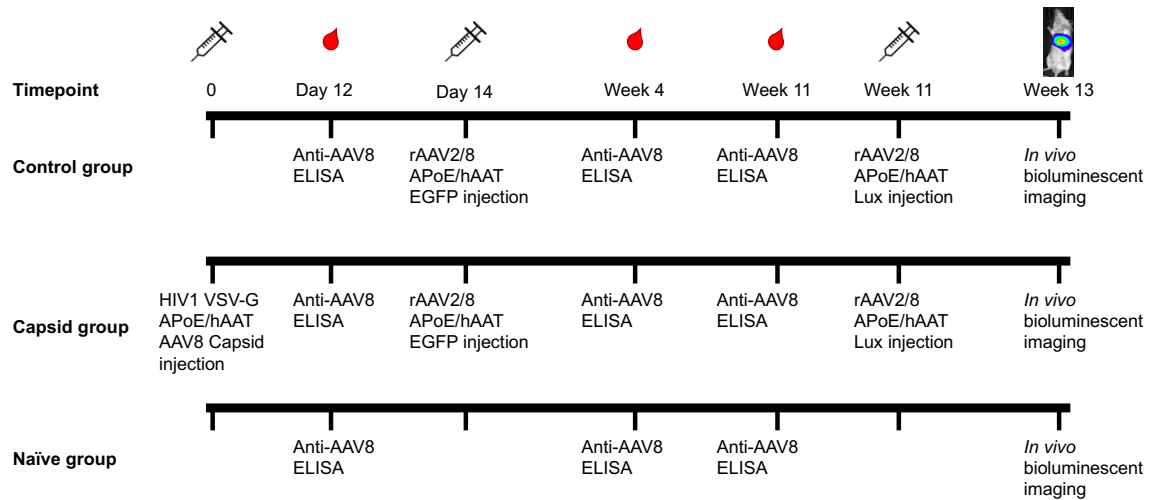


Figure 3.13 | Lentivirus based tolerisation experiment timeline. In the Capsid group, 8×10^9 transducing units of rHIV1 VSV-G APoE/hAAT AAV8 Capsid diluted in 100 μ l of PBS was injected IV. 12 days later serum was taken from all groups and subjected to anti-AAV8 antibody ELISA. 14 days post the first injection, 1×10^{10} GC of rAAV2/8 APoE/hAAT EGFP was injected IV (100 μ l in PBS) to the Control and Capsid groups. 2 weeks and 2 months post rAAV2/8 injection serum samples were collected and again subjected to anti-AAV8 antibody ELISA. 11 weeks after the first injection both the Control and Capsid groups received another dose of rAAV2/8, this time expressing Lux from the APoE/hAAT promoter (rAAV2/8 APoE/hAAT Lux). 14 days after the second rAAV2/8 injection all groups were examined for Lux expression using in vivo bioluminescent imaging.

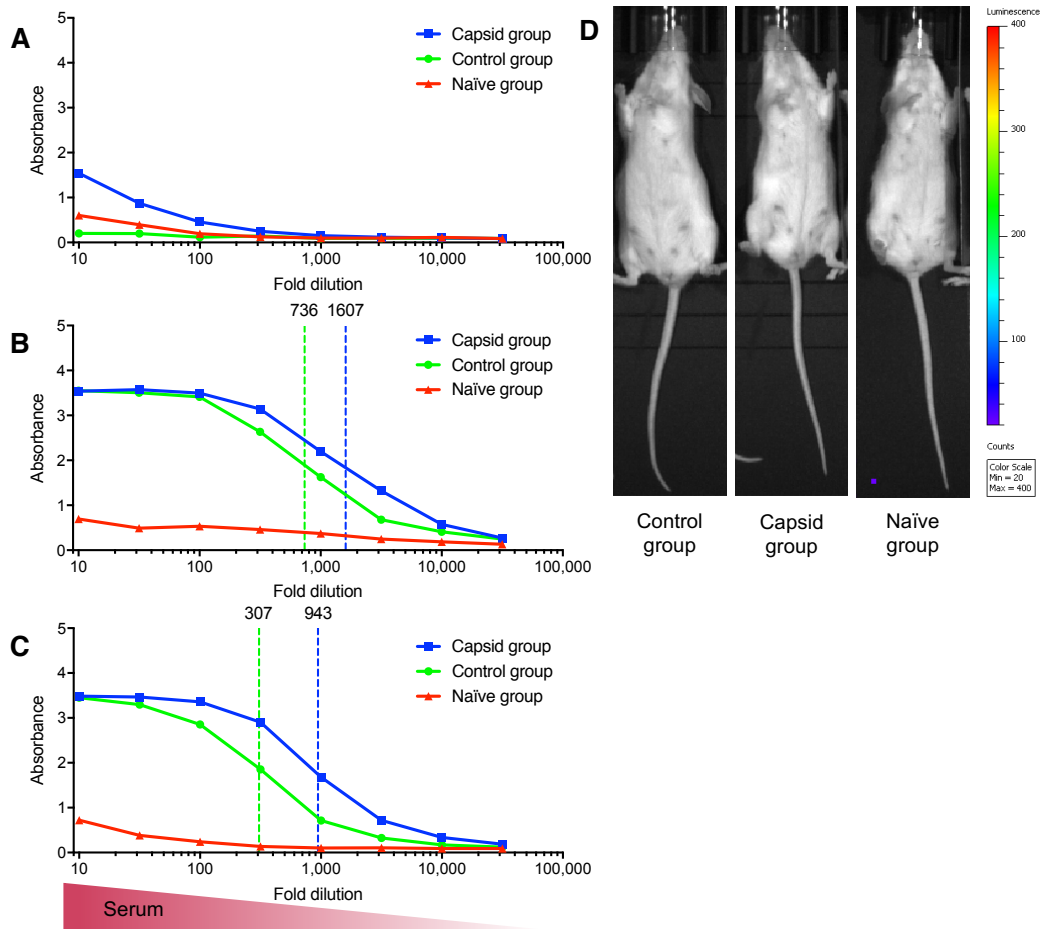


Figure 3.14 | Anti-AAV8 antibodies after lentiviral tolerance induction to AAV8 capsid proteins prevent repeat administration. Female BALB/c mice were injected with VSV-G pseudotyped HIV1 lentivirus expressing AAV8 capsid VP1 from the APoE/hAAT promoter (rHIV1 VSV-G APoE/hAAT AAV8 Capsid, **Table 2.12**, Capsid group). The control group did not receive the tolerising vector (Control group) and the Naïve group received no virus injections (rLV or rAAV) thus serving as a negative control for the anti-AAV8 ELISAs (see schematic in **Figure 3.13**). Serum was collected at 12 days post rLV injection (**A**) and 14 days (**B**) and 2 months (**C**) post the first rAAV2/8 injection (1×10^{10} GC of rAAV2/8 APoE/hAAT EGFP, **B**). The serum was subjected to an indirect ELISA for anti-AAV8 antibodies as before. 13 weeks after the initial rHIV1 VSV-G APoE/hAAT AAV8 Capsid injection the mice in the Control and Capsid group were injected with a second dose of rAAV2/8 expressing Lux (rAAV2/8 APoE/hAAT Lux) and two weeks later all mice were imaged using BLI. Data shown are from serum samples pooled per group (**A-C**). Representative BLI images shown (**D**). $N=5$ per group.

3.3 Discussion

3.3.1 Restricting transgene expression to the liver and inducing immune tolerance

In this chapter, the main aim was to induce immunological tolerance to AAV8 capsid proteins to primarily avert the humoral response that prevents successful re-administration (**Figure 3.10**). The secondary aim of tolerisation study was to dampen the cellular response that was shown to severely limit the expression of therapeutic transgenes in the liver, due to the killing of transduced cells by CD8⁺ cytotoxic T-cells [56]. To achieve immune tolerance, expression of the non-self transgene must be restricted to the liver. To this end the reportedly liver-tropic AAV8 capsid was investigated for its biodistribution when injected IM or IV (**Figure 3.3**). Intramuscular injection did not result in appreciable Lux expression in the liver (determined by BLI), in contrast with other reports [319, 320]. While liver transduction was observed by BLI and qPCR of viral vectors in dissected organs (**Figure 3.6**) after IV injection, a liver-specific promoter (APoE/hAAT) was necessary to prevent expression in tissues beyond the liver (**Figure 3.5**). Liver-specific expression, coupled with a microRNA target sequence that prevents expression in APCs, resulted in cellular tolerance to the non-self transgene luciferase as evidenced by a lack of luciferase-specific interferon- γ producing splenocytes (**Figure 3.7 and 3.8**). Humoral tolerance to Lux could not be established as no antibody response could be measured (data not shown).

3.3.2 Use of rAAV2/8 to express AAV8 capsid protein in the liver does not induce humoral immune tolerance to rAAV2/8

Data in **Figure 3.7 and 3.8** showed that rAAV2/8 APoE/hAAT *Transgene* miR-142-3pT could be used as a tolerogenic vector platform. It was used, therefore, to express the AAV8 VP1 capsid protein in the liver with the aim of preventing an anti-AAV8 antibody response. However, as the data in **Figure 3.11** indicate, this approach did not prevent production of anti-AAV8 antibodies, even after 6 months (data not shown); and thus this approach is highly unlikely to facilitate successful re-administration of rAAV2/8.

The induction of tolerance hinges on three main conditions: 1. Hepatocyte-specific expression and especially avoiding expression in APCs [9], 2. expression above a threshold level [321, 322] and 3. minimal (innate) immune stimulation to prevent immune priming [306]. Transgene tolerance mediated by AAV8 appeared to meet these conditions (**Table 3.2**) but there is an important confounding factor; the viral delivery platform used to induce tolerance (rAAV2/8) also contains the AAV8 protein to which tolerance is attempting to be induced. This means that AAV8 capsid is not only expressed, it is also processed after viral transduction, possibly interfering with tolerogenic antigen processing (**Figure 3.1**). Furthermore, while the presence of the microRNA target sequence in the 3' UTR of the expressed transgene prevents expression of AAV8 capsid in APCs in the liver and elsewhere, inadvertent APC transduction means that AAV8 capsid peptides from the virus itself are still available for display.

Even though Lux-reactive T-cells were detected in the group injected with rAAV2/8 CASI Lux, expression was strong and relatively persistent (**Figure 3.4 and 3.5**), indicating the response was not strong enough to abrogate expression. It was reported that, in BALB/c mice as opposed to C57BL/6, T-cell tolerance is not effective when an immunogenic co-stimulus was provided (adenovirus in this report, [323]). Using rAAV2/8 as the delivery vector in this study could be seen as the co-stimulus that 'breaks' tolerance. In the current study no anti-AAV8 CD8⁺ T-cell response could be measured (**Figure 3.7A**), even after three doses of rAAV2/8 (data not shown). Consequently, the secondary tolerisation aim, cellular tolerisation to AAV8 capsid, was not pursued further.

3.3.3 AAV8 capsid protein expression in the liver with VSV-G pseudotyped lentivirus does not induce humoral immune tolerance to rAAV2/8

To tackle the confounding factor of using rAAV2/8 to deliver AAV8 capsid to the liver, a different gene delivery platform was used to express AAV8 capsid in the liver. The platform, using a VSV-G pseudotyped HIV1-based lentivirus, has previously been used to induce

tolerance to FIX when coupled with a liver-specific promoter and a miR-142-3pT [299]. The only difference between the vector used in this chapter, and the vector used by Annoni et al., is that the latter utilised a mouse transthyretin promoter combined with a synthetic enhancer rather than APoE/hAAT. Lentiviruses could ostensibly also be used to eradicate pre-existing antibodies to AAV as those should not affect the transduction capacity of the lentiviral vector. Using the HIV1 VSV-G-pseudotyped lentiviral vector, Lux was expressed in the liver using BLI (**Figure 3.12**), albeit at ~150-fold lower levels compared with rAAV2/8.

Unexpectedly, there was a small anti-AAV8 antibody response measurable in mice injected with the lentivirus expressing AAV8 capsid (rHIV1 VSV-G APoE/hAAT AAV8 Capsid, Capsid group) even before rAAV2/8 administration (**Figure 3.14A**). This can perhaps be explained by 1. secretion of capsid protein that the cell cannot process due to overexpression, 2. by the enveloping of capsid proteins in lentiviral particles expressed from the genome plasmid in producer cells during manufacturing or 3. by capsid proteins released after cell destruction following a cellular immune reaction to lentivirus infected cells.

Use of the alternative lentiviral delivery platform did not induce tolerance to AAV8 capsid, as administration of rAAV2/8 caused a strong anti-AAV8 antibody response in mice in both the control and AAV8 capsid expressing groups (**Figure 3.14B**) which effectively prevented repeated rAAV2/8 administration as measured by Lux expression of the second vector (**Figure 3.14D**). While lentiviruses were previously successfully used to induce tolerance to different transgenes in genome configurations similar to those used in this study [9, 299-301], it is of note that lentiviruses do not meet all three conditions of tolerisation (**Table 3.2**). While it is difficult to compare expression levels from the vector used in this study with others, compared with rAAV2/8, lentivirus-mediated Lux expression in the liver was around 150-fold lower. It is conceivable that this low expression may not have reached the minimal

level required for tolerance [321, 322] which combined with a less favourable immunogenic profile could explain the lack of tolerisation.

3.3.4 Potential adverse effects of capsid expression

One hypothesis that could negatively impact the potential of capsid tolerisation irrespective of the delivery vector is that capsid expression in the liver could potentially be toxic. The observation that capsid expression is toxic was made in the development of viral packaging cell lines for rAAV (i.e. cells that express rep/cap and the helper proteins either constitutively or through chemical induction) where even low residual expression could induce cell death (Amit Nathwani, personal communication). One could hypothesize from this that capsid expression in the liver of mice as attempted in this chapter could be equally toxic, with cell death and loss of expression as a result. This would effectively prevent successful tolerisation. The expression of AAV8 capsid proteins in the liver of mice was extrapolated from luciferase expression studies (**Figure 3.5**) with an equivalent viral vector genome (APoE/hAAT *Transgene* miR-142-3pT), serotype (rAAV2/8) and production method compared to the tolerisation study in **Section 3.2.5**, with the only difference being the expressed cDNA (AAV8 capsid versus luciferase). While the viral genome plasmids expressing AAV8 capsid were confirmed to effectively express AAV8 capsid VP1 protein in cell culture through western blot (data not shown), unfortunately no confirmatory test was performed on the liver of mice transduced with rAAV2/8 or rLV expressing AAV8 capsid. This precludes validation that AAV8 capsid can be effectively expressed in the murine liver long-term. Thus, the hypothesis that the lack of tolerisation in these experiments is caused by capsid mediated toxicity and cessation of expression could not be excluded, a clear limitation of this study.

3.3.5 Conclusion

The experiments described in this chapter aimed to establish a liver-specific gene delivery platform that was capable of inducing immune tolerance to AAV8 capsid proteins. Both the

rAAV2/8 and lentivirus-based delivery vectors were capable of the liver-specific expression necessary to achieve this and, in addition, the rAAV2/8 platform was capable of inducing cellular immune tolerance to the non-self transgene luciferase. However, both platforms failed to prevent an anti-AAV8 antibody response thus preventing successful repeat administration. With repeat administration in the murine liver not currently feasible, another approach was needed; and long-term therapeutic transgene expression by means of genome editing was considered as an alternative.

Table 3.2 | Comparison of rAAV and rLV delivery platforms for tolerisation requirements

	Hepatocyte specific expression	High-level expression in hepatocytes	Minimal immune induction
rAAV	+++ AAV8 tropism, APoE/hAAT and miR-142-3pT contribute to specific expression	+++ Strong and long-term expression	++ No T-cell response but reported TLR9 to viral ssDNA
rLV	++ APoE/hAAT and miR-142-3pT contribute to specific expression but VSV-G tropism is ubiquitous	+ 150-fold lower than AAV, long term (2 months in this study)	-- Strong immune response to viral proteins (e.g. VSV-G and p24) and ssRNA genome or reverse transcribed proviral DNA [324]

Chapter 4 – Development of a reporter cell line to demonstrate HITI permutations

4.1 Introduction

4.1.1 Genome editing for large constructs

Studies described in Chapter 3 were aimed at long-term therapeutic transgene expression via the induction of immunological tolerance to AAV8 capsid proteins allowing successful repeat administration. As this was not successful, a new strategy was investigated using targeted transgene integration into the genome. As described in **Section 1.2.4**, homology-independent targeted integration (HITI) is a genome editing technique that relies on cNHEJ rather than HDR or MMEJ to integrate exogenous DNA into the cellular genome (**Figure 1.6**). This method allows for the precise integration of dsDNA into a targeted location in the genome and has great potential as a therapeutic genome editing tool. Facilitating targeted integration of large sequences with limited loss of integration efficiency with increasing insert size [136] could allow the insertion of the entire protein coding sequence of an affected gene. In this way, one could potentially correct all mutations of a given gene, a pan-mutation approach that avoids the need for re-optimisation to correct each different mutation.

There are many methods available to achieve genomic insertion of a large stretch of DNA (summarised in **Table 4.1**). Notably, there is a distinction between methods that integrate randomly (such as retroviruses and transposases) and those that integrate in a specific location using a guiding system such as Cas9 or homology arms. The former requires the use of a promoter sequence within the insert to drive expression of the transgene and, although the risk is minimal, such random insertion events can potentially cause genotoxicity [325]. Targeted insertion, however, usually requires an endonuclease (except for GeneRide and RNA-guided transposases, discussed later), which also has mutagenic

potential. Strong, long-term expression of any endonuclease can increase the chance of an off-target dsDNA break, causing mutations when repair is insufficiently accurate.

Endonuclease-HDR is a popular method, mainly because of its high repair fidelity, and is favoured over MMEJ due to the potential of the latter for large deletions around the integration site (**Section 1.2.2**). Moreover, HDR creates the ability to create 'scar-less' repair, meaning there is the potential for repair without any evidence of editing i.e. the resulting sequence is identical to wild-type. Large DNA insertion mediated via HDR can reach impressive rates *ex vivo* [156] but *in vivo* efforts have not achieved the same level [326].

Precise DNA integration without the use of an endonuclease causing a dsDNA break is possible with GeneRide and RNA-guided transposases. GeneRide is an *in vivo*, endonuclease free targeting system [327]. It combines a liver-tropic AAV delivery vector with an HDR targeting vector that aims for integration in-frame within the albumin protein coding sequence. The low integration levels (<1%) are offset by the clever use of the highly expressed endogenous albumin gene, which can drive the expression of the integrated transgene to potentially therapeutic levels in haemophilia B, where a clotting level of 5% of normal can turn severe haemophilia B into the mild form where spontaneous bleeding is less likely [328]. This approach is only applicable to secreted genes and only liver targeting has been attempted so far. Another innovation combines Cas9 targeting with transposase activity to create RNA-guided transposases [329, 330]. This method does not rely on dsDNA breaks and subsequent repair (the Cas9 used is a catalytically 'dead' or inactive variant, dCas9) but instead relies on the dCas9 conveying a transposase towards a specific site in the genome and allowing transposition near to that location. While promising, the tool has only been used in prokaryotes and suffers from high (up to 50%) off target integrations [329-331].

Table 4.1 | Current methods for genomic insertion of large sequences.

Method	Summary	Advantages	Disadvantages	Selected References
Retroviral insertion	Retroviruses reverse transcribe their RNA genome and insert it randomly into the genome using the viral integrase	- DNA integration is a feature of the virus so no need for auxiliary enzymes/mechanisms - Delivery vector can be pseudotyped	- Random integration with associated chance of insertional mutagenesis - Promoter is required in insert - Limited to ~9kb	[332]
Transposases (PiggyBac, Sleeping Beauty)	A transposase integrates a dsDNA sequence flanked by inverted repeats into a quasi-random site in the genome	- Enzymes of eukaryotic origin - Large insert size possible (up to 196kb reported)	- Random integration with associated chance of insertional mutagenesis - Transposase expressed in trans - Promoter is required in insert	[333, 334]
HITI	The genome and donor sequences are cut with the same endonuclease. cNHEJ repair will ligate the donor into the dsDNA break in the genome	- Targeted insertion - No promoter necessary in insert sequence - Large insert size possible (up to 8kb reported) - Active in all cell cycle stages	- Endonuclease expressed in trans - cNHEJ insertion is not mutation free	[136]
Endonuclease-HDR	HDR mediated repair after a dsDNA break by an endonuclease integrates a dsDNA sequence flanked by large (>100bp) homology arms	- Targeted insertion - No promoter necessary in insert sequence	- Endonuclease expressed in trans - Homology arm required limits insert size - Low efficacy in non-dividing cells - Efficacy lower for larger inserts	[156]

Endonuclease-MMEJ	MMEJ mediated repair after a dsDNA break by an endonuclease integrates a dsDNA sequence flanked by short (>20bp) homology arms	• Targeted insertion • No promoter necessary in insert sequence	• Endonuclease expressed in trans • Low efficacy in non-dividing cells • Efficacy lower for larger inserts • Large mutations around insertion sites	[137]
GeneRide	HDR arms targeting Albumin insert a 2A-peptide and cDNA behind the albumin gene for co-expression	• Targeted insertion • No endonuclease required • No promoter in insert sequence necessary	• Low insertional efficacy (<1%) • Low efficacy in non-dividing cells • Limited scope for non-secreted genes	[327]
RNA-guided transposase	A transposase is coupled with a dCas9 to guide the site of insertion by the transposase	• Targeted insertion • No promoter in insert sequence necessary	• High (up to 50%) off-target insertion • Only proven to work in prokaryotes so far	[329, 330]

cNHEJ, classical non-homologous end-joining; dCas9,'dead' (nuclease inactivated) Cas9; HDR, homology directed repair; HITI, homology-independent targeted integration; MMEJ, microhomology-mediated end-joining;

4.1.2 Aims

This chapter describes studies to investigate genome editing approaches using HITI with a reporter cell line. This chapter also highlights and addresses an unintended side-effect of trying to integrate an entire cDNA into a genome using rAAV2 based constructs; expression from the ITR.

4.2 Results

4.2.1 Reporter cell line construct

While genome editing strategies are ostensibly best confirmed on a genome level, for instance by PCR or sequencing, this process can be cumbersome and costly. At the same time, positive evidence of editing on a genomic level might not always indicate that the genome editing method can attain the desired effect (e.g. transgene expression). To alleviate this, a HITI reporter cell line was designed to allow for a straightforward readout of results of genome editing using HITI, and to also test different permutations of HITI mediated genome editing.

Figure 4.1 shows the system used, which contains a targeting cassette to insert the whole construct into the safe harbour locus in the genome (AAVS1) together with a HITI targeting cassette that expresses mCherry from the CMV promoter (**Figure 4.1A**). The mCherry gene, a red fluorescent protein [335], is flanked by two SaCas9 gRNA target sites derived from the murine *Rosa26* locus (NCBI Gene ID: 14910). These gRNA targets were selected for efficient cleavage in a T7 endonuclease assay (**Figure 4.1B**, protocol **Section 2.10**) and in-silico screening (**Section 2.6.2**) predicted that there were no off-target sites in the human genome with fewer than four mismatches (CRISPOR, [252]). The entire construct was targeted to the first intron of PPP1R12C (the site of the AAVS1 safe harbour locus) using HITI with a single gRNA target site on both sides of the targeting construct. Effective insertion of the construct into the AAVS1 locus of HEK293T cells was determined by clonal growth after puromycin selection (**Figure 4.1C**) that when grown out resulted in >99% of

cells expressing mCherry in cells maintained in puromycin and >96% when puromycin was omitted from the culture medium, indicating that the integration was stable (**Figure 4.1D**). All subsequent experiments were performed in the absence of puromycin selection to prevent potential toxicity during transfection due to increased intracellular puromycin concentrations caused by the transfection reagents. Sanger sequencing of the products of in-and-out PCRs at the insertion junctions (**Section 2.7**) showed seamless insertion into the correct locus (**Figure 4.1E**). The possible permutations of HITI that were tested, and their predicted effects on reporter gene expression in this cell line, are shown in **Figure 4.2**.

4.2.2 5' and 3' integration of mNeonGreen with a transcription blocker or IRES

Firstly, the three permutations that require only a single cut in the genome (**Figure 4.2 A-C**) were evaluated for their feasibility in the HITI reporter cell line. Cells were transfected (**Section 2.6.9**) with plasmids based on the AAV2 genome plasmid (**Table 2.2 and 2.5**) in a molar ratio of 1:2 for SaCas9 (**Figure 4.3A**) and HITI donor (**Figure 4.3B**) respectively. The gRNA target sites in the HITI donor plasmid were the same on either end of mNeonGreen (a green fluorescent reporter protein [336]) and are identical to the SaCas9 target site targeted around mCherry but in reverse-complement orientation. Cells were assessed for mCherry/mNeonGreen expression resulting from each integration permutation (**Figure 4.4**); this was performed 3 weeks post-transfection to minimise residual plasmid causing false-positive expression and thus only capturing expression from integrated DNA (see **section 4.2.5**). Integration of mNeonGreen at the 5'-end of mCherry resulted in expression of mNeonGreen (**Figure 4.4 A and B**). The use of a transcription blocker (TB) prevented mCherry expression (**Figure 4.4A**) whereas an IRES allowed simultaneous expression of both mCherry and mNeonGreen (**Figure 4.4B**). Similarly, targeting the 3'-end of mCherry with an IRES in front of mNeonGreen resulted in effective co-expression of mNeonGreen and mCherry (**Figure 4.4C**).

All three single-cut permutations also resulted in double-negative cells, suggesting that somehow mCherry expression could be abolished and no mNeonGreen inserted (8-14% in permutation A-C, flow cytometry plots of **Figure 4.4 A-C**). One explanation could be that the HITI donor plasmid had not been cut at the 3'-gRNA target site, meaning that the plasmid backbone was inserted in front of mNeonGreen, thereby preventing expression of both mCherry and mNeonGreen. Similarly, if only the backbone had been inserted, which is also present in the cell after the HITI donor is cut, then mCherry would also cease to be expressed. Another explanation is that the loss of mCherry is simply caused by cNHEJ such that a large deletion is created, causing the disruption of the mCherry expression cassette.

Sanger sequencing of single sequences of in-and-out PCR products around the 5'- and 3'- integration junctions of permutation A (**Figure 4.4A**) indicated that seamless insertion occurred in ~90% of cases (18 out of 20) and that any mutations that did occur during integration were no more than 2bp in length (**Figure 4.5**, mutations in red). These results suggest that cNHEJ mediated repair of dsDNA breaks, which is the pathway upon which HITI integration relies, is less error prone than is sometimes suggested [132].

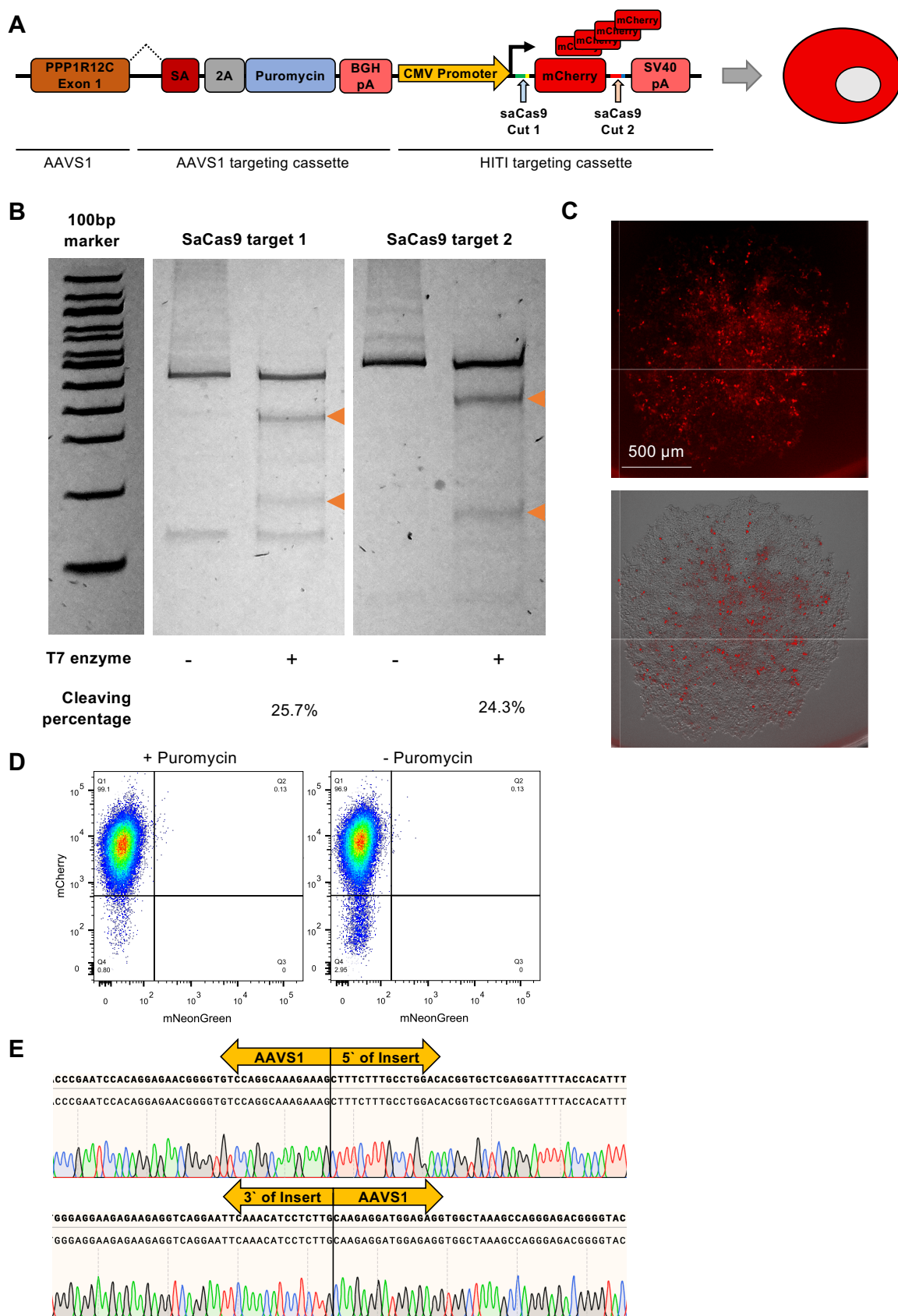


Figure 4.1 | HITI reporter cell line construct and verification. The targeting cassette contains a splice acceptor that will accept a splicing donor from the PPP1R12C gene when the entire construct is inserted into the first intron of PPP1R12C. This in turn will result in expression of the selection marker puromycin, through a 2A cleaving peptide (**A**). The SaCas9 gRNA target sites were derived from the murine Rosa26 locus. Effective targeting by SaCas9 was tested by plasmid transfection into Hepa16 cells (a mouse liver carcinoma cell line; ATCC CRL-1830) and T7 endonuclease I assay on nested PCR products (**B**, cleavage products indicated with orange arrowheads and percentage of cleaved product indicated below, protocol **Section 2.10**). HEK293T cells were transfected (protocol **Section 2.6.9**) with an SaCas9 expression plasmid containing a gRNA to the AAVS1 locus (pGM559, **Table 2.3**) and the targeting plasmid for the HITI Reporter cell line (pGM584, **Table 2.6**). Cells were maintained in HEK293T growth medium containing 1 mg/ml puromycin to select for cells with correct integration. Single cell clones of puromycin selected cells (**C**, mCherry top, mCherry + brightfield bottom) were cultured with or without puromycin for 2 weeks to test stability of integration (**D**, cells grown in 1 mg/ml puromycin left panel, without puromycin right panel). Sanger sequencing was performed on in-and-out PCR of the 5'-and 3'-integration junctions (**Section 2.9**) to confirm seamless integration. SA, splice acceptor; 2A, P2A cleaving peptide; BGH pA, bovine growth hormone poly A, SV40 pA, simian virus 40 poly A.

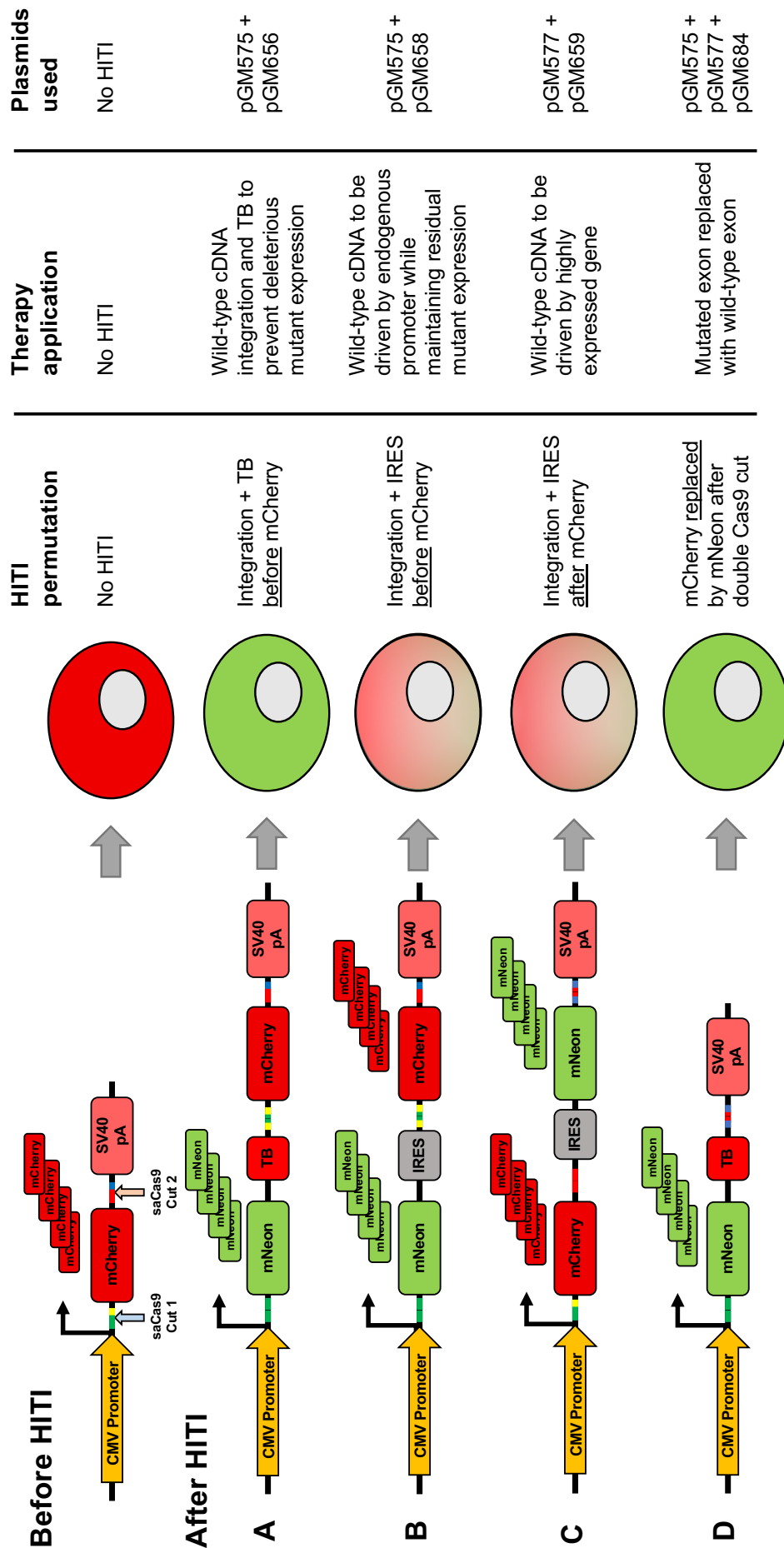


Figure 4.2 | Diagram of possible HITI permutations in the reporter construct. The use of different gRNA targeting sites on the 5' or 3' end of the mCherry sequence and integrating different constructs lead to different genome editing outcomes. The two SaCas9 gRNA target sites are indicated with green and red guide sequences and yellow and blue PAM sequences respectively. Following HITI mediated integration, the gRNA target sites will be in gRNA-gRNA and PAM-PAM orientation as in **Figure 1.6C**. In permutation A the 5' gRNA is targeted and mNeonGreen is inserted. This mNeonGreen will be expressed from the CMV promoter and the transcription blocker 3' to it will prevent mCherry expression, turning correctly targeted cells from red to green. Both permutations B and C aim to induce expression of both mNeonGreen and mCherry simultaneously. They do so by using an internal ribosome entry site (IRES) that allows the translation of two proteins (mCherry and mNeonGreen in this case) from a single mRNA, here driven by CMV. Permutations B and C differ in the position of their target gRNA (5' and 3' of mCherry respectively) which requires the IRES to be on the alternate sides of mNeonGreen to function. In permutation D both gRNA target sites are cut at the same time; and it is anticipated that mCherry will be cut out of the genome and replaced by mNeonGreen carrying a transcription blocker. Each permutation has a theoretical therapeutic application when applied to a genomic target. The plasmids used in this chapter to achieve the desired genome editing permutation is indicated on the far right and can be found in **Table 2.5**. IRES, internal ribosome entry site; mNeon, mNeonGreen; TB, transcription blocker; SV40 pA, simian virus 40 poly A.

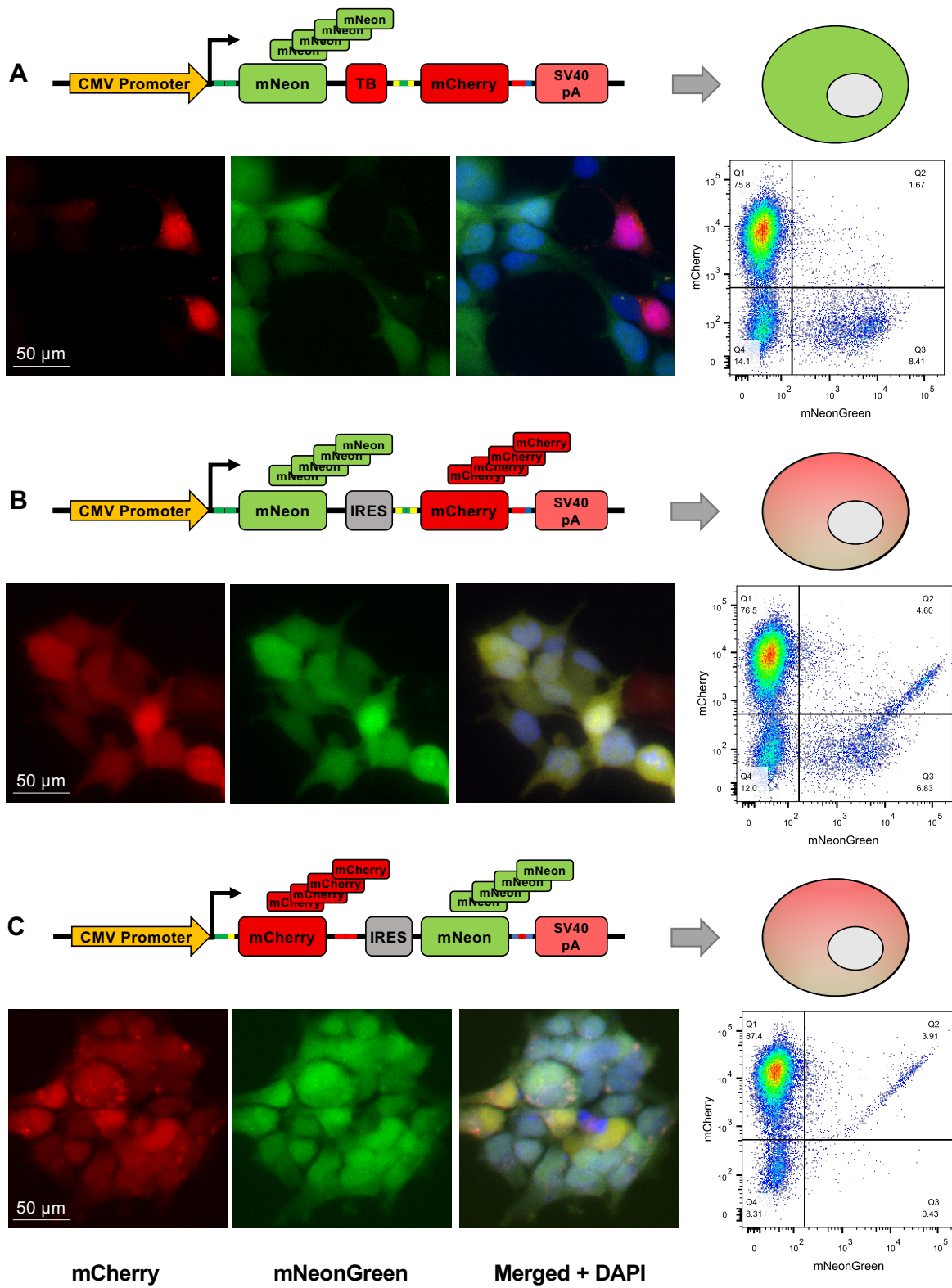


Figure 4.4 | Single cut HITI permutations allow selective mNeonGreen and/or mCherry expression. Fluorescent microscopy images and flow cytometry analysis of HITI permutations A-C (**Figure 4.2**) are shown in **A-C** respectively. Cells were transfected with 1:2 ratio of SaCas9 and HITI donor plasmid respectively (protocol **Section 2.6.9**) and cultured in the absence of puromycin for 3 weeks before imaging and flow cytometry analysis (**Section 2.5**) to confirm loss of the plasmids. In permutation A, transfecting the SaCas9 plasmid targeting the 5' gRNA (pGM575, **Table 2.2**) and corresponding HITI Donor with mNeonGreen and a transcription blocker (pGM656, **Table 2.5**) resulted in loss of mCherry expression due to the transcription blocker in cells that were positive for mNeonGreen expression (**A**, the left panel shows mCherry, middle mNeonGreen and right panel shows a merged image of both with a nuclear stain, DAPI). Inserting mNeonGreen in the gRNA target site 5' of mCherry and followed by an IRES resulted in simultaneous mCherry and mNeonGreen expression (**B**) (using the same SaCas9 targeting plasmid as in A (pGM575) combined with the IRES containing HITI donor (pGM658, **Table 2.5**)). A HITI donor plasmid in which the IRES is placed in front of mNeonGreen (pGM659, **Table 2.5**) combined with a SaCas9 plasmid targeting the gRNA target site on the 3'-end of mCherry (pGM577, **Table 2.2**) likewise resulted in cells positive for both mCherry and mNeonGreen (**C**). The green/yellow and red/blue bars indicate gRNA/PAM sequences in a manner equivalent to **Figure 4.2**.

