



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

CRISPR-Based Gene Drives for Pest Control

Gus Rowan McFarlane



Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

May 2021

“Technology is a useful servant but dangerous master”

Christian Lange, 1921

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Gus Rowan McFarlane, May 2021

Abstract

Invasive pests impact the environment, economy and society. Current control methods are costly and largely inadequate, and they often lead to unwanted suffering in target and non-target species. Gene drives that enable super-mendelian inheritance of a transgene may offer a more cost-effective, humane and species-specific alternative to current methods. By harnessing gene drives to distort the sex-ratio of the breeding population it may be possible to control a population's reproductive success. Using CRISPR-Cas gene editing technology, this PhD project aimed to design, model and engineer a safeguard gene drive, known as a split gene drive, in mice that could spread female infertility through a laboratory-contained mouse population. Three gene drive strategies were designed and *in silico* modelled in wild mouse populations. Reagents were generated to engineer two of the three split gene drive strategies using mouse embryonic stem cell technology. Both these approaches aim to disrupt an essential female fertility gene (*OOEP*) to confer a recessive female-infertility phenotype. Split gene drive harbouring mouse embryonic stem cells were engineered using plasmid donor-DNA and a combination of SpCas9 ribonucleoprotein or plasmid-based AsCas12a endonuclease. Engineered cells were screened through a pipeline, which included analyses by PCR, droplet digital PCR, Sanger sequencing and functional testing of the integrated transgenic systems. These validated split gene drive embryonic stem cells and the corresponding regulatory approval for animal testing now allows for two split gene drive mouse models to be generated by blastocyst injection of the engineered cells. It is hoped that the findings from this PhD project will help guide the future development of safe gene drive systems for vertebrate pest management.

Lay summary

Invasive pests are a global concern that impact our environment, economy and society. Examples of damaging invasive pests include rabbits and feral cats in Australia, mink and grey squirrel in the United Kingdom, wild boar in the United States, and the widespread infestation of rats and mice around the globe. It is estimated that in the United States alone, introduced rats cost the economy more than US\$27 billion per year. To counter the impact of invasive pests, control measures are used, such as shooting, trapping and poisoning, which are brutal, costly and often lead to unwanted animal suffering in both targeted pests and non-targeted native animals. Gene drive technology may offer a humane, cost-effective and species-specific alternative. In most cases a particular gene has a 50% chance of being transmitted from parent to offspring. Gene drives are genes that beat the odds by being passed on to more than 50% of the offspring – that could be 50.01% or 100% – and this advantage enables them to spread through a population. Using a gene editing technology, known as CRISPR, scientists can now develop gene drives that can spread beneficial traits through wild populations. This PhD project set out to design, model and engineer a safe version of a gene drive, as a proof-of-concept, that could spread female infertility through a laboratory-contained population of mice. By spreading female infertility through a population, it is possible to reduce the population size due to the lack of fertile females available for breeding. If most females in the population are infertile, the population will reduce in size but if all females are infertile the population will collapse. This approach could be used as a humane pest management tool. Within this PhD project, gene drive strategies for this purpose were designed and computationally simulated in mouse populations. Two safeguarded gene drive approaches were selected for further development. These two safeguarded strategies will only work in laboratory mice and not in wild mice, protecting wild mice from unintentional release. CRISPR gene editing tools were developed and used to introduce the gene drives into the DNA of mouse cells. These modified mouse cells that carry the gene drives in their DNA can now be used to generate two gene drive mouse models for further study in laboratory-contained mouse populations. The results from this PhD project will help guide the development of safe gene drives as humane and practical pest management tools.

Acknowledgements

This PhD journey has been one of my most challenging endeavours and it would not have been possible without the support of my colleagues, friends and family. To these people I extend my deepest gratitude.

To my supervisors, Prof Bruce Whitelaw and Dr Simon Lillico, thank you for giving me this opportunity and the continued guidance throughout. Bruce, your optimism and advice steadied the ship in rough seas. Simon, your knowledge, logic and light-hearted humour has kept me sane and on track.

To Claire Neil and Dr Chris Proudfoot. Claire, without your nurturing and laboratory magic I would not have progressed to this point. Chris, my unofficial third supervisor, you have ensured that I have dotted the i's and crossed the t's each step of the way.

To my fellow PhD students, past and present, especially those in the Whitelaw group, Sarah, Dina, Akshay and Hamish, thank you for the reassurance and encouragement.

To Dr Gregor Gorjanc and Nicky Faber, for your mind-boggling mathematical and computation brilliance. It was a pleasure collaborating with you both.

To the numerous other people who have helped me throughout this project, including Dr Jo Stevens, Dr Derya Ozdemir, Dr Peter Hohenstein, Bob Fleming and Dr Ben Davies.

To the Commonwealth Scholarship Committee and the BBSRC for their financial support.

To my family, especially my mother and sister, despite the geographic distance you have always been my biggest fans, cheering me on from the sideline.

To my wife, Wiebke, thank you for your support and patience over the last four years. You're my rock and you fill my life with happiness, love and laughter. Tilo and I are so lucky to have you and I can't wait to see what the next chapter holds for us.

Finally, I dedicate this thesis to my father who lost his courageous fight with motor neuron disease during my PhD. You have shaped me into the man I am today, and I am forever proud, grateful and fortunate to call you my father.

Contents

Declaration	i
Abstract	ii
Lay summary	iii
Acknowledgements	iv
Contents	v
List of Figures	xi
List of Tables	xvi
List of Abbreviations	xvii
1 Introduction	1
1.1 Invasive vertebrate pests	1
1.2 Selfish genetic elements	2
1.2.1 Homing endonuclease genes	4
1.2.2 Meiotic drive	4
1.3 Harnessing evolution	5
1.4 What are gene drives?	5
1.5 Classical genetic engineering	6
1.6 Nuclease-assisted genetic engineering	6
1.7 ZFNs and TALENs	8
1.8 CRISPR-Cas	10
1.8.1 SpCas9	10
1.8.2 AsCas12a	12
1.8.3 CRISPR uptake and innovation	13
1.8.4 CRISPR-based gene drives	13
1.9 Driving population suppression	14

1.9.1	Homing-based gene drives	15
1.9.2	X-Shredder	17
1.10	Containing the spread	18
1.10.1	Daisy-chain system	19
1.10.2	Daisy-field system	20
1.11	Mice as a model organism	21
1.12	Public perception	22
1.13	Project aims	23
2	Materials and methods	25
2.1	General molecular biology	25
2.1.1	Genomic DNA extraction	25
2.1.2	Polymerase chain reaction (PCR)	25
2.1.3	Agarose gel electrophoresis	26
2.1.4	Purification of DNA from agarose gels and PCR reactions	26
2.1.5	Nucleic acid quantification	27
2.1.6	DNA sequencing	27
2.1.7	Droplet digital PCR	27
2.2	Molecular cloning	28
2.2.1	Restriction enzyme digestion	28
2.2.2	Annealing DNA oligos	28
2.2.3	Restriction cloning	29
2.2.4	Gibson assembly	29
2.2.5	Topoisomerase (TOPO) cloning	29
2.3	Bacterial molecular biology	30
2.3.1	Transformation protocol	30
2.3.2	Small scale preparation of plasmid DNA	30
2.3.3	Large scale preparation of plasmid DNA	31
2.4	CRISPR reagent development	32
2.4.1	gRNA design	32
2.4.2	Preparation and cloning of gRNAs in plasmids	34
2.4.3	Preparation of SpCas9 ribonucleoprotein complex	34
2.4.4	Screening gRNAs	35
2.5	Mammalian cell culture	36

2.5.1	Mouse embryonic stem (ES) cells	36
2.5.2	Human embryonic kidney 293T (HEK) cells	41
2.6	Bacteria, cells and mice	42
2.7	Reagents and buffers	42
2.7.1	Bacterial culture	42
2.7.2	Cell culture media and reagents	43
2.7.3	Plasmids	44
2.7.4	Other buffers	47
2.8	Computational tools	47