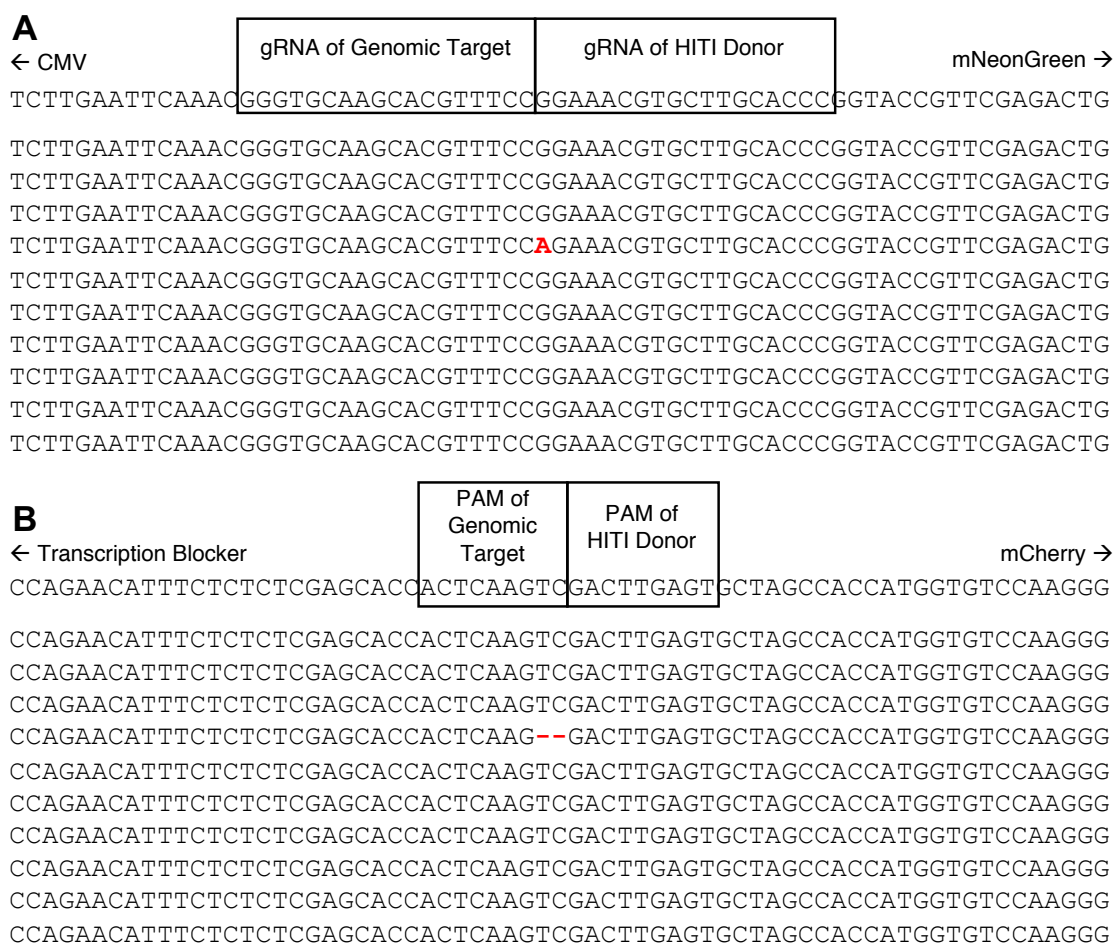


Figure 4.5 | Sanger sequencing of HITI integration junctions shows high integration fidelity. Individual sequences derived from PCR products around the 5' (A) and 3' (B) integration junctions of a HITI permutation A experiment (10 sequences per side, see Figure 4.4A) were subjected to Sanger sequencing and aligned an expected genome sequence after integration. Small (1-2) bp insertions and deletions can be observed (indicated in red in alignments shown) in 10% of clones, meaning integration is seamless in ~90% (18 out of 20) of clones. Protocols for isolation and sequencing of individual sequences from PCR products can be found in Section 2.6.4 and 2.8 respectively.

4.2.3 Replacement of mCherry with mNeonGreen

After evaluating the feasibility of the three HITI permutations requiring only a single cut in the genome, HITI permutation D was also investigated (**Figure 4.2D**). Permutation D relies on the genomic DNA being simultaneously cut at both the 5' and 3' ends of mCherry, so that mCherry is completely removed from the genomic DNA and mNeonGreen can be inserted in its place. A similar plasmid configuration to that used for permutation A (**Figure 4.3B1**) was used, but with the respective 5' and 3' gRNA target sites of the mCherry target cassette on the left and right of mNeonGreen respectively, again in reverse complement orientation. Cells of the HITI reporter line were transfected with this HITI donor combined with two SaCas9 plasmids to target both sides around mCherry, pGM575 and pGM577 (**Table 2.2**) to demonstrate permutation D. Imaging and flow cytometry analysis (**Figure 4.6 A and B**) showed that this approach was capable of replacing mCherry with mNeonGreen. However, the observation of double-negative cells (indicated with white arrowheads in **Figure 4.6A**) revealed that this happens in conjunction with complete loss of mCherry expression without mNeonGreen expressing in its place. To investigate the genome configurations produced, PCR was performed spanning the gRNA target sites which suggested three permutations were commonly obtained: 1. integration of mNeonGreen (1183bp band), 2. no change (894bp band) and 3. deletion of mCherry (133bp band) (**Figure 4.6C**). Sanger sequencing of the 5'-end of sequences of the 1183bp band indicated correct replacement and again showed that ~90% of clones displayed seamless integration (**Figure 4.6D**).

Figure 4.6 | Replacement of mCherry with mNeonGreen results in mNeonGreen expression without mCherry expression. Cells of the HITI reporter cell line were transfected with both SaCas9 plasmids targeting the gRNA sites on both ends of mCherry (pGM575 and pGM577, **Table 2.2**) and a HITI donor with mNeonGreen and TB between the two gRNA target sites corresponding to those targeted around mCherry (pGM684, **Table 2.5**) in a 1:1:4 ratio (permutation D, **Figure 4.2D**). Cells were cultured in the absence of puromycin for 3 weeks before imaging and flow cytometry analysis. Fluorescent microscopy images (**A**) show targeting where cells expressing mNeonGreen have lost mCherry expression as intended (indicated channels for left and middle panel, merged channels plus nuclear DAPI stain, right panel). White arrowheads indicate cells that have lost expression of mCherry without the integration of mNeonGreen. Flow cytometry analysis of cells not transfected (Control, left), cells transfected with both SaCas9 plasmids but no donor (middle panel) and cells transfected with both SaCas9 plasmids and the HITI donor (right panel) (**B**, protocol **Section 2.5**). This shows that mCherry deletion is common with ~23% of double-negative cells (neither expressing mCherry nor mNeonGreen) observed in cells transfected with both SaCas9 plasmids versus 1% in controls. Cells that also received the donor sequence also contain a population of mNeonGreen-only cells (right panel), corroborating the fluorescent microscopy images in **A**. PCR amplification outside of the gRNA target sites confirmed all three predicted genome arrangements (deletion of mCherry, 133bp; no change, 894bp; insertion of mNeonGreen, 1183bp) after HITI targeting (**C**). Sanger sequencing of the 5'-integration junction after correct mNeonGreen integration shows similar integration fidelity (**D**, ~90% seamless 1 out of 10, red dash indicating a deletion) compared with single cut HITI permutations A-C (**Figure 4.5**). The green/yellow and red/blue bars indicate gRNA/PAM sequences in a manner equivalent to **Figure 4.2**.

4.2.4 Development of a HITI reporter mouse line

Experiments using the reporter cell line above confirmed that all four of the HITI permutations were feasible, as assessed by imaging, flow cytometry and sequencing. As a next step, the aim was to combine these HITI permutations with *in vivo* delivery methods through the derivation of a mouse model with a reporter construct equivalent to the HITI reporter cell line. A construct was designed (**Figure 4.7A**) in which a ubiquitous promoter (elongation factor 1 alpha, Ef1 α) could express the reporter gene tdTomato, a red fluorescent protein with high stability and bright fluorescent signal, *in vivo* [335]. The tdTomato cDNA is flanked by gRNA target sites (**Figure 4.7B**) for both SaCas9 and SpCas9 on both ends of the cDNA to allow potential genome editing with both Cas9 orthologs. Homology arms targeting the murine safe harbour locus *Rosa26* flank the sequence to allow Cas9-mediated HDR integration in the targeted locus. After the construct was designed it was manufactured using a combination of gene synthesis and molecular cloning in collaboration with the WIMM Genome Engineering facility (<https://www.imm.ox.ac.uk/research/facilities/genome-engineering-facility>). The subsequent steps in the derivation of this mouse model were performed by the WIMM Genome Engineering and Transgenics Core Facilities (<https://www.imm.ox.ac.uk/research/facilities/mouse-transgenic-core-facility>). The project was started in February 2018 and at time of thesis preparation had progressed to the second stage where correct integration was determined before blastocyst injection (**Figure 4.7C**). As a result of slower than anticipated progress, the mouse model was not available to be used in experiments before the completion of this study.

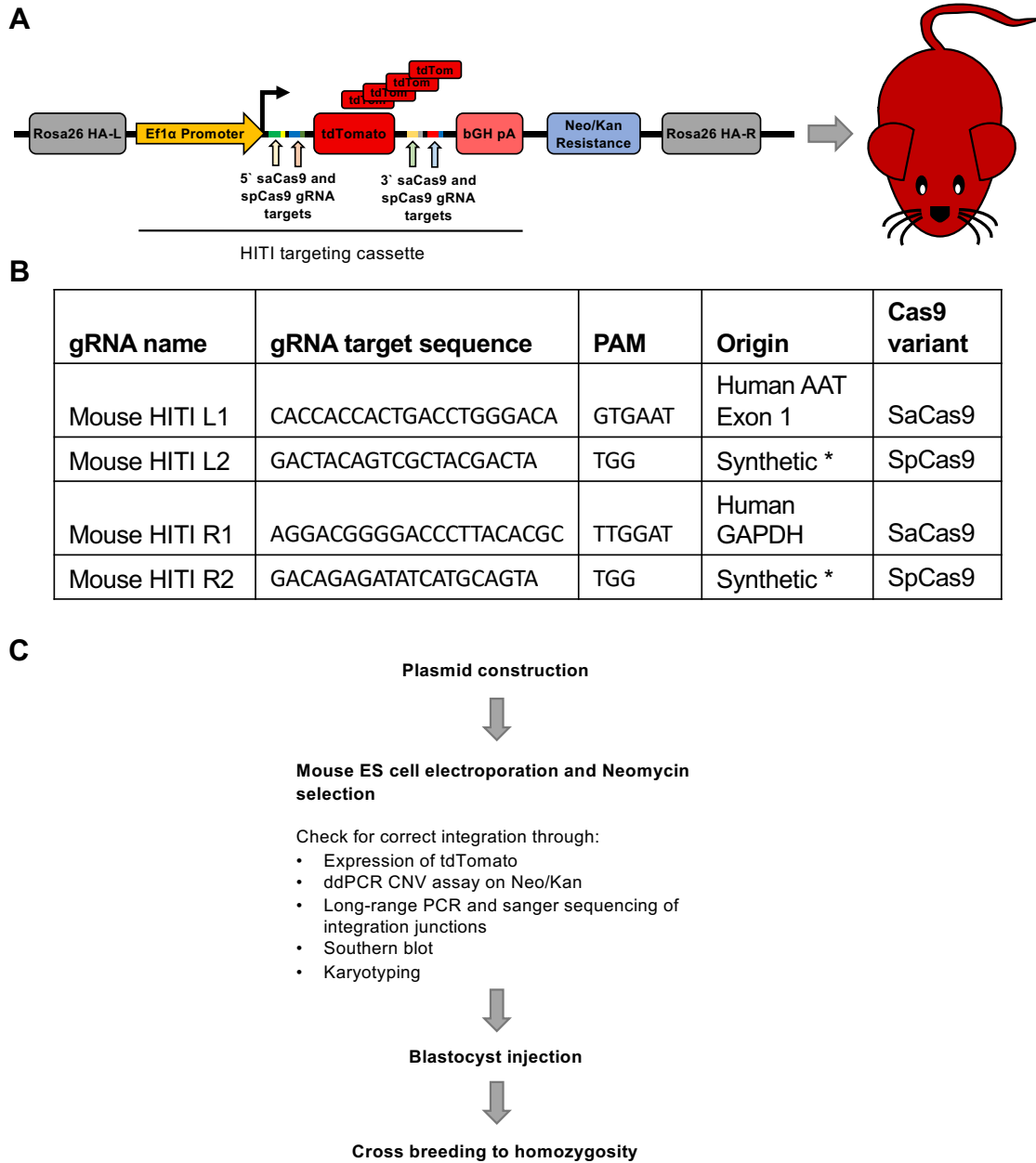


Figure 4.7 | Design of HITI reporter mouse for in vivo testing of genome editing permutations. A schematic of the Rosa26 targeted mouse HITI reporter construct is shown in **A**. The gRNA target sites for both SpCas9 and SaCas9 on both sides of tdTomato are shown in **B**. Following plasmid construction, the major steps in the derivation of the mouse line are outlined. Rosa16 HA-L and HA-R, murine Rosa26 specific homology arms for left and right side respectively; EF1α, elongation factor 1 alpha; BGH pA, bovine grown hormone poly A; Neo/Kan resistance, expression cassette conveying resistance to the antibiotics neomycin and kanamycin. PAM, protospacer adjacent motif; CNV, copy number variation; ddPCR, digital droplet PCR. * obtained from [337].

4.2.5 Strategies to prevent AAV2 ITR expression

One observation that was made early in the HITI cell line experiments (**Section 4.2.2**) was that transfection of HITI donor plasmid alone resulted in significant expression of the encoded EGFP or mNeonGreen fluorescent transgene (**Figure 4.8A**). Interestingly, expression was notably lower when the HITI donor was co-transfected with Cas9 targeting gRNA target sites left and right of the cDNA (**Figure 4.8A and B**). Coincidentally, this also confirmed that Cas9 is capable of cutting the HITI donor. The implication is that the dsDNA sequence remaining after cutting does not exert transcriptional activity itself, which leaves the AAV2 ITR sequences as likely to be responsible for the strong expression. The bacterial backbone present is unlikely to be the cause of the unwanted expression as it is doubtful that transcription could proceed through the secondary structure of the AAV ITRs flanking the transgene. Crucially, however, the ITR of AAV2 has been shown previously to act as a promoter in viral vector context [338].

Experiments in cell culture, as described above, are not negatively affected by the ITR expression since the plasmid, and accompanying expression, should be diluted out by cell division, while integrated DNA, will faithfully pass to daughter cells and over time become the only source of transgene expression. *In vivo* however, high levels of cell turnover are not expected, as evidenced by long-term Luciferase expression in mice (over one year, internal observation, data not shown) and factor IX in humans [262] from un-integrated AAV expression transgene cassettes. Thus, transgene expression *in vivo* via the ITR promoter, rather than the integrated sequence, could lead to false positive results and confound interpretation of HITI studies. To remedy this, strategies were explored to restrict expression from the ITR.

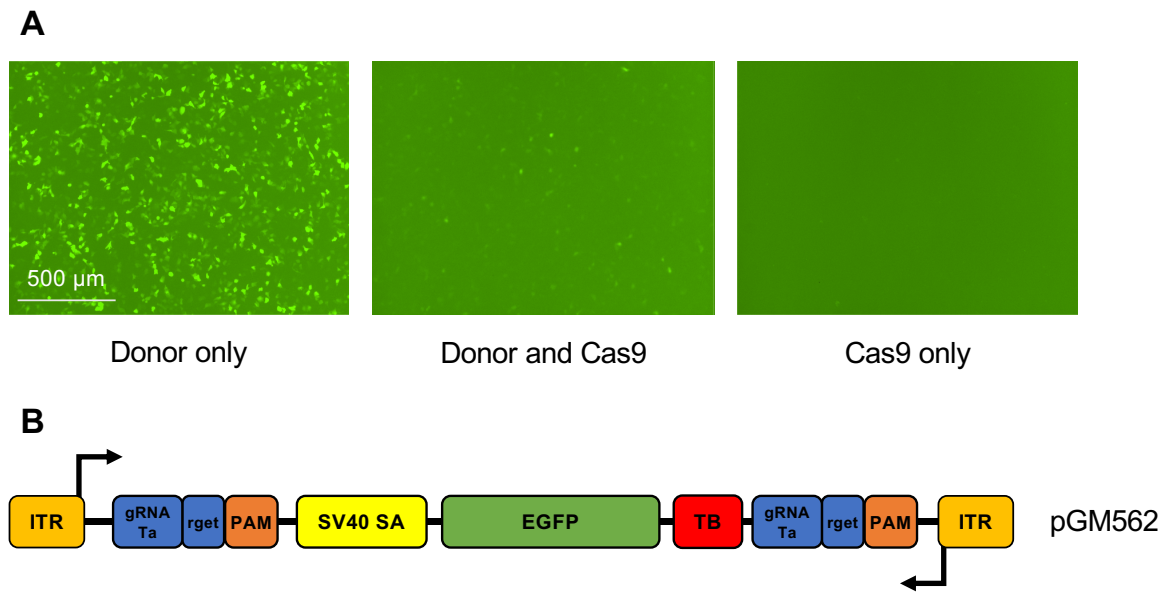


Figure 4.8 | Expression of the fluorescent reporter gene in HITI Donor plasmid. Huh7 cells (400.000 in 6-well plate) were transfected with 1.25 μ g of each plasmid and imaged 2 days later. **(A)** Strong fluorescent reporter gene expression was observed after transfection of a HITI donor (pGM565 (with AAT Exon 1 gRNA targets) is based on HITI plasmid pGM561, structure in **B**, **Table 2.5** and **2.4** respectively) plasmid only, compared with marked lower fluorescence when co-transfected with SaCas9 targeting both gRNA target sites around the reporter gene (pGM551, **Table 2.3**). ITR, inverted terminal repeat; PAM, protospacer adjacent motif; SV40 SA, splice acceptor derived from simian virus 40; EGFP, enhanced green fluorescent protein; TB, transcription blocker.

4.2.5.1 Dual transcription blocker

The first strategy explored to limit ITR-dependent transgene expression was the use of a transcription blocker (TB), specifically, a small (154 bp) sequence that contains a synthetic poly A signal from the SV40 polyoma virus and an RNA polymerase II transcriptional pause signal from the human haemoglobin subunit alpha 2 gene. This sequence was adapted from the pGL1 plasmid (Promega) and is reported to strongly inhibit background transcription [339]. The sequence was inserted between the left ITR and the 5' gRNA target site of pGM562 generating a new plasmid (pGM587, Dual TB). In this new configuration, the newly added TB sequence should limit ITR transcriptional activity, without being integrated into the genome in HITI experiments (**Figure 4.9A**). There are two ITR sequences present in the AAV2 genome used (pZac2.1), both with promoter activity. It is not expected that the right-end ITR should lead to measurable EGFP/mNeonGreen expression, as transcription proceeding 5' to 3' should not encounter an ORF for the reporter transgenes. Representative microscopy images in **Figure 4.9B** show that the transcription blocker in pGM587 reduces ITR-driven transgene expression as seen by the lower fluorescence signal in HEK293T cells transfected with the Dual TB plasmid (pGM587) compared with the control plasmid (pGM562). However, expression is not completely abolished, so further modifications were warranted.

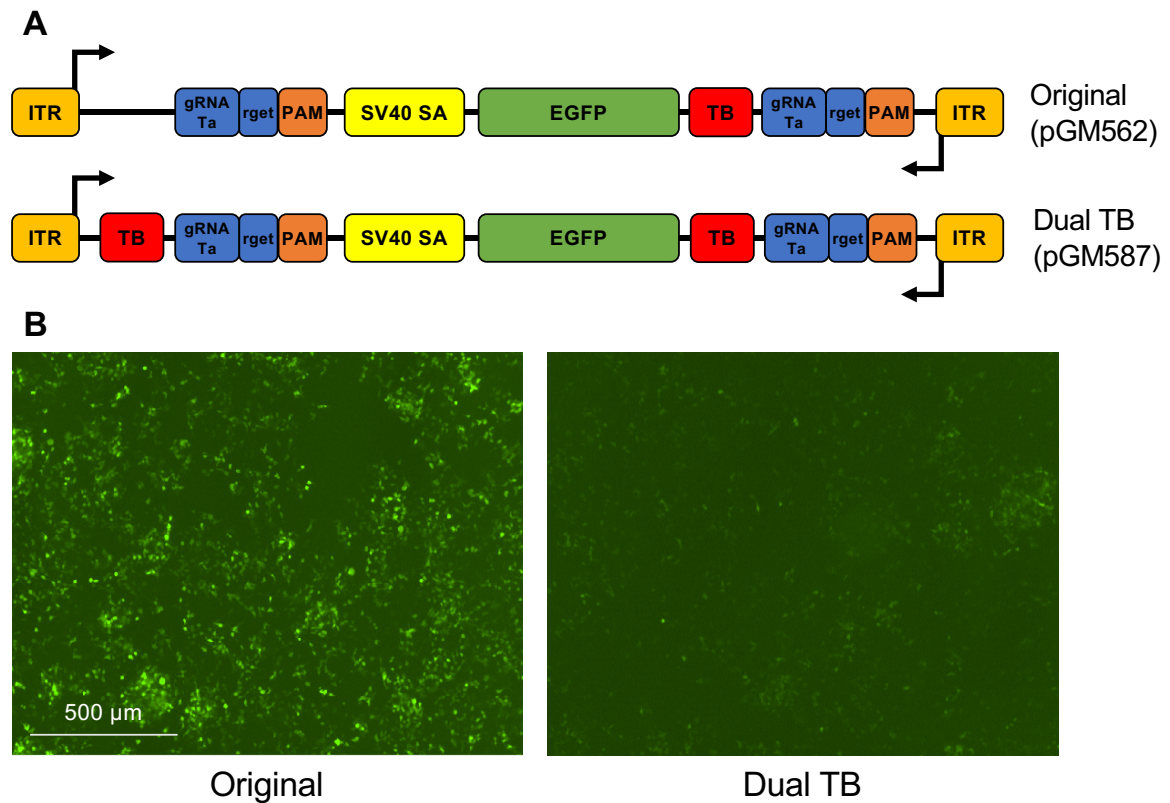


Figure 4.9 | Dual transcription blocker in HITI donor constructs to reduce ITR transgene expression. A transcription blocker (TB; red) was introduced between the left AAV2 ITR and the 5'-gRNA target site acceptor (**A**). HEK293T cells (500,000 in 6-well plate) were transfected with 1.25μg of the original and dual TB HITI donor plasmids (pGM562 and pGM587 respectively) and imaged after 2 days. (**B**) Fluorescence from the EGFP transgene was lower, but not completely absent. ITR, inverted terminal repeat; PAM, protospacer adjacent motif; SV40 SA, splice acceptor derived from simian virus 40; EGFP, enhanced green fluorescent protein; TB, transcription blocker.

4.2.5.2 HITI SpliceOut

A second strategy investigated was to place the entire sequence for insertion (including the gRNA target sites and transgene) into a synthetic intron (**Figure 4.10**). In this mechanism, termed SpliceOut, the mNeonGreen transgene should be spliced out when pre-mRNA is transcribed from the active ITR promoter. In this way, the mature mRNA that is normally translated does not contain mNeonGreen. At the same time, the synthetic intron sequence will not be integrated into the target site in the genome because the gRNA target sites are located within the intron. Note that this configuration precludes the use of a splice acceptor or synthetic intron in the insert sequence as this would interfere with the SpliceOut mechanism. Different SpliceOut plasmids were generated including two that incorporated the dual transcription blocker (Dual TB; **Section 4.2.5.1**) and two with a U6 gRNA expression cassette between the ITR and insert (**Figure 4.11A**). The latter strategy was based on the hypothesis that RNA polymerase III activity from the U6 promoter might interfere with the RNA polymerase II machinery of the ITR promoter. When comparing the fluorescence intensity generated by each plasmid configuration with and without the SpliceOut intron, it was clear that the SpliceOut mechanism did not have a significant influence on ITR-dependent transgene expression (**Figure 4.11B**). Similarly, the U6 promoter did not affect expression from the ITR, therefore, of the configurations evaluated, the Dual TB configuration was best at reducing unwanted ITR expression under these circumstances.

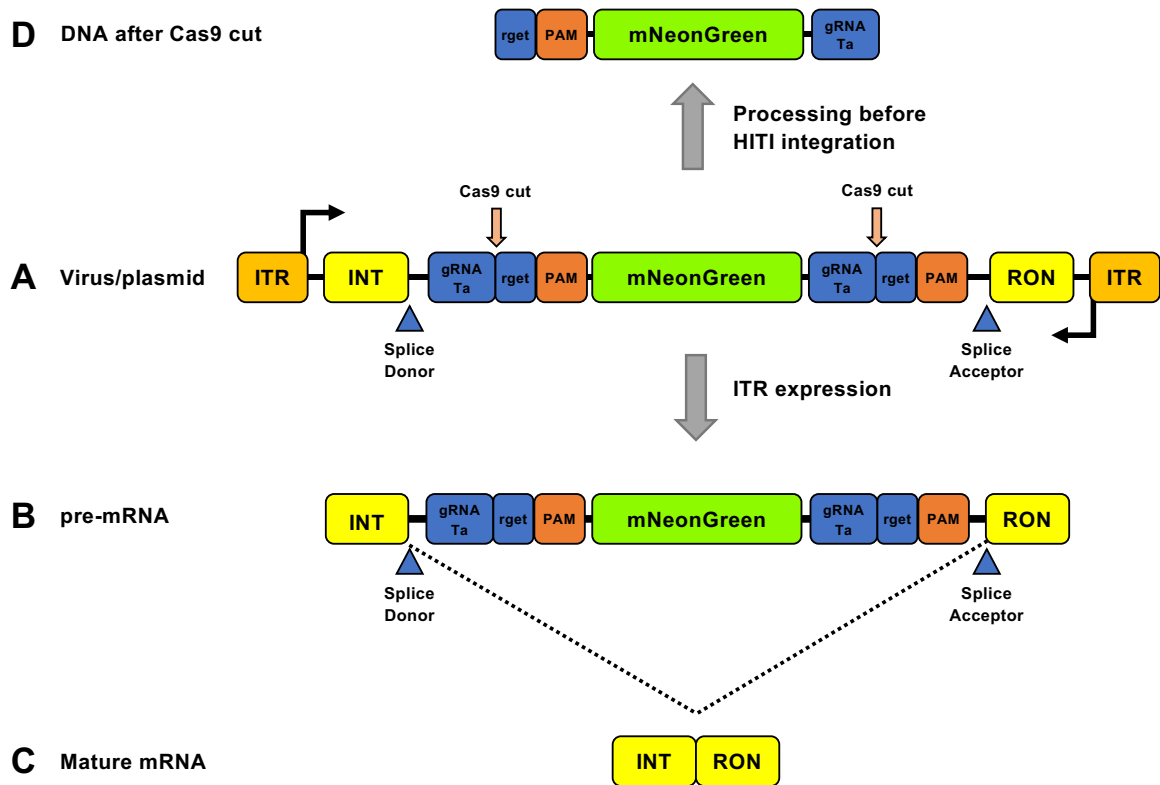
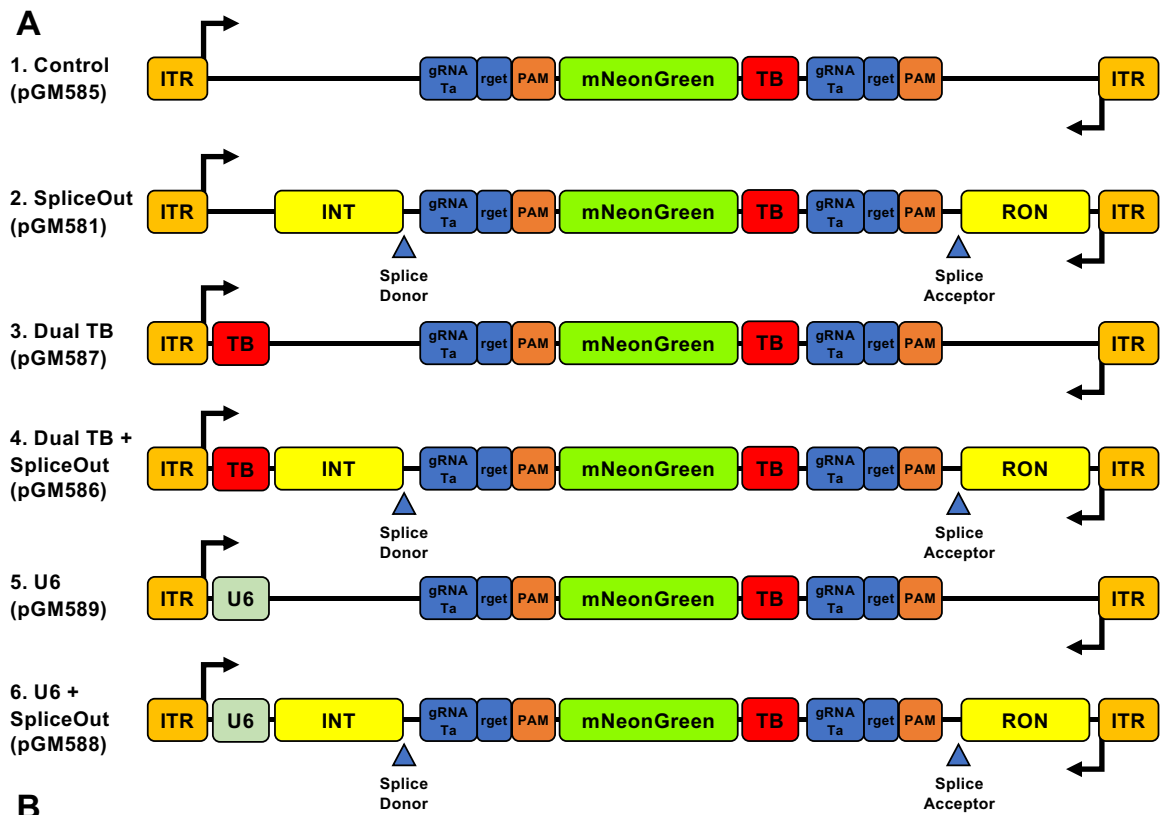


Figure 4.10 | Schematic showing the SpliceOut mechanism. The HITI SpliceOut mechanism was devised as an attempt to prevent transgene expression from HITI donor plasmids containing the AAV2 ITR. A synthetic intron was constructed by using the SV40 small *t* antigen intron. The intron was cut in two parts, the splice acceptor part (INT) and the splice donor part (RON), taking care to keep the branch point sequence in the donor part. (A) After transfection of the SpliceOut HITI donor plasmid (pGM581, **Table 2.4**), any transcription from the 5' ITR should result in a pre-mRNA that encompasses the mNeonGreen transgene in a synthetic intron (B) which should be spliced out and thus not be translated into protein (C). The dsDNA of the donor plasmid or virus will be cut at the gRNA target sites that are placed inside the synthetic intron and as such the intron sequences should not be integrated into the target genome site (D). ITR, inverted terminal repeat; PAM, protospacer adjacent motif; INT, splice donor derived from SV40; RON, splice acceptor derived from SV40.



B

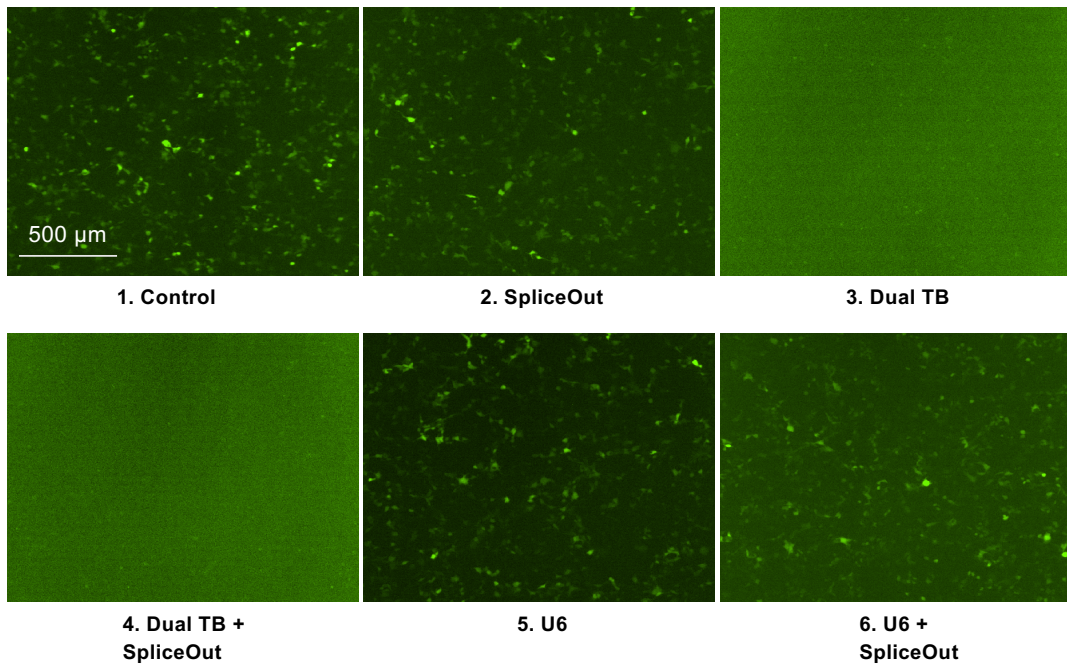


Figure 4.11 | Configuration and expression of SpliceOut plasmids. Different configurations of potential SpliceOut HITI donor plasmids were generated **(A)** based on the original (Control), the transcription blocker (Dual TB) and U6 promoter (U6) plasmids, with and without the synthetic SpliceOut intron, to allow comparison of the effect of the intron on expression within each plasmid pair. Huh-7 (400.000 cells per well) were transfected with 1.25 µg of plasmid and imaged 2 days later. **(B)** Fluorescence microscopy images show that there is no effect of the SpliceOut mechanism on transgene expression. All plasmids can be found in **Table 2.4**. ITR, inverted terminal repeat; PAM, protospacer adjacent motif; TB, transcription blocker; INT, splice donor derived from simian virus 40; RON, splice acceptor derived from simian virus 40; U6, U6 promoter with gRNA expression construct of SaCas9.

4.2.5.3 pX601 ITR

When comparing ITR expression between two different cell lines, HEK293T (human embryonic kidney cells) and Huh-7 (human liver carcinoma cells), it was observed that there was more residual ITR expression in Huh-7 cells, even when using the Dual TB configuration (**Figure 4.12A**). This is consistent with the recent discovery that the 3' UTR of wild-type AAV2 contains a transcription factor binding site for the liver specific transcription factor HNF1A [52]. Upon closer inspection, it transpired that both ITR regions of the pZac2.1 genome plasmid contained this bidirectional binding site, whereas the SaCas9 expression plasmid pX601 does not (**Figure 4.3A** for pX601 plasmid schematic and **4.12B**). The HNF1A binding site is within the region annotated as ITR in pZac2.1, but this plasmid map is most likely mis-annotated (cross-referenced with the wild-type AAV2 ITR sequences) and the HNF1A site is more likely located within the 5' and 3' UTR regions of transcripts originating from the ITR sequences (**Figure 4.12B**, actual ITR region is underlined); this is in line with the aforementioned report [52]. Moving the HITI donor sequences from the pZac2.1 plasmid backbone into the pX601 backbone had a modest but noticeable effect on ITR expression, but again does not completely abolish it (**Figure 4.12C**). While comparison with the pZac2.1 Dual TB is difficult as expression is already low, owing to the TB, replacing the backbone to the one from pX601 combined with the Dual TB configuration appeared to rid expression altogether (**Figure 4.12C**).

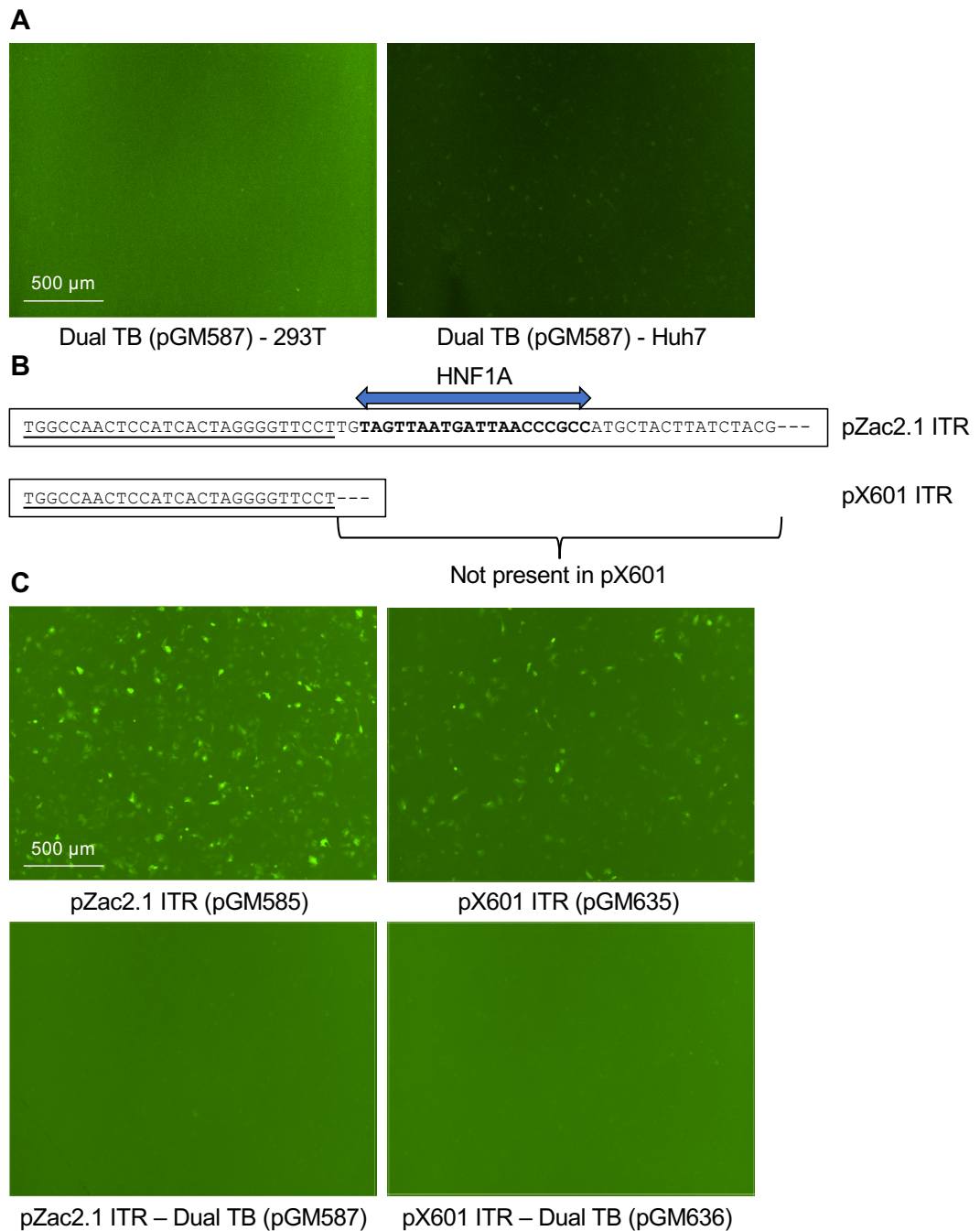


Figure 4.12 | HNF1a transcription factor enhances ITR transcription activity. HEK293T and Huh-7 cells were transfected with 1.25 μg plasmid per 500.000 cells and imaged 2 days later. Transgene expression driven from the ITR appeared higher in Huh-7 cells than HEK293T cells (**A**). (**B**) The HNF1A transcription factor binding site is present in the annotated ITR sequence of plasmid pZac2.1 (binding site in bold; --- indicates the remainder of the viral genome). The actual ITR sequence (as defined by alignment to wild-type AAV2) is highlighted by underlining. (**C**) Replacing the AAV2 genome backbone from pZac2.1 with pX601 reduced the ITR-driven transgene expression.

4.3 Discussion

4.3.1 Feasibility of different permutations of HITI permit different therapeutic applications

The experiments described in this chapter were designed to query the utility of HITI for different therapeutic applications depending on the sequence inserted and the insertion location relative to a target gene (**Figure 4.2**). A fluorescent reporter cell line was created to allow straightforward analysis of integration potential of each HITI permutation at the molecular level monitored by transgene expression (**Figure 4.1**). All four permutations turned out to be feasible whether using a ‘single-cut-and-integration’ (**Figure 4.3**) or a double cut ‘excision-and-replace’ approach (**Figure 4.5**). When examining the site of NHEJ repair – the integration junctions – more than 90% of sequences derived from unique integration events examined were devoid of any detectable mutations, with all observed indel mutations being 3bp or smaller (**Figure 4.5 and 4.6D**). There is a tendency to categorise cNHEJ as an error-prone repair mechanism [132], probably due to it being used in cNHEJ knock-out experiments where Cas9 will continue to cut and re-cut until an indel is produced, but the results produced here are in line with published reports [136]. The small number and, importantly, the small size of these mutations are unlikely to negatively influence projected HITI genome editing outcomes. However, it is important to keep in mind that the method used to test for mutation formation in this study is limited to small mutations due to the PCR amplification. This means that larger deletions, which have been reported in cNHEJ knock-out [143] and Cas9 nickase [144] experimental approaches, are unlikely to be detected. It has been observed that large deletions are associated with microhomologies between repaired ends [144]. While the gRNA-gRNA and PAM-PAM sequences at the integration junction might look like microhomologies, their palindromic nature means they are unlikely to be a target for MMEJ-mediated repair.

4.3.2 HITI reporter mouse model combines confirmation of HITI permutations with *in vivo* targeting for translational purposes

With the different HITI permutation established, focus shifted towards combining the different HITI approaches with *in vivo* delivery methods, which is important to for assessing the potential for future clinical translation. One of the most important aspects of making the translational leap is the need for efficient delivery, something the more established gene therapy field has grappled with for decades [340]. Many genome editing groups have adopted delivery platforms that have been assessed in gene therapy clinical trials or have made their way to the clinic, providing the *in vivo* genome editing field with a jump start [341]. As expected however, similar limitations are being experienced, such as restricted vector size, which can require resorting to a dual vector approach with each of the vectors expressing one half of the Cas9 protein [312] or each of the ZFN halves [341]. In many cases, including HITI, this also requires a third vector carrying the donor DNA, further reducing the probability that all components necessary to induce genome editing are delivered to the same cell at the same time. The mouse line described in **Section 4.2.4** would be well suited to evaluate delivery methods, for instance different ratios of endonuclease and donor. Due to the ubiquitous tdTomato expression in the HITI reporter mouse line, it will allow not only for assessing the targeting and integration efficacy of an organ or interest, but also the detection of ectopic insertion. If this happens it implies that the vector has both delivered the donor and endonuclease to that tissue but also that the endonuclease was able to express and induce a dsDNA break and cNHEJ repair.

4.3.3 The AAV2 ITR is a potent promoter

Along with the aim to integrate a complete transgene cDNA into the genome comes an unintended side effect, transgene expression from the delivery vector. This was observed in early *in vitro* HITI experiments in which an AAV2 genome plasmid was transfected into cells (**Figure 4.8**). With expression of the transgene being the eventual aim of HITI as a therapeutic modality, this unintentional expression can be observed as a false positive. To

remedy this, several approaches were pursued in an effort to prevent this. These include the use of a transcription blocker (**Figure 4.9**), placing the transgene inside an intron (**Figure 4.10 and 4.11**) and modifying the AAV2 genome backbone such that it no longer contains a liver-specific transcription factor binding site (**Figure 4.12**). The combination of a TB and the AAV2 backbone from the pX601 plasmid proved to be most efficacious in limiting ITR expression (**Figure 4.12C**).

A TB will not prevent transcription from the ITR in itself, but instead prevents the transgene that follows it to be incorporated transcripts originating before the TB. Any inadvertent transcription start sites after the TB that are still influenced by ITR promoter activity could still lead to expression. *Ex vivo* approaches have the advantage of allowing expansion before insertion, diluting out any delivery vector-based expression even when a promoter is present in the insert [156], a luxury not shared by *in vivo* genome editing. ITR expression is not something that is often brought up because it is not usually an issue. For instance, when HDR is used to repair a point mutation or to insert part of a cDNA, there will not be a start codon or Kozak sequence present to start translation of any transcript that may be produced. Furthermore, when using HDR, the long homology arms might prevent transcription reaching the insert.

Even though the described *in vitro* experiments still display some residual expression (**Figure 4.12**) it was decided that further optimisation of transgene expression repression was not warranted. First, even plasmids in which the AAV2 genome plasmid backbone was replaced with a promoter-less non-viral plasmid backbone, i.e. no ITR should be present to drive expression, some residual expression comparable with the pX601 Dual TB could be observed (data not shown). This indicates that further optimisation might not be realistic or even measurable using the methods used. In addition, it is important to acknowledge the differences between *in vitro* plasmid transfection, in which many thousands of plasmids are

transfected per cell, and *in vivo* AAV mediated transduction, where the mean copy number per cell is between one and three (**Figure 3.5**).

4.3.4 Conclusion

The work in this chapter set out to test different permutations of HITI by means of a HITI reporter cell line that allows quick readout of feasibility of each permutation. By successfully confirming each permutation the therapeutic scope of HITI is greatly expanded. While ITR-driven expression has the potential to lead to false-positive results, the combined use of a transcription blocker and ITR sequences devoid of liver-specific transcription factor binding sites is capable of significantly reducing this.

One HITI permutation in particular, permutation A, was taken forward for testing of therapeutic feasibility in an applicable disease, alpha-1 antitrypsin deficiency.

Chapter 5 – Investigation into HITI as a treatment for AAT deficiency

5.1 Introduction

5.1.1 Genome editing for alpha-1 antitrypsin deficiency

As discussed in **Section 1.3**, alpha-1 antitrypsin (AAT) is a highly expressed secreted protein, expressed mainly from the liver and to a lesser extent lung-resident monocytes and lung epithelial cells [189-191]. The many mutations in the SERPINA1 gene that encodes the AAT protein can cause AAT deficiency (AATD), either by creating a dysfunctional protein, a protein that cannot be secreted, or resulting in no protein at all (**Table 1.7 and 1.8**). The most common mutation that causes AATD, named PI*Z, is highly prevalent in northern Europe (**Figure 1.8**) and causes deficiency by polymerisation and accumulation in hepatocytes [196], resulting in insufficient functional protein in the blood stream (~15% of normal). This polymerisation also leads to ER stress in the hepatocytes and is the most likely cause for increased liver disease (liver cirrhosis) in PI*Z affected patients (**Table 1.8**) [197]. While there are potential strategies to overcome deficiency of AAT protein by addition of functional AAT directly into the bloodstream via enzyme replacement therapy (ERT) or by expression from a delivered copy of the gene (gene addition; **Section 1.1**) these strategies will not be able to address the liver phenotype.

A therapy that can alleviate both the lung and liver conditions of PI*Z patients must achieve therapeutic levels of wild-type (PI*M) protein whilst simultaneously preventing the production of the toxic PI*Z variant. One approach is to combine an rAAV mediated gene addition approach to deliver a PI*M expression cassette to the liver that also includes a miRNA expression construct that targets the PI*Z mRNA [239]. In pre-clinical experiments in mice harbouring several copies of the human PI*Z locus (termed PI*Z knock-in (PI*Z KI) mice, also used in this study, see **Chapter 6**), Mueller et al. [239] used rAAV2/9 to deliver both the PI*M cDNA, driven by a hybrid chicken β -actin promoter, and triplets of miRNAs

against three separate sites on the AAT mRNA, placed within the intron and/or poly-A region of the PI*M expression cassette. This resulted in around 80% knock-down of PI*Z mRNA whilst achieving supra-physiological levels of PI*M AAT. Several silent mutations were introduced in the PI*M cDNA to prevent the miRNAs targeting PI*Z from interfering with PI*M expression. While promising, this approach suffers from dilution of vector genome in growing livers of young children (**Section 3.1.1**). Integrating the expression cassette into the cellular genome has been attempted with *in vivo* retroviral delivery to the liver, but failed to generate therapeutic levels of AAT in the serum [223, 224]. Lentiviral vectors seem not to have been picked up for liver directed AATD therapy so far although preliminary evidence in haemophilia suggests it could be a possibility [342].

Genome editing has the capability of life-long expression because genomic alterations will be carried into daughter cells after cell division and, uniquely, this approach has the ability to 'revert' a patient's genome back to wild-type by correcting the genetic mutation directly. In the case of PI*Z AATD, this would correct both the lung and liver phenotype at the same time, which has made AATD a prime candidate for genome editing [343]. Early *in vitro* experiments proved the feasibility of correction of the PI*Z mutation using HDR [344, 345]. Using the GeneRide technology [327] delivered by rAAV2/8 to 1-day-old PI*Z KI mice, Borel et al., [346] integrated the miRNA cassettes targeting PI*Z (and de-targeted from PI*M AAT, same constructs as in [239]) 3' to the mouse *Albumin* locus to both express PI*M and prevent PI*Z expression. Interestingly, while only 0.2% of cells were effectively targeted at 4 months post-treatment, after 8 months this increased to 7.2%, indicating that hepatocytes in which PI*Z had been silenced by the miRNAs possessed a growth advantage over hepatocytes still expressing the toxic PI*Z protein. Another report targeted PI*Z more directly by disrupting the AAT locus through cNHEJ [240]. The rAd type 5 vector expressing SpCas9 and a gRNA against human SERPINA1 was injected IV into 9-week-old PI*Z KI mice and disruption of PI*Z measured. Immunohistochemistry against PI*Z found a significant reduction in cells expressing PI*Z correlated with decreased circulating AST and

ALT concentrations, which are indicative of liver damage [240]. Importantly however, this approach does not alleviate deficiency of AAT and thus would not positively impact the lung phenotype. In fact, given that PI*Z patients still secrete 15% of AAT it could be counterproductive. Another report used rAd to deliver SpCas9 to the liver and used a second rAd to also deliver a homology donor targeting the mouse *Rosa26* locus to integrate a PI*M AAT expression cassette driven by the elongation factor 1 alpha (EF1 α) promoter [347]. Stable integration was achieved with long-term (200 days) PI*M expression in the serum of normal C57BL/6J mice. However, long-term AAT expression was also found in mice injected with the donor rAd only (no rAd SpCas9), indicating that the rAd genome itself could contribute to the observed AAT in the serum [347].

Two back-to-back publications aimed to address both the deficiency and PI*Z toxicity by targeted correction of the PI*Z mutation [241, 242]. Shen et al., [241] used two rAAV2/8 vectors to deliver the SaCas9 expression construct driven by the (liver-specific) thyroxine binding globulin promoter and an HDR donor containing the gRNA expression cassette and homology donor sequence. Analysis by deep RNA sequencing revealed that around 4% of AAT mRNA transcripts contained the intended correction. Conversely, most of the SaCas9 editing events resulted in cNHEJ induced indels rather than HDR mediated correction, with a concomitant reduction in PI*Z expression [241]. The other report by Song et al., used rAAV2/9 to deliver SaCas9, this time expressed through the mouse U1a snRNA promoter, and rAAV2/8 to deliver the gRNA and homology donor sequence [242]. In 1-week-old and 8-week-old PI*Z KI mice, the number of cells positive for corrected AAT was 14% and 18% respectively. The resulting PI*M AAT serum expression was reported to be 8-12% of the therapeutic target. Again, the majority of edits were cNHEJ induced indels [241, 242].

A more radical approach, that relies on targeted editing of RNA rather than DNA, showed correction of the PI*Z mutation in mRNA transcripts by mutating the AAG codon to IAG (I for inosine) *in vitro* [348]. Inosine is decoded as guanosine, so the codon is effectively

changed from AAG (lysine) to GAG (glutamine), correcting the mutation. It would be interesting to see how this *in vitro* proof-of-concept translates to *in vivo* models, perhaps rather than an antisense oligonucleotide [348], using Cas proteins to target the mutation site in the mRNA [103, 104].

5.1.2 Liver organoids

Organoids are three-dimensional *in vitro* models of tissue or organs derived from induced pluripotent stem cells, pluripotent embryonic stem cells, or the adult stem cells found in primary tissue [349]. They aim to replicate the physiology of the tissue they represent whilst keeping the genetic profile of the organism from which they are derived. These characteristics make them ideal for use in multiple applications including disease modelling, high-throughput screening and development of small molecule drugs (excellently reviewed in [349] and [350]). Organoids also have potential for gene therapy and genome editing. For instance, gut organoids from CF patients were used to assess genome editing of CFTR reported in a CFTR-dependent ‘swelling-assay’ [351], and were also used to confirm expression of CFTR in a lentiviral gene addition protocol [352]. Similarly, human retinal organoids have been used for AAV vector tropism assessment [353].

Liver organoids can be derived from human induced pluripotent stem cells [354, 355] and primary liver stem cells from humans or mice [251, 356-358] and can serve as a stepping-stone between *in vitro* cell based assays and *in vivo* experiments. For instance, they have been used as a validation for a base editing approach to treat phenylketonuria before progressing to animal experiments [359]. As such organoids have the potential to limit the number of animals used in preclinical testing. It was reported that liver organoids derived from patients carrying the AATD PI*Z mutation accurately mimic the liver disease phenotype with regards to PI*Z accumulation and reduced AAT secretion [357]. Given that the organoids have an identical genome to the patient from which the tissue is obtained, they could make an ideal model for AATD genome editing.

5.1.3 Aims

The studies in this chapter aim to determine the applicability of HITI mediated genome editing to AATD deficiency and explores liver organoids as a model for liver genome editing.

5.2 Results

5.2.1 mNeonGreen expression in Huh-7 cells from the endogenous AAT locus prevents AAT expression

Any approach that aims to treat the PI*Z mutant variant of AATD needs to achieve alleviation of both the lung and liver phenotype, and one of the HITI permutations demonstrated in Chapter 4 (**Figure 4.2**) could be beneficial in achieving this. Permutation A in particular, in which experiments using the HITI reporter cell line (**Section 4.2.2**) showed the ability to induce mNeonGreen and diminish mCherry expression, was chosen for AAT in an effort to both express wild-type AAT (the PI*M variant) through the endogenous AAT promoter and prevent expression of the endogenous PI*Z.

A 'double-hit' strategy was designed in which the PI*M cDNA and transcription blocker (TB) would be inserted into the first exon of the genomic AAT locus (**Figure 5.1A**). This would ensure both PI*M expression and loss of endogenous AAT as the TB was anticipated to prevent AAT mRNA transcription (**Figure 5.1B**). To investigate this strategy, gRNA targets in AAT exon 1 were designed using CRISPOR [252] and tested for cleavage activity through a T7 assay where gRNA target 2 was taken forward (**Figure 5.2A**). The Huh-7 liver carcinoma cell line, which expresses PI*M AAT and was confirmed to not contain the PI*Z mutation by Sanger sequencing (data not shown), was transfected with an SaCas9 plasmid targeting AAT exon 1 gRNA 2 (pGM551, **Figure 5.3 and Table 2.3**) and a HITI donor with corresponding gRNA target sites sporting an mNeonGreen cDNA (pGM653, **Figure 5.3 and Table 2.5**) to allow easy determination of integration and expression from the endogenous AAT promoter. In-and-out PCR performed on both integration sides (5' and 3' of the insert) demonstrated that integration had occurred (**Figure 5.2B and C**). Fluorescent imaging of

mNeonGreen (native) and AAT (stained with an anti-human AAT antibody) showed that integration happened in both a homozygous (**Figure 5.4A**, as indicated by loss of AAT expression) and heterozygous manner (**Figure 5.4B**, as indicated by expression of both mNeonGreen and AAT in the same cells). The Huh-7 cells have two copies of AAT as measured by ddPCR CNV assay (data not shown, protocol in **Section 2.12.2**). Loss of AAT expression was also observed in some cells where no concomitant mNeonGreen expression was observed (**Figure 5.4C**), indicating that no integration had occurred, but that AAT expression had been inhibited, perhaps via a large deletion or backbone insertion. The total number of mNeonGreen positive cells was measured at 12%, with a corresponding transfection efficacy of 30%.

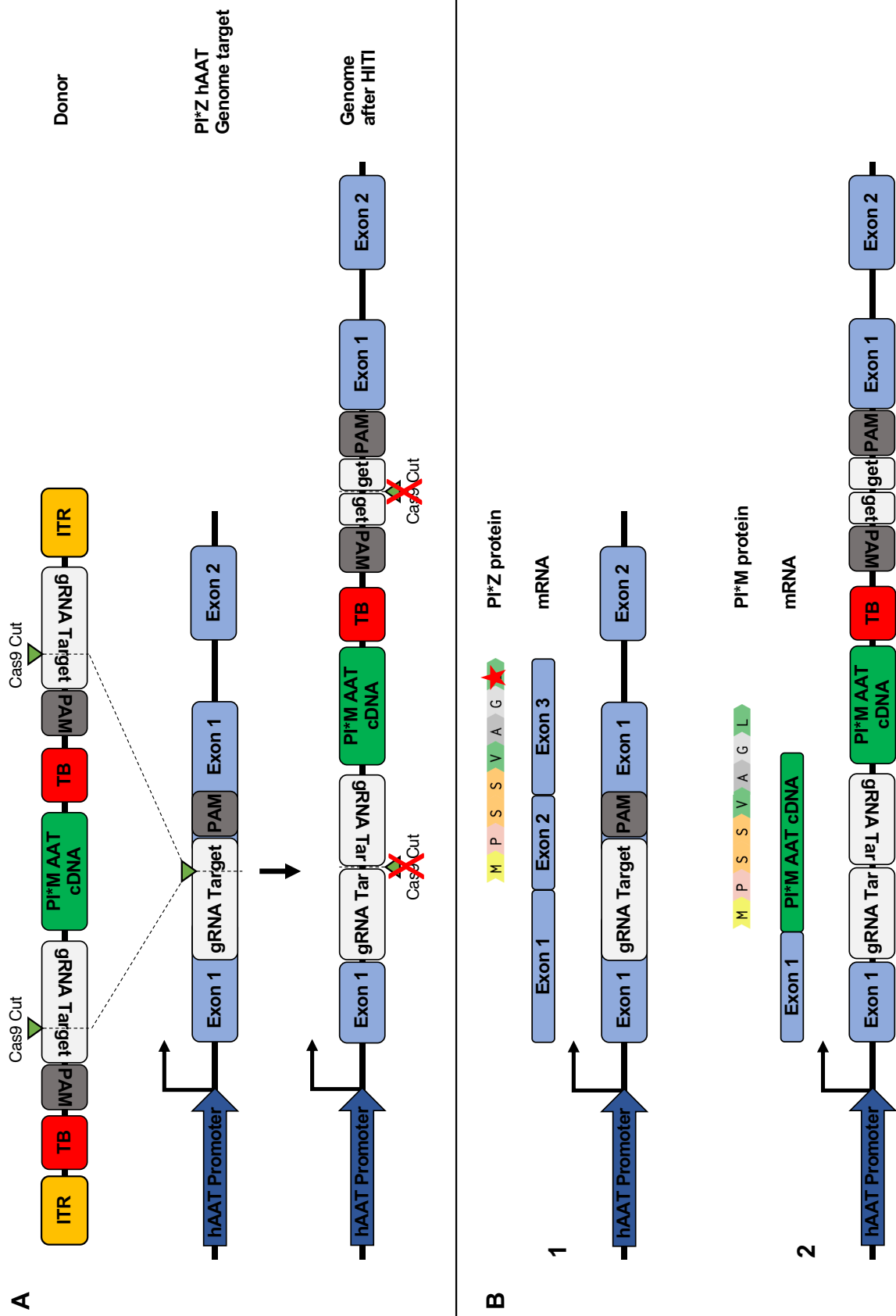


Figure 5.1 | AAT HITI double-hit integration mechanism. The double-hit strategy for PI*Z AATD targets HITI in the first exon of the AAT gene locus. A HITI donor is provided in the form of an rAAV vector genome or AAV2 genome plasmid (**A**) which includes a transcription blocker (TB) between the ITR and integration cassette to prevent ITR driven expression of the cDNA leading to false-positives (**Section 4.2.5.3**, plasmid based on HITI donor plasmid pGM651, **Table 2.4**). gRNA targets in the HITI donor are in reverse orientation to the target in AAT exon 1 to prevent re-cutting after HITI mediated cNHEJ ligation (**Figure 1.6C**). AAT translation starts from exon 2 (**B1**). After integration, AAT PI*Z expression will be abrogated because of the transcription blocker placed behind the PI*M AAT cDNA (**B2**). This ensures that the only mRNA produced is the one containing the wild-PI*M AAT cDNA. This way both functional AAT protein will be produced and the toxic PI*Z protein production will be abolished. hAAT, human AAT; TB, transcription blocker; ITR, inverted terminal repeat; PAM, protospacer adjacent motif; cDNA, complementary DNA.

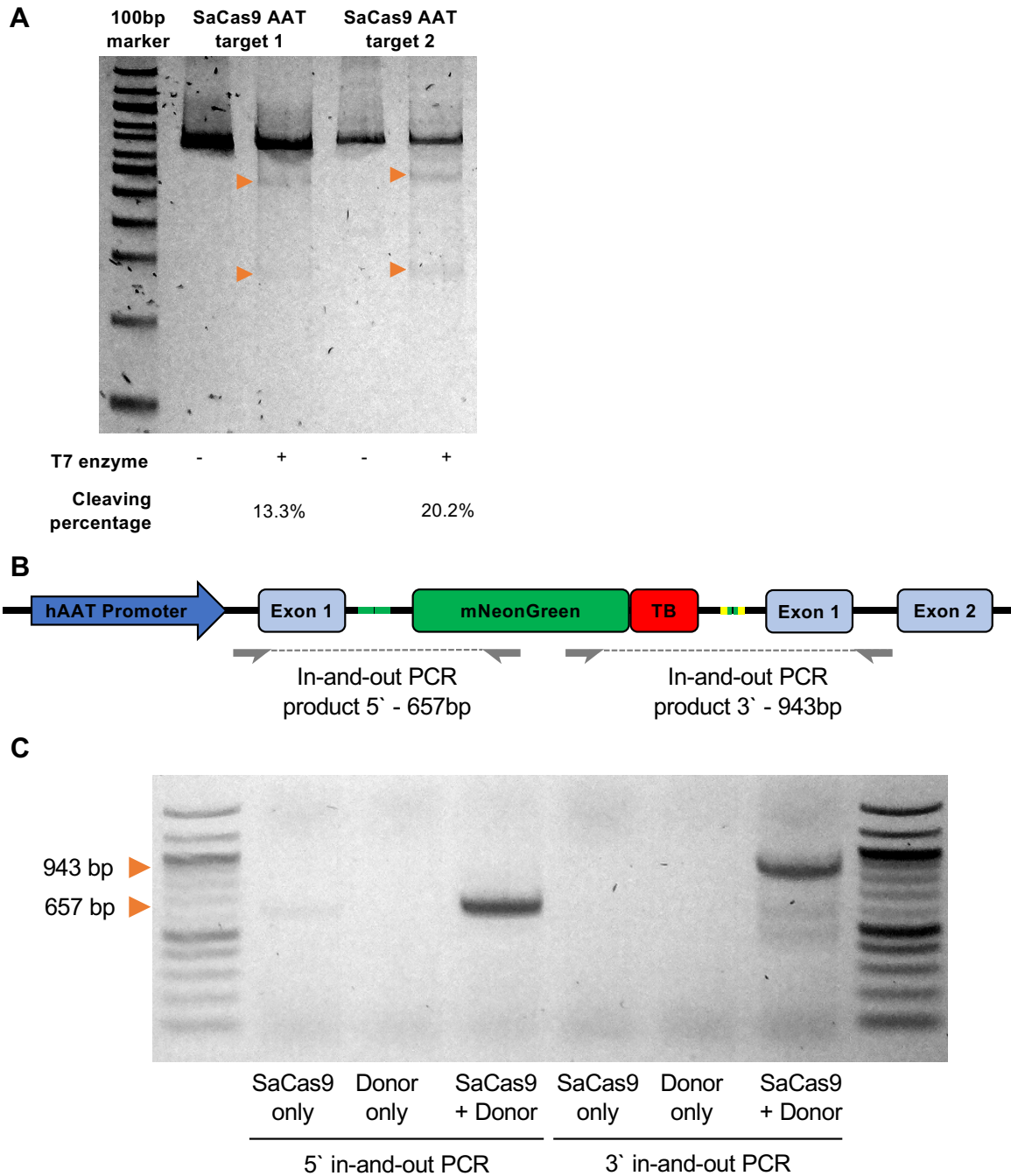


Figure 5.2 | AAT gRNA target site cleavage assessment and in-and-out PCR integration assay. Huh-7 cells transfected with SaCas9 plasmids targeting gRNA sites in exon 1 of the AAT locus were assessed for cleavage activity by T7 assay (**A**, protocol **Section 2.10**). Orange arrowheads indicate cleavage products and cleavage percentages are indicated below. In-and-out PCR products were designed around the integration junctions of the integrated mNeonGreen insert. Huh-7 cells were transfected with SaCas9 targeting the AAT Exon 1 gRNA target 2 (pGM551, **Figure 5.3 and Table 2.3**) and/or HITI donor with the corresponding gRNA target sites and mNeonGreen cassette (pGM653, **Figure 5.3 and Table 2.5**). In-and-out PCR of **B** was performed on DNA extracts of those cells after 3 weeks of culturing to dilute out the transfected plasmids. The agarose gel in **C** shows that only cells transfected with both donor and SaCas9 contain the in-and-out PCR target sequences, indicating correct integration into AAT exon 1. TB, transcription blocker.

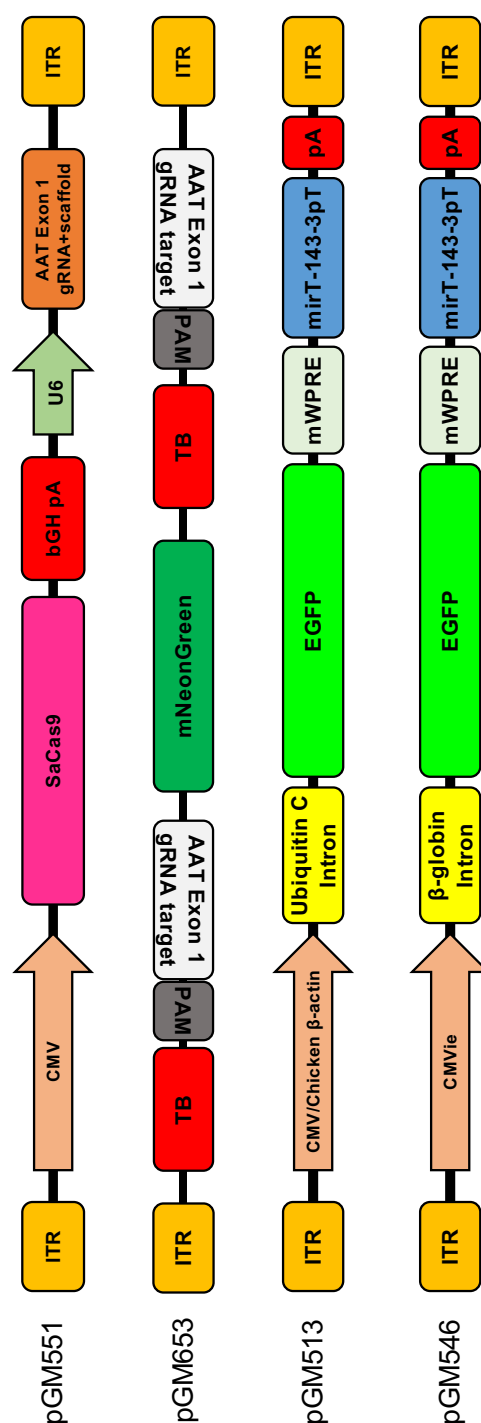


Figure 5.3 | Schematic of AAV2 based viral constructs used in Chapter 5. Plasmids used are based on an AAV2 genome plasmid (see **Table 2.2**). EGFP, enhanced green fluorescent protein; mWPRE, mutant woodchuck hepatitis virus posttranscriptional regulatory element, TB, transcription blocker; PAM, protospacer adjacent motif; bGH, bovine growth hormone; pA, poly A; ITR, inverted terminal repeat; U6, U6 promoter; CMVie, CMV immediate early promoter.

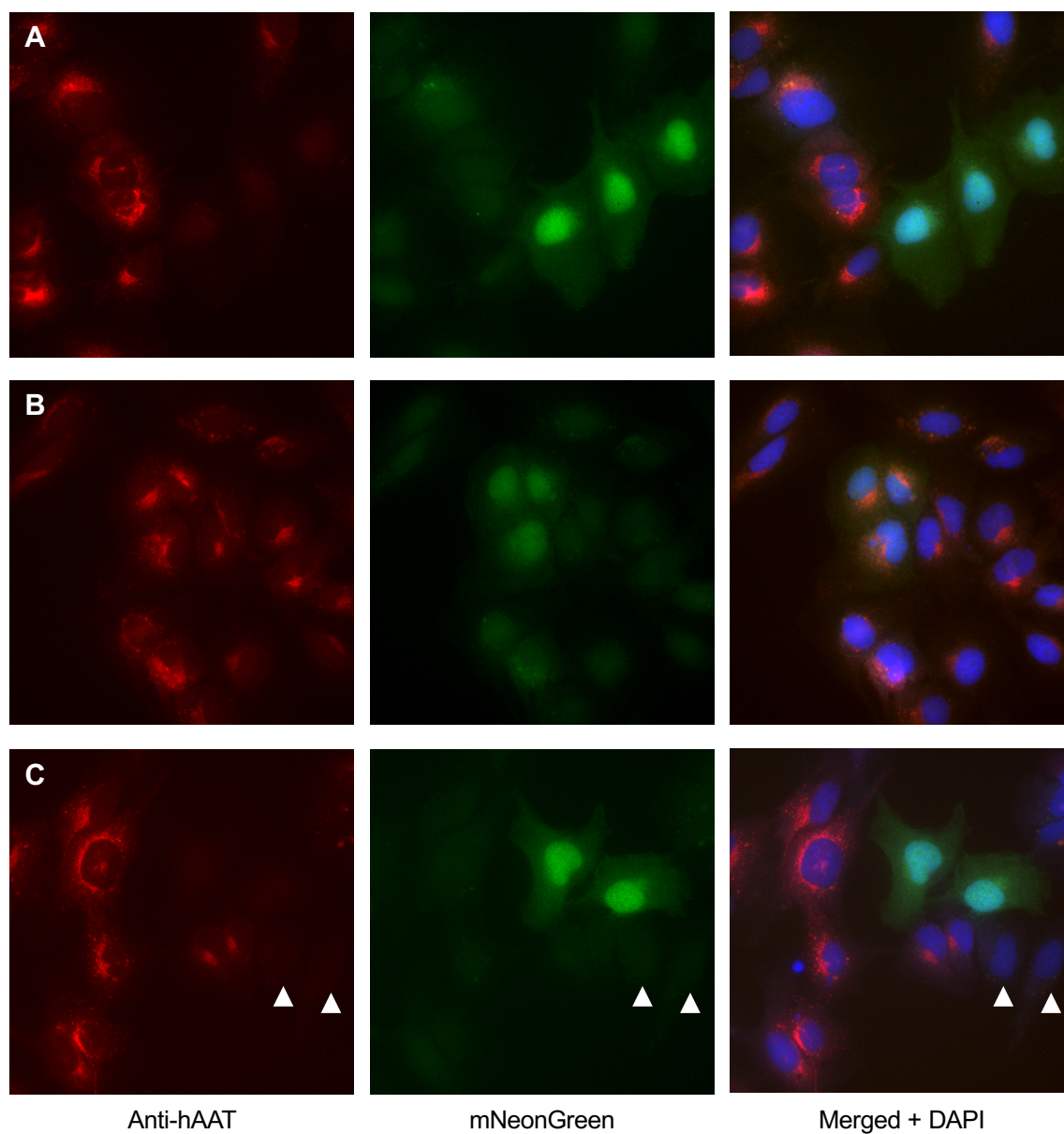


Figure 5.4 | Fluorescent microscope images of AAT HITI integrated cells. Huh-7 cells were transfected with an AAT exon 1 gRNA 2 targeting SaCas9 plasmid (pGM551) and HITI donor pGM653 containing mNeonGreen and transcription blocker (**Figure 5.3**). Cells were passaged for 3 weeks to dilute out donor plasmid to minimise potential false-positive mNeonGreen expression. Cells were grown on a coverslip and stained with a rabbit anti-human AAT antibody to visualize AAT expression (red) alongside native mNeonGreen (green) (**Section 2.4.1.1**). Cell nuclei were stained with DAPI (blue). Groups of cells (potentially clonal) were observed that expressed mNeonGreen but had lost AAT expression (**A**), indicating that the transcription blocker was effective in preventing endogenous AAT expression (**Figure 5.1**) and that homozygous integration had been achieved. Cells expressing both AAT and mNeonGreen were likely to be heterozygous for HITI integration (**B**). Some clones were observed that had lost AAT expression but did not express mNeonGreen (**C**, white arrowheads). These cells could contain the plasmid backbone integrated in the genomic DSB (**Figure 5.5**), have the insert integrated the wrong way around or have a large cNHEJ mediated deletion disrupting AAT expression. mNeonGreen expression observed in the nucleus could be the result of large focal planes in the microscope also capturing mNeonGreen signal above and below the nucleus. The total number of mNeonGreen positive cells was 12% after culturing to dilute out the donor plasmid. Transfection efficiency of Huh-7 cells was around 30% (GFP positive cells using pGM546 (**Table 2.2**) as a transfection control using the same transfection protocol, data not shown), measured using the MiniMax 300 Imaging Cytometer (**Section 2.4.2**). Representative images shown from three independent experimental repeats. DAPI, 4',6-diamidino-2-phenylindole; hAAT, human AAT.

5.2.2 Characterising HITI integration in the AAT locus

The fact that plasmids are used in the above experiments raises the possibility of unwanted integration of the plasmid backbone. After all, Cas9 cutting the plasmid twice (on each side of the insert) will result in two dsDNA molecules with blunt ends available for cNHEJ ligation into the DSB in the genome. To determine whether this was the case, PCR primers were designed to amplify the integration junctions in the event of backbone insertion (**Figure 5.5A and B**). They were designed to be specific to the direction of the integration, i.e. backbone inserted when the gRNA and PAM sequences of backbone and genome create the HITI characteristic palindromes (5'-gRNA-gRNA and 3'-PAM-PAM, **Figure 5.5A**) and where the PAM and gRNA recreate targetable gRNA sites (5'-gRNA-PAM and 3'-gRNA-PAM, **Figure 5.5B**). As the amplicons in **Figure 5.5C** show, there is evidence that the backbone had inserted in the AAT exon 1 target site in both directions. However, this assay does not account for integration of both the backbone and the insert at the same time, for example when Cas9 cuts only a single gRNA target site in the donor. To assess the fidelity of HITI mediated integration of the insert sequence in Huh-7 cells, the PCR amplicons of the 5' and 3' insert junctions of **Figure 5.2** were sequenced. The observed indels were restricted to 3 bp or less with seamless integration in 87.5% of cases (**Figure 5.6**), similar to earlier findings in HEK293T cells (**Figure 4.5**).

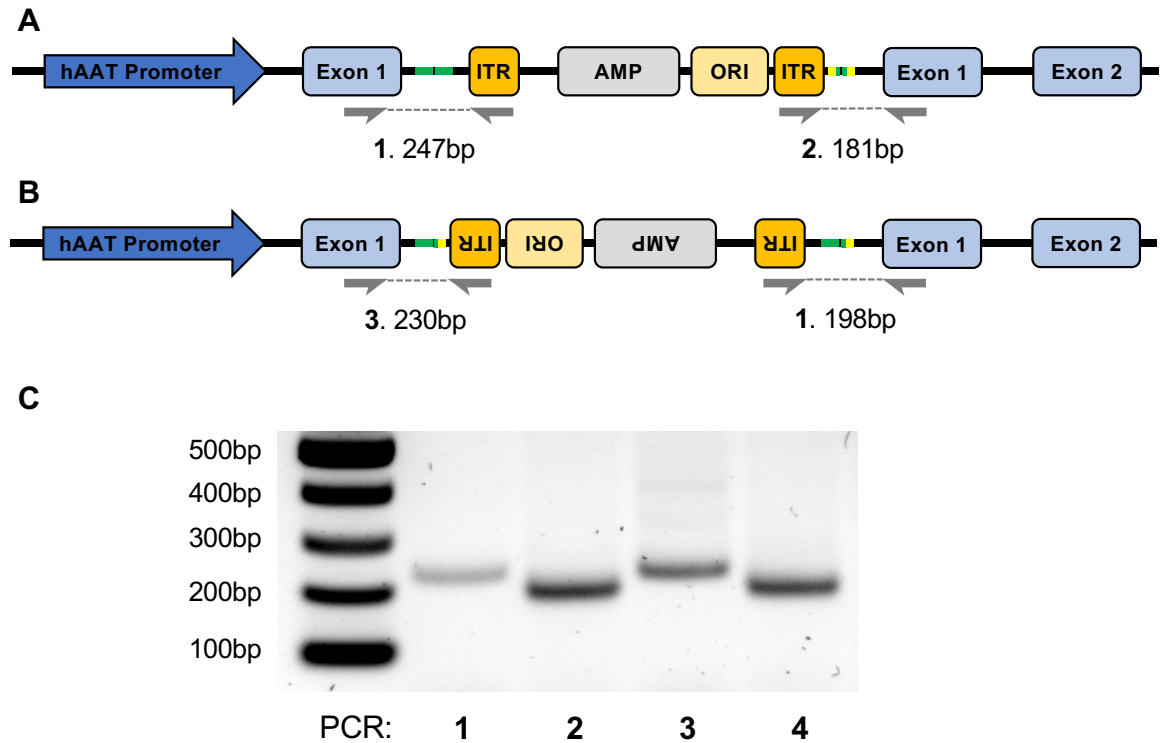


Figure 5.5 | Investigation into AAV2 plasmid backbone integration. To address the possibility of plasmid backbone integration, PCR was used to amplify regions around the integration junction after gRNA-gRNA and PAM-PAM oriented integrated backbone (**A**) and gRNA-PAM and gRNA-PAM oriented backbone integration (**B**). PCR amplicons were designed so as to not include the entire ITR sequences, as the repetitive nature of the ITR would likely prevent efficient amplification, rather primers were designed on the edges of the ITR sequences. As seen in **C**, products were amplified for both sides of the integration junctions for both orientations, indicating that backbone integration is a possible outcome of HITI using plasmids. Note that this assay does not preclude the insertion of backbone and insert simultaneously, as might happen after the plasmid has only been cut once. The DNA used here was extracted from cells shown in **Figure 5.4**. ORI, origin of replication; AMP, ampicillin resistance gene; ITR, inverted terminal repeat; hAAT, human AAT.

5.2.3 Liver organoids can be transduced and transfected

Cell lines, whatever their origin, rarely recapitulate all the characteristics of the tissue from which they were derived. In addition, cells may display variable karyotypes within a population; Huh-7 cells, for instance, display a wide variety of chromosome counts, with a mode of 60 chromosomes per cell [360]. Furthermore, vast differences in genomic and phenotypic profiles have been reported in cell lines of the same origin after passaging and between labs [361, 362]. Organoids on the other hand display genomic stability over time [357] and display many of the phenotypic hallmarks of the tissue of origin [251, 356], offering a more physiologically and genetically relevant *in vitro* model than cultured cells.

To address whether liver organoids could be used to investigate liver-targeted gene therapy and genome editing strategies, liver organoid lines were derived from murine (female BALB/c) and human liver tissue, according to the protocol of Broutier et al. [251]. A brief overview of steps taken in liver organoid isolation is shown in **Figure 5.7**. Two human lines were derived from healthy liver tissue (as determined by a pathologist) from surgical resections (**Section 2.3**). The BALB/c derived mouse line was assessed for its amenability to plasmid transfection (rAAV2 CASI EGFP, pGM513 see **Table 2.2**) using Lipofectamine 2000, Lipofectamine 3000 and PEIpro for gene transfer to organoids digested into single cell suspensions (**Figure 5.8**). The results showed that Lipofectamine 3000 was the most successful in transfection resulting in EGFP expression, but the efficiency was low at <1% of EGFP positive organoids. Both the murine and human-derived liver organoid lines were also investigated for transduction with several AAV serotypes (**Figure 5.9**). Vectors were produced as detailed in **Section 2.11.1** using the same genome plasmid (AAV2 CMVie EGFP, pGM546 see **Table 2.2**) for each serotype. Single cell suspensions of liver organoids were transduced with the indicated vectors at an MOI of 50,000 (**Section 2.11.4**). While there was no discernible EGFP expression with any of the serotypes tested in the human organoids (**Figure 5.9 B and C**), both AAV1 and AAV6.2 transduced murine liver organoids at 32% and 39% of organoids containing EGFP positive cells respectively (**Figure 5.9A**).

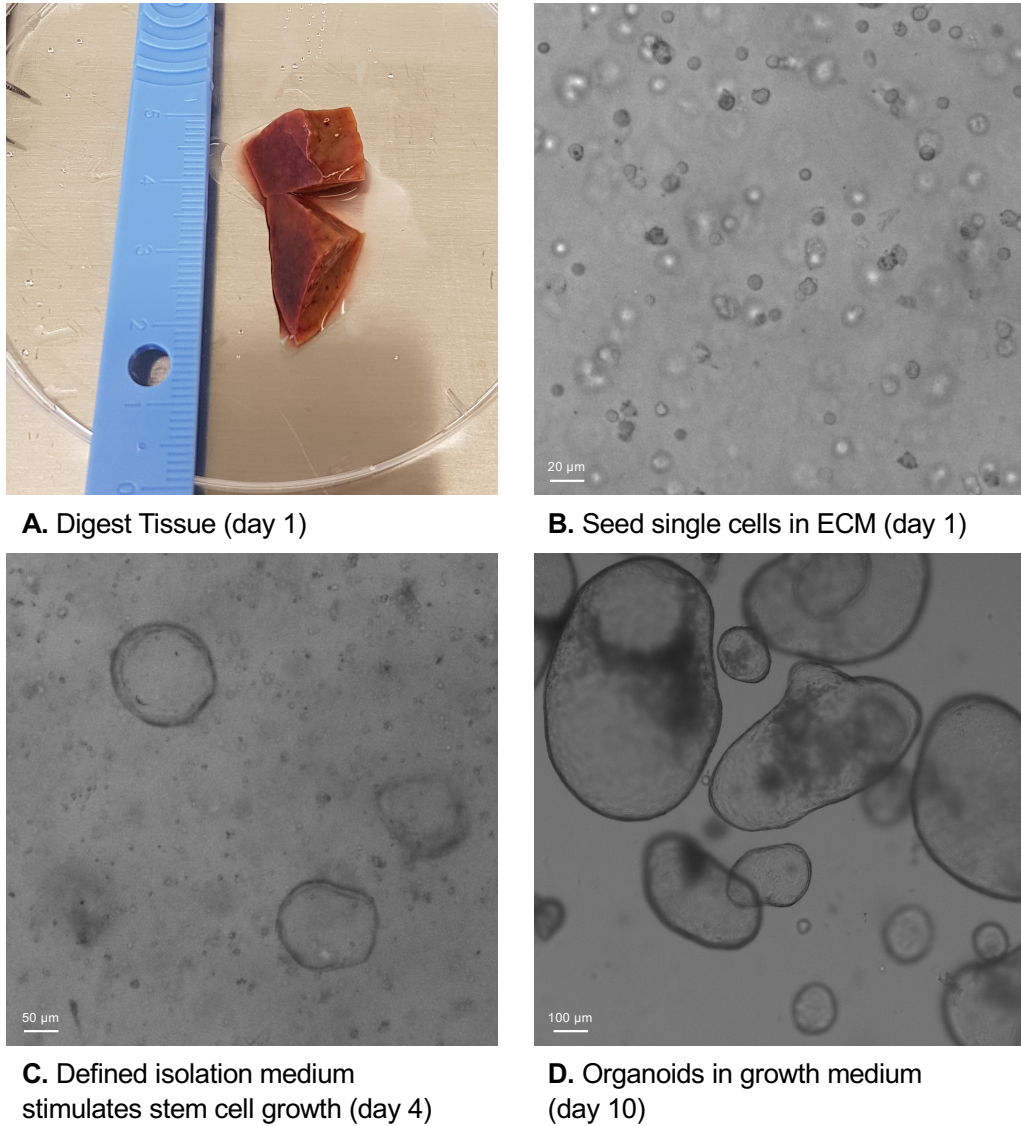


Figure 5.7 | Derivation of (human) liver organoids. Fresh liver tissue from human or mouse origin was processed (**Section 2.3** and Broutier et al. [251]). Human liver tissue from surgical resection (**A**) or harvested mouse liver tissue was digested into single cells which were washed and seeded in the appropriate extracellular matrix (ECM, **B**). After incubation with defined isolation medium, small (<100 µm) organoids started to form (**C**) before growing out into organoids with a single-cell layer thick wall measuring in between 100-800 µm in diameter (**D**).