3 Gene drive population modelling, strategy formulation and experimental design 49

3.1	Introduction	49
3.1.1	Homing-based gene drive designs	49
3.1.2	<i>In silico</i> population modelling	52
3.1.3	Safeguarding gene drive research	52
3.1.4	Construct design and <i>in vitro</i> gene drive systems	53
3.1.5	Experimental design for animal research	54
3.1.6	Aims	55
3.2	Additional materials and Methods	55
3.3	Results	56
3.3.1	Candidate gene drive strategies	56
3.3.2	<i>In silico</i> population simulations of proposed gene drive strategies	59
3.3.3	Candidate female-fertility genes	63
3.3.4	Safeguarding against spread	64
3.3.5	Gene drive constructs	66
3.3.6	<i>In vitro</i> system for gene drive testing	67
3.3.7	Experimental design of animal experiments	68
3.4	Discussion	70
3.4.1	Strategy A is highly invasive	70
3.4.2	Strategy B and C have low invasive potential	71
3.4.3	Limitations of the model	72
3.4.4	Expression and design	72
3.4.5	Safeguard strategies	72

3.4.6	<i>In vitro</i> gene drive strategy	73
3.4.7	Experimental design and regulatory approval	75
3.4.8	Conclusions	75
4	Reagent development and validation	76
4.1	Introduction	76
4.1.1	Cell transfection methods	76
4.1.2	SpCas9 vs AsCas12a	77
4.1.3	<i>In silico</i> gRNA design	77
4.1.4	gRNA screening	78
4.1.5	Construct design	79
4.1.6	Molecular cloning techniques	83
4.1.7	Aims	86
4.2	Additional methods	87
4.2.1	gRNA screening workflow	87
4.3	Results	88
4.3.1	<i>In silico</i> gRNA design	88
4.3.2	Primer design and optimisation	91
4.3.3	gRNA screening	93
4.3.4	SpCas9 ribonucleoprotein screening	100
4.3.5	Construct development	101
4.4	Discussion	110
4.4.1	Designing gRNAs	110
4.4.2	Screening gRNAs	112
4.4.3	Benefits of Cas ribonucleoprotein	115
4.4.4	Cloning	116
4.4.5	Conclusions	118
5	Genome engineering split gene drives in mouse embryonic stem cells	119
5.1	Introduction	119
5.1.1	ES cell technology	119
5.1.2	Optimal ES cell culture	120
5.1.3	Media formulation	121

5.1.4	Targeting strategies	122
5.1.5	Genotyping ES cells	123
5.1.6	Karyotyping	126
5.1.7	Functional analysis of ES cells <i>in vitro</i>	127
5.1.8	Aim	128
5.2	Additional methods	128
5.2.1	ES cell genome engineering pipeline	128
5.2.2	Expansion of positive ES cell clones	129
5.3	Results	129
5.3.1	Transfections	129
5.3.2	Primer design and PCR screening	130
5.3.2	ddPCR primer testing	132
5.3.3	ddPCR results	133
5.3.4	Reconfirming PCRs	134
5.3.5	Wild-type allele sequencing	137
5.3.6	Subcloning and reconfirmation of G10 polyclonal line	139
5.3.7	Karyotyping	141
5.3.8	Morphology of selected ES clones	143
5.3.9	Cardiomyocyte differentiation	144
5.3.10	Floxed removal of <i>PAC</i> cassette	144
5.3.11	Functional validation of SpCas9 and gRNA expression	146
5.4	Discussion	150
5.4.1	ES cell technology	150
5.4.2	Non-model species	151
5.4.3	Genotyping	151
5.4.4	Droplet digital PCR (ddPCR)	152
5.4.5	New generation screening technologies	153
5.4.6	Ribonucleoprotein (RNP_	154
5.4.7	Techniques for increasing HDR efficiency	155
5.4.8	Karyotyping	156
5.4.9	Morphology and <i>in vitro</i> differentiation	158
5.4.10	Functional validation	158
5.4.11	Conclusions	159

6	Discussion	160
6.1	Ambitions and realities	160
6.2	Planned <i>in vivo</i> testing	160
6.3	Value of fundamental research	163
6.4	Interdisciplinary nature of gene drive research	164
6.5	Phased testing pathway	164
6.6	Risk assessment	165
6.7	Governance and regulations	165
6.8	Engagement	166
6.9	Human values	168
6.10	Other vertebrate applications	169
6.11	Final remarks	170
	References	171
A	Appendices	192
A.1	Publication: CRISPR-Based Gene Drives for Pest Control	192
A.2	PCR primers	198
A.3	PCR thermocycling conditions	200
A.4	Primers used for Sanger sequencing	200
A.5	ddPCR probes	201
A.6	Gene fragments and DNA oligos for cloning	201
A.7	gRNA DNA oligos	202
A.8	crRNAs for SpCas9 ribonucleoprotein	203
A.9	Project License for animal work	204
A.10	CRISPOR data outputs for <i>in silico</i> gRNA design	212
A.11	Preliminary PCR screens of engineered ES cells	214
A.12	ddPCR quality control data	220
A.13	Publication: On-farm genome editing of livestock	222
A.14	Publication: Accelerating Evolution	230

List of Figures

Figure 1.1 Invasive vertebrate pests are a global concern that impacts the environment, economy and society.	2
Figure 1.2 Examples of the three known primary mechanisms by which selfish genetic elements achieve 'drive'.	3
Figure 1.3 Mendelian and gene drive inheritance.	6
Figure 1.4 Nuclease-assisted gene targeting.	7
Figure 1.5 A schematic representation of zinc finger nuclease (ZFN) structure.	9
Figure 1.6 A schematic representation of transcription activator-like effector nuclease (TALEN) structure.	9
Figure 1.7 A schematic representation of <i>Streptococcus pyogenes</i> Cas9 (SpCas9).	11
Figure 1.8 A schematic representation of <i>Acidaminococcus</i> sp. Cas12a (AsCas12a).	12
Figure 1.9 CRISPR-based homing gene drive (GD).	15
Figure 1.10 X-Shredder (XS).	17
Figure 1.11 Daisy chains drive system.	20
Figure 1.12 Daisy-field system.	21
Figure 2.1 Example of the CRISPOR input screen for SpCas9 guide RNA (gRNA) design in the mouse reference genome.	33
Figure 2.2 Example of the CRISPOR output screen for SpCas9 guide RNA (gRNA) design in the mouse reference genome.	33
Figure 2.3 Embryoid body (EB) formation and plated EBs.	40
Figure 2.4 Plasmid maps.	46
Figure 3.1 Schematic diagrams of the three gene drive constructs considered for development.	56

Figure 3.2 Mechanism of inheritance and suppression for Strategy A.	57
Figure 3.3 Mechanism of inheritance and suppression for Strategy B.	58
Figure 3.4 Mechanism of inheritance and suppression for Strategy C.	59
Figure 3.5 <i>In silico</i> simulation of Strategy A in a population of 50,000 wild mice.	60
Figure 3.6 <i>In silico</i> population simulation of Strategy B and C with 3 gRNAs in a population of 50,000 wild mice.	62
Figure 3.7 <i>NLRP5</i> and <i>OOEP</i> candidate female-fertility genes from the mouse reference genome (GRCm38.p6).	63
Figure 3.8 A split gene drive (SGD) separates the gRNA/s and Cas onto separate chromosomes.	65
Figure 3.9 <i>ROSA26</i> from the mouse reference genome (GRCm38.p6).	65
Figure 3.10 Location of the gene drive constructs in mouse genome reference (GRCm38.p6).	66
Figure 3.11 Schematic designs of the split gene drive (SGD) constructs.	67
Figure 3.12 Workflow for <i>in vitro</i> testing gene drive functionality and homing efficiency in ES cells.	69
Figure 3.13 Possible genome editing outcomes from germline Cas expression (Strategy B) and constitutive Cas expression (Strategy C).	74
Figure 4.1 Cre- <i>LoxP</i> system.	82
Figure 4.2 Workflow for screening plasmid expressed CRISPR-Cas in mouse E14 embryonic stem (ES) cells.	87
Figure 4.3 SpCas9 (red) and AsCas12a (green) gRNAs targeting exon 1 and 2 of <i>OOEP</i> .	89
Figure 4.4 SpCas9 (red) and AsCas12a (green) gRNAs targeting exon 7 of <i>NLRP5</i> .	90
Figure 4.5 SpCas9 (red) and AsCas12a (green) gRNAs targeting intron 1 of the <i>ROSA26</i> locus.	91
Figure 4.6 Primer pair (pp) locations.	92
Figure 4.7 Gradient PCR of primer pair (pp) <i>ROSA26</i> , ppOOEP and ppNLRP5.	93
Figure 4.8 T7 endonuclease I (T7EI) assays of SpCas9 (top) and AsCas12a (bottom) gRNAs at the <i>OOEP</i> locus.	95