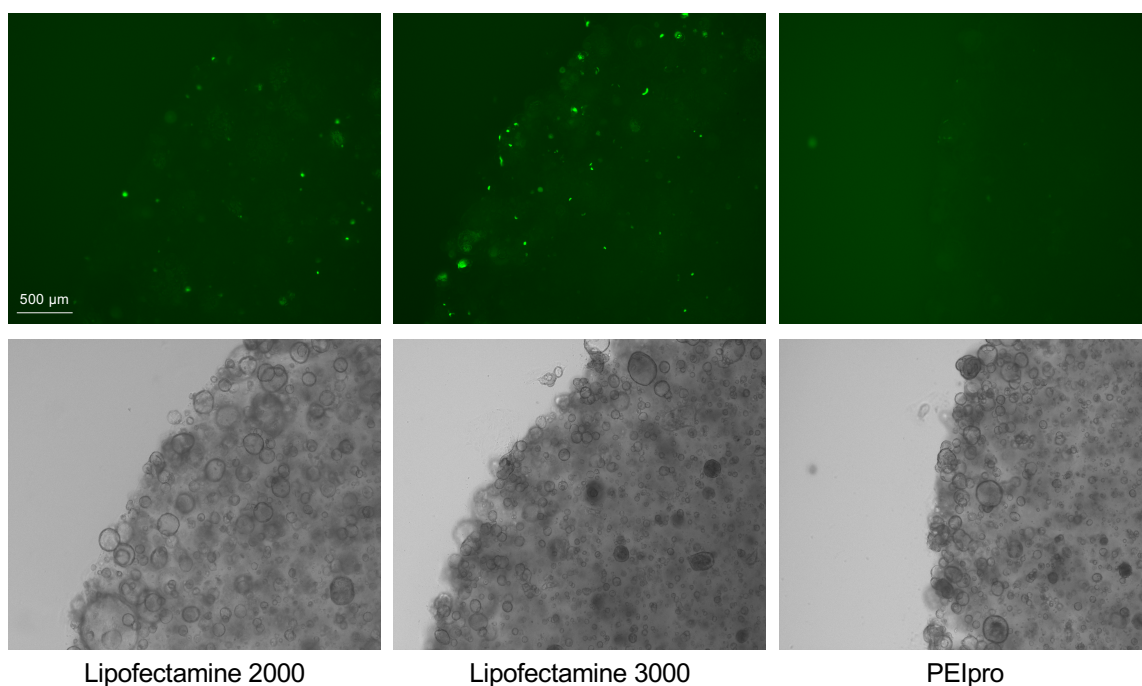
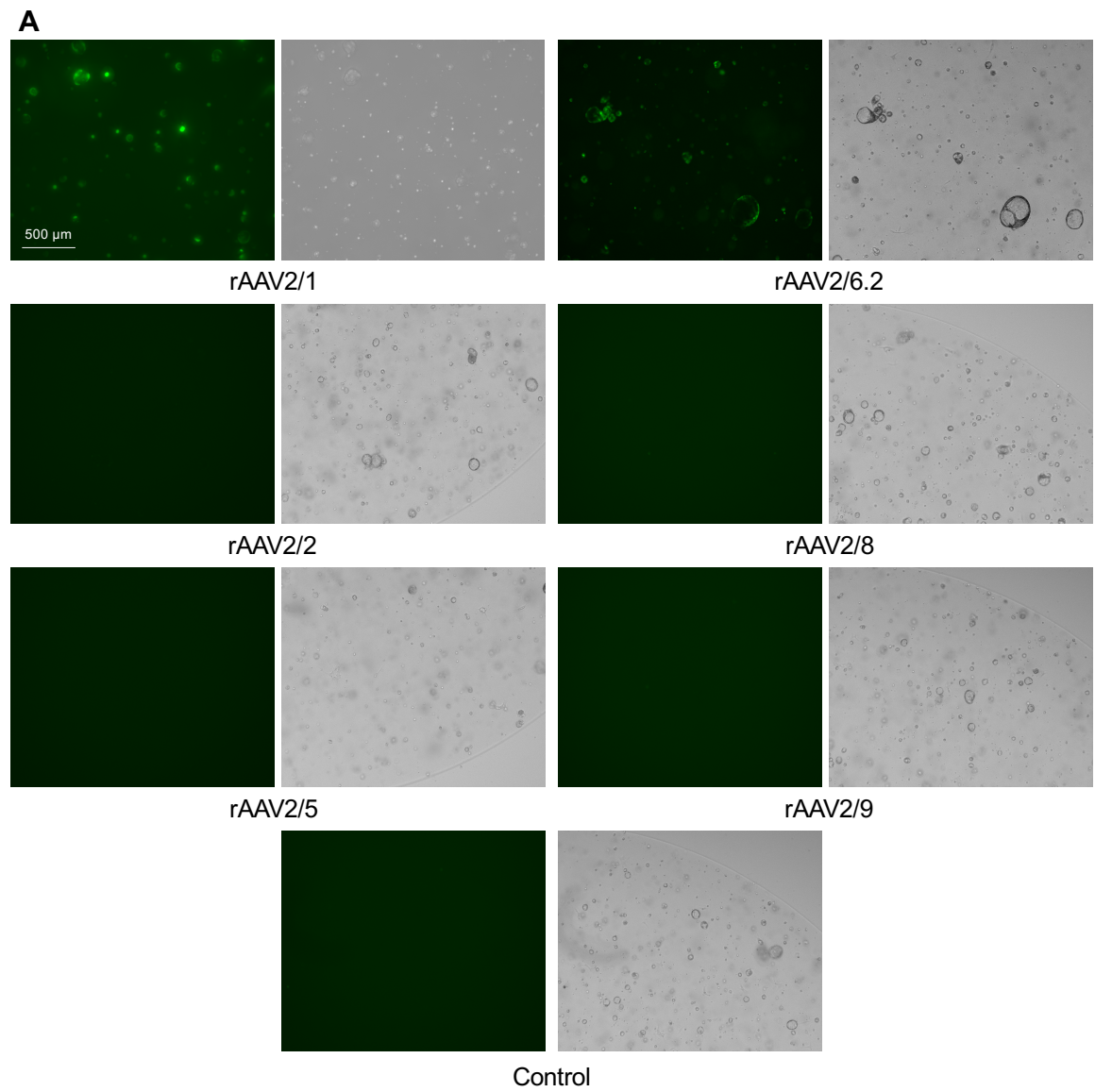
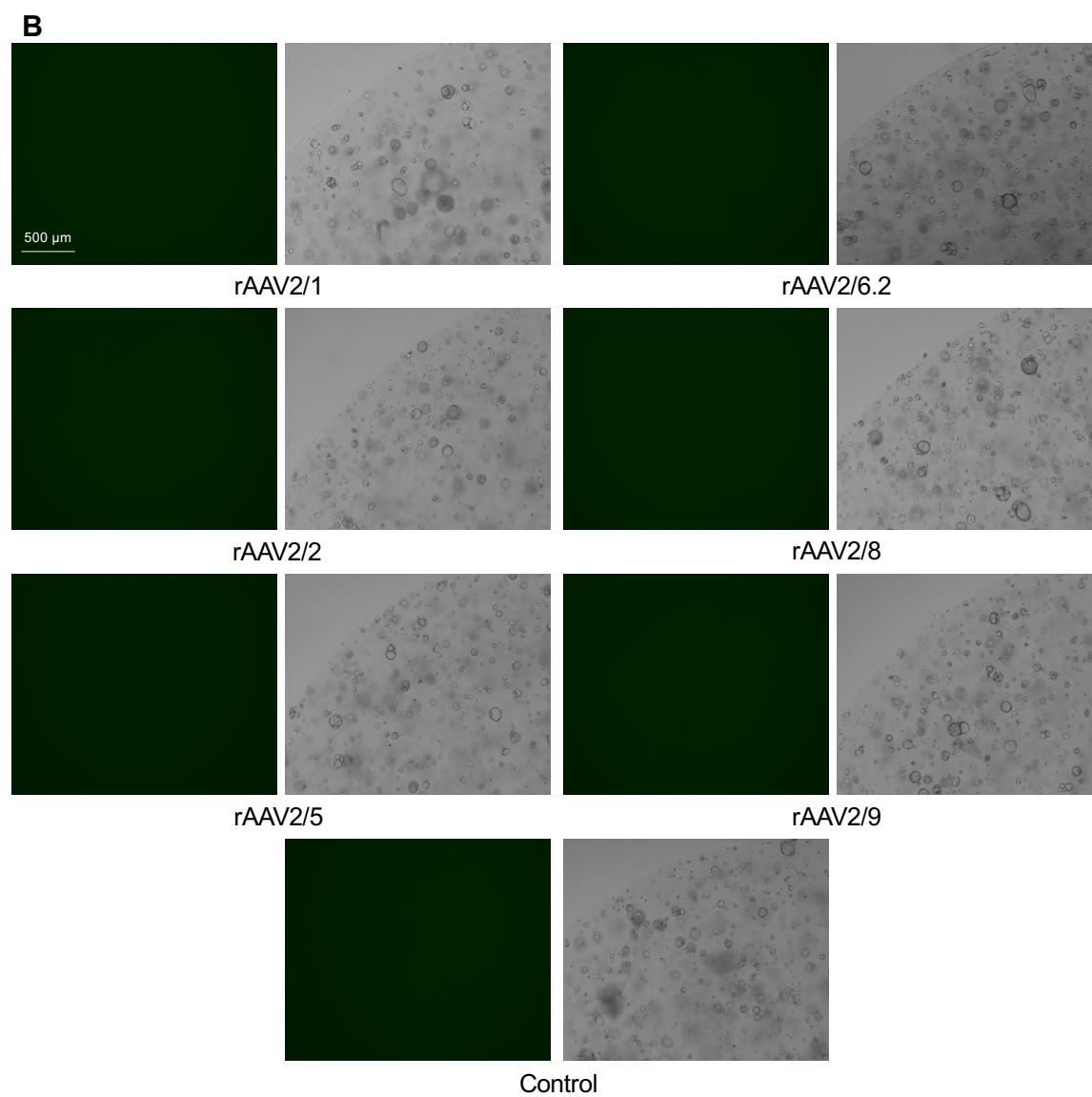


Figure 5.8 | Plasmid transfection of murine liver organoids. Mouse liver organoids derived from BALB/c mice were generated (**Section 2.3**) and 500,000 single cells were transfected with 0.8 μg of AAV2 CASI EGFP plasmid (pGM513, **Table 2.2**) using the gene transfer agents shown in 50 μl total volume of Optimem (**Section 2.6.10**). Organoids were cultured for 6 days and imaged for native EGFP fluorescence (top) and bright field (bottom). Representative images shown, $n=2$ per transfection reagent.





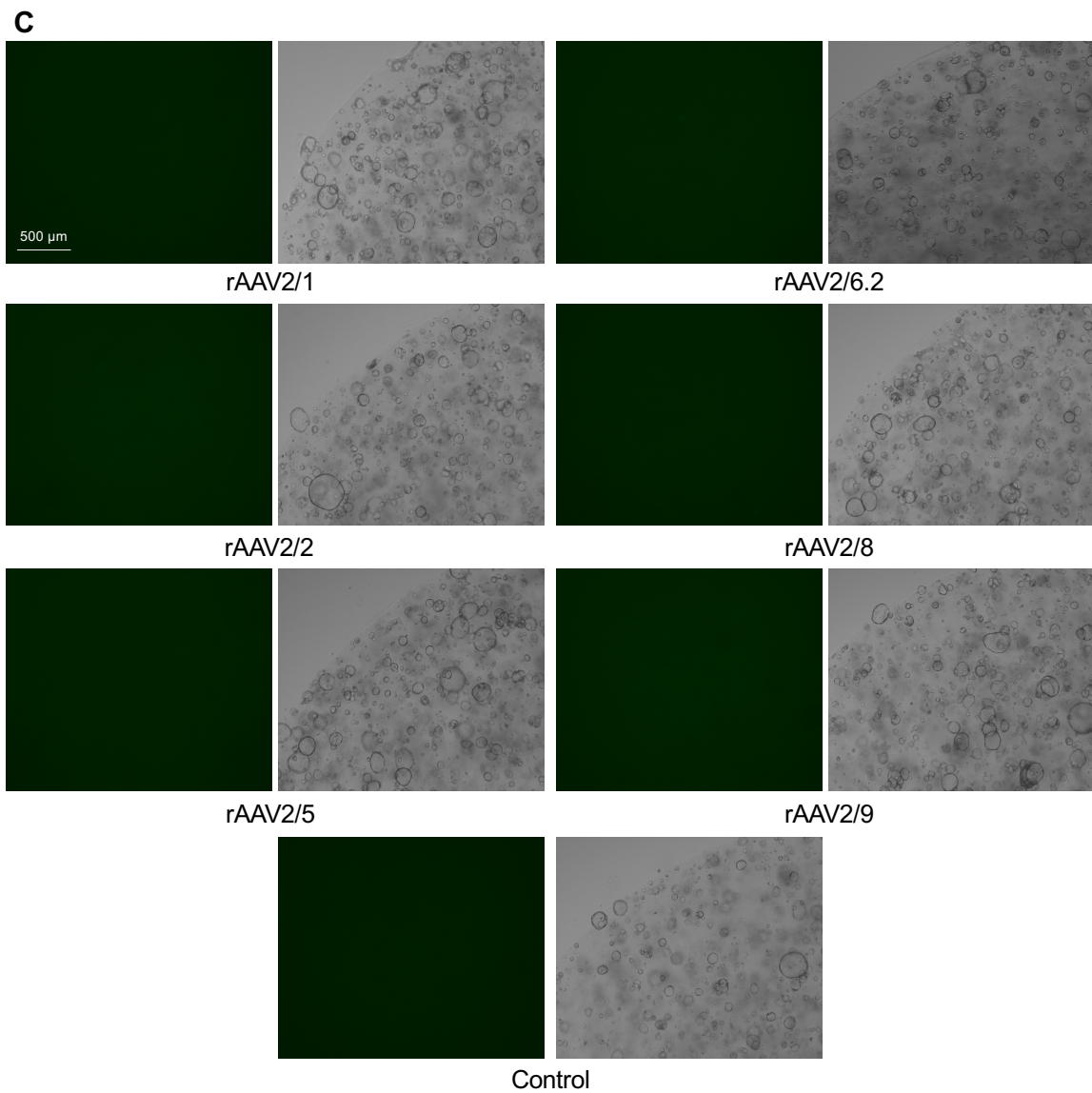


Figure 5.9 | rAAV transduction of murine and human liver organoids. Murine liver organoids derived from BALB/c mice (**A**), and human liver organoids derived from two separate donors (**B** and **C**), were transduced with rAAV vectors based on the AAV2 CMVie EGFP genome plasmid (pGM546, **Table 2.2**) and with the indicated serotypes according to the protocol in **Section 2.11.4** (MOI = 50,000). Organoids were cultured for 6 days and imaged for native EGFP fluorescence. EGFP positive organoids (at least one green cell per organoid) were counted manually. rAAV2/1 and rAAV2/6.2 showed positive signal in mouse liver organoids at 32% and 39% respectively (**A**). No EGFP positive organoids were observed in either human organoid line for any serotype tested (**B** and **C**). Representative images shown with EGFP channel (left) and bright field (right) for each serotype, n=2 per organoid line and serotype, two independent experimental repeats.

5.3 Discussion

5.3.1 The AAT ‘double-hit’ genome editing strategy is viable for PI*Z AATD

The prevalent PI*Z mutation variant of AATD requires a strategy that addresses both the lung and liver phenotype simultaneously. The lung phenotype can be alleviated by augmentation of functional AAT by ERT or gene addition. The liver phenotype however requires the reverse; inhibition of expression of the PI*Z protein that causes liver cirrhosis. The studies described in this chapter aimed to apply the HITI methods demonstrated in **Chapter 4**, and in particular the HITI permutation A (**Figure 4.2**), to AATD in order to both express the functional variant of AAT (PI*M) whilst simultaneously preventing PI*Z expression. This strategy, termed AAT double-hit, used HITI to integrate a cDNA of the PI*M variant of AAT into the first exon of the genomic AAT locus to allow endogenously controlled PI*M AAT expression (**Figure 5.1A**). Exon 1 is part of the untranslated region of the AAT mRNA and therefore should any small indels arise around the Cas9 cut site during editing (i.e. in cells where no HITI donor is present) this would not be expected to negatively affect residual AAT activity. As shown in Chapter 4, for permutation A (**Figure 4.1A**) the cDNA was followed by a transcription blocker, so in the double-hit strategy no PI*Z mRNA is expected to be produced and thus expression of the toxic PI*Z protein should be abolished (**Figure 5.1B**). Using mNeonGreen as a surrogate reporter gene for the PI*M AAT cDNA, plasmid transfection into the Huh-7 liver cell line was effective in integrating into the intended target site as evidenced by PCR (**Figure 5.2**). Fluorescence readout of mNeonGreen expression and anti-AAT staining confirmed that suggested homozygous integration abolished endogenous AAT expression (**Figure 5.4A**) and heterozygous integration showed both mNeonGreen and AAT expression (**Figure 5.4B**). Loss of AAT expression without concomitant mNeonGreen expression was also observed (**Figure 5.4C**), which is possibly due to large deletions caused by cNHEJ DSB repair or by integration of the plasmid backbone. The potential for backbone insertion was confirmed by PCR on backbone-specific sequences (**Figure 5.5**) although this does not preclude possible integration of the entire plasmid after a single cut of SaCas9 of the HITI donor.

Unfortunately, no single-cell clones could be grown from the transfected Huh-7 cells to investigate singular integration events in more detail to check for whole plasmid integrations. The implication of backbone integration is that, in a setting where rAAV would be used rather than AAV2 genome plasmids (see **Chapter 6**), the AAV ITR sequence can integrate into the intended target site (and potentially at off-target sites). Evidence for this was reported recently using a cNHEJ-mediated exon excision approach for DMD where AAV integration occurred in the cut site, with the majority of integrations comprising AAV2 ITR only [363]. Using SaCas9 to ‘release’ the ITR sequences could potentially contribute to ITR integration into the genome. In Huh-7 cells the fidelity of on-target integration of the insert sequence displayed similarly high levels of seamless integration as observed previously (**Figure 5.6**)

5.3.2 Liver organoids provide a promising tool for gene therapy and genome editing research

Organoids derived from multiple tissues constitute an *in vitro* model showing greater phenotypic similarities to the tissues from which they were derived compared with cell lines, whilst also keeping a stable genomic composition that is identical to the (patient) donor. As such, they are used profitably in many fields [349, 350]. In this chapter, liver organoids from one mouse and two human donors were successfully isolated and propagated (**Figure 5.7**), and subjected to standard *in vitro* gene therapy tools such as plasmid transfection and rAAV transduction to test their efficiency. Plasmid transfection efficiency in mouse organoids was low with the reagents used (**Figure 5.8**) and was deemed insufficient for potential genome editing experiments. Transduction with rAAV was more successful, with both rAAV2/1 and rAAV2/6.2 able to transduce murine organoids with some efficiency (32-39%, **Figure 5.9A**). The AAV1 and AAV6 serotypes (from which AAV6.2 is derived, they differ only in a single amino acid [364]) are both found in Clade A of the AAV phylogenetic tree and only differ in six amino acids. The AAV1 and AAV6 serotypes can bind to both alpha2,3-linked or alpha2,6-linked sialic acid [365], whereas AAV6 (and AAV6.2) can also bind to heparan

sulphate proteoglycan, most likely as a co-receptor since it is not sufficient for internalisation [364]. These shared characteristics may explain why these two serotypes displayed similar transduction properties, and perhaps highlights sialic acid as the receptor through which they transduce murine liver organoids. Human liver organoids on the other hand were not efficiently transduced by any of the serotypes tested (**Figure 5.9B and C**).

Of note, is that the organoids used for transfection and transduction in this study were immature cells, or so-called bipotent progenitor cells as defined by their co-expression of ductal (e.g. cytokeratin 19) and hepatocyte (Hepatocyte Nuclear Factor 4) markers [357]. Keeping cells in the progenitor stage, through the use of a defined growth medium that suppresses differentiation, allows for long-term culture and an increased supply of organoids for experimentation. Use of a differentiation medium leads to differentiation and maturation of the liver organoids, now termed mature organoids, which are terminally differentiated hepatocytes expressing genes such as albumin and AAT. The methods used here for transfection and transduction rely on digestion of immature organoids into single cells, however (**Section 2.6.10 and 2.12.4**), and using a similar approach for mature organoids would likely be ineffective as mature organoids do not re-form following digestion into single cells. On the other hand, the disadvantage with immature organoids is that they do not accurately recapitulate the mature liver and as a result they might not express the same receptors on their surface. This could be why rAAV2/8, the serotype of choice for *in vivo* gene transfer of the murine liver [308], was not able to transduce the immature liver organoids (**Figure 5.9A**) and why none of the tested serotypes transduced human liver organoids (**Figure 5.9B and C**). One possible solution to the poor levels of human organoid transfection/transduction is microinjection into the lumen of mature, differentiated liver organoids, a procedure which has been reported for lung organoids [366]; this would require a microinjection system in an aseptic environment, but unfortunately this was not available at the time of the study. Similarly, the ideal liver organoids to be used for the studies described here would be derived from tissue obtained from an AATD patient, in particular a

patient homozygous for the PI*Z mutation. Despite continued efforts, such tissue could not be procured within the timeframe of the study due to the rarity of the disease and low diagnosis rate.

5.3.3 Conclusion

The aim of the studies in this chapter was to assess the application of HITI to edit the PI*Z mutation of AATD. The pan-mutation ‘double-hit’ integration strategy, using HITI permutation A (described in Chapter 4), was targeted to the first exon of AAT and was shown to be effective in expressing a reporter gene from the endogenous AAT promoter and at the same time preventing endogenous AAT expression. The practical difficulties in obtaining suitable (fresh rather than fixed) liver tissue from a PI*Z AATD patient in order to make organoids, together with the difficulties in transducing human liver organoids, meant that additional, complementary studies were needed to investigate the double-hit strategy for AAT. Such studies were performed in a mouse model of PI*Z AATD and are described in Chapter 6.

Chapter 6 – *In vivo* HITI mediated integration for alpha-1 antitrypsin deficiency

6.1 Introduction

6.1.1 Animal models of alpha-1 antitrypsin deficiency

In **Chapter 5** a novel ‘double hit’ strategy for treatment of AATD was explored to use HITI to incorporate a cDNA into the AAT locus and allow expression of the cDNA by the endogenous AAT promoter with simultaneous inhibition of the PI*Z mutant of AAT (**Figure 5.1**). A proof-of-concept study was performed in a liver carcinoma cell line (Huh-7) and showed that this was a feasible approach (**Figure 5.3 and 5.4**). Due to the rarity of AATD, it proved difficult to procure liver tissue from AATD patients, and so focus turned to animal models. Whilst patient derived liver organoids would provide a more genetically accurate model, they do not recapitulate a key part of *in vivo* gene therapy or genome editing: delivery. Only by combining a genome editing method with effective delivery in a relevant model can one make a rational hypothesis on therapeutic applicability of a potential treatment.

Several animal models of AATD have been developed. Some approaches have focused on the lung phenotype of AATD (i.e. emphysema and COPD) [367]. While these models are useful to test symptomatic treatments of the effects of AATD they are not appropriate for investigating AATD genome editing because they lack the target AATD mutation. Other models do recapitulate AATD due to AAT mutations and are regularly used in AAT gene therapy and genome editing research [240-242, 346]. The mouse model most often used is a transgenic mouse line harbouring the full human SERPINA1 locus containing the PI*Z mutation [255]. As this knock-in mouse line (hereafter named PI*Z KI) still carries the mouse homologues of AAT, it does not suffer from the deficiency aspect of AATD and thus does not display the lung phenotype. The presence of the PI*Z mutation in this model is an attractive target for approaches aiming to knock-down [239], knock-out [240] or correct [241,

242] the human PI*Z mutant of AAT (see **Section 5.1.1**). The characteristics of the PI*Z KI mouse line are discussed later in the results section of this chapter.

To overcome the main limitation of the PI*Z KI mouse model, the lack of lung phenotype, efforts were made to knock-out all five orthologues of mouse AAT, *Serpina1a* to *Serpina1e* [368]. This embryonic genome editing tour-de-force resulted in complete abrogation of murine AAT expression and the mice were shown to develop lung emphysema with age. Of note is that previous attempts at knocking-out *Serpina1a* [369] and *Serpina1b* [370] resulted in embryonic lethality. Why this was not observed in the study that knocked out all five orthologues [368] is not clear. A recent conference abstract described a ferret model of PI*Z AATD, in which the ferret homologue of AAT was targeted to create the equivalent PI*Z mutation resulting in age-progressive lung emphysema [371]. No evidence of a liver phenotype was reported in this model, but it will be interesting to see whether this model completely recapitulates the effects of human PI*Z AATD. Of course, since this model contains the ferret version of the mutation, any therapeutic genome editing approaches would have to be verified in a model in which the gRNA matches the human target.

6.1.2 Aims

This chapter describes studies to investigate the feasibility of using HITI-mediated genome editing to tackle alpha-1 antitrypsin deficiency *in vivo*. The genomic composition of the PI*Z KI mouse line was first characterised before using it to evaluate the HITI genome editing approach described in **Chapter 5**.

6.2 Results

6.2.1 Characterisation of PI*Z KI mouse line

6.2.1.1 Breeding patterns and ddPCR genotyping

The PI*Z KI mouse line (**Section 2.12.2**) was selected as the optimal mouse model available. Although the mouse line does not display a detectable lung phenotype it does contain the entire locus of human AAT PI*Z. Furthermore, this model displays the characteristic PI*Z polymer formation [255], which would allow for the evaluation of the version of HITI that incorporates a transcription blocker aimed at preventing PI*Z expression (**Figure 5.1B and 5.4A**).

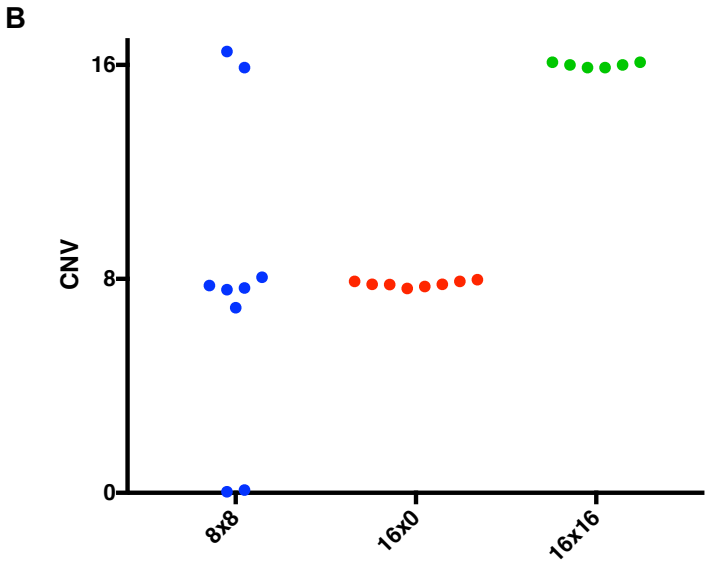
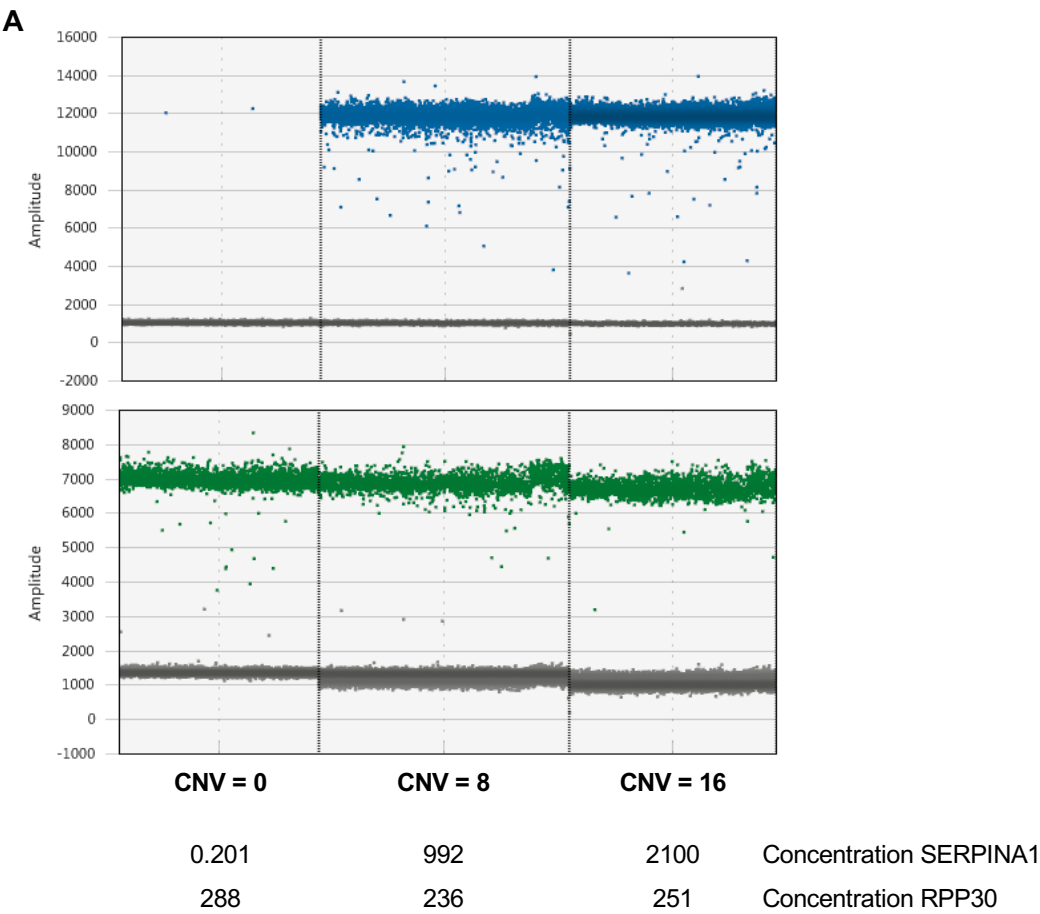
The PI*Z KI mouse line was originally made by microinjection of a linearized copy of the human PI*Z AAT gene plus 2.0 and 2.3 kb of genomic sequences 5' and 3' respectively, into single cell embryos of ICR females and C57BL/6 male mice [255]. This untargeted approach of genomic insertion resulted in random integration into an unknown location in the mouse genome. There have been contradictory reports on the number of copies incorporated onto the genome of this mouse line. Copy number per genome (copy number variation or CNV) were reported as 10 measured by southern blot band densitometry [255], >11 by targeted locus amplification (TLA, further detailed in **Section 7.1**) [240] and eight by ddPCR [240]. The founder mice, were reported to have a CNV of three to four per chromosome (as determined using ddPCR), achieved by selective cross-breeding with wild-type C57BL/6J mice (Chris Mueller, personal communication). This was determined to be the lowest stable CNV, hypothesising that there was a repeat of three copies of PI*Z in tandem in a single genomic location. Upon breeding and genotyping of litters using ddPCR (**Section 2.12.2**), however, it transpired that the true copy number of these mice was closer to eight per allele. Subsequent breeding of heterozygous mice resulted in litters with copy numbers of 0, 8 and 16 in a Mendelian ratio (**Figure 6.1A and B, blue**). Breeding of homozygous mice (i.e. with CNV = 16) with mice where no PI*Z was detected resulted in litters containing only mice with CNV = 8 (**Figure 6.1B, red**); and breeding two homozygous

mice resulted in litters with CNV = 16 only (**Figure 6.1B, green**). Staining cryosections using an antibody against human AAT showed strong expression of AAT mice with CNV = 8 (**Figure 6.2C and E**) with marked increase in AAT in mice with CNV = 16 (**Figure 6.2D and F**). In both mice with CNV = 8 and 16 the morphology of AAT staining resembles the distinct globular cellular inclusions characteristic of PI*Z polymerisation [255].

6.2.1.2 Nanopore sequencing of the PI*Z locus

6.2.1.2.1 Background

While the ddPCR results of **Section 6.2.1.1** provided a clue to the genomic arrangement of the PI*Z transgenic locus in the PI*Z mice, the incongruence between the CNV results obtained (eight per allele) in this study and the reported CNV of the lab of Chris Mueller (three to four per allele) warranted further investigation. To investigate the true genomic organisation of the PI*Z locus a method was required to reveal the structure of the entire transgenic integration locus. To achieve this, sequencing of the entire PI*Z locus could indicate both the number of copies in that region as well as the orientation and potentially structural variations and mutations in the PI*Z genes that may have arisen over time. Targeted locus amplification [372, 373] was reported in a previous publication [240] to accurately determine the location of integration but the characteristics of the assay (further described in **Section 7.1**) do not allow for accurate quantification of copy numbers when repeated sequences are present. In that report for instance, TLA indicated a CNV of >10 per allele whereas ddPCR found the number to be eight. Another method named Cas-mediated PCR-free enrichment (henceforth named CasEnrich), developed by the sequencing company Oxford Nanopore (Oxford, UK), provides the potential for sequencing the entire PI*Z locus in a single sequencing read. A schematic of the CasEnrich method is shown in **Figure 6.2**, and a peer-reviewed application of the method has been published [374].



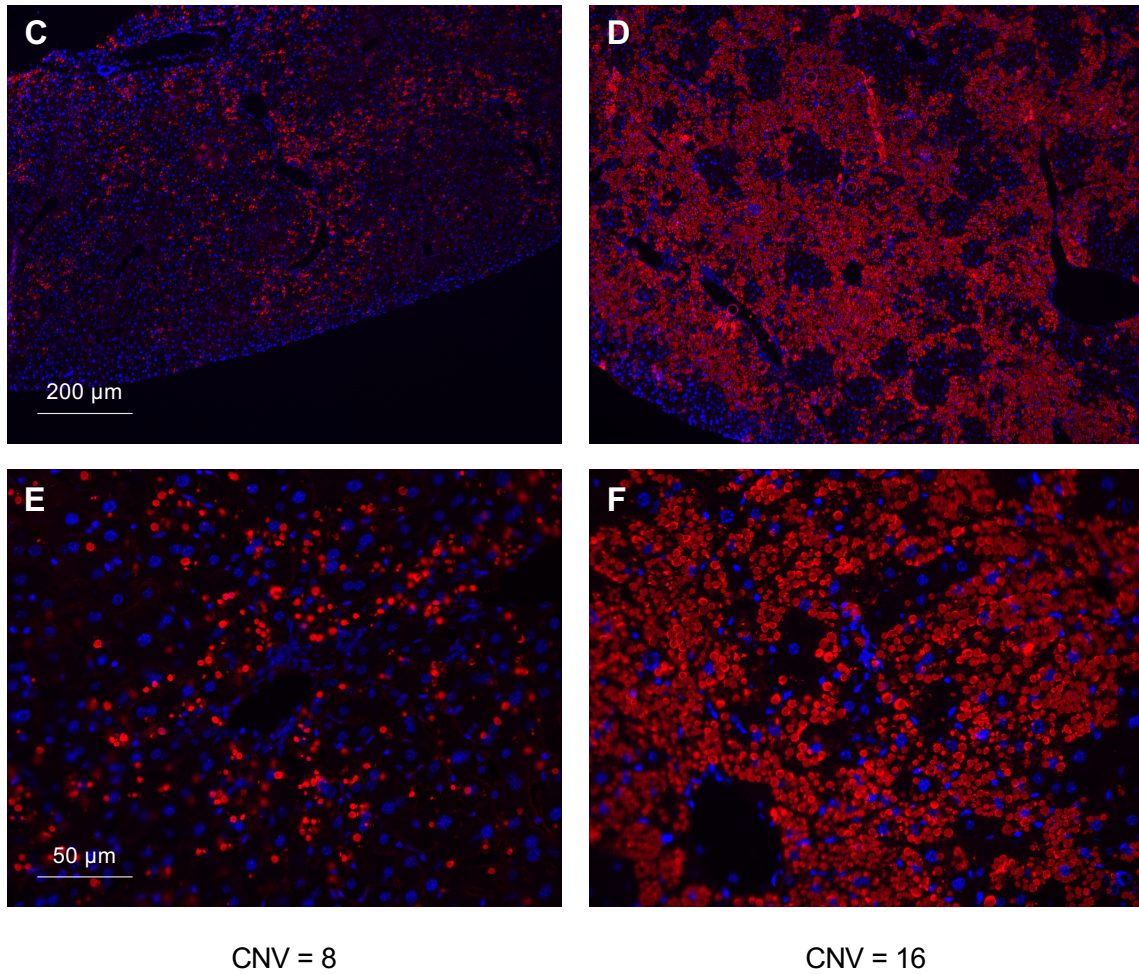


Figure 6.1 | Mendelian inheritance of the PI^*Z locus results in mice with 8 or 16 copies of PI^*Z . A one-dimensional ddPCR plot is shown in **A**, where the blue and green dots indicate the number of droplets positive for PI^*Z and the reference gene *Rpp30* respectively. The concentrations shown below were used to determine the CNV for each mouse. **B** shows the CNV of three litters from crossbreeding of parents with different CNV combinations indicated by the axis labels. When stained for anti-human AAT (**Section 2.4.1.2**) livers from female mice with heterozygous (CNV = 8; **C** and **E**) and homozygous (CNV = 16; **D** and **F**) genotypes, show higher AAT expression in homozygous mice (compare **C** with **D**, 4x; **E** with **F**, 20x). Greater magnification (20x) shows the PI^*Z polymer inclusions characteristic of PI^*Z liver disease [255] (**E** and **F**).

To sequence the PI*Z locus, high-molecular weight (HMW) DNA was extracted from the liver of a female PI*Z KI mouse using the 'salting out' method [375]. High molecular weight DNA was required so that the eight copies of PI*Z together with several kb of mouse genomic DNA either side could be obtained. The entire PI*Z locus with eight copies is estimated as ~110kb (**Figure 6.3A**) and the reported size of DNA extracted using the salting out method is ~200kb [376]. The size of extracted DNA could not be assessed for this study. TLA assessment of PI*Z KI mice located the insertion site of the PI*Z copies to the *Kcnh7* locus (potassium voltage-gated channel subfamily H member 7) [240]. This allowed for the design of gRNA target sites purported to be around the PI*Z integration site. As a heterozygous mouse (i.e. CNV = 8) would have one allele which contained PI*Z inside *Kcnh7* and one with wild-type *Kcnh7*, a homozygous mouse was chosen for this experiment to prevent wasting sequencing capacity on sequencing of the wild-type allele. Four gRNAs, two either side, were designed (**Section 2.7.2**) to target the PI*Z locus (**Figure 6.3A and B**) to allow for redundancy of cutting and ensure sequencing of the target DNA.

6.2.1.2.2 Method

The extracted HMW DNA (10 µg) was subjected to dephosphorylation with calf intestinal alkaline phosphatase (NEB) and heat inactivated before addition of the RNPs. Each gRNA (as annealed tracrRNA and crRNA, both purchased from IDT, Leuven, Belgium) was complexed with HiFi Cas9 (a high-fidelity Cas9 variant with reported low off-target cuts [187] commercialised by IDT) in an equimolar ratio. A total of 6.2 µM of complexed RNP was used to cut the four gRNA target sites per reaction. After cutting, the RNPs were heat inactivated and free ends were A-tailed using Taq polymerase (NEB). Adapters from the Ligation Sequencing Kit (SQK-LSK109, Nanopore) were ligated onto the A-tailed ends and the reaction was sequenced using a R9.4.1 flow cell on a MinION device (following the manufacturer's protocol for ligation-based sequencing). Base-calling was performed using Guppy version 3.1.5 for Windows with a Phred quality cut-off of >9.

To analyse the sequencing results, a theoretical DNA reference sequence was generated to resemble the PI*Z locus structure. This combined information from the original publication [255], which provided information on the size of the insert (whole PI*Z AAT locus plus 2.2 and 2.3 kb of genomic sequences 5' and 3' respectively), in-house ddPCR data indicating there were eight copies of PI*Z AAT and the integration location (*Kcnh7* with a 22 kb deletion caused by the insertion) from the TLA experiment in Bjursell et al. [240]. Phred quality score filtered sequencing reads were size-selected (custom python script, **Appendix 1**) and aligned to the reference sequence using Minimap2 [377], with the option to prevent multiple mapping events as the sequence is highly repetitive.

6.2.1.2.3 Results

Figure 6.4 shows the results of the CasEnrich sequencing experiment into the arrangement of the PI*Z locus. Screenshots (**Figure 6.4A and B**) from the IGV genome alignment software [378] of the left (5') and right (3') ends of the sequencing alignment are shown. The coverage plots (grey, on top) show a step-wise increase from each end inwards, indicating that the sequencing reads become more numerous from those sites, and coincide with the designed L1/L2 and R1/R2 Cas9 cut sites. This suggests that the designed Cas9 cut site methodology was able to facilitate Nanopore sequencing adapter ligation, and that the Nanopore device selectively started sequencing at the anticipated sites. The aligned reads, in blue and red below the coverage plot in grey (**Figure 6.4A and B**), show that sequencing was mostly directional, with red reads going from left to right and the blue reads vice versa. The fact that the majority of reads starting from the cut sites (~95%, 125 out of 131) are read in the direction as anticipated by the Cas9 gRNA designs suggests that the Cas9 complex was largely still bound to the other end of the cut DNA molecule (**Figure 6.2**, step 4) preventing sequencing from that end, explaining the observed preferred directionality. Sequencing coverage plots of the whole locus are shown in **Figure 6.4 C and D**. The reads are size-selected to show reads over 1000bp (**C**, to reduce calculation requirements and ensure alignment to *Kcnh7*) and 100,000bp (**D**, to only have reads that

potentially cover the entire locus) using a custom python script (**Appendix 1**). Per-base coverage numbers were calculated using the Bedtools genomecov function [379] and further binned in 100bp bin sizes. The coverage plot of the 1000bp size-selected reads shows a characteristic pattern, with the highest coverage close to the Cas9 cut site (average coverage is 36.4). This is expected as 1. not all DNA molecules will be the size of the entire locus and 2. Nanopore sequencing can be disrupted after an indeterminate sequence length, which together will tend to reduce coverage in the middle of the intended region of interest. Coverage plots for both the 1000bp and 100,000bp size-selected reads showed regular 'dips' in coverage. Upon further inspection, it transpired that these were small (40bp) deletions compared to the hypothesised PI*Z reference sequence. During the isolation of PI*Z AAT DNA (methods of [380] applied in [255]), the DNA was linearized with *EcoRI*. To achieve this, a pair of internal *EcoRI* sites present in the PI*Z gene were deleted from AAT intron 1 and the location of these coincided exactly with the 40bp deletions observed in the Nanopore sequencing data. Interestingly, the linearization preceding the microinjection step that was used to establish the transgenic line, was performed with a single restriction enzyme (*EcoRI*); this simplified preparation of DNA for the microinjection step, but could have contributed to the concatemerisation of PI*Z copies at integration, with the *EcoRI* digestion overhangs providing a ligation site on each end of the linear DNA molecules. In total, twelve Nanopore sequencing reads were identified that spanned the entire PI*Z locus. Mapping these reads to hypothesised reference sequences containing seven or nine copies of PI*Z resulted in large insertions and deletions in the read alignment. Manual inspection revealed that all twelve reads contained eight of the 40bp deletions noted above and that the PI*Z gene can be aligned to those reads a total of eight times. This provides strong evidence that the true copy number of the PI*Z locus is eight in a concatemeric fashion. Furthermore, the sequencing evidence suggests that all PI*Z genes are concatenated in the same direction.

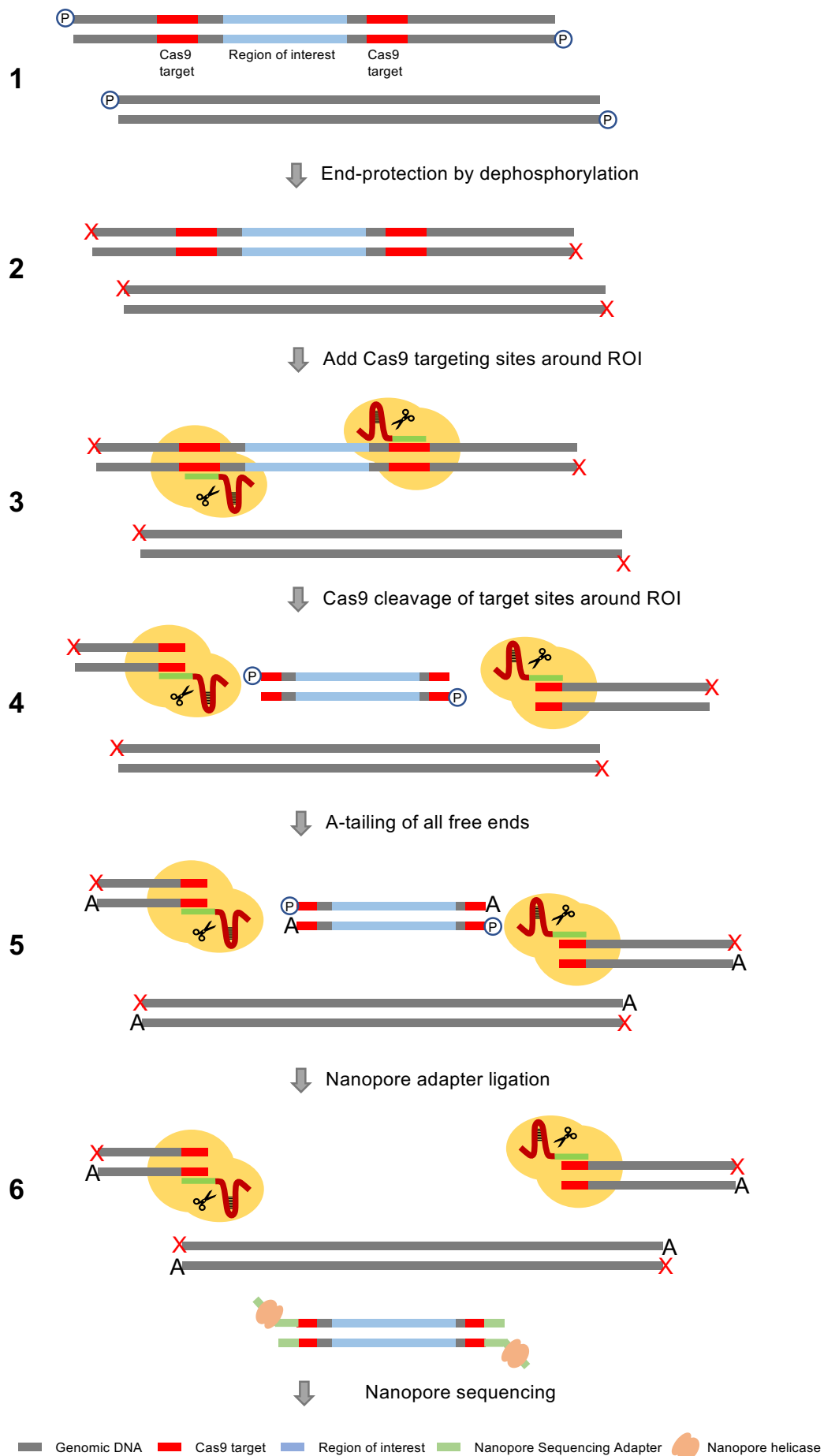


Figure 6.2 | Schematic of Cas9-mediated enrichment of *PI*Z* locus for Nanopore sequencing. High molecular weight DNA extracted from a tissue or organism of interest (1) is end-protected by de-phosphorylation (removes phosphate from blunt dsDNA ends), effectively preventing dsDNA ligation to those sites (2). Cas9 in RNP form is added to the end-protected DNA targeting sites around the region of interest (ROI) (3). The orientation of the gRNA target sites is of importance, as after Cas9 cutting the RNA will remain on the cut end where the 17bp of gRNA remains, i.e. the opposite site of where the PAM is located (4). The free blunt dsDNA ends are A-tailed, with the exception of the ends where Cas9 is still bound (5). Adapters for Nanopore sequencing are ligated to the ends that contain both a phosphate and adenosine overhang (6) ready for sequencing on a MinION device.

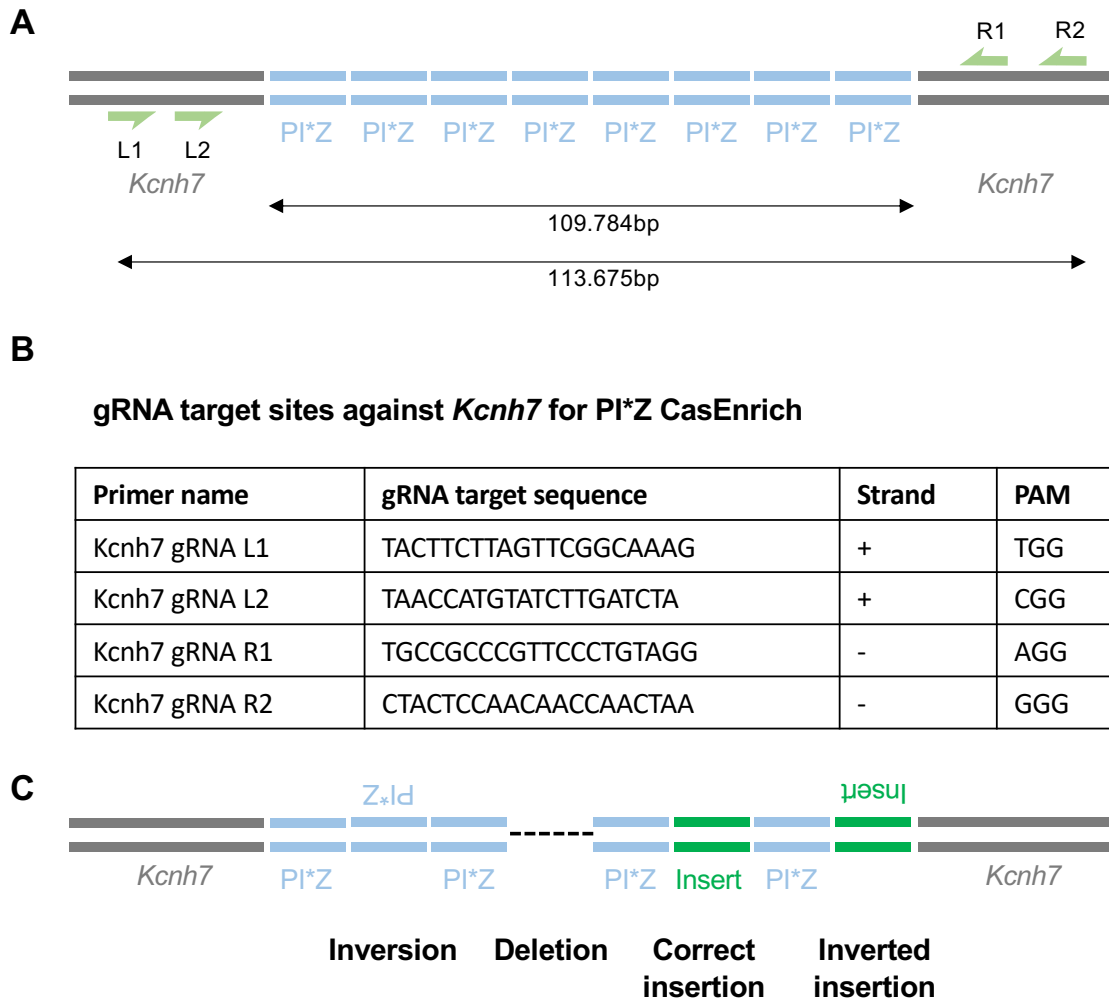


Figure 6.3 | PI*Z reference sequence and CasEnrich gRNA design. A schematic of the presumed PI*Z locus, as predicted by evidence from ddPCR and a report applying targeted locus amplification to the PI*Z locus [240] is shown in **A**. The gRNA target sites for CasEnrich of the PI*Z locus are shown in **B**, and visualised in the PI*Z construct using green arrows in **A**. Potential results of genome editing in the highly repetitive PI*Z locus are hypothesised in **C**. *Kcnh7*, potassium voltage-gated channel subfamily H member 7 locus; L1, L2, R1 and R2 stand for their respective gRNA target sites as shown in **B**.

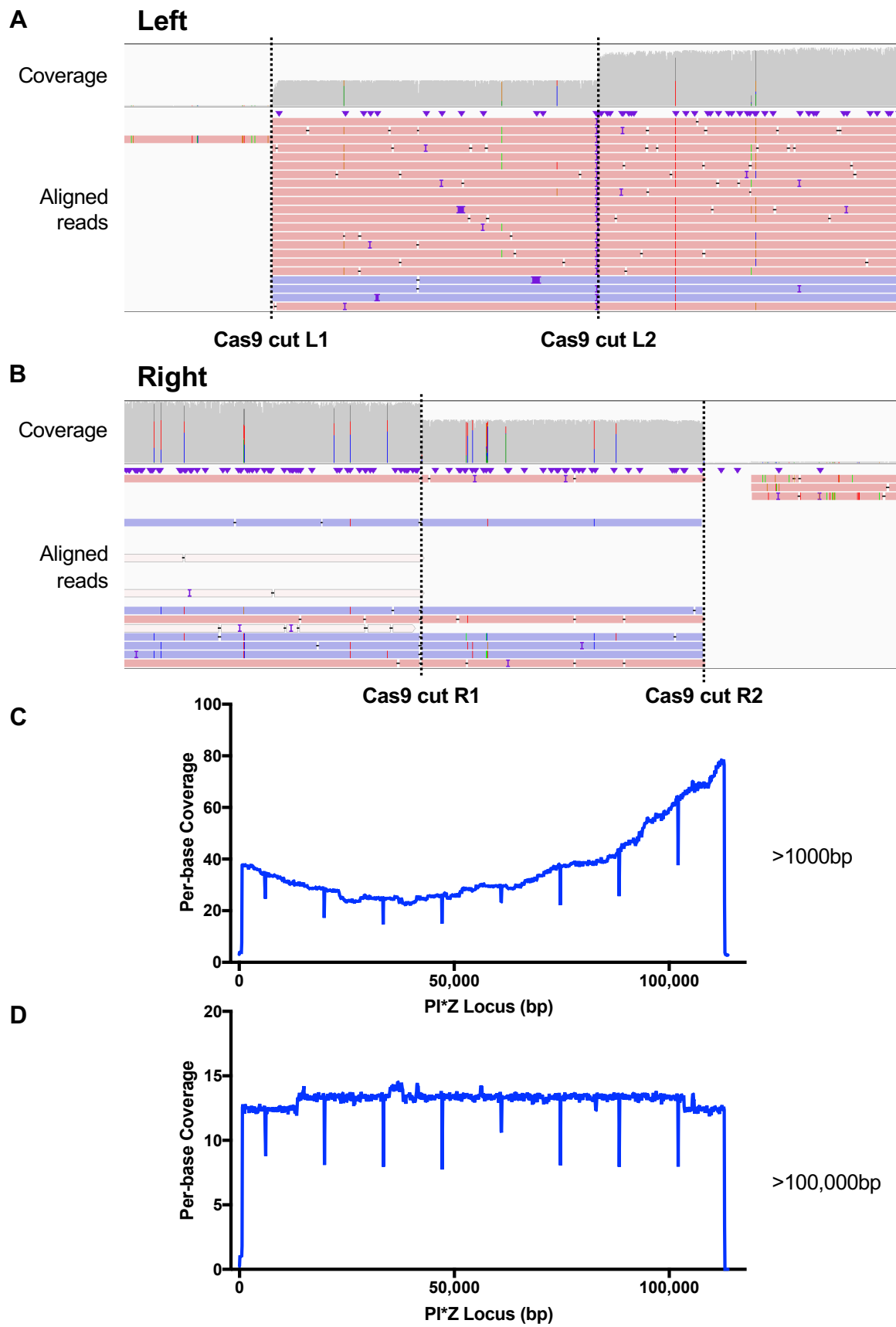


Figure 6.4 | CasEnrich mediated Nanopore sequencing of the *PI*Z* locus. A mouse with a CNV of 16 was sacrificed and high molecular weight DNA was extracted using the salting-out method [375]. This DNA was subjected to Cas9 mediated enrichment (**Section 6.2.1.2.2**) using the gRNAs shown in **Figure 6.3**. Coverage plots (screenshot from the genomic visualisation software IGV [378]) show the ‘stepping’ of coverage increases that coincides with the Cas9 cut sites on both the 5` (**A**) and 3` (**B**) sides of the *PI*Z* locus. The read direction indicated by blue (right to left) and red (left to right) colours show that Cas9 cutting induces directionality to the reads. Coverage plots of the entire locus using size-selected reads are graphed in **C** (reads >1000bp) and **D** (reads >100,000bp). The graph in **C** shows the CasEnrich characteristically high coverage close to the Cas9 cut site where the sequencing starts and where short reads are over-represented. When mapping the long reads only, a more even coverage plot is generated (**D**). The dips visible in both **C** and **D** are the result of a non-coding deletion in the first exon of the *PI*Z* sequence induced during the generation of the mouse line [255, 380]. Eight dips are observed, supporting the ddPCR data in **Figure 6.1B** that suggested that the mouse line has eight copies of *PI*Z* per chromosome. L1, L2, R1 and R2 stand for their respective gRNA target sites as shown in **Figure 6.4B**.

6.2.2 In vivo HITI-mediated mNeonGreen integration in PI*Z

With the genomic configuration of the PI*Z KI mice confirmed at eight copies of PI*Z per allele, the double-hit method described in Chapter 5 (**Figure 5.1**) was applied to the PI*Z KO mouse line. Importantly, the sequencing data of **Section 6.2.1** confirmed that the gRNA targeting exon 1 of AAT was present in the PI*Z locus. Two rAAV2/8 viruses were produced containing a genome expressing SaCas9 from the APoE/hAAT promoter (rAAV2/8 APoE/hAAT SaCas9, plasmid pGM715, **Table 2.2** and **Figure 6.5A**) and a HITI donor vector. The HITI donor used here contained a mNeonGreen cDNA plus transcription blocker placed between the two gRNA target sites for AAT exon 1 gRNA 2. Furthermore, a gRNA expression cassette for AAT exon 1 gRNA 2 and transcription blocker preventing ITR expression were placed before the left gRNA target site (based on Dual TB plasmid, **Section 4.2.5.1**), completing the genome (rAAV2/8 mNeonGreen HITI donor, plasmid pGM702, **Table 2.5** and **Figure 6.5A**).

At 6-11 weeks of age, PI*Z KI mice of both sexes were injected into the tail vein with both viruses, or with the HITI donor only, in the doses outlined in **Table 6.1**. Three weeks post-injection, the mice were sacrificed and indicated organs harvested (**Section 2.12.6**). DNA extracted from the livers of these mice were processed for ddPCR (**Section 2.9 and 2.12.2**) against *Rpp30* as a reference gene and the 5' integration junction as shown in **Figure 6.5B**. This ddPCR target should only amplify correctly integrated HITI donor sequences and the positioning of the probe should prevent signal in amplicons in which there has been a large deletion (although this was not verified). The plot in **Figure 6.5C** shows the result of the ddPCR per group expressed as fraction of integrations per genome (normalised against *Rpp30*), with each dot indicating a single mouse. The results suggest a dose-dependent increase in integration fraction between the 1e10 GC group and 1e11 GC groups, but this was not statistically significant (Kruskal-Wallis). Comparing the two groups of 1e11 GC dosed mice, differing in CNV only, the CNV = 16 group showed almost twice (0.31 to 0.60) the quantity of integrations. This could imply that a similar number of cells were targeted by

the rAAV2/8 vectors but that the greater availability of PI*Z targets allowed for higher integration levels. It is important to note that the fraction of integrations per genome does not necessarily correlate with the number of integrations per cell. This is because at eight PI*Z copies per cell, any cell could have between zero and eight insertions, meaning the fraction of integrations per genome should (at best) be seen as an average number of integrations per unit of DNA. This assumption is corroborated by images of liver cryosections visualising native mNeonGreen fluorescence in green and anti-human AAT in red (**Figure 6.6**). The images of mice in the 1e11 GC dosed groups with CNV = 8 (**A-D**) or 16 (**E-H**) show clear indication of mNeonGreen expression, providing evidence of correct integration and expression from the endogenous AAT promoter. However, while the number of mNeonGreen copies could not be quantified because of the irregular shape of the green fluorescent signal, the number of cells positive for mNeonGreen did not reflect the fraction of insertion per genome (0.31 and 0.60 for CNV = 8 and CNV = 16 respectively, **Figure 6.5C**). Contrary to the experiments in the Huh-7 cell line (**Figure 5.4A**), which contains only two copies of AAT, there is no clear evidence that HITI integration was able to completely abrogate PI*Z expression, as no mNeonGreen positive cells could be identified in which PI*Z was completely absent. Rather, holes in the mNeonGreen fluorescent signal (most clearly visible in the CNV=16 group, **Figure 6.6H**), suggested PI*Z polymers were present in cells successfully targeted by the HITI donor. Complete cellular abrogation of PI*Z expression was perhaps unlikely, as this would require integration of the HITI insert carrying the transcription blocker into all eight or 16 copies of PI*Z (depending on the treatment group) in any cell.

Table 6.1 | In vivo double hit HITI experimental groups

Group name	SaCas9 dose (GC per mouse) *	HITI donor dose (GC per mouse) **	Mice used
Naïve control	-	-	6 mice of CNV = 8, 3 females, 3 males
1e10 GC, CNV = 8	1e10	2.5e10	6 mice of CNV = 8, 3 females, 3 males
1e11 GC, CNV = 8	1e11	2.5e11	6 mice of CNV = 8, 3 females, 3 males
1e11 GC, CNV = 16	1e11	2.5e11	2 mice of CNV = 16, 2 males
Donor only control		2.5e11	2 mice of CNV = 8, 2 males

* rAAV2/8 APoE/hAAT SaCas9, plasmid pGM715, **Table 2.2** and **Figure 6.5A**. ** rAAV2/8 mNeonGreen HITI donor, plasmid pGM702, **Table 2.5** and **Figure 6.5A**. GC, genome copies.

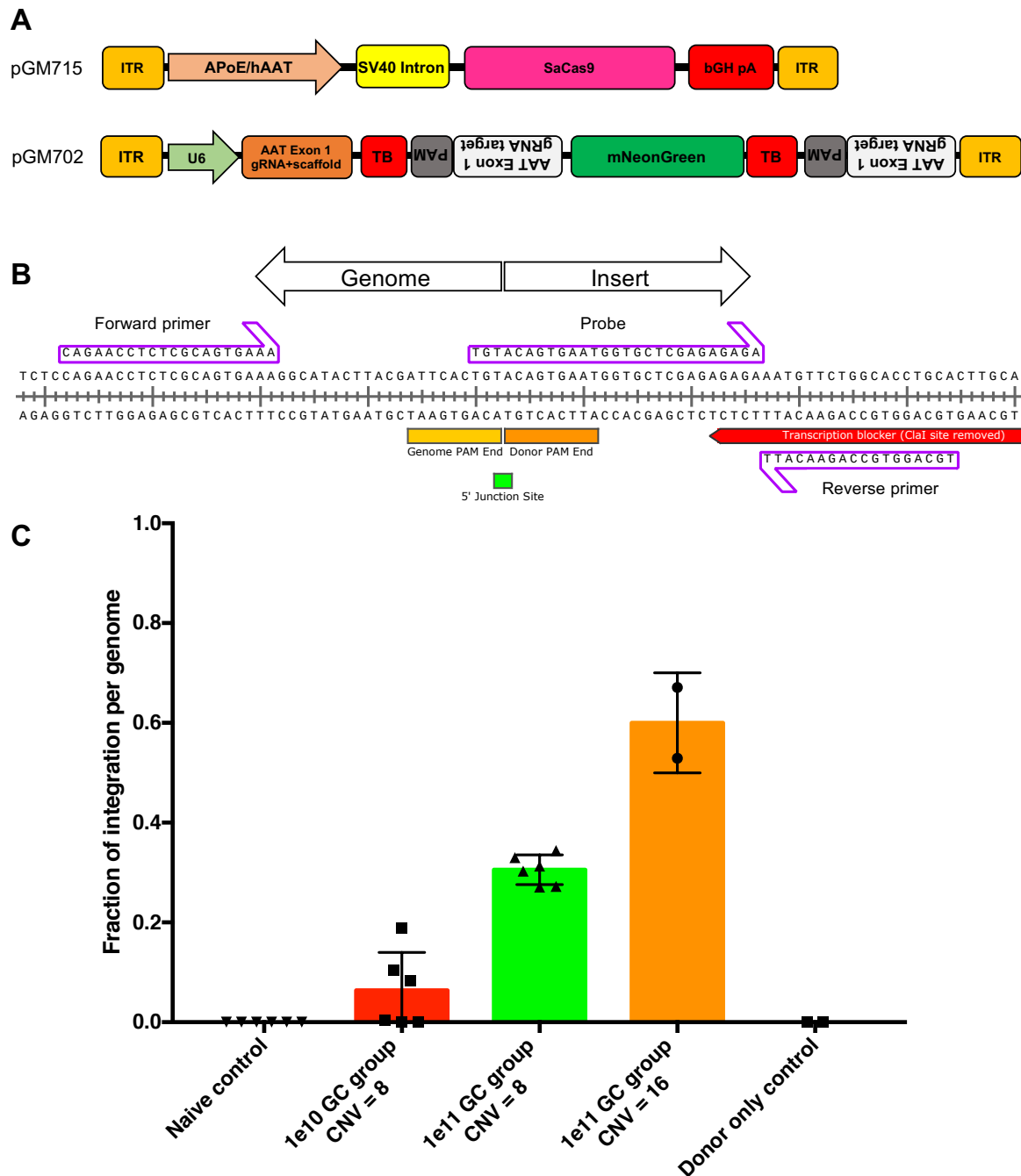
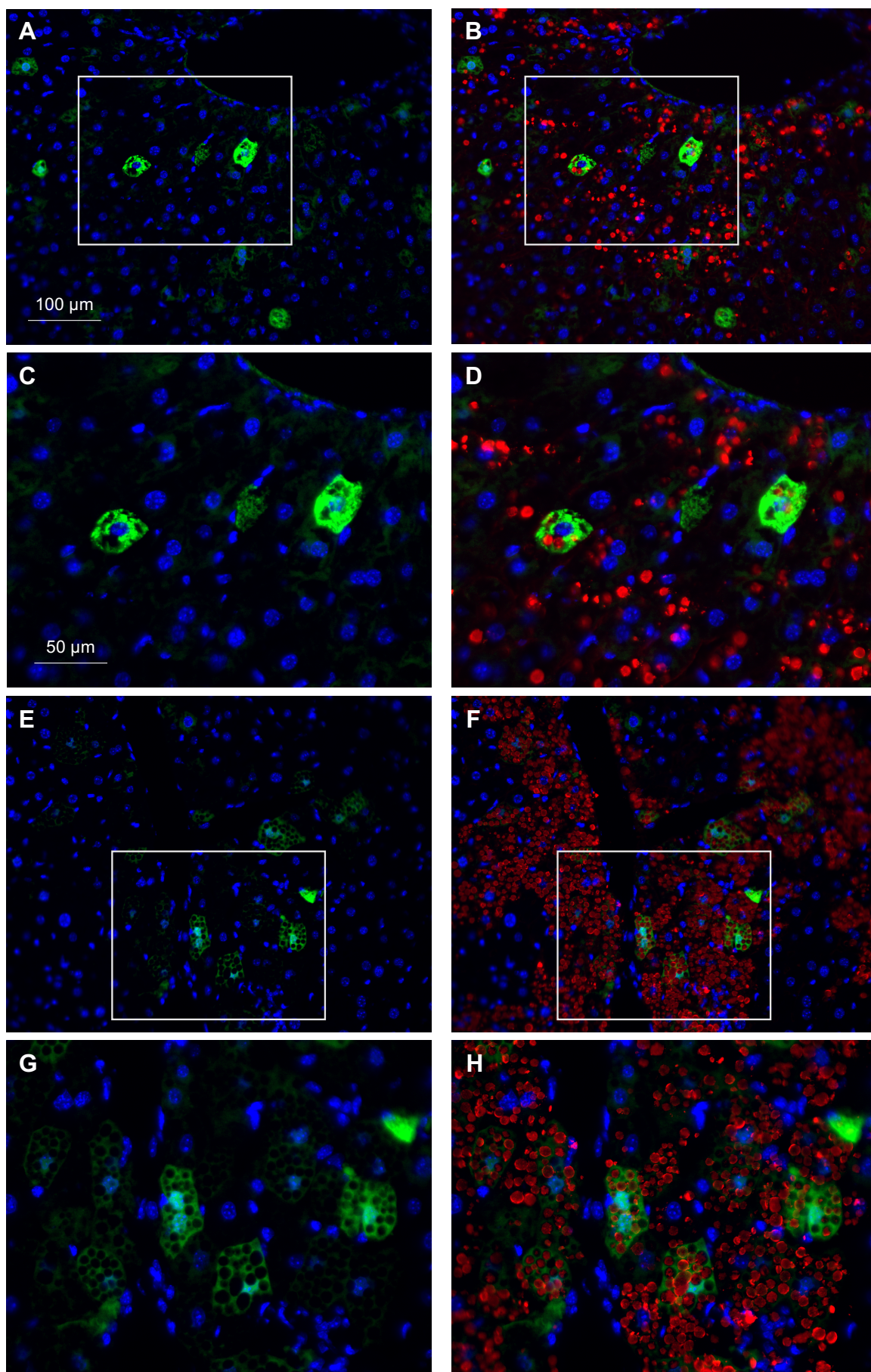


Figure 6.5 | Schematic of PI*Z HITI vector genomes and ddPCR assay for 5' HITI integration junction. Features of the AAV vector genomes used in the in vivo double-hit AAT study are shown in **A** (Table 2.2 and 2.5). The ddPCR assay shown in **B** covers the 5'-integration junction of correctly integrated HITI donor into the first exon of AAT. The ddPCR results of each group as summarised in Table 6.1 are shown in **C**, expressed as fraction of integration per genome, normalised against the reference gene *Rpp30* (Section 2.12.2). Values shown are mean per group ± SEM



mNeonGreen only

mNeonGreen and AAT

Figure 6.6 | Fluorescence microscopy images of in vivo HITI genome editing of the AAT locus. The livers of mice from the experiment in **Figure 6.5** were cryosectioned and stained with anti-human AAT (red) and DAPI (nuclei, blue). Green signal is native mNeonGreen expression. Images **A-D** are from representative sections in mice of group 1e11 GC, CNV = 8 and **E-F** show 1e11 GC, CNV = 16 (**Table 6.1**). Images in A-B and E-F were taken at 20x, and images of C-D and G-H are taken at a 40x magnification, where the white squares in the images taken at 20x indicate the location of the image at 40x directly below it. All pairs of images on left and right are of the same location, with the red channel showing AAT disabled in the images on the left for ease of visualisation of mNeonGreen, which can be obfuscated by the high expression of PI*Z AAT. The green signal in mNeonGreen positive cells appear to have 'holes' in it, which is more pronounced in the mice with CNV=16. The location of the holes coincides with the presence of PI*Z globules (red), which is the likely cause of the observed mNeonGreen pattern. Coincidentally, this also confirms that in these cells, PI*Z expression is not totally abrogated.

6.3 Discussion

6.3.1 PI*Z KI mice have eight copies of the AAT PI*Z gene per allele

The main aim of the studies in this chapter was to apply the double-hit HITI mediated genome editing strategy explored in a cell line in Chapter 5 to a mouse model of AATD. Discrepancies between reports about genomic make-up of the chosen model, the PI*Z KI mouse line, necessitated further investigation. Use of ddPCR genotyping indicated that the number of PI*Z copies in this mouse line was eight per allele, propagated in a Mendelian manner, suggesting that they were located in a small region of the genome (**Figure 6.1A and B**). Fluorescence microscopy showed that the PI*Z mouse resembles the human PI*Z liver phenotype [196] with globular inclusions caused by PI*Z polymerisation (**Figure 6.1C-F**). With this in mind a homozygous mouse (PI*Z = 16) was subjected to a Cas-mediated PCR-free enrichment (CasEnrich) protocol developed by Oxford Nanopore to allow sequencing of the entire genomic locus of PI*Z (**Figure 6.2**). The gRNA targets designed to flank the PI*Z locus (**Figure 6.3**) allowed for the full PI*Z locus to be sequenced in single reads. This provided evidence to support the hypothesis that the PI*Z locus comprises eight copies of the PI*Z gene, concatenated in a single direction (**Figure 6.4**).

6.3.2 The double-hit AAT approach using rAAV2/8 is achievable *in vivo*

Following verification of the PI*Z KI mouse line above, the double-hit approach was investigated *in vivo*. Contrary to transfection of Huh-7 cells with plasmid DNA (**Section 5.2.1**), the genome editing of the mouse line utilised the rAAV2/8 vector combined with liver-specific SaCas9 expression from the APoE/hAAT promoter as discussed in Chapter 3 (**Figure 3.5** shows liver-restricted transgene expression from rAAV2/8 APoE/hAAT Lux). This approach aimed to limit genome editing to the liver by restricting expression of SaCas9. Similarly, the gRNA expression cassette normally present in the SaCas9 expression vector was instead placed into the HITI donor (**Figure 6.5A**). In this way, only cells which both the SaCas9 expressing vector and HITI donor vector had transduced would be amenable to

genome editing. This also prevents spurious SaCas9 cutting and cNHEJ induced mutagenesis of exon 1 of AAT when no HITI donor was present.

The delivery method combined with the double hit genome editing approach proved able to integrate the mNeonGreen cassette into the first exon of AAT PI*Z as measured by ddPCR at the 5' integration junction (**Figure 6.5B and C**). There was a trend towards dose dependency with the higher dose (1e11) showing higher integration efficiency compared with the low dose (1e10 GC, both CNV = 8, **Figure 6.5C**). Dosing mice that are homozygous, i.e. with a CNV of 16, increased the integration fraction by roughly twofold (**Figure 6.5C**) although the number of mice used (n = 2) is too low to make any generalised conclusions. Both groups receiving the high dose showed native mNeonGreen expression in cryosections of the liver (**Figure 6.6**). The number of PI*Z copies per haploid genome make complete abrogation of PI*Z expression in a given cell very challenging and, as expected, this was not observed. In order to achieve a potential cure for AATD, a serum AAT concentration of 11 μ M is required (**Section 1.3**), which is roughly 25% of normal. The assumption was made that to reach this level using the endogenous AAT promoter (rather than a strong artificial promoter), 25% of cells must be successfully edited. While the percentage of cells positive for mNeonGreen was difficult to quantify, it was clear that it did not reach 25%. If this experiment were to be repeated with a PI*M AAT cDNA in place of mNeonGreen, the AAT expression would be unlikely to reach the 11 μ M required for therapeutic benefit.

Considering the complexity of the PI*Z locus in PI*Z KI mice, it is expected that genome editing has some unintended consequences. Using an endonuclease such as SaCas9 to cut eight times in close proximity within a stretch of DNA could lead to rearrangements such as deletions and inversions at the locus (**Figure 6.3C**). Similarly, integrating a dsDNA sequence (as attempted here) could lead to inversions, multiple integrations in concatemers and other recombination events. To address this, the prospect of repeating the CasEnrich

method with genome edited mouse liver was raised. Unfortunately, neither the salting-out method used previously, nor the Qiagen MagAttract HMW DNA Kit, proved satisfactory to extract sufficiently high-molecular weight DNA to warrant another CasEnrich experiment to elucidate this. This could be due to the fact that the mouse livers were frozen at -80 °C shortly after harvesting (**Section 2.13.6**) and freeze-thawing affected the integrity of the DNA.

6.3.3 Conclusion

This chapter sought to apply the HITI-mediated integration method to an *in vivo* model of AATD, the PI*Z KI mouse line. This proved successful as evidenced by ddPCR analysis for genomic insertion and expression of a fluorescent reporter gene. However, the currently observed levels of efficiency likely warrant further optimisation to reach therapeutic levels of AAT expression.

Chapter 7 – Analysis of off target sites for the HITI approach

7.1 Introduction

When investigating a therapeutic approach based on genome editing one must anticipate accidental editing at sites other than intended, which is especially important when using targeted endonucleases to induce DSBs to increase the chance of successful editing. The potential for that endonuclease to cut elsewhere in the genome resulting in mutations should not be underestimated [381]. The propensity of these endonucleases to induce off-target cuts, discussed in **Section 1.2.5**, is largely influenced by their enzymatic action and the chemistry involved in binding and cutting the DNA. The detection of off-target cuts in itself, which is widely studied in the field, is beyond the scope of this study. However, since the aim of this study is to integrate a dsDNA sequence into the genome, the potential of off-target integration must be addressed.

The main methods to detect off-target genome editing focus on the detection of nuclease-induced dsDNA breaks (**Figure 1.7**), but these are not readily amenable to finding off-target integrations, especially with large (>1kb) sequences integrated into the genome using the HITI method. Two methods have been described that could be used, specifically, Uni-Directional Targeted Sequencing (UDiTaS) [382] applied to large inserts in [363] and targeted locus amplification (TLA, [372, 383]). Schematics of these methods are shown in **Figure 7.1**. Both strategies rely on amplification of a genome region adjacent to the insert using PCR primers complementary to a region within the insert. This ensures specific amplification and enrichment of fragments of interest.

7.1.1 Aims

The aim of this chapter is to investigate the potential of two protocols, Capture-seq and CasEnrich, to detect off-target integration of dsDNA sequences inserted by HITI.

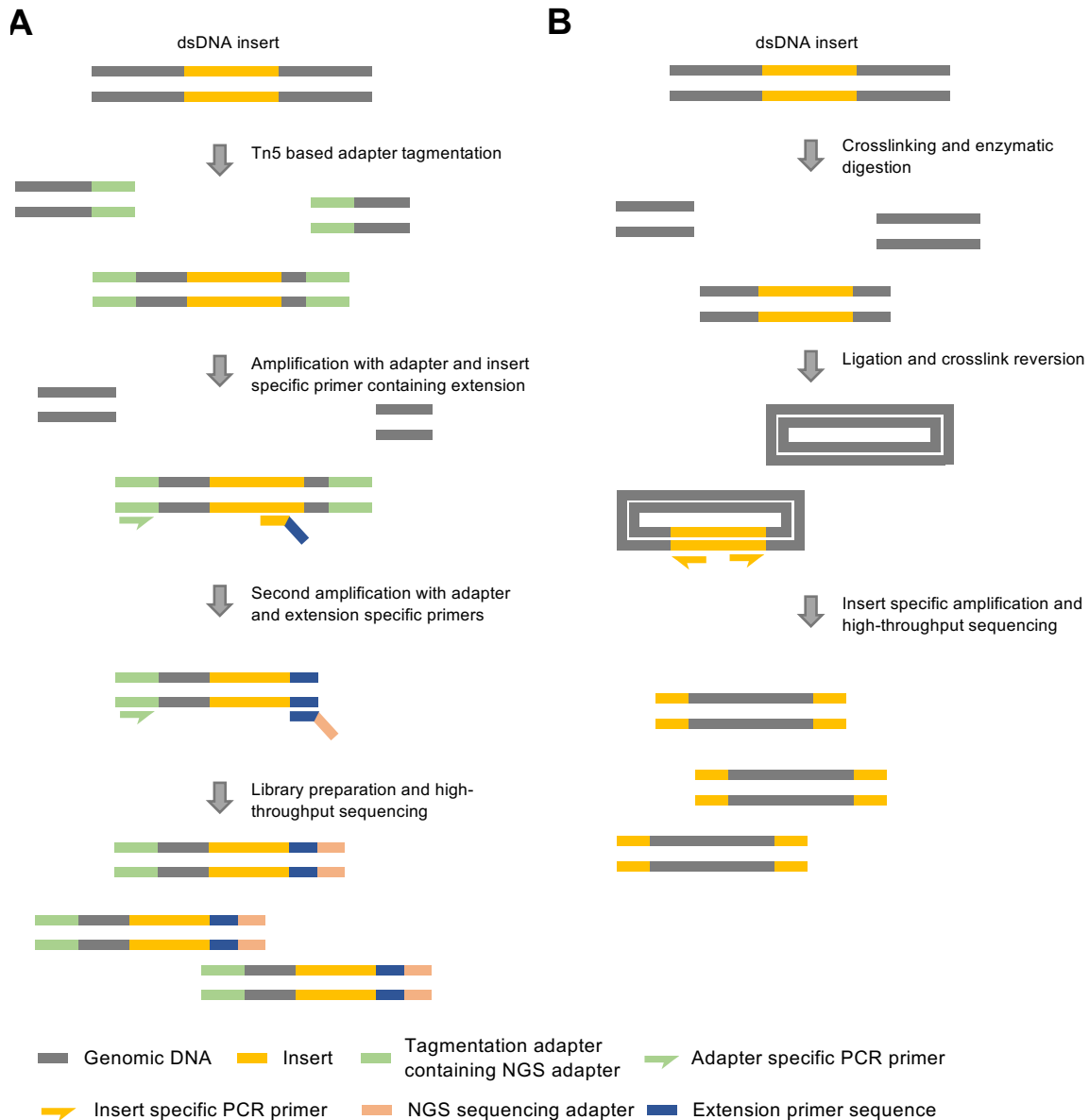


Figure 7.1 | Currently available methods of detection of dsDNA insert location.

UDiTaS and targeted locus amplification (TLA) are used to assess insertion sites of dsDNA inserts [372, 382]. UDiTaS (**A**) works by tagmentation of adapters into the DNA and targeted amplification using primers to the insert and adapter to enrich for sequences containing insert and genome. In TLA (**B**) the DNA is cross-linked and digested with restriction enzymes before ligation into circular dsDNA molecules. PCR primers specific to the insert sequence allow amplification of either side of the circular template creating amplicons that contain genomic sequences with insert sequences on their ends. Tn5, transposase.

7.2 Results

7.2.1 Off-target integration detection by Capture-seq

7.2.1.1 Background

The Capture-seq method uses complementary biotinylated oligos to enrich for sequences of interest (DNA or RNA) using streptavidin-coated magnetic beads and washing away contaminating DNA. In theory, this removes non-specific ‘noise’ from DNA sequencing results and frees up sequencing capacity to provide higher sequencing coverage of the regions of interest [384]. The Capture-seq protocol was performed in collaboration with James Davies (WIMM, University of Oxford, UK) using an adapted Capture-C protocol [385] without preparation of the 3C libraries, and instead using unmanipulated genomic DNA, starting with the sonication step. A schematic highlighting the main steps of Capture-seq for investigation into off-target integrations is shown in **Figure 7.2A**. Of note is the sequence of the biotinylated oligos, which are designed to be complementary to the ends of the HITI donor insert, at the edge of the integration junction (**Figure 7.2B and D**). This should result in the capture of DNA molecules that contain both a piece of the HITI insert and adjacent genome (**Figure 7.2D**) and can then be used to determine the location of the insert, whether on- and off-target.

7.2.1.2 Method

A Huh-7 cell transfection experiment utilising the HITI donor plasmid (pGM565, **Table 2.5**) targeting AAT exon 1 was performed similar to those described (**Section 5.2.1**), except that the cells were cultured for 1.5 months to maximise loss of non-integrated plasmid from the cells by dilution. Cells (500,000) were transfected with SaCas9 plasmid only (pGM551, **Table 2.3**), Donor plasmid only (pGM565, **Table 2.5**), both SaCas9 and Donor plasmids, or left un-transfected. DNA was extracted using a QIAamp kit (Qiagen) and 2 µg from each experimental sample (~2.2e5 genomes assuming a chromosome count of 60 for Huh-7 cells [360]) was taken forward. The DNA was processed according to the protocol [385] using different indexing primers on each sample to allow for multiplexed sequencing. The resulting

captured DNA was pooled, and an Illumina NGS library was prepared for sequencing on a MiSeq machine (Illumina, San Diego, USA) using a 2 x 150bp paired-end read MiSeq Reagent Kit v2 (Illumina) according to the manufacturer's instructions. The sequencing adapters from de-multiplexed sequencing reads were trimmed from the reads using Trim Galore [386] and paired-end reads were joined using FLASH [387]. PCR duplicates arising during amplification of the captured reads (**Figure 7.2A** step 4) were de-duplicated using a custom script written in python (**Appendix 2**). On average, this reduced the number of reads by ~23-fold, indicating the original captured sequences were amplified ~23 times. The sequences were then aligned to the insert sequence using Blat [388]. Because the output of the Blat alignment is a *.psl format file in which information such the sequence and Phred score is removed, a custom perl script (written by James Davies) was employed to recreate a *.fastq format file containing all the sequencing information necessary for the next step; which involved mapping of the sequences which show alignment to the HITI insert (i.e. sequences that contain some of the HITI insert sequence and potentially some of the genome) to the human reference genome (hg19) using Bowtie2 [389]. The resulting *.sam file was indexed for use with alignment software such as IGV [378] using SAMtools [390]. The bioinformatics steps taken, and specific commands are shown in **Appendix 3**. Read counts for a given locus were determined by counting the reads in a 2000bp window using a custom python script (**Appendix 4**).

7.2.1.3 Results

An ideogram of human chromosomes 10 and 14-16 and associated coverage plots (red) shown in **Figure 7.3** represent the quantity of reads containing the HITI insert sequence and adjacent genomic sequence aligning to that part of the human genome. Greater coverage signifies more reads sequenced for that region and is an indication that more of those sequences were present in the original genomes. Since these reads contain both the insert sequence and genomic sequence, it is likely that an insertion has occurred at this location. Three peaks that are considerably higher than any other are visible on

chromosomes 10, 14 and 16, indicated with green boxes A, B and C respectively. Box B corresponds to the SERPINA1 locus, the intended target site of the double-hit HITI strategy of Chapter 5; this indicates that there is a significant quantity of reads in that locus, indicative of correct integration. Box C is in the HBA1 (alpha-globin) locus, and whilst initially identified as a potential off target, it transpired that the 3' capture oligo (**Figure 7.2B and C**) encompasses part of the transcription blocker, which contains the transcription pause signal derived from the alpha-globin locus [339]. Since this sequence will be both in the insert and in the human genome, it is expected to be efficiently captured by the capture oligos, aligned to the insert and subsequently mapped to the genome, providing a useful technical positive control. In contrast, chromosome 10 does feature a potential off-target integration site (**Figure 7.3A**). Looking closely at both the on-target (SERPINA1, **B**) and off-target (intergenic, between HIF1AN and PAX2, **C**) regions shows that both sites have a distinct bi-lobar sequence coverage. The observation that the coverage was asymmetric at both sites could indicate that one of the capture oligos was more efficient in capturing the insert than the other, such that one of the integration junctions has increased representation. Interestingly, while the SERPINA1 site has the SaCas9 target site directly in the middle of the two lobes, the chromosome 10 off-target only has potential SaCas9 target sites just outside of the proposed cut site (orange arrows in **Figure 7.3**, using NNGRR rather than NNGRRT as the SaCas9 PAM, which has been reported to allow cutting but at lower efficiency [102]). One would expect the cut site to be in the middle of the lobes as predicted for the insert sequence (the insert sequence is not present in the reference genome used for the mapping, so the reads are clipped there). Further investigation into potential off-target sites, using read counts in 2000bp windows throughout the genome (using a cut-off of eight reads, which is the average number of reads per window in the no-transfection control) produced the results shown in **Table 7.1**. Here, the number of reads per window was divided by the total number of reads over the entire genome (minus reads at the HBA1 locus) to obtain the percentage of reads per locus. The on-target integration into the SERPINA1 locus accounts for ~60% of the reads counted, while the off-target at

chromosome 10 makes up another ~10%. Other loci positive for integration accounted for 2% of reads or less each (**Table 7.1** shows the 10 loci with the highest read percentages, out of a total of 178 loci). Interestingly, manual analysis of all *in silico* predicted SaCas9 off target sites (using CRISPOR [252]) revealed no evidence of integration at any of those loci (**Appendix 6**).

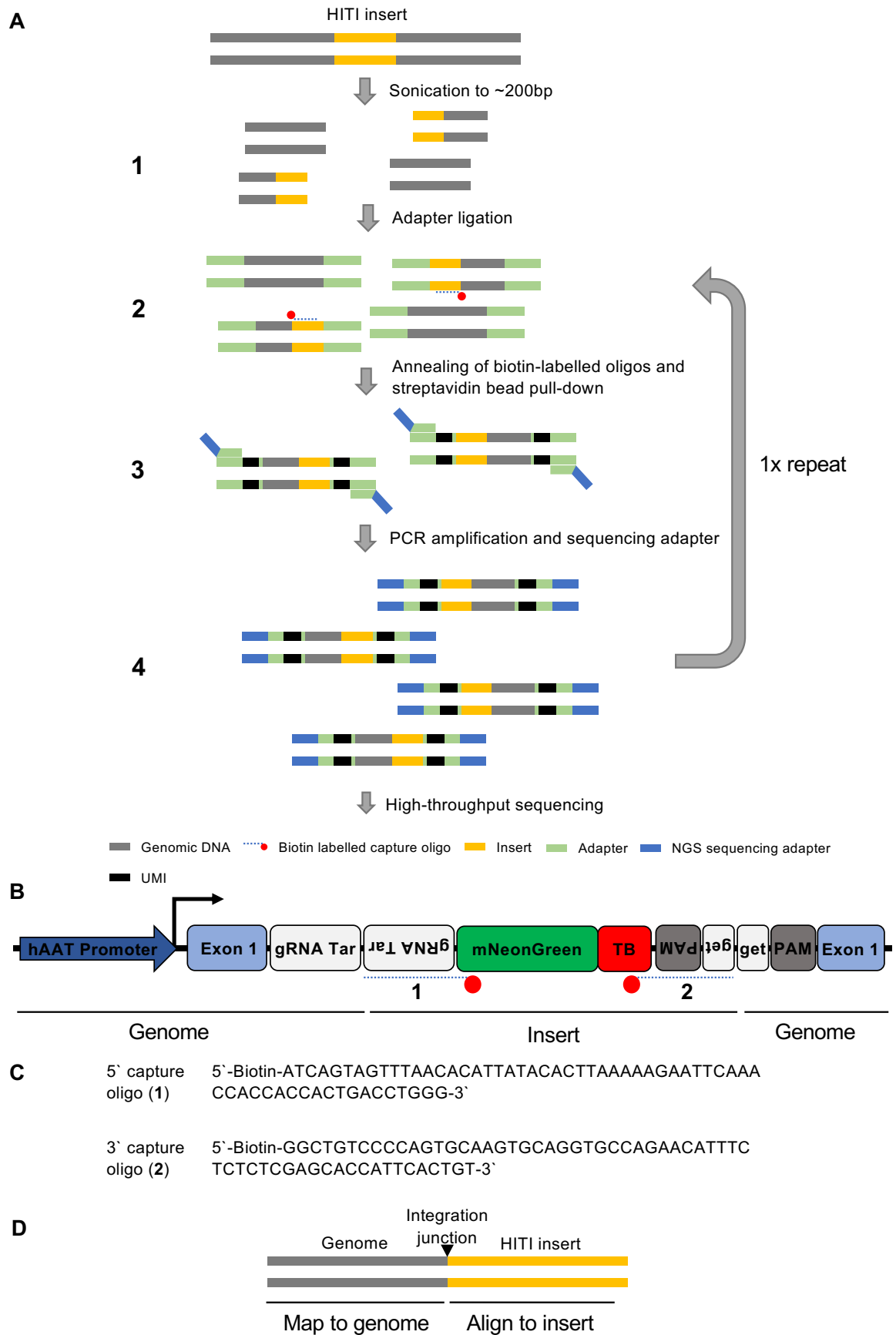


Figure 7.2 | Schematic of Capture-seq protocol for dsDNA inserts. Cellular genomic DNA is extracted from the transfected (e.g. Huh-7) cells and subjected to Capture-seq (**A**). DNA is sonicated to ~ 200bp fragments (**1**), end-repaired, A-tailed, and universal adapters are ligated onto the ends (**2**). These adapters contain a unique molecular identifier (UMI) sequence of 3bp, but the adapter ligation is random both sides and thus are likely to have a different UMI, increasing the UMI size to 6bp. Biotin-labelled capture probes (shown in **B** with sequence in **C**) are hybridised to the target DNA and enriched using streptavidin coated magnetic beads (**3**) ostensibly capturing reads that contain both insert and genomic sequence (**D**). The captured DNA is subjected to PCR using primers specific to the adapter sequence (**4**). The capture and amplification steps are repeated once to increase the number of unique sequences that can be sequenced [385]. After the second round of capture and amplification the DNA is sequenced on a MiSeq machine.