Figure 4.9 T7 endonuclease I (T7EI) assays of SpCas9 (top) and AsCas12a (bottom) at the <i>NLRP5</i> locus.	96
Figure 4.10 T7 endonuclease I (T7EI) assays of SpCas9 (top) and AsCas12a (bottom) at the <i>ROSA26</i> locus.	97
Figure 4.11 Sanger sequence chromatograms from the <i>OOEP</i> target region.	98
Figure 4.12 Sanger sequence chromatograms from the <i>ROSA26</i> target region.	98
Figure 4.13 Inference of CRISPR Edits (ICE) analysis results of SpCas9 gRNAs targeting <i>OOEP</i> .	99
Figure 4.14 Inference of CRISPR Edits (ICE) analysis results of AsCas12a gRNAs targeting <i>ROSA26</i> .	99
Figure 4.15 Sequence comparison of <i>OOEP</i> target site in E14 mouse embryonic stem (ES) cells and C57BL/6NCrl (Bl6) mice.	100
Figure 4.16 T7 endonuclease I (T7EI) assays of <i>OOEP</i> SpCas9 gRNAs 1 and 4 in ribonucleoprotein (RNP) form.	101
Figure 4.17 Development of <i>OOEP_IVT</i> plasmid.	104
Figure 4.18 Functional testing of <i>OOEP_IVT</i> in HEK cells.	105
Figure 4.19 Development of <i>OOEP_CAG</i> plasmid.	106
Figure 4.20 Alignment of the homology arms and gRNAs within <i>OOEP_CAG</i> and <i>OOEP_VASA</i> plasmids.	106
Figure 4.21 Development of <i>OOEP_VASA</i> plasmid.	107
Figure 4.22 Functional testing of <i>OOEP_CAG</i> and <i>OOEP_VASA</i> plasmids in E14 mouse embryonic stem (ES) cells.	108
Figure 4.23 Sanger sequencing traces from E14 mouse embryonic stem (ES) cells (red arrows) aligned to the homology arms in pDonor- <i>ROSA26</i> .	108
Figure 4.24 Development of <i>ROSA26_gRNAs</i> plasmid.	109
Figure 4.25 Alignment of the gRNA and adjusted homology arms within <i>ROSA26_gRNAs</i> plasmid.	110
Figure 5.1 Workflow for genome engineering ES cells.	128
Figure 5.2 Workflow for expanding PCR positive genome engineered ES clones from frozen 96 well plate (WP).	129

Figure 5.3 Schematics representation of the transfection combinations used for genome engineer three ES cell lines: OOEP_CAG, OOEP_VASA and ROSA26_gRNAs.	130
Figure 5.4 Primers and probes for screening clonal cell populations.	131
Figure 5.5 Summary of PCR analysis using 5' and 3' homology arm flanking primers in clonal populations of OOEP_CAG, OOEP_VASA and ROSA26_gRNAs edited clones in 96 well plate format.	132
Figure 5.6 Gel electrophoresis of PCR products to check specificity of ddPCR primers.	133
Figure 5.7 Summary of ddPCR transgene copy number results.	134
Figure 5.8 Reconfirmation of homology arm integration in thawed OOEP_CAG target cells.	135
Figure 5.9 Reconfirmation of homology arm integration in thawed OOEP_VASA targeted cells.	136
Figure 5.10 Reconfirmation of homology arm integration in thawed ROSA26_gRNAs targeted cells.	137
Figure 5.11 Sanger sequencing of the <i>OOEP</i> target region on the homologous wild-type (WT) chromosome in heterozygous OOEP_CAG targeted cells.	138
Figure 5.12 Sanger sequencing of the <i>OOEP</i> target region on the homologous wild-type (WT) chromosome in heterozygous OOEP_VASA targeted cells.	138
Figure 5.13 Sanger sequencing of the <i>ROSA26</i> target region on the homologous wild-type (WT) chromosome in heterozygous ROSA26_gRNAs targeted cells.	139
Figure 5.14 PCR analysis of the subcloned ROSA26_gRNAs G10 subclones.	140
Figure 5.15 Sanger sequencing of the <i>ROSA26</i> target region on the homologous chromosome in heterozygous ROSA26_gRNAs G10 subclones.	141
Figure 5.16 Summary of chromosome counting from metaphase spreads.	142
Figure 5.17 Representative images of metaphase chromosome spreads used for karyotyping of the clonal ES cell populations.	142
Figure 5.17 Morphology of selected ES clones plated on gelatine at low density.	143
Figure 5.18 Images of regions of beating cardiomyocytes, differentiated from selected clonal ES cell lines.	144
Figure 5.19 Primer design for testing the functionality of the <i>LoxP</i> sites flanking the <i>PAC</i> cassette in each cell line.	145

Figure 5.20 Flanking PCRs of the floxed <i>PAC</i> cassette in selected clones in the absence of Cre recombinase.	146
Figure 5.21 Flanking PCRs of the floxed <i>PAC</i> cassette in selected clones in the presence of Cre mRNA.	146
Figure 5.22 Functional analysis of SpCas9 expression using T7 endonuclease I assay on selected cell lines transfected with <i>NLRP5</i> gRNA 4 (gRNA).	148
Figure 5.23 Functional analysis of SpCas9 expression using T7 endonuclease I assay on selected cell lines transfected with SpCas9 expressed from pSL70.	149
Figure 6.1 Schematic overview of planned in vivo work.	161
Figure 6.2 Communities, publics, and stakeholders.	167
Figure A.11.1 Gel electrophoresis of PCRs from DNA extracted from ROSA26_gRNAs targeted cells in clonal 96 well plate format.	214
Figure A.11.2 Gel electrophoresis of PCRs from DNA extracted from ROSA26_gRNAs targeted cells in clonal 96 well plate format.	215
Figure A.11.3 Gel electrophoresis of PCRs from DNA extracted from OOEP_VASA targeted cells in clonal 96 well plate format.	216
Figure A.11.4 Gel electrophoresis of PCRs from DNA extracted from OOEP_VASA targeted cells in clonal 96 well plate format.	217
Figure A.11.5 Gel electrophoresis of PCRs from DNA extracted from OOEP_CAG targeted cells in clonal 96 well plate format.	218
Figure A.12.1 Quality control for ddPCR calls in genome edited cell lines.	220
Figure A.12.2 Quality control graphs for droplet digital PCR experiments.	221

List of Tables

Table 4.1 List of genetic elements, their origin and function used in all targeting constructs developed in this project.	103
Table A.2 PCR primers.	198
Table A.3 PCR thermocycling conditions for each polymerase.	200
Table A.4 Primers used for Sanger sequencing.	200
Table A.5 Probes for ddPCR.	201
Table A.6 Gene fragments (gBlocks) and ssDNA oligos used for cloning plasmid DNA constructs.	201
Table A.7 DNA oligos for cloning crRNAs into pSL70 and pY094 plasmids.	202
Table A.8 crRNAs and tracrRNA used for SpCas9 ribonucleoprotein.	203
Table A.10.1 CRISPOR predicted gRNA characteristics for SpCas9 gRNAs at <i>OOEP</i> , <i>NLRP5</i> and <i>ROSA26</i> loci.	212
Table A10.2 CRISPOR predicted gRNA characteristics for AsCas12a gRNAs at <i>OOEP</i> , <i>NLRP5</i> and <i>ROSA26</i> loci.	213

List of Abbreviations

2i	MEK and GSK3 inhibitors
3Rs	Principles of Replacement, Reduction and Refinement in animal research
AsCas12a	CRISPR associated protein 12a from <i>Acidaminococcus</i> sp.
AWERB	Animal Welfare and Ethical Review Body
Bl6	C57BL/6NCrl mouse strain
bp	Base pair
Cas	CRISPR associated protein
Cas12a	CRISPR associated protein 12a from <i>Acidaminococcus</i> sp.
Cas9	CRISPR associated protein 9 from <i>Streptococcus pyogenes</i>
CL	Containment level
CRISPOR	CRISPOR gRNA design software
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CWD	Chronic wasting disease
ddPCR	Droplet digital PCR
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand DNA break
dsDNA	Double-stranded DNA
dsRED	<i>Discosoma</i> sp. Red Fluorescent Protein
EB	Embryoid Body
EDTA	Ethylenediaminetetraacetic acid