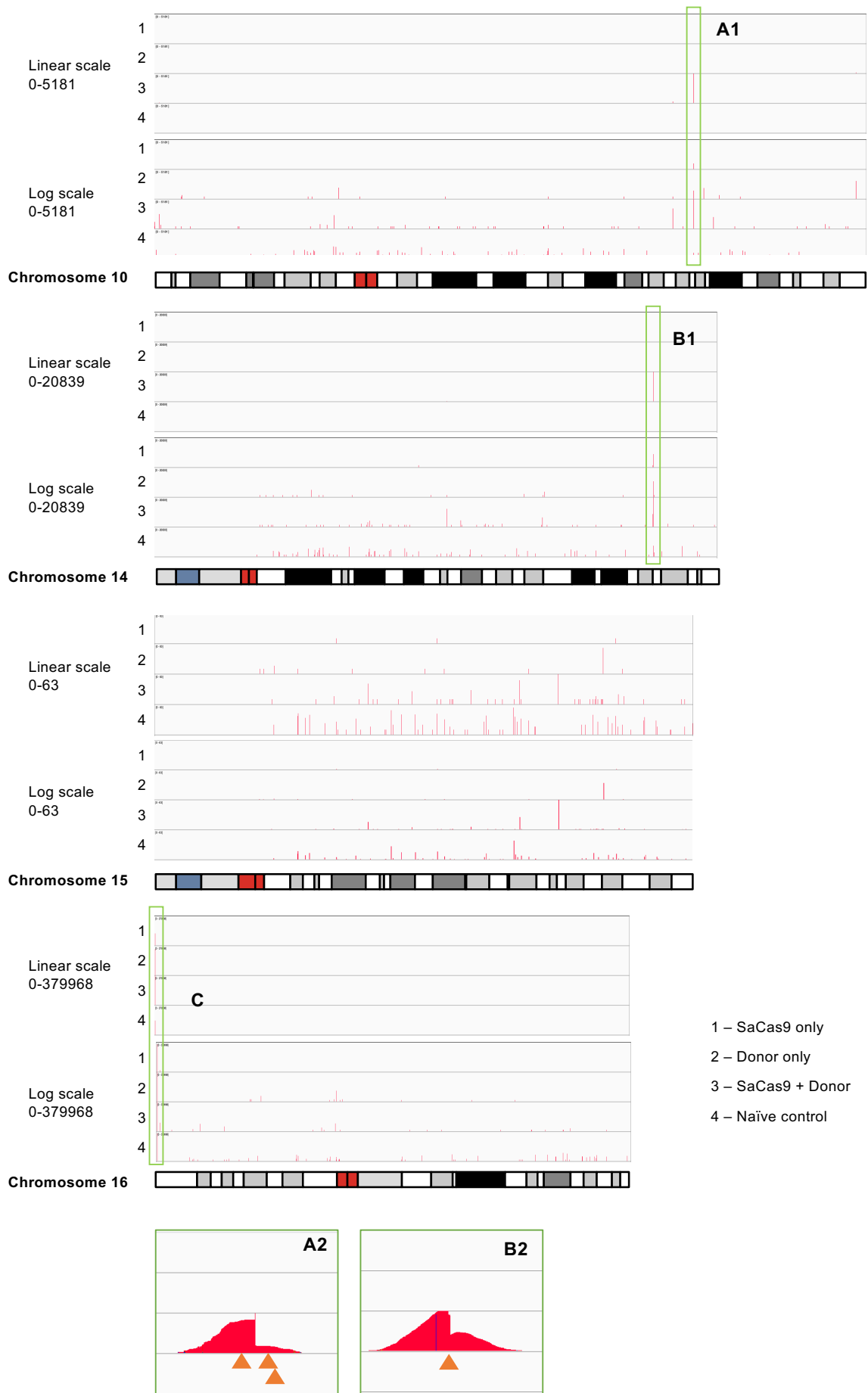


Figure 7.3 | Ideogram of selected chromosomes of Capture-seq. Ideograms of chromosomes 10 and 14-16 are shown with the corresponding coverage plot above them, showing the coverage of the samples processed: SaCas9 only, Donor only, SaCas9 + Donor and naïve, numbered 1-4 respectively, using both logarithmic (lower) and linear (upper) panel scales. The range of the scale differs between chromosomes (indicated left of each coverage plot). Note for example chromosome 15, where no sample reached a coverage of >63, indicating there is little-to-no evidence for off-target integration detectable in this chromosome; while chromosome 10 has a scale of 5181, as there is a high peak, highlighted by the green box (**A1**). When zooming in using the IGV genome alignment software, there is a clear peak visible in the sample that was transfected with both SaCas9 and Donor (**A2**, corresponding to peak in **A1**). Orange arrowheads correspond to potential SaCas9 target sites (NNGRR PAM used). Notice that the SaCas9 target sites are not in the middle of the bi-lobar sequence coverage in a similar way to the on-target SaCas9 target site in SERPINA1 (**B2**, the zoomed in view of the peak in **B1**). The peak in chromosome 16 (**C**) is located in the alpha-globin locus and is caused by capture of the genomic sequences that are also present in the insert as part of the transcription blocker that are present in the 3' capture oligo (**Figure 7.2B and C**). Both captured genomic and insert sequence containing the alpha-globin sequence will map to that region.

Table 7.1 | Capture-Seq integration sites with greatest frequency

Chromosome	Start	End	Number of reads	Percentage of total reads (%)	Location in human genome
chr14	94854001	94856000	31177	60.03	SERPINA1 Exon 1
chr10	102328001	102330000	5241	10.09	Intergenic: HIF1AN-PAX2
chr2	112962001	112964000	928	1.78	Intergenic: FBLN7-ZC3H8
chr6	48688001	48690000	778	1.49	Intergenic: PTCHD4-MUT
chr11	126932001	126934000	676	1.30	Intergenic: KIRREL3-AS3-LOC101929473
chr7	119272001	119274000	611	1.18	LVCAT5 intron 1
chr6	102844001	102846000	598	1.15	Intergenic: GRIK2-HACE1
chr4	52912001	52914000	569	1.10	Intergenic: SGCB-SPATA18
chr7	71136001	71138000	483	0.93	WBSCR17 intron 8
chr8	38292001	38294000	450	0.87	FGFR1 intron 2

7.2.2 Off-target integration detection by CasEnrich

7.2.2.1 Background

The CasEnrich method was used previously in this study to address the genomic configuration of the PI*Z locus in PI*Z KI mice (**Section 6.2.1.2**), which were thought to have eight copies of the PI*Z AAT gene concatenated at a single genomic location, and CasEnrich was used to enrich for sequences that spanned the entire region (~110kb). Following this successful use of CasEnrich, the approach was applied to investigation of off-target integration in HITI experiments. Instead of cutting outside of a region of interest and sequencing inwards (the approach used for the PI*Z AAT locus sequencing), the approach relied upon Cas9 cutting the HITI insert sequence itself and sequencing outwards into the surrounding genomic locus (**Figure 7.4 and 7.5B**). Thus, similar to Capture-seq (**Figure 7.2D**), the anticipated sequencing reads would contain DNA of both the HITI insert and directly adjacent genomic DNA.

7.2.2.2 Methods

The DNA used for this experiment was derived from pooling the DNA from all six mice in the 1e11 GC CNV=8 group described in **Section 6.2.2**. These mice had been dosed with 1e11 GC of rAAV2/8 virus expressing SaCas9 from the APoE/hAAT promoter (pGM715, **Table 2.2 and Figure 6.5A**) and 2.5e11 GC of rAAV2/8 virus containing an mNeonGreen HITI donor with appropriate gRNA target sites (pGM702, **Table 2.5 and Figure 6.5A**). The DNA was pooled and concentrated using AMPure XP beads (Beckman Coulter) by binding to 1:1 volume of DNA:beads, washing twice with 70% ethanol and eluting in a 5x reduced volume. This concentration step was required to generate sufficient DNA (20 µg) for the CasEnrich protocol, a schematic of which is shown in **Figure 7.4**.

A TapeStation trace (Agilent Genomic DNA ScreenTape, Agilent, Waldbronn, Germany) showing the sizes of the DNA (comparable to an agarose gel) was performed to ensure that the extracted DNA was of sufficient length to warrant progressing to the next step (**Figure**

7.5A). As indicated (**Section 6.3.2**), extracting HMW DNA from the frozen livers of these mice was challenging, and the size of the DNA shown in **Figure 7.5A** was at an average of 28kb likely too short for a CasEnrich experiment to encompass the entire PI*Z locus in the PI*Z KI mice (at ~110kb). However, the 28kb average was deemed long enough for the experiment as depicted in **Figure 7.4** as only a small part of the insert, and sufficient genomic DNA to allow successful mapping to the reference genome, would be adequate for assessment of off-target integration. Similar to the CasEnrich methods (**Section 6.2.1.2.2**) the DNA was end-protected with intestinal alkaline phosphatase. Subsequently, alkaline phosphatase activity was inhibited by heat inactivation. In an equimolar ratio, HiFi Cas9 (IDT) was complexed with gRNAs against the insert (**Figure 7.5 B and C**) and incubated with the processed DNA at 37 °C for 60 minutes. Four sites were chosen in mNeonGreen, with two on each side (for redundancy) with the gRNA sites oriented to initiate sequencing outward into the genomic DNA (**Figure 7.5B**). After heat inactivation of the RNP complex, the DNA was A-tailed and Nanopore ligation adapters (Ligation Sequencing Kit, LSK-109, Nanopore Technologies) were ligated onto the A-tailed ends. The reaction was sequenced on a MinION device using a R9.4.1 flow cell (following manufacturer's protocol for ligation-based sequencing) and base-calling performed using Guppy version 3.1.5 for Windows with a Phred quality cut-off of >9.

Analysing the results of the CasEnrich Nanopore sequencing reads proved more complicated than both the Capture-seq and PI*Z CasEnrich experiments, partly due to HITI-mediated insertion into the highly repetitive PI*Z locus. Any combination of insertion, inversion and deletion could, in principle, occur in the repetitive PI*Z locus with or without insertion of the HITI donor (see **Figure 6.3C**). In addition, as the DNA molecules were not long enough for CasEnrich extending from *Kcnh7* into the edited locus as in **Section 6.2.1.2**, there was no reliable genomic context from which to anchor the sequencing. Similar to Capture-seq, the reads were first aligned to the HITI insert sequence using Minimap2 [377] and subsequently reads that contained at least part of the HITI insert sequence were

mapped to the mouse reference genome (mm9). Specific commands used can be found in **Appendix 5**. Unlike the Capture-seq approach, no de-duplication was necessary since no amplification steps are undertaken in the CasEnrich protocol.

7.2.2.3 Results

Similar to the CasEnrich experiment in Chapter 6, the coverage plot of reads aligned to the HITI donor insert showed a stepwise increase in coverage from the Cas9 cut sites, although this time in the opposite direction (compare **Figure 7.5D** with **Figure 6.4 A and B**). This implies that Cas9 has successfully cut the gRNA target sites in mNeonGreen and that, as anticipated, sequencing started from the gRNA cut sites progressing out of the HITI insert sequence towards the encompassing genome. There were no sequence reads containing the HITI insert that mapped to the mouse reference genome, suggesting that no off-target integration had occurred in the mouse experiments described in **Section 6.2.2**. This is in disagreement with the results of the Capture-seq approach which identified that off-target integrations did indeed occur. This suggests that the CasEnrich method is less sensitive in identifying such events. Note that on-target integration will not show up with this analysis approach as the mouse reference genome does not contain the human PI*Z locus. A total of 1685 reads were obtained that aligned to the HITI insert sequence, of which 14 reads were the result of integration into the PI*Z locus; 13 (~93%) in the anticipated orientation, one in reverse orientation. The fact that only 14 reads (< 1%) aligned to the insert were from a HITI insertion could be explained by the observation that the vast majority of the reads appeared to be from the unintegrated HITI donor genome. This observation is consistent with a preliminary Capture-seq experiment (data not shown), where a large proportion of reads were derived from captured plasmid (hence the subsequent incorporation of a prolonged culturing step for the Huh-7 cells to dilute out the HITI plasmids). Thus, in this CasEnrich study many sequencing reads were 'wasted' on sequencing unintegrated viral genomes. Contrary to *in vitro* experiments, there is no straightforward opportunity to dilute out these viral genomes *in vivo*.

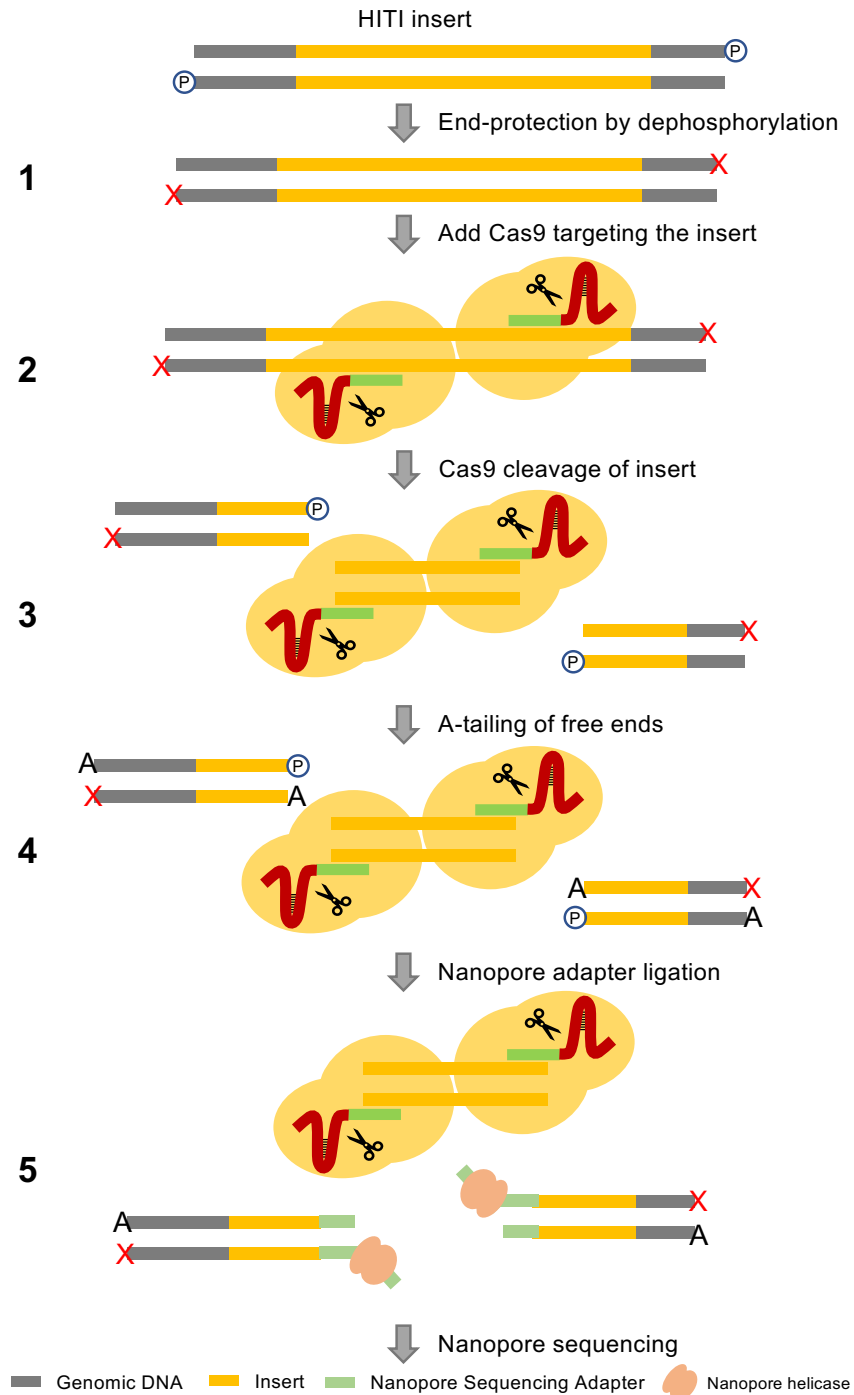


Figure 7.4 | Schematic of CasEnrich method for off-target detection of HITI insertions.

The use of CasEnrich to detect off-target integration is similar to that used to characterise the *PI*Z* locus in *PI*Z* KI mice (**Section 6.2.1.2**). The main difference is that here the Cas9 cut sites are designed in the opposite direction, thus sequencing is directed from the region of interest (the HITI insert) outward into the genome. The pooled DNA from the livers of mice from the 1e11 GC CNV=8 group (see experiment **Section 6.2.2**) was end-protected by dephosphorylation (1) and Cas9 added to the end-protected DNA (2). After cleavage of the insert (3, four gRNAs located in the mNeonGreen cDNA sequence, see **Figure 7.5**) the free DNA ends were A-tailed (4). Since any free ends that were not cleaved by Cas9 should not possess free phosphate groups on their ends they should not be amenable to Nanopore adapter ligation (5) and Nanopore sequencing.

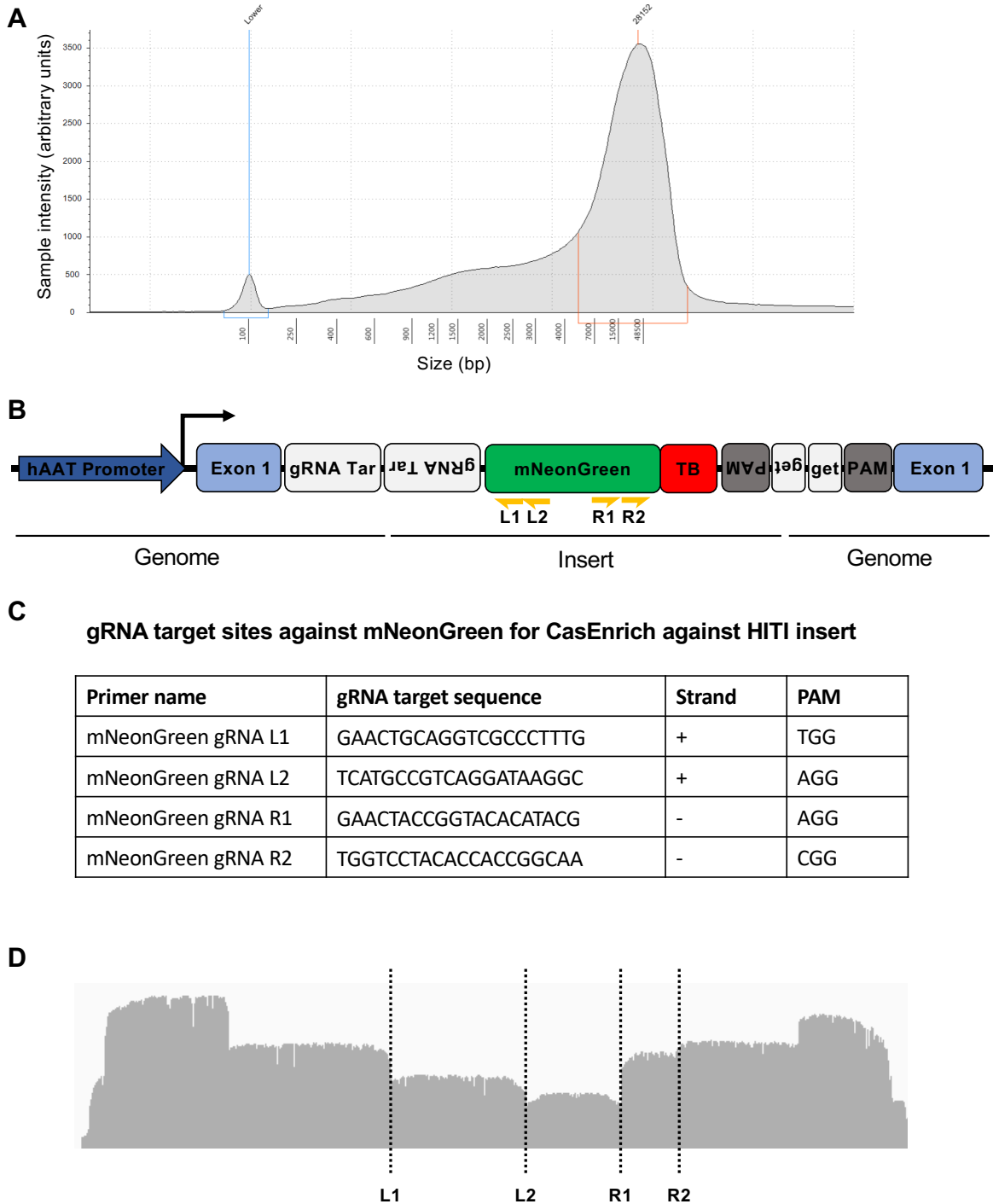


Figure 7.5 | CasEnrich mediated Nanopore sequencing allows directional sequencing from a HITI donor. DNA extracted using a QIAamp kit from mice from group 1e11 GC, CNV = 8 from the experiments discussed in **Section 6.2.2** was pooled and concentrated using Ampure XP magnetic beads. The DNA was analysed on a Genomic Screentape (Agilent) to estimate the size of the DNA fragments after extraction and the average (~28kb) length was deemed sufficient for a CasEnrich experiment targeting insertion sites (**A**). The gRNA target sites in the mNeonGreen cDNA in the HITI donor plasmid (pGM702, **Table 2.5**) are shown as orange arrows in **B**, where the HITI donor is integrated into the AAT locus. The arrows point in the direction of sequencing as determined by Cas9 retention in the CasEnrich protocol (**Figure 7.4**), which should allow sequencing of DNA molecules containing both insert and genome. The gRNA sequences used for this experiment are shown in **C**. The coverage plot of reads that successfully aligned to the insert sequence show the characteristic ‘stepping’ of increased coverage at the Cas9 cut sites (**D**).

7.3 Discussion

7.3.1 Capture-seq enriches for targeted integration and potential off-target sequences

The Capture-seq protocol enriches for sequences of interest by physical capture using hybridisation with complementary biotin-labelled oligonucleotides. This study aimed to use Capture-seq to selectively enrich for sequences containing a HITI donor directly adjacent to the Cas9 cut site that, when integrated in the genome, would also contain a small portion of the genomic sequence (**Figure 7.2**). Aligning reads to the insert to obtain only those containing the insert sequences and mapping to a reference genome was successful in providing a map of integration sites (**Figure 7.3**). Counting the number of reads in a given genomic location generated an approximate of number of integrations at that locus (**Table 7.2**). This indicated that ~60% of inserts were in the intended position (AAT Exon 1) and that there was one prominent off-target site (with ~10% of the reads) in chromosome 10. Guide-seq uses targeted amplification of a dsODN captured in an endonuclease induced DSB to identify off-target integration ([178], **Figure 1.7A**). This dsODN capture is a blunt ended dsDNA ligation process similar to that in HITI. To account for spurious integration of the capture dsODN into DSBs arising spontaneously during cell division, the Guide-seq analysis queries the DNA adjacent to the potential off-target for the presence of a gRNA target site that resembles the on-target gRNA [178, 391]. If a gRNA target site is not found, the read is not considered as being 'off-target'. Omitting this step can cause the on-target cutting efficiency (100% minus all off-targets combined) to be reduced by up to 49% [391]. The analysis of the data presented here did not include a systematic search of gRNA target sites around the potential cut and integration site (the middle of the bi-lobar structure, **Figure 7.3B**). As indicated, the off-target on chromosome 10 (**Figure 7.3A**) did not feature a potential cut site close to the integration site. Similarly, no homology between the sequence around the off-target site and the SaCas9 AAT exon 1 gRNA 2 target gRNA was found within a 100 bp region of the integration site. With endogenous DSBs occurring up to

10 times per cell cycle, per cell, in normal human cells [391, 392], there is the potential that some of the observed integration sites could be due to integration into spontaneous DSBs rather than SaCas9-induced DSBs. Interestingly, no insertions were detected in any of the *in silico* predicted off-target sites (**Table 7.2**). However, one must be wary to dismiss integration sites based on the method of integration; there is no reason to think that insertions in spontaneous DSBs would be less detrimental than those caused by off-target Cas9 cutting, and this is an important consideration in HITI experiments. Contrary to using plasmids or viral genomes as carriers of HDR donors in conventional HDR Cas9 engineering approaches, in HITI, blunt-ended dsDNA intermediates are produced that may more readily insert into spontaneous DSBs. Of note is that the cells used for the Capture-seq experiment were cultured for 1.5 months in order to reduce the amount of plasmid captured by the capture oligos; this created the potential for clonal outgrowth of any given cell, thereby increasing its proportion in the percentage of reads counted in **Table 7.1**.

7.3.2 CasEnrich is not effective in detecting off-target insertion sites

Contrary to the Capture-seq experiment of above, the CasEnrich strategy employed to detect HITI insertion locations (**Figure 7.4**) did not uncover any off-target locations. Given that the protocols were applied to DNA obtained from vastly different sources (Huh-7 liver carcinoma cell line transfected with plasmids and mouse livers transduced with rAAV2/8 for Capture-seq and CasEnrich respectively) it is difficult to compare the methods directly. It is unlikely that no off-target integrations occurred in the mouse livers, however, which means that this finding should more likely be ascribed to the high number of reads obtained from HITI donor genome sequences that had not been integrated, and also to a lack of sensitivity in the CasEnrich protocol. To highlight this, Capture-seq reported 9,489,156 unique (i.e. de-duplicated reads that aligned to the insert (of which 1,208,286 aligned to the reference genome) derived from 2 µg of starting material, versus 1686 reads from CasEnrich from 20 µg of starting DNA. The original number of reads in Capture-seq was roughly 23x higher before de-duplication, increasing the sensitivity of this method even more (amplifying a

single read ultimately increases the chances of it being successfully sequenced). Taken together, while CasEnrich was able to successfully sequence in a directional fashion, and on-target integrations could be observed, the sensitivity of this method was not high enough to detect off-target integrations. Repeating the experiment where non-integrated HITI donor could somehow be excluded (e.g. in a cell line by dilution) and targeting a genome with a more straightforward genomic make-up (e.g. a genome with only two copies of PI*Z) could potentially increase the feasibility of this method.

7.3.3 Conclusion

This chapter applied two novel methods, Capture-seq and CasEnrich, to the detection of off-target HITI integration. It was shown that Capture-seq was able to successfully detect both on- and off-target integration sites, although no distinction could be made between integrations into spontaneous DSBs versus those induced by off-target SaCas9 cuts. CasEnrich was found to lack the sensitivity to detect off-target integration in the model chosen.

Chapter 8 – Discussion

The development and licensing of Advanced Therapeutic Medicinal Products offer groundbreaking opportunities for the treatment of human disease. The recently approved gene therapy drugs for spinal muscular atrophy (Zolgensma) and a form of retinal dystrophy (Luxturna) provide truly life-changing results for patients. Discussion around the general availability and remuneration of these drugs has re-emerged [393]. With prices of up to \$2.1 million for a single administration drug such as Zolgensma, a pay-for-performance payment scheme, first introduced by GSK for their ADS-SCID drug Strimvelis [394], has been promoted as a possible solution [393]. For treatments to be cost-effective for the payer and commercially viable for the pharmaceutical company (see ICER Value Assessment Framework [395]), and to provide lasting benefit to the patient, the treatment must be life-long. Achieving sustained and sustainable benefits is therefore a key challenge for the development of gene therapies. This study aimed to address this challenge, by investigating a strategy to support successful repeat administration of rAAV vectors and, also, to use genome editing to make a single administration last a lifetime.

8.1 Repeat administration of rAAV

Gene addition protocols using rAAV for *in vivo* delivery to the liver have become increasingly successful in long-term expression of therapeutic transgenes at clinically relevant levels [59, 261, 262]. Although genomic insertion of rAAV has been reported, the majority of observed expression is from the episomal form of the recombinant viral genome [396]. The non-dividing nature of several target tissues can explain some of the long-term expression driven by rAAV, for instance up to 15 years in the brains of non-human primates [397], but cells of the liver are slowly dividing [265, 266]. There seems to be no clear understanding of why rAAV can maintain expression over the multiple years of patient follow-up in human trials [262].

The repeat administration strategy described in **Chapter 3** relied on generating immune tolerance to non-self transgenes, induced by the specific expression of those transgenes in the liver [293]. Two liver directed gene therapy vectors were developed, one based on the liver tropic rAAV2/8 and one recombinant lentivirus pseudotyped with VSV-G; both systems were capable of specifically expressing a reporter gene in the liver by employing the liver-specific APoE/hAAT promoter (**Figure 3.5** for rAAV2/8 and **3.12** for rLV). Using AAV8 capsid as the transgene, successful induction of tolerance might prevent the strong antibody response to the capsid proteins seen after rAAV administration (**Section 3.2.4**). Unfortunately, neither rAAV2/8 or VSV-G pseudotyped HIV1 based lentivirus was able to induce the sought-after humoral tolerance (**Section 3.2.5 and 3.2.6**).

The AAV8 serotype, rather than another serotype, was specifically chosen as the transgene, because during rAAV production (**Figure 1.2**) the transgene is expressed from the genome plasmid. Even when the APoE/hAAT promoter is used, while not optimal for expression in the preferred virus producer cells (HEK293T), the high concentration of plasmid transfected still results in significant transgene expression. If an alternative AAV serotype had been used, such as AAV2, the Rep2Cap8 helper plasmid used to produce rAAV2/8 (**Table 2.6**) would express AAV8 VP1, 2 and 3, and the genome plasmid would express AAV2 VP1, leading to chimeric virus particles containing both capsid proteins. It is possible that such chimeric AAV2/AAV8 particles could result in altered tissue tropism, negatively affecting liver-specific expression, and the presence of antibodies cross-reactive between AAV2 and AAV8 [73] that would likely have made any ELISA results difficult to interpret. Transgene expression during production could ostensibly be prevented by using an inducible promoter; **Figure 8.1** shows such a vector, featuring a Tet-ON 3G system for the regulation of expression by addition of doxycycline (Dox). In such a system, virus would be produced in the absence of Dox, with little or no AAV2 capsid expressed during production, and should consist solely of AAV8 capsid proteins. In this way liver tropism should be maintained and the APoE/hAAT promoter regulating expression of the Tet-ON

3G transcription factor ensures liver-specific expression of AAV2 capsid in the presence of Dox. An added advantage with this configuration is that the timing and level of AAV2 capsid expression could be controlled, only inducing tolerance around the time of a therapeutic AAV2 injection. Using such a cross-tolerising rAAV vector also raises the potential to eradicate pre-existing anti-AAV neutralising antibodies. This requires the use of a tolerising delivery virus, for which the patient has no pre-existing nAbs, to express the capsid proteins of the virus for which the patient does have pre-existing nAbs.

As mentioned in **Section 3.1.2**, the most successful repeat administration strategy so far is the approach taken by Meliani and colleagues [292]. The reported method involves temporary immune suppression by means of nanoparticles containing rapamycin (an inhibitor of B- and T-cell activation) given at the same time as the rAAV dose. Immune suppression was demonstrated to be restricted to the antigen given contemporaneously with the nanoparticles and as a result antibodies to the rAAV vector could not be detected, which permitted successful repeat administration. While only two AAV doses have been evaluated with this approach, this already provides a great step forwards compared to previous attempts, potentially allowing a single administration at an early age and another at a later age once the first the effect of the first administration had waned due to dilution.

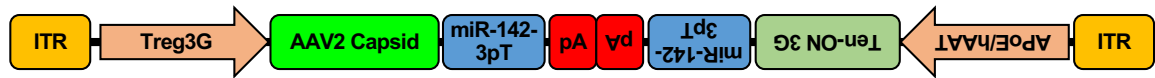


Figure 8.1 | Proposed single-vector, liver-specific, AAV construct with inducible expression. The APoE/hAAT sequence expresses Ten-ON 3G, a transcription factor that specifically recognises the promoter sequence of Treg3G. In this example, Treg3G drives the expression of AAV2 capsid. Both expression cassettes are restricted from APCs by the miR-142-3pT sequence in their 3' UTRs. The expression cassettes are placed in reverse complement orientation with respect to each other to prevent cross-transcription and spurious AAV2 capsid expression. Using a single vector containing both the inducible promoter cassette and the expressed transgene (AAV2 capsid in this example) avoids the need for a dual vector system but does limit transgene size to around 2.5kb, still sufficient for an AAV capsid cDNA.

8.2 Characteristics of HITI mediated integration

When unable to demonstrate successful repeat administration, focus turned to a single administration approach to life-long transgene expression; genome editing and in particular targeted transgene integration. The HITI approach was favoured, a genome editing method that relies on the blunt-end ligation activity of the cNHEJ pathway to insert a blunt-ended dsDNA sequence into a DSB, thereby differentiating itself from other insertional genome editing methods that use homology (**Section 1.2.3**). HITI has been applied in a variety of animal models, and has been investigated for neurological [136] and neuromuscular diseases [398]. In this study, HITI was applied to the liver, both *in vitro* (**Chapter 5**) and *in vivo* (**Chapters 6**) in an effort to provide a long-lasting treatment option for AAT deficiency, specifically for the PI*Z mutant where both expression of the correct version of AAT and repression of the toxic PI*Z protein are required. This ‘double-hit’ strategy proved capable of integration into the first exon of the AAT gene, providing both expression of the reporter gene mNeonGreen and concomitant suppression of the endogenous AAT gene (**Figure 5.4**). Transferring the approach to an *in vivo* model of AATD, the PI*Z KI mouse (characterised in **Section 6.2.1**), it was possible to integrate in the intended site as measured by insertion-specific ddPCR and mNeonGreen expression (**Figure 6.5 and 6.6**). Surveying for off-target effects of the HITI donor insert, using Capture-seq, showed that there was a significant amount of off-target integration (**Section 7.2.1**). This short-read method to find off-targets is more sensitive than the CasEnrich method applied to inserts (**Section 7.2.2**), but it does not illuminate structural alterations that might have occurred in the PI*Z locus after integration.

Ideally, the CasEnrich experiment of **Section 6.3.2** would be applied to *in vivo* genome edited mice, allowing full evaluation of all potential genome editing outcomes. However, considering only 12 reads were identified that encompassed all of the PI*Z locus (**Figure 6.4**), using the CasEnrich method to address the genomic make-up after HITI integration

would be unlikely to lead to sufficient full reads to draw any meaningful conclusions. Another option would be the use of fluorescence *in situ* hybridisation, targeting the PI*Z genes and insert at the same time. This could theoretically provide an image similar to the schematic in **Figure 6.3C** to ascertain the number of inserts on a cell-by-cell basis as well as potentially the location (but not orientation) in respect to the PI*Z genes.

A newly developed mouse line of AATD with, compared to the PI*Z KI mouse, a more straightforward genomic makeup has become available. This mouse line no longer expresses detectable murine AAT as all five mouse orthologues (*Serpina1a* to *e*) have been knocked out using Cas9 [368]. Using this model would allow simpler quantification and characterisation of HITI insertion in one of the chosen murine *Serpina1* genes as each chromosome would only feature a single copy. This model displays a slight age-correlated emphysema phenotype, however it does not feature the liver phenotype since it does not contain the human PI*Z locus; thus this model is not amenable to investigation of the double-hit strategy. Furthermore, any target site in a mouse *Serpina1* gene would require re-verification in the PI*Z model, and probably an *in vitro* model containing a human genome such as liver organoids (**Section 5.2.3**), to detect SaCas9-mediated off-target insertions before being considered as translatable to clinical application. Nevertheless, the *Serpina1a-e* knock out mouse line was obtained for use, but ambiguity about the genotype of this mouse line caused breeding delays such that experiments in this mouse line could not be performed in time of writing.

The fact that no off-target insertions could be detected at predicted SaCas9 off-target cut sites (**Section 7.2.2.3 and Appendix 5**) suggests that most integrations occurred in spontaneous dsDNA breaks. While this would seemingly dismiss the nuclease as the culprit, one of the hallmarks of HITI, creating blunt ended dsDNA breaks to allow cNHEJ ligation into the target site, is also the likely cause for the integration into spontaneous DSBs, and as such is hard to counteract. HITI has been touted as a genome editing

technique superior to HDR in terminally differentiated cells, such as neurons, purportedly due to the low HDR activity in non-dividing cells [136] (**Figure 1.5**). Studies described in **Chapters 5 and 6** assessed whether HITI could be employed in liver cells with low mitotic activity [265, 266]. It would be interesting to compare HITI and HDR genome editing efficiencies, but due to the chosen location of the HITI integration site in the double-hit AAT strategy (SERPINA1 exon 1) HDR was not feasible, because when aiming for direct comparison (and using the same gRNA target), the left homology arm of the purported HDR donor would contain a large part of the AAT promoter and thus generate profound false positives in the much the same way as described in **Section 4.2.5**. While ostensibly the lack of donor cutting in HDR would diminish the integration of donor sequences in spontaneous DSBs, there are reports that whole AAV genomes can integrate in DSBs [399], something which is now actively exploited as a genome editing method (presented at ASGCT 2019, [400]). Complete rAAV integration was not found in this study using the technique most likely to identify it – CasEnrich for inserts (**Section 7.2.2**) – but this method probably lacked sensitivity to do so. The original paper describing the HITI method used in this study [136] only addressed potential off-target cut sites of Cas9 and did not include an in-depth evaluation of off-target integration. As many off-target detection methods, including those described in **Section 1.2.5.1**, solely focus on the DSB induction by Cas9 proteins (by only addressing sites with potential Cas9 recognition sequences), many potentially deleterious insertions in spontaneous DSBs might have gone undetected in studies performed by others.

On a similar note, integration in a non-target tissue (ectopic integration) should likewise not be underestimated. Use of a liver-specific platform for SaCas9 expression (**Section 3.2.2**) in combination with a HITI donor construct that bears the gRNA expression cassette (**Figure 6.5A**) should ostensibly minimise unwanted ectopic integration, but this was not addressed in this study. A recent report utilising a more sensitive method of measuring biodistribution of rAAV2/8 highlighted transient transduction in tissues such as the kidney, brain and testes

[401]. The HITI reporter mouse line (**Section 4.2.4**) would be ideal to test ectopic insertions as the tdTomato cassette is expressed ubiquitously and integrations could be assessed by simply looking for mNeonGreen positive cells.

8.3 Expanding HITI applications

As highlighted in **Figure 4.2**, there are several feasible permutations of HITI that differ in the location of integration and integrated sequence with corresponding therapeutic applications. HITI permutation A (**Figure 4.2A**) differentiates itself from the others by blocking the expression of the targeted gene, achieved by inserting a poly A or a transcription blocker (**Figure 4.4A**), thus benefitting situations where genetic mutations cause deleterious effects. One example is the P1*Z mutation in AATD (**Section 1.3**, applied in **Chapters 5 and 6**) where polymers of mutant protein are formed that cause toxicity in the liver. One interesting twist on permutation A is the possible integration of part of a cDNA rather than the whole protein coding sequence. This could be beneficial for genes in which the protein coding sequence exceeds the packaging capacity of a chosen donor delivery system such as AAV (**Figure 8.2A**), which is known to be the case with the CFTR cDNA in cystic fibrosis or dystrophin cDNA in DMD. To alleviate this, the last few exons, previously termed ‘superexon’ [402], can be packaged into the HITI donor, as outlined in **Figure 8.2B**. In this way, the gene is expressed from the endogenous promoter and the first exons are derived from the genomic sequence, whereas the last exons originate from the inserted sequence. While this method is no longer strictly pan-mutation, as only the mutations located in the exons contained within the HITI donor are addressed, this could still be applicable for a substantial number of affected patients depending on the gene and how much of the protein coding sequence can be accommodated in the HITI donor delivery vector. The superexon method has been tested *in vitro* for cystic fibrosis, using HDR as the genome editing method, with some success [402].

Permutations B and C both preserve the expression of the targeted gene by means of the IRES sequence, but there are some differences between these strategies. Firstly, the use of the IRES has been reported to reduce the amount of expressed protein from the cDNA placed downstream (3' of the IRES) by up to 50% [403]. This is also evident when comparing mNeonGreen levels from permutations B and C, where expression appears lower when mNeonGreen is downstream of the IRES in permutation C (**Figure 4.4**). If maximal transgene expression is required then the transgene should ideally be placed 5' of the targeted gene, although a decrease in endogenous gene expression is to be expected. On the other hand, there is less of a chance that inadvertent insertion of any backbone or viral genome sequence might interfere with endogenous gene expression if the transgene is inserted 3' of gene in the genome. As seen in **Figure 4.4B**, there are cells that are positive for mNeonGreen that are also negative for mCherry, indicating that there is disruption of mCherry expression despite the use of the IRES. Importantly, this effect is absent in permutation C (**Figure 4.4C**).

Permutation D (**Figure 4.2D and 4.6**) is an entirely new way of performing HITI and genome editing more generally. By cutting both sides of the mCherry cDNA, the sequence can be completely removed and subsequently replaced by mNeonGreen. While in the reporter cell line both the cut sequence and insert are whole cDNAs, the real advantage of this system is in diseases where the whole cDNA does not fit into the HITI donor delivery vehicle, such as the size-limited rAAV (**Figure 8.2A**). In this case, the HITI donor would carry selected exon(s) in which common mutation(s) are located, with appropriate splice acceptor and donor sequences; **Figure 8.2C** shows the location of the SaCas9 cut in the introns. As seen in the HITI reporter cell line permutation D experiments (**Figure 4.6**), the targeted exon(s) would often be deleted, which in many diseases would be detrimental. One prime candidate for the exon-replace strategy, however, is DMD, where excising specific exons generates a truncated version of dystrophin which may result in a milder phenotype (Becker muscular dystrophy) [147-149]. An exon-replacement strategy offers the best of both worlds; effective

integration restores a fully functioning dystrophin protein while inadvertent deletion is also therapeutic (**Figure 8.2C**). Preliminary reports of *in vitro* data are encouraging [398].

One application where the HITI mouse model (**Section 4.2.4**) could be invaluable is the assessment of genome editing delivery platforms that offer transient endonuclease expression. Other than in the case of gene addition, where long-term expression from the viral genome is required, long term transgene expression in genome editing such as HITI arises from the integrated sequence. Extended expression of an endonuclease can increase the frequency of off-target cleavage (**Section 1.2.5.1**). Non-viral vectors and integrase deficient lentiviruses (IDLV) could both offer transient expression [404, 405] and have greater capacity than rAAV, potentially allowing all genome editing components to be placed in a single all-in-one vector (**Figure 8.3**). Non-viral gene transfer vectors also carry the added advantage of potential repeat administration [406], which could increase targeting efficacy or a return to the therapeutic window when expression level drops over time due to cell death.

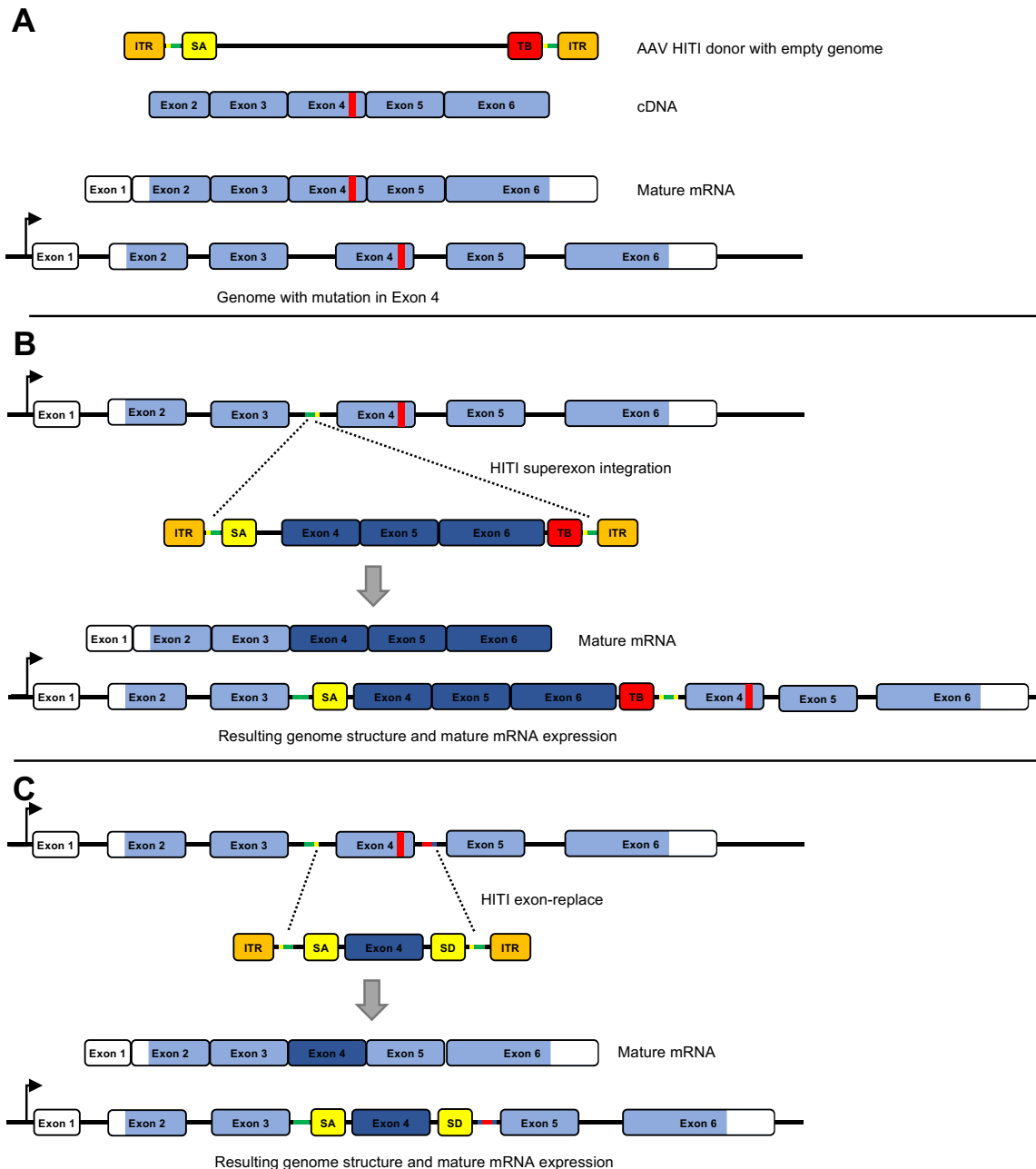


Figure 8.2 | HITI strategies for genes exceeding the rAAV packaging capacity. The indicated (hypothetical) gene (**A**), with a mutation in exon 4, is too large for the entire cDNA to be packaged by an rAAV-based HITI donor vector. The ‘superexon’ approach (**B**) alleviates this by integrating only the last three exons (dark blue) of the gene into the third intron, making the resulting mRNA a mix between endogenous and inserted exons (**B**). The exon-replace mechanism (**C**) is derived from HITI permutation D, but now exon 4 is excised and replaced with a wild-type vector encoded in the HITI donor. SA, splice acceptor; SD, splice donor, TB, transcription blocker; ITR, inverted terminal repeat.

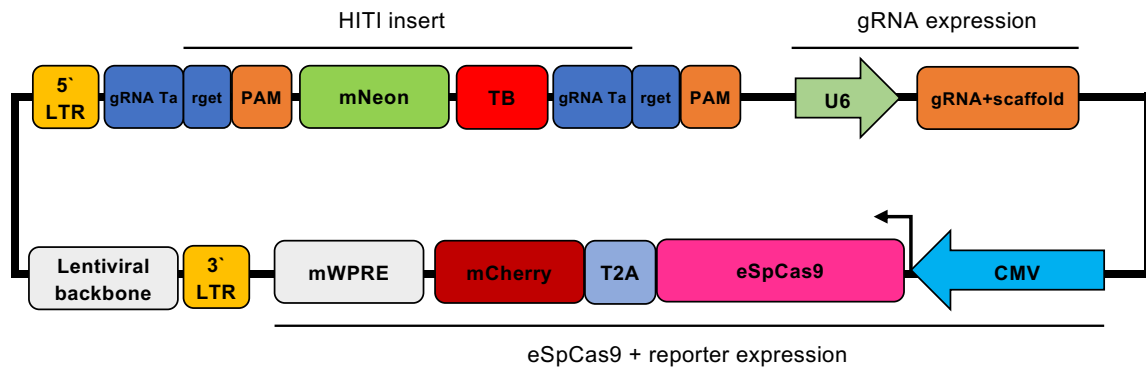


Figure 8.3 | Diagram of integrase deficient Lentiviral vector (IDLV) ‘all-in-one’ HITI construct. The IDLV construct, or a non-viral plasmid genome, contains the endonuclease and gRNA expression cassettes (for Cas9 in this case but ZFNs or TALENs are equally possible). In addition, the HITI insert sequence, including gRNA target sites are on the same plasmid, alleviating the need for co-administration of separate vectors carrying Cas9 and HITI donor. LTR, long terminal repeat; PAM, protospacer adjacent motif; mNeon, mNeonGreen; TB, transcription blocker; mWPRES, mutated version of Woodchuck Hepatitis virus posttranscriptional regulatory element; T2A, *Thosea asigna* virus 2A derived cleaving peptide.

8.4 Translational potential

Arguably, the main reason that has delayed successful AATD-targeted gene therapy is the high therapeutic threshold (11 μ M) which, so far, has not been achieved in gene therapy clinical trials (**Section 1.3.1**). This level, which roughly equates to 25% of normal, ostensibly requires 25% of target cells to be correctly edited since the endogenous AAT promoter is targeted. Application of the double-hit strategy to PI*Z KI mice, did not achieve this threshold, as judged by the number of positive cells observed in microscopy images (**Figure 6.6**). While experiments replacing the mNeonGreen cDNA with a cDNA of human AAT were not undertaken, there is little reason to expect the required serum AAT concentration could be reached in this way. Of note is that *in vivo* gene therapy protocols used a comparatively higher dose of rAAV than used in this study. The clinical trial for Zolgensma for instance used a dose of 2e14 GC per kilogram injected IV [264], which equates to roughly 4e12 GC for a mouse of 20 grams. Based on the safety of the dose given to patients, the maximum dose used in the PI*Z KI mice could be increased at least 10-fold; and given that there appeared to be a dose-dependent increase between the 1e10 GC and the 1e11 GC groups as measured by ddPCR (**Figure 6.5C**), it is possible that a 1e12 GC dose group could reach the therapeutic threshold.

Both a humoral and cellular immune response against SpCas9 has been identified in mice after intramuscular delivery of rAAV2/9 [312]. While no significant T-cell induced cellular damage was reported in this study, results of a long-term genome editing study for DMD revealed that Cas9-specific CD8⁺ cytotoxic T-cells could have a detrimental effect on *in vivo* genome editing efforts [363]. In the study by Nelson et al. mice were injected with SaCas9 targeting the DMD locus in both neonates (2-day-old) and adults (8-week-old). Contrary to the neonates, where no immune response was detected and genome editing was sustained, in adults there was a humoral and cellular response to SaCas9 with a significant decrease in the percentage of edited cells [363]. While no direct causation could

be established, the correlation is striking and resembles the loss in transgene expression and associated immune response to rAAV2/8 gene delivery in the liver [56]. The addition of a transient immunosuppression regimen upon detection of liver damage could repress this effect, with increased expression as a result [59]. This approach could potentially circumvent any T-cells from attacking transduced cells when Cas9 expression is transient (even when T-cells are pre-existing, as they are in humans for SaCas9 and SpCas9 [407, 408]). It would be unlikely to provide lasting benefit in Cas9 delivered by rAAV, however, as Cas9 expression from this vector is long-term. The successful protocol described in Chapter 3 (**Section 3.2.3**) to induce cellular tolerance against a non-self transgene in mice (luciferase in this case) could provide a solution. Inducing tolerance to Cas9 proteins using a similar protocol could allow long-term survival of edited cells. This could be assessed straightforwardly by genome editing a secreted transgene such as AAT in a knock-out mouse and following expression and anti-Cas9 responses over time.

8.5 Conclusion and future directions

Life-long expression of a therapeutic transgene is a prerequisite for most genetic diseases and the field has made great strides to achieve this goal. However, administration in early life, with the accompanying dilution of the episomal rAAV vector, requires effective repeat administration. In this thesis, attempts to address this challenge were unsuccessful, and so a genome editing strategy was employed to ensure expression following a single administration persists. The strategy, based on the genome editing technique HITI, proved successful *in vitro* in targeted integration using a variety of permutations, each with a different potential therapeutic application. One permutation was selected for assessment of applicability in AATD and showed, *in vitro* and *in vivo*, targeted integration resulting in expression of a reporter gene from the endogenous AAT promoter whilst preventing expression of the targeted gene. Long-term experiments using the therapeutic transgene will be required to determine whether the therapeutic threshold can be reached and whether

anti-Cas9 immune responses can be avoided. Of course, careful examination of any potentially deleterious effects in non-target tissues and regions of the genome is essential when aiming to translate these approaches to the clinic. The tools developed in this study will allow for further investigation in this area. In particular, the detection of inadvertently targeted ectopic tissues in HITI reporter mice, and off-target genome editing in human liver organoids, should significantly contribute to understanding both the long-term efficiency and safety of the HITI genome editing approach.

References

1. Whicher, D., S. Philbin, and N. Aronson, *An overview of the impact of rare disease characteristics on research methodology*. Orphanet J Rare Dis, 2018. **13**(1): p. 14.
2. Zincarelli, C., et al., *Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection*. Mol Ther, 2008. **16**(6): p. 1073-80.
3. Griesenbach, U., et al., *Assessment of F/HN-pseudotyped lentivirus as a clinically relevant vector for lung gene therapy*. Am J Respir Crit Care Med, 2012. **186**(9): p. 846-56.
4. Kobayashi, M., et al., *Pseudotyped lentivirus vectors derived from simian immunodeficiency virus SIVagm with envelope glycoproteins from paramyxovirus*. J Virol, 2003. **77**(4): p. 2607-14.
5. Huang, Y., *Preclinical and Clinical Advances of GalNAc-Decorated Nucleic Acid Therapeutics*. Mol Ther Nucleic Acids, 2017. **6**: p. 116-132.
6. Himeda, C.L., X. Chen, and S.D. Hauschka, *Design and testing of regulatory cassettes for optimal activity in skeletal and cardiac muscles*, in *Muscle Gene Therapy*. 2011, Springer. p. 3-19.
7. Nathwani, A.C., et al., *Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver*. Blood, 2006. **107**(7): p. 2653-61.
8. Sarcar, S., et al., *Next-generation muscle-directed gene therapy by in silico vector design*. Nat Commun, 2019. **10**(1): p. 492.
9. Brown, B.D., et al., *Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer*. Nat Med, 2006. **12**(5): p. 585-91.
10. Qiao, C., et al., *Liver-specific microRNA-122 target sequences incorporated in AAV vectors efficiently inhibits transgene expression in the liver*. Gene Ther, 2011. **18**(4): p. 403-10.
11. Thule, P.M., et al., *Long-term glycemic control with hepatic insulin gene therapy in streptozotocin-diabetic mice*. J Gene Med, 2015. **17**(8-9): p. 141-52.
12. Barrett, J.A., et al., *Regulated intratumoral expression of IL-12 using a RheoSwitch Therapeutic System((R)) (RTS((R))) gene switch as gene therapy for the treatment of glioma*. Cancer Gene Ther, 2018. **25**(5-6): p. 106-116.
13. Chaveroux, C., et al., *Regulating the expression of therapeutic transgenes by controlled intake of dietary essential amino acids*. Nat Biotechnol, 2016. **34**(7): p. 746-51.
14. Zhong, G., et al., *Rational design of aptazyme riboswitches for efficient control of gene expression in mammalian cells*. Elife, 2016. **5**: p. e18858.
15. Chung, H.K., et al., *Tunable and reversible drug control of protein production via a self-excising degron*. Nat Chem Biol, 2015. **11**(9): p. 713-20.
16. Smith, K.O. and J.L. Melnick, *Adenovirus-Like Particles from Cancers Induced by Adenovirus-12 but Free of Infectious Virus*. Science, 1964. **145**(3637): p. 1190-2.
17. Atchison, R.W., B.C. Casto, and W.M. Hammon, *Adenovirus-Associated Defective Virus Particles*. Science, 1965. **149**(3685): p. 754-6.
18. Hermonat, P.L., et al., *Chromosomal latency and expression at map unit 96 of a wild-type plus adeno-associated virus (AAV)/Neo vector and identification of p81, a new AAV transcriptional promoter*. J Hum Virol, 1999. **2**(6): p. 359-68.
19. Cao, M., H. You, and P.L. Hermonat, *The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication*. PLoS One, 2014. **9**(8): p. e104596.
20. Maurer, A.C., et al., *The Assembly-Activating Protein Promotes Stability and Interactions between AAV's Viral Proteins to Nucleate Capsid Assembly*. Cell Rep, 2018. **23**(6): p. 1817-1830.

21. Earley, L.F., et al., *Adeno-associated Virus (AAV) Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11*. J Virol, 2017. **91**(3): p. e01980-16.
22. Sonntag, F., K. Schmidt, and J.A. Kleinschmidt, *A viral assembly factor promotes AAV2 capsid formation in the nucleolus*. Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10220-5.
23. Kelsic, E., P. Ogden, and G. Church. *Discovery of a Novel Frameshifted ORF within the AAV Capsid Gene*. in ASGCT. 2019. CELL PRESS 50 HAMPSHIRE ST, FLOOR 5, CAMBRIDGE, MA 02139 USA.
24. Samulski, R.J., L.S. Chang, and T. Shen, *A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication*. J Virol, 1987. **61**(10): p. 3096-101.
25. Mizukami, H., N.S. Young, and K.E. Brown, *Adeno-associated virus type 2 binds to a 150-kilodalton cell membrane glycoprotein*. Virology, 1996. **217**(1): p. 124-30.
26. Summerford, C. and R.J. Samulski, *Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions*. J Virol, 1998. **72**(2): p. 1438-45.
27. Flotte, T., et al., *A phase I study of an adeno-associated virus-cftr gene vector in adult cf patients with mild lung disease*. Johns hopkins children's center, baltimore, maryland. Human gene therapy, 1996. **7**(9): p. 1145-1159.
28. Duan, D., et al., *Polarity influences the efficiency of recombinant adenoassociated virus infection in differentiated airway epithelia*. Hum Gene Ther, 1998. **9**(18): p. 2761-76.
29. Pillay, S., et al., *An essential receptor for adeno-associated virus infection*. Nature, 2016. **530**(7588): p. 108-12.
30. Dudek, A.M., et al., *An Alternate Route for Adeno-associated Virus (AAV) Entry Independent of AAV Receptor*. J Virol, 2018. **92**(7).
31. Hamilton, B.A., et al., *Polarized AAVR expression determines infectivity by AAV gene therapy vectors*. Gene Ther, 2019. **26**(6): p. 240-249.
32. Bartlett, J.S., R. Wilcher, and R.J. Samulski, *Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors*. J Virol, 2000. **74**(6): p. 2777-85.
33. Duan, D., et al., *Dynamin is required for recombinant adeno-associated virus type 2 infection*. J Virol, 1999. **73**(12): p. 10371-6.
34. Sanlioglu, S., et al., *Endocytosis and nuclear trafficking of adeno-associated virus type 2 are controlled by rac1 and phosphatidylinositol-3 kinase activation*. J Virol, 2000. **74**(19): p. 9184-96.
35. Nonnenmacher, M. and T. Weber, *Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway*. Cell Host Microbe, 2011. **10**(6): p. 563-76.
36. Nonnenmacher, M. and T. Weber, *Intracellular transport of recombinant adeno-associated virus vectors*. Gene Ther, 2012. **19**(6): p. 649-58.
37. Weinberg, M.S., et al., *Recombinant adeno-associated virus utilizes cell-specific infectious entry mechanisms*. J Virol, 2014. **88**(21): p. 12472-84.
38. Grimm, D. and S. Zolotukhin, *E Pluribus Unum: 50 Years of Research, Millions of Viruses, and One Goal--Tailored Acceleration of AAV Evolution*. Mol Ther, 2015. **23**(12): p. 1819-31.
39. Samulski, R.J. and N. Muzyczka, *AAV-Mediated Gene Therapy for Research and Therapeutic Purposes*. Annu Rev Virol, 2014. **1**(1): p. 427-51.
40. Dong, B., H. Nakai, and W. Xiao, *Characterization of genome integrity for oversized recombinant AAV vector*. Mol Ther, 2010. **18**(1): p. 87-92.
41. Wu, Z., H. Yang, and P. Colosi, *Effect of genome size on AAV vector packaging*. Mol Ther, 2010. **18**(1): p. 80-6.
42. McClements, M.E. and R.E. MacLaren, *Adeno-associated Virus (AAV) Dual Vector Strategies for Gene Therapy Encoding Large Transgenes*. Yale J Biol Med, 2017. **90**(4): p. 611-623.

43. Tiffany, M. and M.A. Kay, 253. *Expanded Packaging Capacity of AAV by Luminal Charge Alteration*. Molecular Therapy, 2016. **24**: p. S99-S100.
44. Huang, L.Y., S. Halder, and M. Agbandje-McKenna, *Parvovirus glycan interactions*. Curr Opin Virol, 2014. **7**: p. 108-18.
45. Pillay, S. and J.E. Carette, *Host determinants of adeno-associated viral vector entry*. Curr Opin Virol, 2017. **24**: p. 124-131.
46. Asokan, A., D.V. Schaffer, and R.J. Samulski, *The AAV vector toolkit: poised at the clinical crossroads*. Molecular Therapy, 2012. **20**(4): p. 699-708.
47. Lisowski, L., S.S. Tay, and I.E. Alexander, *Adeno-associated virus serotypes for gene therapeutics*. Current opinion in pharmacology, 2015. **24**: p. 59-67.
48. Srivastava, A., *In vivo tissue-tropism of adeno-associated viral vectors*. Curr Opin Virol, 2016. **21**: p. 75-80.
49. Chandler, R.J., et al., *Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy*. J Clin Invest, 2015. **125**(2): p. 870-80.
50. Donsante, A., et al., *AAV vector integration sites in mouse hepatocellular carcinoma*. Science, 2007. **317**(5837): p. 477.
51. Nault, J.C., et al., *Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas*. Nat Genet, 2015. **47**(10): p. 1187-93.
52. Logan, G.J., et al., *Identification of liver-specific enhancer-promoter activity in the 3' untranslated region of the wild-type AAV2 genome*. Nat Genet, 2017. **49**(8): p. 1267-1273.
53. Gil-Farina, I., et al., *Recombinant AAV Integration Is Not Associated With Hepatic Genotoxicity in Nonhuman Primates and Patients*. Mol Ther, 2016. **24**(6): p. 1100-1105.
54. Berns, K.I., et al., *Adeno-Associated Virus Type 2 and Hepatocellular Carcinoma?* Hum Gene Ther, 2015. **26**(12): p. 779-81.
55. Xiao, X., J. Li, and R.J. Samulski, *Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus*. J Virol, 1998. **72**(3): p. 2224-32.
56. Manno, C.S., et al., *Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response*. Nat Med, 2006. **12**(3): p. 342-7.
57. Mingozzi, F., et al., *CD8(+) T-cell responses to adeno-associated virus capsid in humans*. Nat Med, 2007. **13**(4): p. 419-22.
58. Mays, L.E., et al., *Adeno-associated virus capsid structure drives CD4-dependent CD8+ T cell response to vector encoded proteins*. J Immunol, 2009. **182**(10): p. 6051-60.
59. Nathwani, A.C., et al., *Adenovirus-associated virus vector-mediated gene transfer in hemophilia B*. N Engl J Med, 2011. **365**(25): p. 2357-65.
60. Kulinski, J.M., V.L. Tarakanova, and J. Verbsky, *Regulation of antiviral CD8 T-cell responses*. Crit Rev Immunol, 2013. **33**(6): p. 477-88.
61. Halbert, C.L., et al., *Capsid-expressing DNA in AAV vectors and its elimination by use of an oversize capsid gene for vector production*. Gene Ther, 2011. **18**(4): p. 411-7.
62. Chadeuf, G., et al., *Evidence for encapsidation of prokaryotic sequences during recombinant adeno-associated virus production and their in vivo persistence after vector delivery*. Molecular Therapy, 2005. **12**(4): p. 744-753.
63. Lecomte, E., et al., *Advanced Characterization of DNA Molecules in rAAV Vector Preparations by Single-stranded Virus Next-generation Sequencing*. Mol Ther Nucleic Acids, 2015. **4**(10): p. e260.
64. Wang, Z., et al., *Elimination of contaminating cap genes in AAV vector virions reduces immune responses and improves transgene expression in a canine gene therapy model*. Gene Ther, 2014. **21**(4): p. 363-70.
65. Manning, W.C., et al., *Transient immunosuppression allows transgene expression following readministration of adeno-associated viral vectors*. Hum Gene Ther, 1998. **9**(4): p. 477-85.

66. Sumner-Jones, S.G., D.R. Gill, and S.C. Hyde, *Lack of repeat transduction by recombinant adeno-associated virus type 5/5 vectors in the mouse airway*. J Virol, 2007. **81**(22): p. 12360-7.
67. Afione, S.A., et al., *In vivo model of adeno-associated virus vector persistence and rescue*. J Virol, 1996. **70**(5): p. 3235-41.
68. Rapti, K., et al., *Neutralizing antibodies against AAV serotypes 1, 2, 6, and 9 in sera of commonly used animal models*. Molecular Therapy, 2012. **20**(1): p. 73-83.
69. Nieto, K., et al., *Intranasal vaccination with AAV5 and 9 vectors against human papillomavirus type 16 in rhesus macaques*. Hum Gene Ther, 2012. **23**(7): p. 733-41.
70. Wang, L., et al., *Impact of pre-existing immunity on gene transfer to nonhuman primate liver with adeno-associated virus 8 vectors*. Hum Gene Ther, 2011. **22**(11): p. 1389-401.
71. Louis Jeune, V., et al., *Pre-existing anti-adeno-associated virus antibodies as a challenge in AAV gene therapy*. Hum Gene Ther Methods, 2013. **24**(2): p. 59-67.
72. Scallan, C.D., et al., *Human immunoglobulin inhibits liver transduction by AAV vectors at low AAV2 neutralizing titers in SCID mice*. Blood, 2006. **107**(5): p. 1810-7.
73. Boutin, S., et al., *Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors*. Hum Gene Ther, 2010. **21**(6): p. 704-12.
74. Fitzpatrick, Z., et al., *Influence of Pre-existing Anti-capsid Neutralizing and Binding Antibodies on AAV Vector Transduction*. Mol Ther Methods Clin Dev, 2018. **9**: p. 119-129.
75. Falese, L., et al., *Strategy to detect pre-existing immunity to AAV gene therapy*. Gene Ther, 2017. **24**(12): p. 768-778.
76. Anguela, X.M. and K.A. High, *Oracle or false prophet? Can we predict AAV efficacy based on preexisting antibody titers?* Res Pract Thromb Haemost, 2019. **3**(2): p. 149-151.
77. Hinderer, C., et al., *Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN*. Hum Gene Ther, 2018. **29**(3): p. 285-298.
78. Hordeaux, J., et al., *The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice*. Mol Ther, 2018. **26**(3): p. 664-668.
79. Novartis. *AveXis receives FDA approval for Zolgensma®, the first and only gene therapy for pediatric patients with spinal muscular atrophy (SMA)*. 2019 [cited 2019 07/07/2019]; Available from: <https://www.novartis.com/news/media-releases/avexis-receives-fda-approval-zolgensma-first-and-only-gene-therapy-pediatric-patients-spinal-muscular-atrophy-sma>.
80. Beasley, D. *Second death in Novartis gene therapy trials under investigation*. 2019 [cited 2019 07/07/2019]; Available from: <https://www.reuters.com/article/us-novartis-genetherapy-death/second-death-in-novartis-gene-therapy-trials-under-investigation-idUSKCN1RW005>.
81. Flotte, T.R. and H. Buning, *Severe Toxicity in Nonhuman Primates and Piglets with Systemic High-Dose Administration of Adeno-Associated Virus Serotype 9-Like Vectors: Putting Patients First*. Hum Gene Ther, 2018. **29**(3): p. 283-284.
82. Calcedo, R., et al., *Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses*. J Infect Dis, 2009. **199**(3): p. 381-90.
83. Veron, P., et al., *Humoral and cellular capsid-specific immune responses to adeno-associated virus type 1 in randomized healthy donors*. J Immunol, 2012. **188**(12): p. 6418-24.
84. Chirmule, N., et al., *Immune responses to adenovirus and adeno-associated virus in humans*. Gene Ther, 1999. **6**(9): p. 1574-83.

85. Mingozi, F., et al., *Prevalence and pharmacological modulation of humoral immunity to AAV vectors in gene transfer to synovial tissue*. Gene Ther, 2013. **20**(4): p. 417-24.
86. Rouet, P., F. Smih, and M. Jasin, *Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease*. Mol Cell Biol, 1994. **14**(12): p. 8096-106.
87. Bibikova, M., et al., *Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases*. Genetics, 2002. **161**(3): p. 1169-1175.
88. Kim, Y.G., J. Cha, and S. Chandrasegaran, *Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain*. Proc Natl Acad Sci U S A, 1996. **93**(3): p. 1156-60.
89. Boch, J., et al., *Breaking the code of DNA binding specificity of TAL-type III effectors*. Science, 2009. **326**(5959): p. 1509-12.
90. Christian, M., et al., *Targeting DNA double-strand breaks with TAL effector nucleases*. Genetics, 2010. **186**(2): p. 757-61.
91. Makarova, K.S., et al., *A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action*. Biol Direct, 2006. **1**(1): p. 7.
92. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
93. Knott, G.J. and J.A. Doudna, *CRISPR-Cas guides the future of genetic engineering*. Science, 2018. **361**(6405): p. 866-869.
94. Doench, J.G., et al., *Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation*. Nat Biotechnol, 2014. **32**(12): p. 1262-7.
95. Listgarten, J., et al., *Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs*. Nat Biomed Eng, 2018. **2**(1): p. 38-47.
96. Chuai, G., et al., *DeepCRISPR: optimized CRISPR guide RNA design by deep learning*. Genome Biol, 2018. **19**(1): p. 80.
97. Doench, J.G., et al., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9*. Nat Biotechnol, 2016. **34**(2): p. 184-191.
98. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems*. Science, 2013. **339**(6121): p. 819-23.
99. Mali, P., et al., *RNA-guided human genome engineering via Cas9*. Science, 2013. **339**(6121): p. 823-6.
100. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases*. Nat Biotechnol, 2013. **31**(9): p. 827-32.
101. Kim, E., et al., *In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni*. Nat Commun, 2017. **8**: p. 14500.
102. Ran, F.A., et al., *In vivo genome editing using Staphylococcus aureus Cas9*. Nature, 2015. **520**(7546): p. 186-91.
103. Abudayyeh, O.O., et al., *RNA targeting with CRISPR-Cas13*. Nature, 2017. **550**(7675): p. 280-284.
104. Cox, D.B.T., et al., *RNA editing with CRISPR-Cas13*. Science, 2017. **358**(6366): p. 1019-1027.
105. Gootenberg, J.S., et al., *Nucleic acid detection with CRISPR-Cas13a/C2c2*. Science, 2017. **356**(6336): p. 438-442.
106. Gootenberg, J.S., et al., *Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6*. Science, 2018. **360**(6387): p. 439-444.
107. Chen, J.S., et al., *CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity*. Science, 2018. **360**(6387): p. 436-439.
108. Esvelt, K.M., et al., *Orthogonal Cas9 proteins for RNA-guided gene regulation and editing*. Nat Methods, 2013. **10**(11): p. 1116-21.
109. Rousseau, B.A., et al., *Programmable RNA Cleavage and Recognition by a Natural CRISPR-Cas9 System from Neisseria meningitidis*. Mol Cell, 2018. **69**(5): p. 906-914 e4.

110. Karvelis, T., et al., *crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus*. RNA Biol, 2013. **10**(5): p. 841-51.
111. Zetsche, B., et al., *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System*. Cell, 2015. **163**(3): p. 759-771.
112. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems*. Mol Cell, 2015. **60**(3): p. 385-97.
113. Burstein, D., et al., *New CRISPR-Cas systems from uncultivated microbes*. Nature, 2017. **542**(7640): p. 237-241.
114. Liu, J.-J., et al., *CasX enzymes comprise a distinct family of RNA-guided genome editors*. Nature, 2019. **566**(7743): p. 218.
115. Abudayyeh, O.O., et al., *C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector*. Science, 2016. **353**(6299): p. aaf5573.
116. Smargon, A.A., et al., *Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28*. Mol Cell, 2017. **65**(4): p. 618-630 e7.
117. Shmakov, S., et al., *Diversity and evolution of class 2 CRISPR-Cas systems*. Nat Rev Microbiol, 2017. **15**(3): p. 169-182.
118. Konermann, S., et al., *Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors*. Cell, 2018. **173**(3): p. 665-676 e14.
119. Gaudelli, N.M., et al., *Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage*. Nature, 2017. **551**(7681): p. 464-+.
120. Komor, A.C., et al., *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage*. Nature, 2016. **533**(7603): p. 420-+.
121. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression*. Cell, 2013. **152**(5): p. 1173-83.
122. Chavez, A., et al., *Highly efficient Cas9-mediated transcriptional programming*. Nat Methods, 2015. **12**(4): p. 326-8.
123. Hilton, I.B., et al., *Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers*. Nat Biotechnol, 2015. **33**(5): p. 510-7.
124. Stepper, P., et al., *Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase*. Nucleic acids research, 2016. **45**(4): p. 1703-1713.
125. Tanenbaum, M.E., et al., *A protein-tagging system for signal amplification in gene expression and fluorescence imaging*. Cell, 2014. **159**(3): p. 635-46.
126. Tsui, C., et al., *dCas9-targeted locus-specific protein isolation method identifies histone gene regulators*. Proc Natl Acad Sci U S A, 2018. **115**(12): p. E2734-E2741.
127. Morgan, S.L., et al., *Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping*. Nature Communications, 2017. **8**: p. 15993.
128. Chang, H.H.Y., et al., *Non-homologous DNA end joining and alternative pathways to double-strand break repair*. Nat Rev Mol Cell Biol, 2017. **18**(8): p. 495-506.
129. Miyaoka, Y., et al., *Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing*. Sci Rep, 2016. **6**: p. 23549.
130. Beucher, A., et al., *ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2*. EMBO J, 2009. **28**(21): p. 3413-27.
131. Woodbine, L., A.R. Gennery, and P.A. Jeggo, *The clinical impact of deficiency in DNA non-homologous end-joining*. DNA Repair (Amst), 2014. **16**: p. 84-96.
132. Bétermier, M., P. Bertrand, and B.S. Lopez, *Is non-homologous end-joining really an inherently error-prone process?* PLoS genetics, 2014. **10**(1): p. e1004086.
133. Seol, J.H., E.Y. Shim, and S.E. Lee, *Microhomology-mediated end joining: Good, bad and ugly*. Mutat Res, 2018. **809**: p. 81-87.
134. Bae, S., et al., *Microhomology-based choice of Cas9 nuclease target sites*. Nat Methods, 2014. **11**(7): p. 705-6.
135. Taleei, R. and H. Nikjoo, *Biochemical DSB-repair model for mammalian cells in G1 and early S phases of the cell cycle*. Mutat Res, 2013. **756**(1-2): p. 206-12.

136. Suzuki, K., et al., *In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration*. Nature, 2016. **540**(7631): p. 144-149.
137. Nakade, S., et al., *Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9*. Nat Commun, 2014. **5**: p. 5560.
138. Danner, E., et al., *Control of gene editing by manipulation of DNA repair mechanisms*. Mamm Genome, 2017. **28**(7-8): p. 262-274.
139. Orthwein, A., et al., *Mitosis inhibits DNA double-strand break repair to guard against telomere fusions*. Science, 2014. **344**(6180): p. 189-93.
140. Lin, S., et al., *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery*. Elife, 2014. **3**: p. e04766.
141. Beck, M., et al., *The quantitative proteome of a human cell line*. Mol Syst Biol, 2011. **7**(1): p. 549.
142. Richardson, C.D., et al., *CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway*. Nat Genet, 2018. **50**(8): p. 1132-1139.
143. Kosicki, M., K. Tomberg, and A. Bradley, *Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements*. Nature Biotechnology, 2018. **36**: p. 765.
144. Owens, D.D.G., et al., *Microhomologies are prevalent at Cas9-induced larger deletions*. Nucleic Acids Res, 2019. **47**(14): p. 7402-7417.
145. Allen, F., et al., *Predicting the mutations generated by repair of Cas9-induced double-strand breaks*. Nat Biotechnol, 2018. **37**: p. 64.
146. Shen, M.W., et al., *Predictable and precise template-free CRISPR editing of pathogenic variants*. Nature, 2018. **563**(7733): p. 646-651.
147. Tabebordbar, M., et al., *In vivo gene editing in dystrophic mouse muscle and muscle stem cells*. Science, 2016. **351**(6271): p. 407-411.
148. Nelson, C.E., et al., *In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy*. Science, 2016. **351**(6271): p. 403-7.
149. Long, C., et al., *Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy*. Science, 2016. **351**(6271): p. 400-3.
150. Sheridan, C., *Go-ahead for first in-body CRISPR medicine testing*. Nature Biotechnology, 2018.
151. Iyer, S., et al., *Precise therapeutic gene correction by a simple nuclease-induced double-stranded break*. Nature, 2019. **568**(7753): p. 561-565.
152. Kan, Y., et al., *Mechanisms of precise genome editing using oligonucleotide donors*. Genome Res, 2017. **27**(7): p. 1099-1111.
153. Richardson, C.D., et al., *Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA*. Nat Biotechnol, 2016. **34**(3): p. 339-44.
154. Aird, E.J., et al., *Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template*. Commun Biol, 2018. **1**(1): p. 54.
155. Maruyama, T., et al., *Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining*. Nat Biotechnol, 2015. **33**(5): p. 538-42.
156. Dever, D.P., et al., *CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells*. Nature, 2016. **539**(7629): p. 384-389.
157. Ledford, H., *First test of in-body gene editing shows promise*. Nature, 2018.
158. Bak, R.O. and M.H. Porteus, *CRISPR-Mediated Integration of Large Gene Cassettes Using AAV Donor Vectors*. Cell Rep, 2017. **20**(3): p. 750-756.
159. Orlando, S.J., et al., *Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology*. Nucleic Acids Res, 2010. **38**(15): p. e152.
160. Cristea, S., et al., *In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration*. Biotechnol Bioeng, 2013. **110**(3): p. 871-80.
161. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity*. Cell, 2013. **154**(6): p. 1380-9.

162. Moreno-Mateos, M.A., et al., *CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing*. Nat Commun, 2017. **8**(1): p. 2024.
163. Wang, Y., et al., *Systematic evaluation of CRISPR-Cas systems reveals design principles for genome editing in human cells*. Genome Biol, 2018. **19**(1): p. 62.
164. Maresca, M., et al., *Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining*. Genome Res, 2013. **23**(3): p. 539-46.
165. Auer, T.O., et al., *Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair*. Genome Res, 2014. **24**(1): p. 142-53.
166. Kimura, Y., et al., *Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering*. Sci Rep, 2014. **4**: p. 6545.
167. Li, J., et al., *Intron-based genomic editing: a highly efficient method for generating knockin zebrafish*. Oncotarget, 2015. **6**(20): p. 17891-4.
168. Kesavan, G., et al., *CRISPR/Cas9-Mediated Zebrafish Knock-in as a Novel Strategy to Study Midbrain-Hindbrain Boundary Development*. Front Neuroanat, 2017. **11**: p. 52.
169. Yamamoto, Y., J. Bliss, and S.A. Gerbi, *Whole Organism Genome Editing: Targeted Large DNA Insertion via ObLiGaRe Nonhomologous End-Joining in Vivo Capture*. G3 (Bethesda), 2015. **5**(9): p. 1843-7.
170. Shi, Z., et al., *Targeted integration of genes in Xenopus tropicalis*. Genesis, 2017. **55**(1-2).
171. Wang, Y., et al., *Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew*. Nat Biotechnol, 2014. **32**(9): p. 947-51.
172. Fei, J.F., et al., *Efficient gene knockin in axolotl and its use to test the role of satellite cells in limb regeneration*. Proc Natl Acad Sci U S A, 2017. **114**(47): p. 12501-12506.
173. Kumagai, H., et al., *CRISPR/Cas-mediated knock-in via non-homologous end-joining in the crustacean Daphnia magna*. PLoS One, 2017. **12**(10): p. e0186112.
174. Yu, M., et al., *Highly Efficient Transgenesis in Ferrets Using CRISPR/Cas9-Mediated Homology-Independent Insertion at the ROSA26 Locus*. Sci Rep, 2019. **9**(1): p. 1971.
175. Kempton, H.R. and L.S. Qi, *When genome editing goes off-target*. Science, 2019. **364**(6437): p. 234-236.
176. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Nat Biotechnol, 2013. **31**(3): p. 233-9.
177. Fu, Y., et al., *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells*. Nat Biotechnol, 2013. **31**(9): p. 822-6.
178. Tsai, S.Q., et al., *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases*. Nat Biotechnol, 2015. **33**(2): p. 187-197.
179. Lazzarotto, C.R., et al., *Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq*. Nat Protoc, 2018. **13**(11): p. 2615-2642.
180. Tsai, S.Q., et al., *CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets*. Nat Methods, 2017. **14**(6): p. 607-614.
181. Wienert, B., et al., *Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq*. Science, 2019. **364**(6437): p. 286-289.
182. Guilinger, J.P., D.B. Thompson, and D.R. Liu, *Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification*. Nat Biotechnol, 2014. **32**(6): p. 577-582.
183. Tsai, S.Q., et al., *Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing*. Nat Biotechnol, 2014. **32**(6): p. 569-76.
184. Chen, J.S., et al., *Enhanced proofreading governs CRISPR-Cas9 targeting accuracy*. Nature, 2017. **550**(7676): p. 407-410.
185. Kleinstiver, B.P., et al., *High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects*. Nature, 2016. **529**(7587): p. 490-5.

186. Slaymaker, I.M., et al., *Rationally engineered Cas9 nucleases with improved specificity*. Science, 2016. **351**(6268): p. 84-8.
187. Vakulskas, C.A., et al., *A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells*. Nat Med, 2018. **24**(8): p. 1216-1224.
188. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs*. Nat Biotechnol, 2014. **32**(3): p. 279-284.
189. Brantly, M., T. Nukiwa, and R.G. Crystal, *Molecular basis of alpha-1-antitrypsin deficiency*. Am J Med, 1988. **84**(6A): p. 13-31.
190. Mason, D.Y., et al., *Alpha 1-antitrypsin is present within the primary granules of human polymorphonuclear leukocytes*. Am J Pathol, 1991. **139**(3): p. 623-8.
191. Cichy, J., J. Potempa, and J. Travis, *Biosynthesis of alpha1-proteinase inhibitor by human lung-derived epithelial cells*. J Biol Chem, 1997. **272**(13): p. 8250-5.
192. Pandey, K.C., S. De, and P.K. Mishra, *Role of Proteases in Chronic Obstructive Pulmonary Disease*. Front Pharmacol, 2017. **8**: p. 512.
193. Stockley, R.A. and A.M. Turner, *alpha-1-Antitrypsin deficiency: clinical variability, assessment, and treatment*. Trends Mol Med, 2014. **20**(2): p. 105-15.
194. Green, C.E., et al., *PiSZ alpha-1 antitrypsin deficiency (AATD): pulmonary phenotype and prognosis relative to PiZZ AATD and PiMM COPD*. Thorax, 2015. **70**(10): p. 939-45.
195. Curiel, D., et al., *Alpha 1-antitrypsin deficiency caused by the alpha 1-antitrypsin Nullmattawa gene. An insertion mutation rendering the alpha 1-antitrypsin gene incapable of producing alpha 1-antitrypsin*. J Clin Invest, 1989. **83**(4): p. 1144-52.
196. Lomas, D.A., et al., *The mechanism of Z alpha 1-antitrypsin accumulation in the liver*. Nature, 1992. **357**(6379): p. 605-7.
197. Kelly, E., et al., *Selenoprotein S/SEPS1 modifies endoplasmic reticulum stress in Z variant alpha1-antitrypsin deficiency*. J Biol Chem, 2009. **284**(25): p. 16891-7.
198. de Serres, F.J., *Alpha-1 antitrypsin deficiency is not a rare disease but a disease that is rarely diagnosed*. Environ Health Perspect, 2003. **111**(16): p. 1851-4.
199. Stoller, J.K., et al., *Delay in diagnosis of alpha1-antitrypsin deficiency: a continuing problem*. Chest, 2005. **128**(4): p. 1989-94.
200. Campos, M.A., et al., *Trends in the diagnosis of symptomatic patients with alpha1-antitrypsin deficiency between 1968 and 2003*. Chest, 2005. **128**(3): p. 1179-86.
201. Chu, A.S., K.B. Chopra, and D.H. Perlmutter, *Is severe progressive liver disease caused by alpha-1-antitrypsin deficiency more common in children or adults?* Liver Transpl, 2016. **22**(7): p. 886-94.
202. Rudnick, D.A., et al., *Analyses of hepatocellular proliferation in a mouse model of alpha-1-antitrypsin deficiency*. Hepatology, 2004. **39**(4): p. 1048-55.
203. de Serres, F.J., *Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys*. Chest, 2002. **122**(5): p. 1818-29.
204. de Serres, F.J. and I. Blanco, *Prevalence of alpha1-antitrypsin deficiency alleles PI*S and PI*Z worldwide and effective screening for each of the five phenotypic classes PI*MS, PI*MZ, PI*SS, PI*SZ, and PI*ZZ: a comprehensive review*. Ther Adv Respir Dis, 2012. **6**(5): p. 277-95.
205. South, A., *rworldmap: A New R package for Mapping Global Data*. R Journal, 2011. **3**(1): p. 35-43.
206. Lewis, J.H., et al., *Antithrombin Pittsburgh: an alpha1-antitrypsin variant causing hemorrhagic disease*. Blood, 1978. **51**(1): p. 129-37.
207. DeMeo, D.L. and E.K. Silverman, *Alpha1-antitrypsin deficiency. 2: genetic aspects of alpha(1)-antitrypsin deficiency: phenotypes and genetic modifiers of emphysema risk*. Thorax, 2004. **59**(3): p. 259-64.
208. Edgar, R.G., et al., *Treatment of lung disease in alpha-1 antitrypsin deficiency: a systematic review*. Int J Chron Obstruct Pulmon Dis, 2017. **12**: p. 1295-1308.