EFFG	Essential female-fertility gene
eGFP	Enhanced green fluorescent protein
ES cell	Embryonic stem cell
FACS	Fluorescent-activated cell sorting
FAM	Fluorescein
FCS	Foetal calf serum
FITC	Fluorescein isothiocyanate filter
G	Guanine
gBlock	Synthesised dsDNA fragment
GC	Guanine-cytosine
gDNA	Genomic DNA
GFP	Green fluorescent protein
gRNA	Guide RNA
HA	Homology arm
HDR	Homology-directed repair
HEG	Homing endonuclease gene
HEK cells	Human embryonic kidney 293T cells
Het	Heterozygous
HEX	Hexachlorofluorescein
HFFG	Haplosufficient female-fertility gene
Hom	Homozygous
HR	Homologous recombination
ICE	Inference of CRISPR Edits
IDT	Integrated DNA Technologies
IGMM	Institute of Genetics & Molecular Medicine
Indels	Insertion or deletional mutations in the genome
kb	Kilobase pairs
LB	Luria-Bertani
LIF	Mouse leukaemia inhibitory factor
LSL	LoxP–Stop codon–LoxP

MCS	Multiple cloning site
MHC	Major histocompatibility complex
MMEJ	Microhomology-mediated end joining
NEB	New England Biolabs
NGS	Next generation sequencing
NHEJ	Non-homologous end-joining
NLS	Nuclear localisation signal
nt	Nucleotide
ONT	Oxford Nanopore Technologies
PAC	Puromycin-n-acetyl-transferase gene
PAM	Protospacer-adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PITCh	Precise Integration into Target Chromosome
pp	Primer pair
PPL	Project license from UK Home Office
PPL	UK Home Office granted project license for animal research
qPCR	Quantitative Polymerase Chain Reaction
RE	Restriction enzyme
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RVD	Repeat-variable di-residue
SGD	Split gene drive
SMC	Subcortical maternal complex
SOB	Super Optimal Broth
SOC	Super Optimal broth with Catabolite repression
SpCas9	CRISPR associated protein 9 from <i>Streptococcus pyogenes</i>
ssDNA	Single-stranded DNA
ssODN	Single-stranded oligodeoxynucleotides

STOP	Stop codon (or termination codon)
T7EI	T7 endonuclease I assay
TAE	Tris-acetate-EDTA
TALE	Transcription activator-like effector
TALEN	TALE nucleases
TE	Tris-EDTA
TLA	Targeted locus amplification
tracrRNA	Trans-activating crRNA
TxRed	Texas red filter
UK	United Kingdom of Great Britain
UNCBD	United Nations Convention on Biological Diversity
US	United States of America
UV	Ultraviolet
WGS	Whole genome sequencing
WHO	World Health Organisation
WP	Well plate
WT	Wild type
XS	X-Shredder
ZFN	Zinc finger nucleases
ZFP	Zinc finger protein

1 Introduction

1.1 Invasive vertebrate pests

In agriculture and wildlife, invasive vertebrate species are well known around the globe. Although their spread can have beneficial aspects, invasive species can disrupt ecosystems, impede biodiversity and threaten livelihoods. Damaging invasive species are classified as invasive pests and examples include: rabbits (*Oryctolagus cuniculus*) and feral cats (*Felis catus*) in Australia, American mink (*Neovison vison*) and grey squirrel (*Sciurus carolinensis*) in the United Kingdom (UK), European starling (*Sturnus vulgaris*), Burmese pythons (*Python bivittatus*) and wild boar (*Sus scrofa*) in the United States of America (USA), and the omni-present infestation of rats and mice around the globe (Krull et al., 2014; Robertson et al., 2017; Witmer & Fuller, 2011). The burden of invasive pests has been worsened in recent times by the increases in global trade, travel and climate change (Early et al., 2016). Current pest management tools for vertebrate pests are crude, costly, inhumane and often inadequate to cope with the scale of the problem. The ability to humanely control damaging invasive pests has long been a goal of scientists.

Of the numerous invasive vertebrate pests, mice and rats (Muridae) are recognised as some of the most prolific. At an economic level, the worldwide losses caused by mice and rats are enormous. It has been conservatively estimated that in Asia, each year, rats consume over 30 million tons of rice. That is enough rice to feed 180 million people per year (Singleton, 2003). In outbreaks years, domestic damage to Australian wheat reduced the country's total agricultural production by 3 to 4% (Singleton, 1997). It is estimated that in the US alone, introduced rats cost the economy more than US\$27 billion per year (Pimentel, 2007). In addition to agricultural impacts, a number of mice and rats serve as reservoirs of disease of humans and livestock. For example, in North America the white-footed mouse (*Peromyscus leucopus*) is a major reservoir for Lyme disease (Snow, 2019). In many countries and oceanic islands, rodents are also a major cause of reduction in bird populations due to their predation of native bird eggs (Howal et al., 2007; Towns, Atkinson, & Daugherty, 2006). In response, mitigation and eradication strategies are required.

Current efforts to eradicate rats and mice include the use of traps, poisons, and biological controls, such as the introduction of predators or diseases (Figure 1.1). Application of rodenticides can be expensive due to costs associated with regulation compliance, dispersal methods, size of the treated area, and the toxicant itself (Meerburg, Brom, & Kijlstra, 2008). Mechanical traps are often considered more humane than rodenticides because they do not involve the use of chemicals that could adversely affect non-target animals, humans or overall ecosystem health (Lorvelec & Pascal, 2005; Witmer & Fuller, 2011). However, placing traps and collecting the caught animals is labour intensive, traps do not discriminate between target and non-target organisms, and traps are insufficient to fully eradicate a rodent population without the use of other methods (Lorvelec & Pascal, 2005). Scientists may now be able to utilise the growing body of knowledge on selfish genetic elements and recent advances in genome engineering technologies to develop species-specific, humane and cost-effective genetic pest management tools.

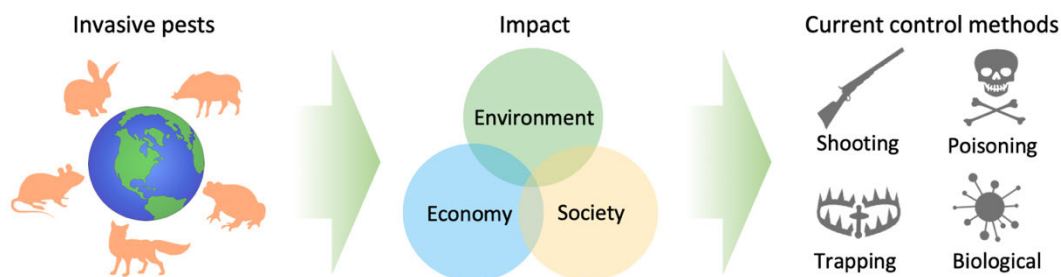


Figure 1.1 Invasive vertebrate pests are a global concern that impacts the environment, economy and society. As a result, control measures are deployed, which are considered to be inhumane, costly, and often inadequate to deal with the scope of the problem. Image from McFarlane, Whitelaw, and Lillico (2018).

1.2 Selfish genetic elements

Selfish genetic elements are described as an exception to the conventional rules of inheritance. First described in 1866 by a monk named Gregor Mendel, the conventional rules of inheritance, also known as Mendelian inheritance, dictate offspring have a 50% chance of inheriting a gene from one of their parents. With Mendelian inheritance, not all offspring will inherit the gene, and so the frequency of that gene in future generations will be similar to the frequency of that gene in the parents' generation. With selfish genetic elements, offspring have more than a 50% chance of inheriting a genetic element from a

parent, and so a specific genotype will increase in the population over time. Such an element is said to ‘drive’ within a population (Burt & Trivers, 2006; Esvelt et al., 2014).