209. Sieluk, J., et al., *Costs of Medical Care Among Augmentation Therapy Users and Non-Users with Alpha-1 Antitrypsin Deficiency in the United States*. Chronic Obstr Pulm Dis, 2018. **6**(1): p. 6-16.
210. Gotzsche, P.C. and H.K. Johansen, *Intravenous alpha-1 antitrypsin augmentation therapy for treating patients with alpha-1 antitrypsin deficiency and lung disease*. Cochrane Database Syst Rev, 2016. **9**: p. CD007851.
211. Gotzsche, P.C. and H.K. Johansen, *Intravenous alpha-1 antitrypsin augmentation therapy for treating patients with alpha-1 antitrypsin deficiency and lung disease*. Cochrane Database Syst Rev, 2010(7): p. CD007851.
212. Griesse, M. and G. Scheuch, *Delivery of Alpha-1 Antitrypsin to Airways*. Ann Am Thorac Soc, 2016. **13 Suppl 4**(Supplement_4): p. S346-51.
213. Clark, V.C., *Liver Transplantation in Alpha-1 Antitrypsin Deficiency*. Clin Liver Dis, 2017. **21**(2): p. 355-365.
214. Carey, E.J., et al., *Outcomes for recipients of liver transplantation for alpha-1-antitrypsin deficiency-related cirrhosis*. Liver Transpl, 2013. **19**(12): p. 1370-6.
215. Kemmer, N., et al. *Alpha-1-antitrypsin deficiency: outcomes after liver transplantation*. in *Transplantation proceedings*. 2008. Elsevier.
216. Teckman, J.H., *Lack of effect of oral 4-phenylbutyrate on serum alpha-1-antitrypsin in patients with alpha-1-antitrypsin deficiency: a preliminary study*. J Pediatr Gastroenterol Nutr, 2004. **39**(1): p. 34-7.
217. Burrows, J.A., L.K. Willis, and D.H. Perlmutter, *Chemical chaperones mediate increased secretion of mutant α 1-antitrypsin (α 1-AT) Z: a potential pharmacological strategy for prevention of liver injury and emphysema in α 1-AT deficiency*. Proceedings of the National Academy of Sciences, 2000. **97**(4): p. 1796-1801.
218. Kaushal, S., et al., *Rapamycin reduces intrahepatic alpha-1-antitrypsin mutant Z protein polymers and liver injury in a mouse model*. Exp Biol Med (Maywood), 2010. **235**(6): p. 700-9.
219. Chiuchiolo, M.J. and R.G. Crystal, *Gene Therapy for Alpha-1 Antitrypsin Deficiency Lung Disease*. Ann Am Thorac Soc, 2016. **13 Suppl 4**(Suppl 4): p. S352-69.
220. Crespo, J., et al., *Long-term expression of the human α 1-antitrypsin gene in mice employing anionic and cationic liposome vectors*. Biochemical pharmacology, 1996. **51**(10): p. 1309-1314.
221. Canonico, A.E., et al., *Aerosol and intravenous transfection of human alpha 1-antitrypsin gene to lungs of rabbits*. Am J Respir Cell Mol Biol, 1994. **10**(1): p. 24-9.
222. Alino, S.F., A. Crespo, and F. Dasi, *Long-term therapeutic levels of human alpha-1 antitrypsin in plasma after hydrodynamic injection of nonviral DNA*. Gene Ther, 2003. **10**(19): p. 1672-9.
223. Kolodka, T.M., M. Finegold, and S.L.C. Woo, *Hepatic Gene-Therapy - Efficient Retroviral-Mediated Gene-Transfer into Rat Hepatocytes in-Vivo*. Somatic Cell and Molecular Genetics, 1993. **19**(5): p. 491-497.
224. Kay, M.A., et al., *Hepatic gene therapy: persistent expression of human α 1-antitrypsin in mice after direct gene delivery in vivo*. Human gene therapy, 1992. **3**(6): p. 641-647.
225. Paul-Smith, M.C., et al., *The murine lung as a factory to produce secreted intrapulmonary and circulatory proteins*. Gene Ther, 2018. **25**(5): p. 345-358.
226. Rosenfeld, M.A., et al., *Adenovirus-mediated transfer of a recombinant alpha 1-antitrypsin gene to the lung epithelium in vivo*. Science, 1991. **252**(5004): p. 431-4.
227. Mastrangeli, A., et al., *Diversity of airway epithelial cell targets for in vivo recombinant adenovirus-mediated gene transfer*. J Clin Invest, 1993. **91**(1): p. 225-34.
228. Cerullo, V., et al., *Antigen-Specific Tolerance of Human α 1-Antitrypsin Induced by Helper-Dependent Adenovirus*. Human gene therapy, 2007. **18**(12): p. 1215-1224.
229. Conlon, T.J., et al., *Efficient hepatic delivery and expression from a recombinant adeno-associated virus 8 pseudotyped alpha1-antitrypsin vector*. Mol Ther, 2005. **12**(5): p. 867-75.

230. Limberis, M.P., et al., *Transduction efficiencies of novel AAV vectors in mouse airway epithelium in vivo and human ciliated airway epithelium in vitro*. Mol Ther, 2009. **17**(2): p. 294-301.
231. Lu, Y., et al., *Therapeutic level of functional human alpha 1 antitrypsin (hAAT) secreted from murine muscle transduced by adeno-associated virus (rAAV1) vector*. J Gene Med, 2006. **8**(6): p. 730-5.
232. Chiuchiolo, M.J., et al., *Intraleural administration of an AAVrh.10 vector coding for human alpha1-antitrypsin for the treatment of alpha1-antitrypsin deficiency*. Hum Gene Ther Clin Dev, 2013. **24**(4): p. 161-73.
233. De, B.P., et al., *High levels of persistent expression of alpha1-antitrypsin mediated by the nonhuman primate serotype rh.10 adeno-associated virus despite preexisting immunity to common human adeno-associated viruses*. Mol Ther, 2006. **13**(1): p. 67-76.
234. Gruntman, A.M., et al., *Bridging from Intramuscular to Limb Perfusion Delivery of rAAV: Optimization in a Non-human Primate Study*. Mol Ther Methods Clin Dev, 2019. **13**: p. 233-242.
235. Gruntman, A., et al., *rAAV9 airway delivery results in effective knockdown of mutant alpha 1-antitrypsin in the liver while upregulating wildtype alpha 1-antitrypsin in the lung*, in 2012 UMass Center for Clinical and Translational Science Research Retreat. 2012: Worcester.
236. Milstein, S., et al., *Methods to Identify and Characterize siRNAs Targeting Serpin A1 In Vitro and In Vivo Using RNA Interference*. Methods Mol Biol, 2017. **1639**: p. 115-126.
237. Cruz, P.E., et al., *In vivo post-transcriptional gene silencing of alpha-1 antitrypsin by adeno-associated virus vectors expressing siRNA*. Lab Invest, 2007. **87**(9): p. 893-902.
238. Guo, S., et al., *Antisense oligonucleotide treatment ameliorates alpha-1 antitrypsin-related liver disease in mice*. J Clin Invest, 2014. **124**(1): p. 251-61.
239. Mueller, C., et al., *Sustained miRNA-mediated knockdown of mutant AAT with simultaneous augmentation of wild-type AAT has minimal effect on global liver miRNA profiles*. Mol Ther, 2012. **20**(3): p. 590-600.
240. Bjursell, M., et al., *Therapeutic Genome Editing With CRISPR/Cas9 in a Humanized Mouse Model Ameliorates alpha1-antitrypsin Deficiency Phenotype*. EBioMedicine, 2018. **29**: p. 104-111.
241. Shen, S., et al., *Amelioration of Alpha-1 Antitrypsin Deficiency Diseases with Genome Editing in Transgenic Mice*. Hum Gene Ther, 2018. **29**(8): p. 861-873.
242. Song, C.Q., et al., *In Vivo Genome Editing Partially Restores Alpha1-Antitrypsin in a Murine Model of AAT Deficiency*. Hum Gene Ther, 2018. **29**(8): p. 853-860.
243. Brigham, K.L., et al., *Transfection of nasal mucosa with a normal alpha1-antitrypsin gene in alpha1-antitrypsin-deficient subjects: comparison with protein therapy*. Hum Gene Ther, 2000. **11**(7): p. 1023-32.
244. Brantly, M.L., et al., *Sustained transgene expression despite T lymphocyte responses in a clinical trial of rAAV1-AAT gene therapy*. Proc Natl Acad Sci U S A, 2009. **106**(38): p. 16363-8.
245. Brantly, M.L., et al., *Phase I trial of intramuscular injection of a recombinant adeno-associated virus serotype 2 alpha1-antitrypsin (AAT) vector in AAT-deficient adults*. Hum Gene Ther, 2006. **17**(12): p. 1177-86.
246. Flotte, T.R., et al., *Phase 2 clinical trial of a recombinant adeno-associated viral vector expressing alpha1-antitrypsin: interim results*. Hum Gene Ther, 2011. **22**(10): p. 1239-47.
247. Mueller, C., et al., *5 Year Expression and Neutrophil Defect Repair after Gene Therapy in Alpha-1 Antitrypsin Deficiency*. Mol Ther, 2017. **25**(6): p. 1387-1394.
248. Chiuchiolo, M.J., et al., *Phase I/II study of intraleural administration of a serotype rh.10 replication-deficient adeno-associated virus gene transfer vector expressing the human alpha1-antitrypsin cDNA to individuals with alpha1-antitrypsin deficiency*. Hum Gene Ther Clin Dev, 2014. **25**(3): p. 112-33.

249. Biotechnologies, A. *Adverum Biotechnologies Provides Program Updates*. 2018 [cited 2019 17/07/2019]; Available from: <http://investors.adverum.com/news-releases/news-release-details/adverum-biotechnologies-provides-program-updates>.
250. Turner, A.M., et al., *Hepatic-targeted RNA interference provides robust and persistent knockdown of alpha-1 antitrypsin levels in ZZ patients*. J Hepatol, 2018. **69**(2): p. 378-384.
251. Broutier, L., et al., *Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation*. Nat Protoc, 2016. **11**(9): p. 1724-43.
252. Haeussler, M., et al., *Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR*. Genome Biol, 2016. **17**(1): p. 148.
253. Dull, T., et al., *A third-generation lentivirus vector with a conditional packaging system*. J Virol, 1998. **72**(11): p. 8463-71.
254. Aurnhammer, C., et al., *Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences*. Hum Gene Ther Methods, 2012. **23**(1): p. 18-28.
255. Carlson, J.A., et al., *Accumulation of Piz alpha 1-antitrypsin causes liver damage in transgenic mice*. J Clin Invest, 1989. **83**(4): p. 1183-90.
256. Sabatino, D.E., et al., *Identification of mouse AAV capsid-specific CD8+ T cell epitopes*. Mol Ther, 2005. **12**(6): p. 1023-33.
257. Limberis, M.P., C.L. Bell, and J.M. Wilson, *Identification of the murine firefly luciferase-specific CD8 T-cell epitopes*. Gene Ther, 2009. **16**(3): p. 441-7.
258. Gambotto, A., et al., *Immunogenicity of enhanced green fluorescent protein (EGFP) in BALB/c mice: identification of an H2-Kd-restricted CTL epitope*. Gene Ther, 2000. **7**(23): p. 2036-40.
259. Martino, A.T., et al., *Measuring immune responses to recombinant AAV gene transfer*. Methods Mol Biol, 2011. **807**: p. 259-72.
260. Janetzki, S., et al., *Guidelines for the automated evaluation of Elispot assays*. Nat Protoc, 2015. **10**(7): p. 1098-115.
261. George, L.A., et al., *Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant*. N Engl J Med, 2017. **377**(23): p. 2215-2227.
262. Nathwani, A.C., et al., *Long-term safety and efficacy of factor IX gene therapy in hemophilia B*. N Engl J Med, 2014. **371**(21): p. 1994-2004.
263. Russell, S., et al., *Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial*. Lancet, 2017. **390**(10097): p. 849-860.
264. Mendell, J.R., et al., *Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy*. N Engl J Med, 2017. **377**(18): p. 1713-1722.
265. Koniaris, L.G., et al., *Liver regeneration*. J Am Coll Surg, 2003. **197**(4): p. 634-59.
266. Stanger, B.Z., *Cellular homeostasis and repair in the mammalian liver*. Annu Rev Physiol, 2015. **77**: p. 179-200.
267. Fagiuoli, S., et al., *Monogenic diseases that can be cured by liver transplantation*. J Hepatol, 2013. **59**(3): p. 595-612.
268. Wang, L., et al., *Hepatic gene transfer in neonatal mice by adeno-associated virus serotype 8 vector*. Hum Gene Ther, 2012. **23**(5): p. 533-9.
269. Bortolussi, G., et al., *Life-long correction of hyperbilirubinemia with a neonatal liver-specific AAV-mediated gene transfer in a lethal mouse model of Crigler-Najjar Syndrome*. Hum Gene Ther, 2014. **25**(9): p. 844-55.
270. Rangarajan, S., et al., *AAV5-Factor VIII Gene Transfer in Severe Hemophilia A*. N Engl J Med, 2017. **377**(26): p. 2519-2530.
271. Carlisle, R.C., et al., *Coating of adeno-associated virus with reactive polymers can ablate virus tropism, enable retargeting and provide resistance to neutralising antisera*. J Gene Med, 2008. **10**(4): p. 400-11.

272. Le, H.T., et al., *Utility of PEGylated recombinant adeno-associated viruses for gene transfer*. J Control Release, 2005. **108**(1): p. 161-77.
273. Lee, G.K., et al., *PEG conjugation moderately protects adeno-associated viral vectors against antibody neutralization*. Biotechnol Bioeng, 2005. **92**(1): p. 24-34.
274. Li, C., et al., *Single amino acid modification of adeno-associated virus capsid changes transduction and humoral immune profiles*. J Virol, 2012. **86**(15): p. 7752-9.
275. Lochrie, M.A., et al., *Mutations on the external surfaces of adeno-associated virus type 2 capsids that affect transduction and neutralization*. J Virol, 2006. **80**(2): p. 821-34.
276. Maersch, S., et al., *Optimization of stealth adeno-associated virus vectors by randomization of immunogenic epitopes*. Virology, 2010. **397**(1): p. 167-75.
277. Halbert, C.L., et al., *Repeat transduction in the mouse lung by using adeno-associated virus vectors with different serotypes*. J Virol, 2000. **74**(3): p. 1524-32.
278. Limberis, M.P. and J.M. Wilson, *Adeno-associated virus serotype 9 vectors transduce murine alveolar and nasal epithelia and can be readministered*. Proc Natl Acad Sci U S A, 2006. **103**(35): p. 12993-8.
279. Wang, L., et al., *Sustained correction of disease in naive and AAV2-pretreated hemophilia B dogs: AAV2/8-mediated, liver-directed gene therapy*. Blood, 2005. **105**(8): p. 3079-86.
280. Riviere, C., O. Danos, and A.M. Douar, *Long-term expression and repeated administration of AAV type 1, 2 and 5 vectors in skeletal muscle of immunocompetent adult mice*. Gene Ther, 2006. **13**(17): p. 1300-8.
281. Carlon, M.S., et al., *Immunological ignorance allows long-term gene expression after perinatal recombinant adeno-associated virus-mediated gene transfer to murine airways*. Hum Gene Ther, 2014. **25**(6): p. 517-28.
282. Bennett, J., et al., *AAV2 gene therapy readministration in three adults with congenital blindness*. Sci Transl Med, 2012. **4**(120): p. 120ra15.
283. Monteilhet, V., et al., *A 10 patient case report on the impact of plasmapheresis upon neutralizing factors against adeno-associated virus (AAV) types 1, 2, 6, and 8*. Mol Ther, 2011. **19**(11): p. 2084-91.
284. Mingozi, F., et al., *Overcoming preexisting humoral immunity to AAV using capsid decoys*. Sci Transl Med, 2013. **5**(194): p. 194ra92.
285. Corti, M., et al., *B-Cell Depletion is Protective Against Anti-AAV Capsid Immune Response: A Human Subject Case Study*. Mol Ther Methods Clin Dev, 2014. **1**.
286. Halbert, C.L., et al., *Successful readministration of adeno-associated virus vectors to the mouse lung requires transient immunosuppression during the initial exposure*. J Virol, 1998. **72**(12): p. 9795-805.
287. Lorain, S., et al., *Transient immunomodulation allows repeated injections of AAV1 and correction of muscular dystrophy in multiple muscles*. Mol Ther, 2008. **16**(3): p. 541-7.
288. Maldonado, R.A., et al., *Polymeric synthetic nanoparticles for the induction of antigen-specific immunological tolerance*. Proc Natl Acad Sci U S A, 2015. **112**(2): p. E156-65.
289. Mingozi, F., *Overcoming humoral immune responses to AAV vectors*. Human Gene Therapy, 2016. **27**(7): p. A2-A2.
290. Mingozi, F., et al., *Pharmacological modulation of humoral immunity in a nonhuman primate model of AAV gene transfer for hemophilia B*. Mol Ther, 2012. **20**(7): p. 1410-6.
291. Unzu, C., et al., *Transient and intensive pharmacological immunosuppression fails to improve AAV-based liver gene transfer in non-human primates*. J Transl Med, 2012. **10**: p. 122.
292. Meliani, A., et al., *Antigen-selective modulation of AAV immunogenicity with tolerogenic rapamycin nanoparticles enables successful vector re-administration*. Nat Commun, 2018. **9**(1): p. 4098.

293. Mount, J.D., et al., *Sustained phenotypic correction of hemophilia B dogs with a factor IX null mutation by liver-directed gene therapy*. *Blood*, 2002. **99**(8): p. 2670-6.
294. Ganschow, R. and B. Hoppe, *Review of combined liver and kidney transplantation in children*. *Pediatr Transplant*, 2015. **19**(8): p. 820-6.
295. Han, S.O., et al., *Low-Dose Liver-Targeted Gene Therapy for Pompe Disease Enhances Therapeutic Efficacy of ERT via Immune Tolerance Induction*. *Mol Ther Methods Clin Dev*, 2017. **4**: p. 126-136.
296. Doerfler, P.A., et al., *Copackaged AAV9 Vectors Promote Simultaneous Immune Tolerance and Phenotypic Correction of Pompe Disease*. *Hum Gene Ther*, 2016. **27**(1): p. 43-59.
297. Boisgerault, F., et al., *Prolonged gene expression in muscle is achieved without active immune tolerance using microRNA 142.3p-regulated rAAV gene transfer*. *Hum Gene Ther*, 2013. **24**(4): p. 393-405.
298. Mays, L.E., et al., *AAV8 induces tolerance in murine muscle as a result of poor APC transduction, T cell exhaustion, and minimal MHCI upregulation on target cells*. *Mol Ther*, 2014. **22**(1): p. 28-41.
299. Annoni, A., et al., *Liver gene therapy by lentiviral vectors reverses anti-factor IX pre-existing immunity in haemophilic mice*. *EMBO Mol Med*, 2013. **5**(11): p. 1684-97.
300. Matrai, J., et al., *Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk*. *Hepatology*, 2011. **53**(5): p. 1696-707.
301. Brown, B.D., et al., *A microRNA-regulated lentiviral vector mediates stable correction of hemophilia B mice*. *Blood*, 2007. **110**(13): p. 4144-52.
302. Auricchio, A., et al., *Noninvasive gene transfer to the lung for systemic delivery of therapeutic proteins*. *J Clin Invest*, 2002. **110**(4): p. 499-504.
303. Le Meur, G., et al., *Safety and Long-Term Efficacy of AAV4 Gene Therapy in Patients with RPE65 Leber Congenital Amaurosis*. *Mol Ther*, 2018. **26**(1): p. 256-268.
304. Chicoine, L.G., et al., *Plasmapheresis eliminates the negative impact of AAV antibodies on microdystrophin gene expression following vascular delivery*. *Mol Ther*, 2014. **22**(2): p. 338-347.
305. Keeler, G.D., D.M. Markusic, and B.E. Hoffman, *Liver induced transgene tolerance with AAV vectors*. *Cell Immunol*, 2019. **342**: p. 103728.
306. LoDuca, P.A., B.E. Hoffman, and R.W. Herzog, *Hepatic Gene Transfer as a Means of Tolerance Induction to Transgene Products*. *Current Gene Therapy*, 2009. **9**(2): p. 104-114.
307. Tiegs, G. and A.W. Lohse, *Immune tolerance: what is unique about the liver*. *J Autoimmun*, 2010. **34**(1): p. 1-6.
308. Davidoff, A.M., et al., *Comparison of the ability of adeno-associated viral vectors pseudotyped with serotype 2, 5, and 8 capsid proteins to mediate efficient transduction of the liver in murine and nonhuman primate models*. *Mol Ther*, 2005. **11**(6): p. 875-88.
309. Jiang, H., et al., *Multiyear therapeutic benefit of AAV serotypes 2, 6, and 8 delivering factor VIII to hemophilia A mice and dogs*. *Blood*, 2006. **108**(1): p. 107-15.
310. Nakai, H., et al., *Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice*. *J Virol*, 2005. **79**(1): p. 214-24.
311. Balazs, A.B., et al., *Antibody-based protection against HIV infection by vectored immunoprophylaxis*. *Nature*, 2012. **481**(7379): p. 81-U88.
312. Chew, W.L., et al., *A multifunctional AAV-CRISPR-Cas9 and its host response*. *Nat Methods*, 2016. **13**(10): p. 868-74.
313. Mowat, F.M., et al., *Tyrosine capsid-mutant AAV vectors for gene delivery to the canine retina from a subretinal or intravitreal approach*. *Gene Ther*, 2014. **21**(1): p. 96-105.
314. Petrs-Silva, H., et al., *High-efficiency transduction of the mouse retina by tyrosine-mutant AAV serotype vectors*. *Mol Ther*, 2009. **17**(3): p. 463-71.

315. He, Y., et al., *Kinetics of adeno-associated virus serotype 2 (AAV2) and AAV8 capsid antigen presentation in vivo are identical*. Hum Gene Ther, 2013. **24**(5): p. 545-53.
 316. Paneda, A., et al., *Effect of adeno-associated virus serotype and genomic structure on liver transduction and biodistribution in mice of both genders*. Hum Gene Ther, 2009. **20**(8): p. 908-17.
 317. Li, H., et al., *Pre-existing AAV capsid-specific CD8+ T cells are unable to eliminate AAV-transduced hepatocytes*. Mol Ther, 2007. **15**(4): p. 792-800.
 318. Somanathan, S., et al., *AAV vectors avoid inflammatory signals necessary to render transduced hepatocyte targets for destructive T cells*. Mol Ther, 2010. **18**(5): p. 977-82.
 319. Greig, J.A., et al., *Intramuscular administration of AAV overcomes pre-existing neutralizing antibodies in rhesus macaques*. Vaccine, 2016. **34**(50): p. 6323-6329.
 320. Greig, J.A., et al., *Intramuscular injection of AAV8 in mice and macaques is associated with substantial hepatic targeting and transgene expression*. PLoS One, 2014. **9**(11): p. e112268.
 321. Cooper, M., et al., *Improved induction of immune tolerance to factor IX by hepatic AAV-8 gene transfer*. Hum Gene Ther, 2009. **20**(7): p. 767-76.
 322. Kumar, S.R.P., et al., *The Balance between CD8(+) T Cell-Mediated Clearance of AAV-Encoded Antigen in the Liver and Tolerance Is Dependent on the Vector Dose*. Mol Ther, 2017. **25**(4): p. 880-891.
 323. Breous, E., S. Somanathan, and J.M. Wilson, *BALB/c mice show impaired hepatic tolerogenic response following AAV gene transfer to the liver*. Mol Ther, 2010. **18**(4): p. 766-74.
 324. Annoni, A., et al., *Modulation of immune responses in lentiviral vector-mediated gene transfer*. Cell Immunol, 2019. **342**: p. 103802.
 325. Howe, S.J., et al., *Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients*. J Clin Invest, 2008. **118**(9): p. 3143-50.
 326. Li, H., et al., *In vivo genome editing restores haemostasis in a mouse model of haemophilia*. Nature, 2011. **475**(7355): p. 217-21.
 327. Barzel, A., et al., *Promoterless gene targeting without nucleases ameliorates haemophilia B in mice*. Nature, 2015. **517**(7534): p. 360-4.
 328. Konkle, B.A., H. Huston, and S. Nakaya Fletcher, *Hemophilia B*, in *GeneReviews((R))*, M.P. Adam, et al., Editors. 1993, University of Washington, Seattle
- University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.: Seattle (WA).
329. Klompe, S.E., et al., *Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration*. Nature, 2019. **571**(7764): p. 219-225.
 330. Strecker, J., et al., *RNA-guided DNA insertion with CRISPR-associated transposases*. Science, 2019. **365**(6448): p. 48-53.
 331. Kwon, J.B. and C.A. Gersbach, *Jumping at the chance for precise DNA integration*. Nat Biotechnol, 2019.
 332. Milone, M.C. and U. O'Doherty, *Clinical use of lentiviral vectors*. Leukemia, 2018. **32**(7): p. 1529-1541.
 333. Kebriaei, P., et al., *Gene Therapy with the Sleeping Beauty Transposon System*. Trends Genet, 2017. **33**(11): p. 852-870.
 334. Rostovskaya, M., et al., *Transposon-mediated BAC transgenesis in human ES cells*. Nucleic Acids Res, 2012. **40**(19): p. e150.
 335. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein*. Nat Biotechnol, 2004. **22**(12): p. 1567-72.
 336. Shaner, N.C., et al., *A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum*. Nat Methods, 2013. **10**(5): p. 407-9.

337. McKenna, A., et al., *Whole-organism lineage tracing by combinatorial and cumulative genome editing*. Science, 2016. **353**(6298): p. aaf7907.
338. Flotte, T.R., et al., *Expression of the cystic fibrosis transmembrane conductance regulator from a novel adeno-associated virus promoter*. J Biol Chem, 1993. **268**(5): p. 3781-90.
339. Enriquez-Harris, P., et al., *A pause site for RNA polymerase II is associated with termination of transcription*. The EMBO Journal, 1991. **10**(7): p. 1833-1842.
340. van Haasteren, J., S.C. Hyde, and D.R. Gill, *Lessons learned from lung and liver in-vivo gene therapy: implications for the future*. Expert Opin Biol Ther, 2018. **18**(9): p. 959-972.
341. Harmatz, P., et al., *Update on phase 1/2 clinical trials for MPS I and MPS II using ZFN-mediated in vivo genome editing*. Molecular Genetics and Metabolism, 2018. **123**(2): p. S59-S60.
342. Cantore, A., et al., *Liver-directed lentiviral gene therapy in a dog model of hemophilia B*. Sci Transl Med, 2015. **7**(277): p. 277ra28.
343. Flotte, T.R., *Alpha-1 Antitrypsin Deficiency as a Candidate for Gene Editing*. Hum Gene Ther, 2018. **29**(8): p. 843-844.
344. Smith, C., et al., *Efficient and allele-specific genome editing of disease loci in human iPSCs*. Mol Ther, 2015. **23**(3): p. 570-7.
345. Yusa, K., et al., *Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells*. Nature, 2011. **478**(7369): p. 391-4.
346. Borel, F., et al., *Survival Advantage of Both Human Hepatocyte Xenografts and Genome-Edited Hepatocytes for Treatment of alpha-1 Antitrypsin Deficiency*. Mol Ther, 2017. **25**(11): p. 2477-2489.
347. Stephens, C.J., et al., *Targeted in vivo knock-in of human alpha-1-antitrypsin cDNA using adenoviral delivery of CRISPR/Cas9*. Gene Ther, 2018. **25**(2): p. 139-156.
348. Merkle, T., et al., *Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides*. Nat Biotechnol, 2019. **37**(2): p. 133-138.
349. Clevers, H., *Modeling Development and Disease with Organoids*. Cell, 2016. **165**(7): p. 1586-1597.
350. Lancaster, M.A. and M. Huch, *Disease modelling in human organoids*. Dis Model Mech, 2019. **12**(7): p. dmm039347.
351. Schwank, G., et al., *Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients*. Cell Stem Cell, 2013. **13**(6): p. 653-8.
352. Alton, E.W., et al., *Preparation for a first-in-man lentivirus trial in patients with cystic fibrosis*. Thorax, 2017. **72**(2): p. 137-147.
353. Gonzalez-Cordero, A., et al., *Assessment of AAV Vector Tropisms for Mouse and Human Pluripotent Stem Cell-Derived RPE and Photoreceptor Cells*. Hum Gene Ther, 2018. **29**(10): p. 1124-1139.
354. Takebe, T., et al., *Vascularized and functional human liver from an iPSC-derived organ bud transplant*. Nature, 2013. **499**(7459): p. 481-4.
355. Shinozawa, T., H.Y. Yoshikawa, and T. Takebe, *Reverse engineering liver buds through self-driven condensation and organization towards medical application*. Dev Biol, 2016. **420**(2): p. 221-229.
356. Huch, M., et al., *In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration*. Nature, 2013. **494**(7436): p. 247-50.
357. Huch, M., et al., *Long-term culture of genome-stable bipotent stem cells from adult human liver*. Cell, 2015. **160**(1-2): p. 299-312.
358. Hu, H., et al., *Long-Term Expansion of Functional Mouse and Human Hepatocytes as 3D Organoids*. Cell, 2018. **175**(6): p. 1591-1606 e19.
359. Villiger, L., et al., *Treatment of a metabolic liver disease by in vivo genome base editing in adult mice*. Nat Med, 2018. **24**(10): p. 1519-1525.
360. Kasai, F., et al., *HuH-7 reference genome profile: complex karyotype composed of massive loss of heterozygosity*. Hum Cell, 2018. **31**(3): p. 261-267.
361. Ben-David, U., et al., *Genetic and transcriptional evolution alters cancer cell line drug response*. Nature, 2018. **560**(7718): p. 325-330.

362. Liu, Y., et al., *Multi-omic measurements of heterogeneity in HeLa cells across laboratories*. Nat Biotechnol, 2019. **37**(3): p. 314-322.
363. Nelson, C.E., et al., *Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy*. Nat Med, 2019. **25**(3): p. 427-432.
364. Wu, Z., et al., *Single amino acid changes can influence titer, heparin binding, and tissue tropism in different adeno-associated virus serotypes*. J Virol, 2006. **80**(22): p. 11393-7.
365. Wu, Z., et al., *Alpha2,3 and alpha2,6 N-linked sialic acids facilitate efficient binding and transduction by adeno-associated virus types 1 and 6*. J Virol, 2006. **80**(18): p. 9093-103.
366. Zhou, J., et al., *Differentiated human airway organoids to assess infectivity of emerging influenza virus*. Proc Natl Acad Sci U S A, 2018. **115**(26): p. 6822-6827.
367. Alam, S. and R. Mahadeva, *Animal Models of Alpha-1-Antitrypsin Deficiency*, in *Alpha-1-antitrypsin Deficiency*, N. Kalsheker and R. Stockley, Editors. 2017, Academic Press: Boston. p. 179-195.
368. Borel, F., et al., *Editing out five Serpina1 paralogs to create a mouse model of genetic emphysema*. Proc Natl Acad Sci U S A, 2018. **115**(11): p. 2788-2793.
369. Wang, D., et al., *Deletion of Serpina1a, a murine alpha1-antitrypsin ortholog, results in embryonic lethality*. Exp Lung Res, 2011. **37**(5): p. 291-300.
370. Kushi, A., et al., *Disruption of the murine alpha1-antitrypsin/PI2 gene*. Exp Anim, 2004. **53**(5): p. 437-43.
371. He, N., et al., *Generation of Alpha-1 Antitrypsin Knockout and PI*ZZ Ferrets Using Crispr/Cas9. A Genetic Model of Emphysema*. Annals of the American Thoracic Society, 2018. **15**(Supplement_4): p. S292-S293.
372. de Vree, P.J., et al., *Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping*. Nat Biotechnol, 2014. **32**(10): p. 1019-25.
373. Hottentot, Q.P., et al., *Targeted Locus Amplification and Next-Generation Sequencing*. Methods Mol Biol, 2017. **1492**: p. 185-196.
374. Watson, C.M., et al., *Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications*. Lab Invest, 2019.
375. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells*. Nucleic Acids Res, 1988. **16**(3): p. 1215.
376. Genomics, X. *Sample Preparation Demonstrated Protocol Salting Out Method for DNA Extraction from Cells*. 2017 [cited 2019 15-07-2019]; Available from: https://assets.ctfassets.net/an68im79xiti/375eRsfeU2YqMsSCc2IMy/476e800d7f3e0dd425b7bb54ad495877/CG000116_SamplePrepDemonstratedProtocol_-_DesaltingDNAExtractionCells_RevA.pdf.
377. Li, H., *Minimap2: pairwise alignment for nucleotide sequences*. Bioinformatics, 2018. **34**(18): p. 3094-3100.
378. Robinson, J.T., et al., *Integrative genomics viewer*. Nat Biotechnol, 2011. **29**(1): p. 24-6.
379. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features*. Bioinformatics, 2010. **26**(6): p. 841-2.
380. Sifers, R.N., et al., *Tissue specific expression of the human alpha-1-antitrypsin gene in transgenic mice*. Nucleic Acids Res, 1987. **15**(4): p. 1459-75.
381. Carroll, D., *Collateral damage: benchmarking off-target effects in genome editing*. Genome Biol, 2019. **20**(1): p. 114.
382. Giannoukos, G., et al., *UDiTaS, a genome editing detection method for indels and genome rearrangements*. BMC Genomics, 2018. **19**(1): p. 212.
383. Hottentot, Q.P., et al., *Targeted locus amplification and next-generation sequencing*, in *Genotyping*. 2017, Springer. p. 185-196.
384. Teer, J.K., et al., *Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing*. Genome Res, 2010. **20**(10): p. 1420-31.
385. Davies, J.O., et al., *Multiplexed analysis of chromosome conformation at vastly improved sensitivity*. Nat Methods, 2016. **13**(1): p. 74-80.

386. Krueger, F., *Trim galore*. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, 2015.
387. Magoc, T. and S.L. Salzberg, *FLASH: fast length adjustment of short reads to improve genome assemblies*. Bioinformatics, 2011. **27**(21): p. 2957-63.
388. Kent, W.J., *BLAT - The BLAST-like alignment tool*. Genome Research, 2002. **12**(4): p. 656-664.
389. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Methods, 2012. **9**(4): p. 357-9.
390. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
391. Nobles, C.L., et al., *iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity*. Genome Biol, 2019. **20**(1): p. 14.
392. Vilenchik, M.M. and A.G. Knudson, *Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer*. Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12871-6.
393. *Gene therapy's next installment*. Nat Biotechnol, 2019. **37**(7): p. 697.
394. Regalado, A., *Gene-therapy cure has money-back guarantee*. MIT Technology Review, 2016.
395. Review, I.f.C.a.E. *Value Assessment Methods for "Single or Short-Term Transformative Therapies" (SSTs)*. 2019 [cited 2019 08-08-2019]; Available from: https://icer-review.org/wp-content/uploads/2019/08/ICER_SST_ProposedAdaptations_080619-2.pdf.
396. Nakai, H., et al., *Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo*. J Virol, 2001. **75**(15): p. 6969-76.
397. Sehara, Y., et al., *Persistent Expression of Dopamine-Synthesizing Enzymes 15 Years After Gene Transfer in a Primate Model of Parkinson's Disease*. Hum Gene Ther Clin Dev, 2017. **28**(2): p. 74-79.
398. Sun, S., et al. *Excise-and-Replace Strategy Using AAV-Mediated Delivery of CRISPR/Cas9 and HITI Donor Restores Full-Length Dystrophin Expression in Human Cardiomyocyte Model of Duchenne Muscular Dystrophy*. in ASGCT. 2019. CELL PRESS 50 HAMPSHIRE ST, FLOOR 5, CAMBRIDGE, MA 02139 USA.
399. Miller, D.G., L.M. Petek, and D.W. Russell, *Adeno-associated virus vectors integrate at chromosome breakage sites*. Nature genetics, 2004. **36**(7): p. 767.
400. Huang, H., et al. *CRISPR/Cas9-Mediated Targeted Insertion of Human F9 Achieves Therapeutic Circulating Protein Levels in Mice and Non-Human Primates*. in ASGCT. 2019. CELL PRESS 50 HAMPSHIRE ST, FLOOR 5, CAMBRIDGE, MA 02139 USA.
401. Lang, J.F., et al., *Standard screening methods underreport AAV-mediated transduction and gene editing*. Nat Commun, 2019. **10**(1): p. 3415.
402. Bednarski, C., et al., *Targeted Integration of a Super-Exon into the CFTR Locus Leads to Functional Correction of a Cystic Fibrosis Cell Line Model*. PLoS One, 2016. **11**(8): p. e0161072.
403. Mizuguchi, H., et al., *IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector*. Mol Ther, 2000. **1**(4): p. 376-82.
404. Li, L., S. Hu, and X. Chen, *Non-viral delivery systems for CRISPR/Cas9-based genome editing: Challenges and opportunities*. Biomaterials, 2018. **171**: p. 207-218.
405. Ortinski, P.I., et al., *Integrase-Deficient Lentiviral Vector as an All-in-One Platform for Highly Efficient CRISPR/Cas9-Mediated Gene Editing*. Mol Ther Methods Clin Dev, 2017. **5**: p. 153-164.
406. Alton, E., et al., *Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial*. Lancet Respir Med, 2015. **3**(9): p. 684-691.
407. Charlesworth, C.T., et al., *Identification of preexisting adaptive immunity to Cas9 proteins in humans*. Nat Med, 2019. **25**(2): p. 249-254.

408. Wagner, D.L., et al., *High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population*. Nat Med, 2019. **25**(2): p. 242-248.

Appendix

Appendix 1 – size selection script (python3)

```
#!/usr/bin/env python3
import os
from Bio.SeqIO.QualityIO import FastqGeneralIterator
from Bio.SeqIO.FastaIO import SimpleFastaParser
from Bio.Seq import Seq
from Bio.Alphabet import generic_dna
from Bio import SeqIO
import glob
count = 0
total_len = 0

f = open('OUTPUT_FASTA.fa', 'w+' )
file_array = sorted(glob.glob("INPUT_FAST_FILES_*.fastq"))

for fastq_file in file_array:
    with open(fastq_file) as in_handle:
        for title, seq, qual in FastqGeneralIterator(in_handle):
            length = len(seq)
            if length > 1000:
                f.write(">" + title + "\n" + seq + "\n" )
```

Appendix 2 – de-duplication script (python3)

```
#!/usr/bin/env python3
import sys
print(sys.version)
from Bio.SeqIO.QualityIO import FastqGeneralIterator
from Bio.Seq import Seq
from Bio.Alphabet import generic_dna
from Bio import SeqIO
import numpy as np
import pandas as pd

pd.set_option('display.max_colwidth', 1000)
count = 0
total_len = 0
h = open('OUTPUT_DEDUPLICATED_READS.fastq', 'w+')
dt = np.dtype(object)
alist = []
blist = []
clist = []
dlist = []

with open("ORIGINAL_READS_combined.fastq") as in_handle1:
    for title, seq, qual in FastqGeneralIterator(in_handle1):
        alist.append("@" + title )
        blist.append(seq )
        clist.append("+")
        dlist.append(qual )

df = pd.DataFrame()
df['titles'] = alist
df['seq'] = blist
df['plus'] = clist
df['qual'] = dlist
```

```

print (df.shape)
df_copy = df.copy()
df_copy.drop_duplicates(subset=['seq'], inplace=True)
listx = []
listx = df_copy.values.tolist()
output2 = []
for hhh in listx:
    for ggg in hhh:
        output2.append(ggg)
h.write("\n".join(output2))

```

Appendix 3 – Capture-seq bioinformatics commands (bash)

```

module load trim_galore
trim_galore -paired ORIGINAL_READS_R1.fq JO_21_R2.fq

module load flash
./flash .\ORIGINAL_READS_R1_val_1.fq .\ORIGINAL_READS_val_2.fq -o
ORIGINAL_READS_combined.fastq

RUN DEDUPLICATION SCRIPT on ORIGINAL_READS_combined.fastq

sed -n '1~4s/^@/>/p;2~4p' DEDUPLICATED_READS.fastq >
DEDUPLICATED_READS.fa

module load blat
/package/blat/35/bin/blat -minScore=20 -minIdentity=25 -maxIntron=10000 -
tileSize=11 INSERT_SEQUENCE.fa DEDUPLICATED_READS.fa
DEDUPLICATED_READS_insert_aln.psl

perl /home/molhaem2/jvanhaas/00scripts/MNase_psl_multi7.pl -f
DEDUPLICATED_READS.fastq -p DEDUPLICATED_READS_insert_aln.psl -r
DEDUPLICATED_READS_insert_aln -all

module load bowtie2
bowtie2 -local -p 4 -X 1000 -x /databank/indices/bowtie2/hg19/hg19
DEDUPLICATED_READS_insert_aln.fastq -S DEDUPLICATED_READS_insert_aln.sam

samtools view -S -b -o DEDUPLICATED_READS_insert_aln.bam
DEDUPLICATED_READS_insert_aln.sam
samtools sort DEDUPLICATED_READS_insert_aln.bam -o
DEDUPLICATED_READS_insert_aln.sorted.bam
samtools index DEDUPLICATED_READS_insert_aln.sorted.bam

module load bedtools
bamCoverage -b DEDUPLICATED_READS_insert_aln.sorted.bam -o
DEDUPLICATED_READS_insert_aln.sorted.bam.bw

```

Appendix 4 – Binned read counter (python3)

```
#!/usr/bin/env python3
import pysam
samfile = pysam.AlignmentFile("BAM_FILE_OF_READS.bam", "rb")

window = 5000
f = open('COUNT_OUTPUT.tsv', 'w+' )

for i in range(len(samfile.references)):
    refname = samfile.references[i]
    seqlen = samfile.lengths[i]
    for j in range(1, seqlen, window):
        stop = j+window-1 if j+window-1 < samfile.lengths[i] else
samfile.lengths[i]
        region_set = set()
        for read in samfile.fetch(refname, j, stop):
            region_set.add(read.query_name)
        #print("{}\t{}\t{}\t{}".format(refname, j, stop,
len(region_set)))
        if len(region_set) > 0:
            f.write("{}\t{}\t{}\t{}\n".format(refname, j, stop,
len(region_set)))
```

Appendix 5 – Commands used for CasEnrich (bash)

```
#align to insert sequence
minimap2 -ax map-ont -N 0 INSERT_SEQUENCE.fa READS_FROM_SEQUENCING.fa >
OUTPUT_ALIGNMENT_TO_INSERT.sam

#extract sequences from sam file
samtools fasta OUTPUT_ALIGNMENT_TO_INSERT.sam -F 4 >
OUTPUT_ALIGNMENT_TO_INSERT.fa

#index reference genome for minimap2
minimap2 -d mm9.mmi mm9.fa

#map sequences to mouse genome
minimap2 -ax map-ont mm9.mmi OUTPUT_ALIGNMENT_TO_INSERT.fa >
OUTPUT_MAPPING_TO_mm9.sam

samtools view -S -b -o OUTPUT_MAPPING_TO_mm9.bam
OUTPUT_MAPPING_TO_mm9.sam
samtools sort OUTPUT_MAPPING_TO_mm9.bam -o
OUTPUT_MAPPING_TO_mm9.sorted.bam
samtools index OUTPUT_MAPPING_TO_mm9.sorted.bam
```


Appendix 6 – CRISPOR predicted off-targets for Serpina1 Exon1 gRNA 2

Chromosome	Start	End	strand	Location in human genome	Insertion detected
chr18	41106648	41106674	+	intergenic: SYT4-RNU6-443P	No
chr17	76454407	76454433	-	intron: DNAH17	No
chr9	116436951	116436977	+	intron: RP11-18B16.2	No
chr3	137373557	137373583	+	intergenic: RP11-575C1.1-NPM1P17	No
chr12	2060633	2060659	-	intron: DCP1B	No
chr3	11879061	11879087	+	intron: TAMM41	No
chr11	117610148	117610174	+	intergenic: RP11-35P15.1-DSCAML1	No
chr2	16126810	16126836	+	intergenic: RN7SL104P-AC010145.4	No
chrX	51897288	51897314	+	intergenic: MAGED4B-MAGED4	No
chr8	34689487	34689513	+	intron: RP11-258J10.1	No
chr5	93505620	93505646	-	intron: KIAA0825	No
chr16	52835944	52835970	-	intergenic: RP11-297L17.6-RP11-467J12.1	No
chr16	25724555	25724581	-	intergenic: HS3ST4-RP11-185O21.1	No
chr9	108778817	108778843	+	intergenic: SLC25A6P5-RP11-6F6.1	No
chrX	47098656	47098682	-	intron: USP11	No
chr18	56722372	56722398	+	intron: OACYLP	No
chr8	96587916	96587942	-	intergenic: RP11-90D11.1-RNU6-690P	No
chr9	131397924	131397950	+	exon: WDR34	No
chr5	31592517	31592543	-	intergenic: C5orf22-RNU6-363P	No
chr10	134928274	134928300	-	intron: GPR123	No