Scientists have known about selfish genetic elements that violate Mendel’s rules since the late 1880s. At a molecular level, they are sequences of DNA, such as genes or their fragments, all or parts of chromosomes, or noncoding DNA, for which inheritance is biased in their favour (Burt & Trivers, 2006). These elements achieve drive through one or more of three key mechanisms (Figure 1.2): 1. Interference: with this strategy a gene gains an advantage by disrupting the transmission of an alternative gene. An example of a selfish element that operates by interference is the t-haplotype in mice (see section 1.2.2). 2. Over-replication: these elements bias their transmission to the next generation by being replicated more than other genes in an animal. A good example is transposable elements. 3. Gonotaxis: where genetic elements bias movement towards the germ line. Some B-chromosomes, which are accessory chromosomes that are not essential for the life of a species, act selfishly by replicating themselves and moving to the germline and avoiding the somatic cells. (Burt and Trivers, 2006).

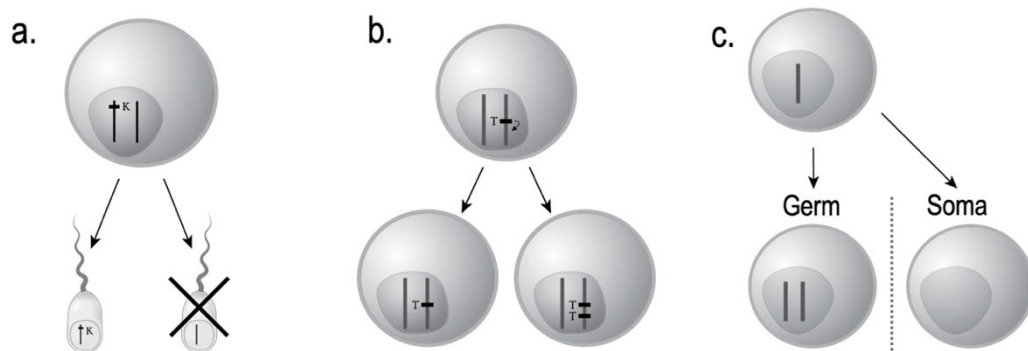


Figure 1.2 Examples of the three known primary mechanisms by which selfish genetic elements achieve ‘drive’. Interference. A killer gene (K) is able to kill non-K bearing sperm. This would benefit the K gene’s chance of fertilising an egg. b. Over-replication. A transposable element (T) makes a second copy of itself and one daughter cell inherits one copy while the other inherits two copies. c. Gonotaxis. A B-chromosome replicates itself and both copies move to the germline and avoid the somatic cells. Image adapted from Burt and Trivers (2006).

One important particularity of selfish genetic elements is that they do not need to make any contribution to the reproductive success of the host organism in order to drive successfully (Burt & Trivers, 2006). In addition to the examples highlighted in Figure 1.2,

homing endonuclease genes and meiotic drivers represent two important examples of natural selfish genetic elements, having evolved several times in different taxa. Due to the significance of meiotic drives and homing endonuclease genes in the development of engineered selfish gene elements, I have briefly described these elements below.

1.2.1 Homing endonuclease genes

Homing endonuclease genes (HEGs) achieve drive through an over-replication mechanism. These genes are situated on a chromosome within a specific DNA sequence that they recognise and cut. They encode an endonuclease that cuts the recognition sequence on the chromosome that is homologous to the one originally containing the HEG. After the sequence is cut, homologous recombination (HR) is used to copy the HEG into the cut in the homologous sequence. When this process occurs in the germline, the proportion of gametes that contain the HEG is greater than 50% and therefore the HEG can drive itself through the population (Fraser, 2012). HEGs are present in eukaryotic organisms, archaea, and bacteria, where their recognition sequences are found at low frequencies in the genome (Jasin, 1996). Well characterised examples of HEGs include, the fungal homing endonuclease I-PpoI from *Physarum polycephaleum* and I-SceI from *Saccharomyces cerevisiae*, which target 15 bp and 18 bp DNA motifs, respectively (Belfort & Bonocora, 2014; Windbichler et al., 2007).

1.2.2 Meiotic drive

Meiotic drive is a method by which a selfish genetic element can achieve drive through an interference mechanism (McDermott & Noor, 2010). In vertebrates, the most studied natural meiotic drive is the t-haplotype in the house mouse (*Mus musculus*; Ardlie, 1998; Silver, 1993). The t-haplotype consists of a series of linked, independent T-complex distorter genes and a T-complex responder gene that are inherited together. When present in the heterozygous (Tt) condition in the male, the wild-type sperm show motility defects and are functionally inactive, so more than 90% of the progeny receive the t-haplotype. This requires the combined action of the distorter genes, which attacks gametes, and a responder gene, which protects gametes carrying the t-haplotype. This action leads to morphological defects in spermatozoa that do not carry a t-haplotype due to excessive activation of the chromosome 17 gene *SMOK* (Bauer et al., 2012; Bauer et al., 2007; Burt & Trivers, 2006; Herrmann & Bauer, 2012).

1.3 Harnessing evolution

Although selfish genetic elements have been recognised by the scientific community since the late 19th century, the idea to using these elements as a means to control natural populations did not surface until the mid-20th century (Burt & Trivers, 2006). In 1960, Craig, Hickey & Vandehey suggested using a breeding program in which a “male-producing factor” that is naturally present in some male mosquitoes could be harnessed to control mosquito populations. Hickey and Craig (1966) went on to identify the genomic region responsible for this phenomenon. Their logic behind this control strategy was that when male mosquitoes with this male producing factor breed, most of their offspring then develop as males. Environmental releases of mosquitoes carrying this male-producing factor could potentially “reduce the number of females below the level required for efficient disease transmission” (Craig, Hickey, & Vandehey, 1960).

In the 1960s, Craig, Hickey and the other early pioneers did not yet have the molecular tools to engineer animals harbouring desirable genes. More than 30 years of basic biological research in genetics and molecular biology took place before potential genetic engineering tools became available. In 2003, Austin Burt proposed using HEGs to drive modified genes through a population (Burt, 2003). It was Burt’s seminal work describing his idea, in combination with advanced knowledge about genetics and modern molecular tools that bolstered the field of inquiry into synthetic ‘gene drives’.

1.4 What are gene drives?

There is a growing acceptance that the term ‘gene drives’ simply refers to engineered selfish genetic elements. As noted above, since the 1960’s researchers have imagined that selfish genetic elements “might serve as the basis for ‘gene drives’ capable of spreading engineered traits through wild populations” (Esvelt et al., 2014). However, the terms ‘gene drive’ and ‘selfish genetic element’ are still used interchangeably. In a recent report by the National Academies of Sciences (2016), they defined gene drives as systems of biased inheritance in which the ability of a genetic element to pass from a parent to its offspring through sexual reproduction is enhanced. Therefore, the result of a gene drive is the preferential increase of a specific genotype from one generation to the next, and potentially spread throughout the population. Figure 1.3 illustrates an idealised difference between Mendelian inheritance and inheritance through a gene drive.

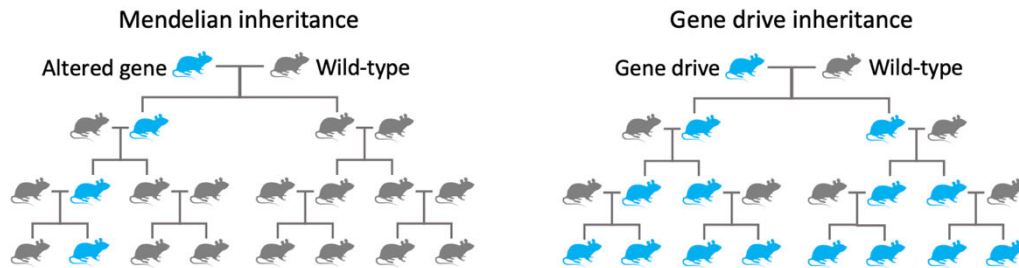


Figure 1.3 Mendelian and gene drive inheritance. Left: Mendelian inheritance of an altered gene. Right: Inheritance of a gene drive. Image from McFarlane et al. (2018).

Although this thesis concentrates on gene drives for controlling populations of invasive vertebrate pests, these elements have been proposed for numerous applications; perhaps the most prominent of these is to reduce populations of malaria-transmitting mosquitoes (Adelman & Tu, 2016; Galizi et al., 2014; Hammond et al., 2016). In this context, gene drives could provide a new means of tackling a disease that still infects 200 million people and causes 400,000 deaths each year (World Health Organization, 2019). Other potential applications outside pest control include genetically immunising populations of animals against disease and improving the sustainability of agriculture by reducing the need for and toxicity of pesticides and herbicides (Esvelt et al., 2014). Synthetic gene drives could have remarkable benefits for the environment, society and the economy.

1.5 Classical genetic engineering

In order to develop synthetic gene drives scientist required precision genetic engineering technologies. Precision genetic engineering focuses on generating site-specific modifications to the genome. Initial successes in the mid 1980's with precision genetic engineering were accomplished by microinjecting DNA sequences with high homology to the targeted genomic sequence into cells (Smithies et al., 1985). These experiments utilised a homologous recombination (HR) dependent approach, which took advantage of the cell's own DNA repair machinery to replace a targeted genetic locus with an introduced DNA sequence. However, the targeted engineering events occurred at very low rates (around 1 in 3×10^4 cells, but this varied between cell types; Capecchi, 1989)

1.6 Nuclease-assisted genetic engineering

A key breakthrough in the field of genetic engineering was the realisation that double-stranded DNA breaks (DSBs) greatly stimulated cellular DNA repair mechanisms (Rouet,

Smih, & Jasin, 1994; Smih et al., 1995). All eukaryotic cells efficiently repair DSBs using one of three primary pathways: non-homologous end-joining (NHEJ), microhomology-mediated end-joining (MMEJ) or a via a form of HR, known as homology directed repair (HDR). NHEJ simply joins the broken ends of the DNA, often creating small insertion or deletion mutations (indels); MMEJ can occur when short microhomologies exist, both upstream and downstream of the DSB, the two microhomologies can be annealed, often resulting in deletions of the intervening sequence; HDR uses a homologous DNA template to replace the broken region with high fidelity (Liu, Rehman, et al., 2019; Sakuma et al., 2016; Yanik et al., 2018). Thus, the induction of directed DSBs at a genomic locus of interest can greatly stimulate HDR-based genetic engineering efficiency (Figure 1.4). As targeted DSBs in the genome could be introduced by nucleases, this method has become known as nuclease-assisted genetic engineering. It is the same premise that HEGs have naturally evolved to use and the logic behind Austin Burt's idea of applying HEGs for developing synthetic gene drives.

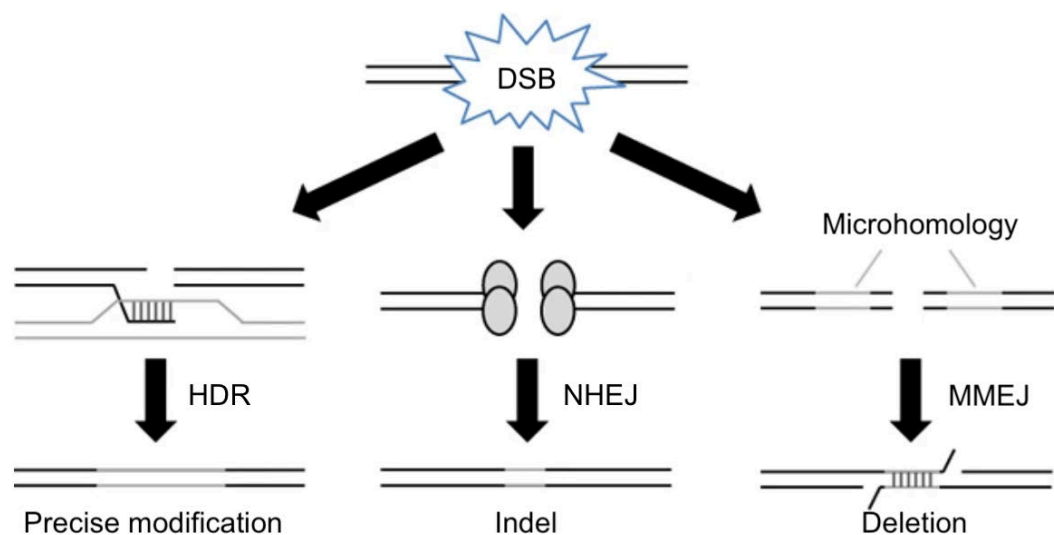


Figure 1.4 Nuclease-assisted gene targeting. A double strand DNA breaks (DSB) from a site-specific nuclease could lead to precise modification through homology directed repair (HDR) in the presence of a DNA repair template, in the form of double stranded DNA or a single strand DNA oligonucleotide, both of which must contain homology arms. DSBs can also be repaired through non-homologous end joining (NHEJ), which frequently leads to small insertions and deletions (indels), or microhomology-mediated end-joining (MMEJ), which often results in a deletion. Image adapted from Sakuma et al. (2016).

Early nuclease-assisted genetic engineering experiments using HEGs demonstrated that DSBs increased the frequency of template integration (Rouet et al., 1994). Working with Austin Burt, Windbichler et al. (2011) did go on to describe the use of a HEG in the creation of a gene drive in mosquitoes. In this instance, a transgenic mosquito was created with a cleavage site near a fluorescence gene, and, upon expression of the HEG from a donor DNA plasmid, the site was cut, allowing for copying of the HEG into the target site through HDR. The increased transmission of the HEG demonstrated these nucleases have potential for developing gene drives. However, the practical utility of HEGs to target any genomic loci at will was limited by the long and highly-specific DNA recognition sequences (14 to 40 bp) of these enzymes. Furthermore, the engineering of HEGs was extremely challenging for the majority of researchers because the DNA recognition sites and cleavage functions of these enzymes are intertwined in a single domain (Burt & Koufopanou, 2004; Rocha-Martins et al., 2015; Sander & Joung, 2014).

Fortunately, the possibility of manipulating genomes at will captivated scientist's attention from all corners of biology and the initial challenges posed by early genetic engineering tools were soon overcome by the development of a new genre of truly programmable nucleases.

1.7 ZFNs and TALENs

Programmable nucleases have transformed the field of genetic engineering. This class of nucleases can be engineered to bind to a specific nucleotide sequence in the genome and subsequently generate a DSB at a user defined site. ZFNs and TALENs were the first classes of programmable nuclease to be developed.

Both these nucleases are engineered DNA binding proteins that facilitate targeted cutting of the genome much like a HEG (Pratt et al., 2012). ZFNs combine a nuclease domain derived from a specific restriction enzyme (typically FokI) with a DNA binding domain mediated by zinc fingers and can be used to target user-defined DNA sequences. Each zinc finger typically binds 3 nt with contact to a fourth. Specificity of each zinc finger is influenced by the neighbouring zinc fingers, which are set in an array of 3-4 for each arm of a ZFN (Figure 1.5). Each arm is targeted against a DNA strand in a tail-to-tail orientation. The ZFNs function as pairs because the enzymatic domains must form dimers in order to cleave DNA (Kim & Kim, 2014; Urnov et al., 2010).

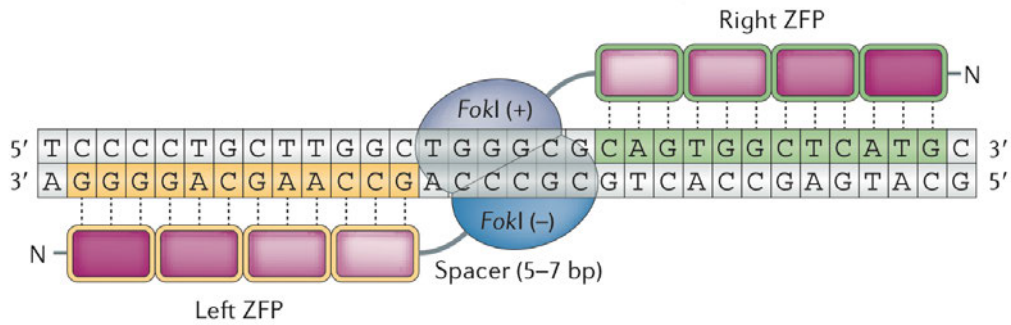


Figure 1.5 A schematic representation of zinc finger nuclease (ZFN) structure. Each ZFN is composed of a zinc-finger protein (ZFP) DNA binding domain and a FokI nuclease domain. Image adapted from Kim and Kim (2014).

Like ZFNs, Transcription Activator-Like Effector Nucleases (TALENs) typically utilize the same nuclease domain from FokI and so function as dimers. In place of the zinc finger, TALENs use an alternative DNA binding domain called a Transcription Activator-Like effector (TALE), derived from the plant pathogenic bacterium *Xanthomonas* (Boch & Bonas, 2010). Each TALE contains a series of 33-35 amino acid repeat domains where each repeat recognizes a single base. The specificity of each TALE is conferred by the repeat-variable di-residue (RVDs), which are two hypervariable amino acids that interact with the nucleotide sequence (Figure 1.6). Like ZFNs, each arm of a TALEN is targeted against a DNA strand in a tail-to-tail orientation, to dimerize the FokI nuclease in order to induce a DSB at the target sequence (Kim & Kim, 2014; Nemudryi et al., 2014).

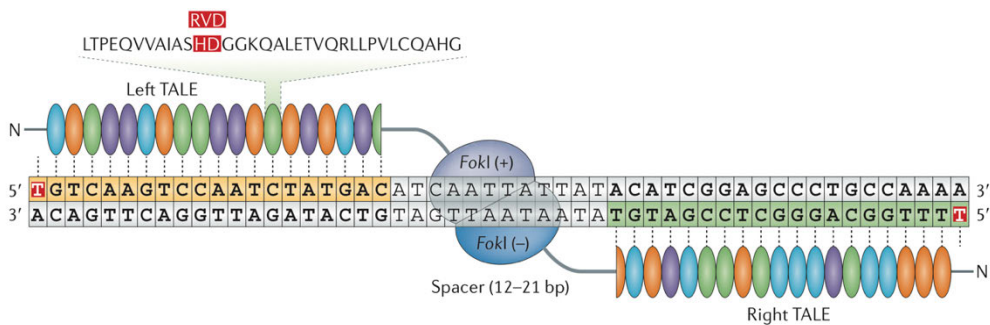


Figure 1.6 A schematic representation of transcription activator-like effector nuclease (TALEN) structure. Each TALEN is composed of DNA binding Transcription Activator-Like effector (TALE) domain and a FokI nuclease domain. Each TALE repeat is comprised of 33–35 amino acids and recognizes a single base pair through the amino acids at positions 12 and 13, which is called the repeat-variable di-residue (RVD; shown in red). Image from Kim and Kim (2014).

Although both ZFNs and TALENs are highly specific genome editors that have successfully modified the genome of many species, including, mice, rats, zebrafish, fruit flies, nematodes, rats, livestock and even monarch butterflies, their creation can be time-consuming and labour-intensive; requiring a new protein pair to be created for every DNA sequence to be edited (Gaj, Gersbach, & Barbas, 2013; Mao et al., 2013). Simoni et al. (2014) showed that ZFNs and TALENs could be used to develop gene drives much like HEGs, with homing frequencies of 34% and 49% to available target loci, respectively, in *Drosophila melanogaster*. In many instances, though, these systems are not transmitted faithfully due to the number of repetitive elements within their design and their subsequent tendency to recombine, leading to their loss of function.

Then, CRISPR arrived (Cong et al., 2013; Jinek et al., 2012; Lander, 2016).

1.8 CRISPR-Cas

The arrival of the RNA-guided CRISPR-Cas (Clustered regularly-interspaced short palindromic repeats-CRISPR associated protein) system has been game changing. The system was first observed in *Escherichia coli* in 1987 by its striking genomic structure (Ishino et al., 1987). Evolved as an adaptive immune system in bacteria and archaea, the system uses a set of Cas genes to incorporate exogenous DNA into the CRISPR locus, and subsequently transcribe them as RNA templates that guide the Cas proteins to cleave invasive DNA or RNA (Horvath & Barrangou, 2010). However, the system was not repurposed into a genome editing tool until 2012 when Jinek et al. published their seminal paper. Three major types of CRISPR-Cas systems have been identified, differing in their targets as well as mechanisms of action. Type I and III CRISPR systems employ an ensemble of Cas genes to carry out RNA processing, targeting and target site cleavage. By contrast, the type II CRISPR-Cas systems uses a single nuclease to cleave target DNA. Among the three types of CRISPR-Cas systems, the type II CRISPR-SpCas9, derived from *Streptococcus pyogenes* and CRISPR-AsCas12a (formerly AsCpf1) from *Acidaminococcus* sp. are presently the most suitable and widely used as a genome editing tools (Mojica & Montoliu, 2016; Yao et al., 2018).

1.8.1 SpCas9

The *S. pyogenes* type II CRISPR-Cas system consists of a SpCas9 nuclease and two RNAs; CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA). The crRNA and tracrRNA hybridise to form a guide RNA (gRNA) before complexing with SpCas9

(Deltcheva et al., 2011). SpCas9 is directed to the target site by 20 nt of the crRNA by Watson-Crick base-pairing with the target DNA sequence (Cong et al., 2013). However, crRNA and tracrRNA can be fused to form a 102 nt single guide RNA (Jinek et al., 2012), which is also referred to as a gRNA within this thesis. Once the gRNA guides the SpCas9 to the target site, cleavage of the target site only occurs if a protospacer adjacent motif (PAM), a defined 3 nt sequence at the 3' end of the crRNA sequence is present. For SpCas9, the PAM sequence is 5'-NGG, although it can recognise other noncanonical 5'-NAG and 5'-NGA PAMs to a lesser extent (Hsu et al., 2013). At an overall structural level, SpCas9 contains two nuclease domains, HNH and RuvC, each of which cleaves one strand of the target DNA (Figure 1.7). In SpCas9, these domains cleave 3 bp upstream from the PAM sequence, leaving blunt ends (Anzalone, Koblan, & Liu, 2020; Garneau et al., 2010; Hsu et al., 2013).

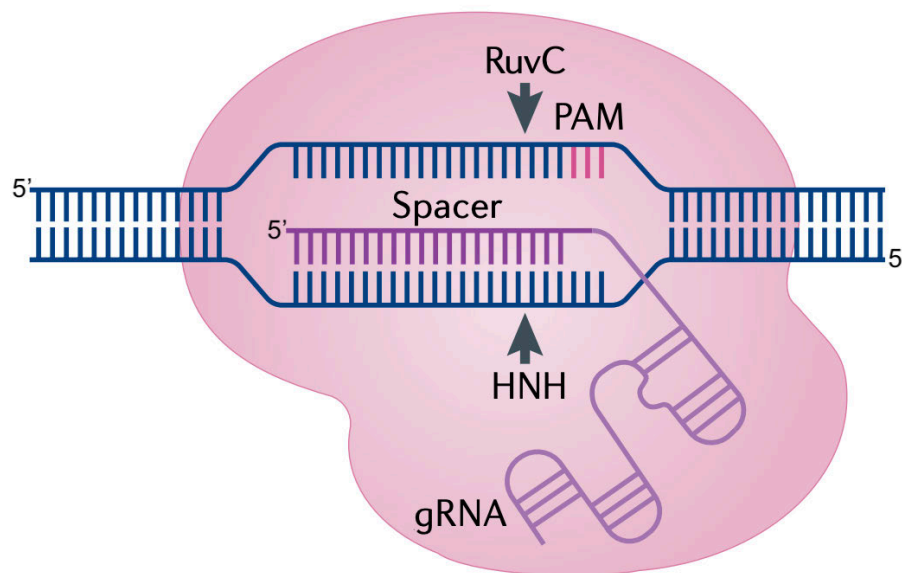


Figure 1.7 A schematic representation of *Streptococcus pyogenes* Cas9 (SpCas9). SpCas9 is programmed here with a single guide RNA (gRNA; purple). The gRNA in this case is a single gRNA comprising a site specific CRISPR RNA (crRNA) and an auxiliary trans-activating crRNA (tracrRNA). Two nuclease domains (RuvC and HNH) both cut one DNA strand 3 bp upstream from the protospacer adjacent motif (PAM; 5'-NGG) to generate a double strand DNA break. Image adapted from Pickar-Oliver and Gersbach (2019).

1.8.2 AsCas12a

Unlike SpCas9, the *Acidaminococcus* sp. Cas12a (AsCas12a) endonuclease only requires a short 42–44 nt crRNA and no tracrRNA. As AsCas12a does not require a tracrRNA, the crRNA is often referred to as the gRNA. The first 19–20 nt corresponding to the repeat sequence and the remaining 21–25 nt to the spacer sequence. AsCas12a recognises a T-rich PAM (5'-TTTN) which is situated upstream of the 5' end of the non-target strand and generates dsDNA breaks with staggered 5' ends at the 19 and 23 nt relative to the PAM sequence (Figure 1.8; Zetsche et al., 2015). It has been shown that AsCas12a processes its own pre-crRNA into mature crRNAs, without the requirement of a tracrRNA, making it a unique effector protein with both endoribonuclease and endonuclease activities (Zetsche et al., 2017). AsCas12a's ability to process crRNA makes it an attractive option for expression of multiple crRNAs from a single transcript. The DNase site within AsCas12a is located in the interface between its RuvC and Nuc domains (Anzalone et al., 2020).

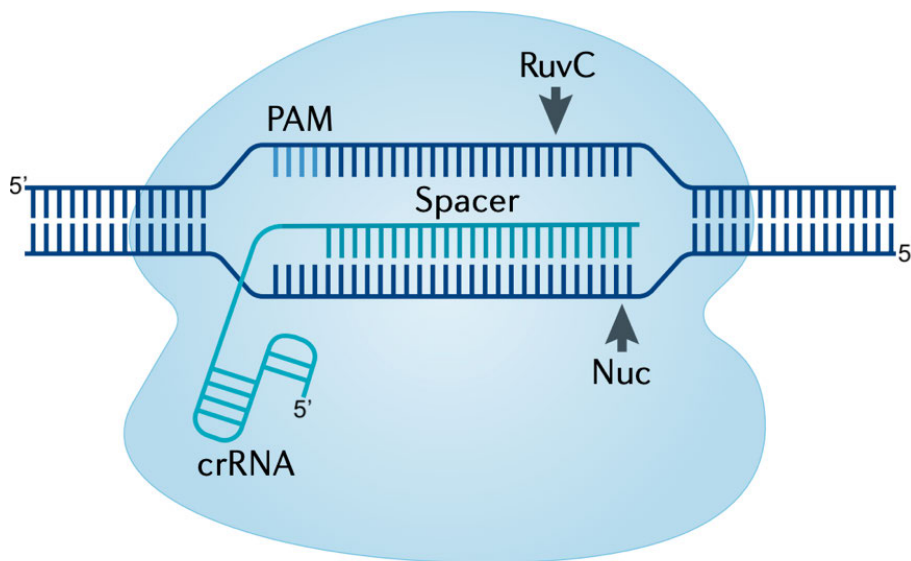


Figure 1.8 A schematic representation of *Acidaminococcus* sp. Cas12a (AsCas12a). AsCas12a nucleases recognise DNA target sequences with complementarity to the CRISPR RNA (crRNA; light blue) spacer positioned next to a 3' PAM (5'-TTTN). Target recognition results in the generation of a staggered DNA double-strand break by a RuvC domain and a putative nuclease (Nuc) domain. Image adapted from Pickar-Oliver and Gersbach (2019).

1.8.3 CRISPR uptake and innovation

In contrast to ZFNs and TALENs, which require the design of proteins that are encoded by large repetitive DNA segments (500 to 1500 bp) for each new target site, CRISPR-Cas can be easily adapted to target almost any genomic sequence by simply exchanging the crRNA sequence. The Cas protein component remains unchanged, alleviating the need for complex protein engineering. Therefore, the reagents are simple, cheap, and quick to design and generate (Gupta & Musunuru, 2014). Since being repurposed into a genome editor in 2012, there has been a flurry of research activity around the technology and an ever expanding toolbox of CRISPR-Cas reagents, including base-editors and nickase systems, which can mediate editing without the need for introducing DSBs (Anzalone et al., 2019; Komor et al., 2016). With these successive innovations, CRISPR-Cas tools have become widely adopted for a multitude of basic and applied research applications, with some translation now progressing into medicine and industry.

1.8.4 CRISPR-based gene drives

CRISPR-Cas facilitates the capability to engineer synthetic gene drives in a standard molecular laboratory. SpCas9-based gene drives have been developed in yeast (DiCarlo et al., 2013), *D. melanogaster* (Gantz & Bier, 2015), two species of mosquitoes (Gantz et al., 2015; Hammond et al., 2016) and most recently in mice (Grunwald et al., 2019). These studies have demonstrated potential mechanisms of how gene drives could be used to achieve two potential outcomes:

1. Population suppression – the spread of a genetic element that causes the number of individuals in a population to decrease.
2. Population replacement – the spread of a genetic element through a population that causes a population's genotype to change.

These outcomes could have remarkable impacts and scientists have already proposed a wide range of applications for gene drives, including to address public health threats, species conservation, agriculture protection and advancing scientific research. Of primary interest to this project are strategies that could utilise CRISPR-Cas to develop gene drives with the goal of achieving species-specific population suppression in vertebrate pests. Additional information on vertebrate population suppression can be found in Appendix A.1, which contains a perspective article I authored during my PhD, titled "CRISPR-Based Gene Drives for Pest Control".

1.9 Driving population suppression

Gene drives could enable humane population suppression by distorting the sex-ratio of a damaging pest population in a species-specific manner. By distorted the sex ratio of breeding population away from the favoured Fisherian ratio of 1:1 male to female, these strategies could manipulate the reproductive performance of a population. In most pest species, including mice and rats, female procreative capacity is responsible for maintaining the overall population size (Champer, Buchman, & Akbari, 2016; Hamilton, 1967). Therefore, an efficient means of population suppression is to bias the sex ratio of the breeding population in favour of males. This could be accomplished by either spreading female infertility through a population or by ensuring all or most offspring born are male (Galizi et al., 2016; Hammond et al., 2016). A grossly male breeding population will result in a population decline, while an all-male breeding population will lead to eradication.

One possibility that has been suggested and is currently under investigation is to engineer the t-haplotype (section 1.2.2) to carry the sex-determining *Sry* gene in genetic females so they develop as males but are sterile because they lack essential Y-linked genes for sperm development (Backus & Gross, 2016; Koopman et al., 1991; Manser et al., 2019). However, this approach is unlikely to gain traction as two laboratory studies on independent mouse populations have found that male mice carrying the t-haplotype only fertilise approximately 20% of offspring when competing against wild-type males (Manser et al., 2017; Sutter & Lindholm, 2015). This can be explained by the relative sperm competitiveness of t-haplotype males compared to wild-type males and is likely why the frequency of the t-haplotype in natural populations is relatively low, typically varying between 5 and 30% (Ardlie, 1998; Manser et al., 2011).

Instead, within this project I choose to examine the potential of two CRISPR-based sex-ratio distorting gene drives – (1) homing-based gene drives and (2) X-shredder – both of which have the potential to distort the sex-ratio of a breeding population in favour of males. To date, these sex-ratio distorting gene drive systems have only been engineered in proof-of-concept studies in mosquitoes with the focus on controlling vector-borne diseases (Galizi et al., 2016; Hammond et al., 2016). However, homing-based gene drives with the alternative aim of population replacement, rather than suppression have been developed in other species; including drosophila, yeast and most recently in mice (DiCarlo et al., 2013; Gantz & Bier, 2015; Grunwald et al., 2019).

1.9.1 Homing-based gene drives

CRISPR-based homing gene drives were theorised and named after Austin Burt's homing endonuclease gene drive approach, which drive through an over-replication mechanism. A homing gene drive works by copying or 'homing' itself into a target site in the genome. To build a CRISPR-based homing gene drive, an animal is engineered with a gene drive cassette that expresses a Cas endonuclease, such as SpCas9 or AsCas12a, and one or more gRNAs. The Cas nuclease and gRNA/s are expressed from one allele to cut at a conserved target site on the sister allele on the homologous chromosome (Figure 1.9). After CRISPR-mediated cleavage, HDR results in the CRISPR machinery and any additional genetic payload (trait of interest) included in the gene drive cassette being copied onto the homologous chromosome. This process ensures homozygosity for the gene drive cassette. When this process occurs in the zygote or germ cells, the proportion of mature gametes in an animal that contains the gene drive cassette is greater than 50%, and therefore the cassette could drive itself through the population, spreading the trait of interest (Burt, 2003; Burt & Trivers, 2006; Champer et al., 2016; McFarlane et al., 2018).

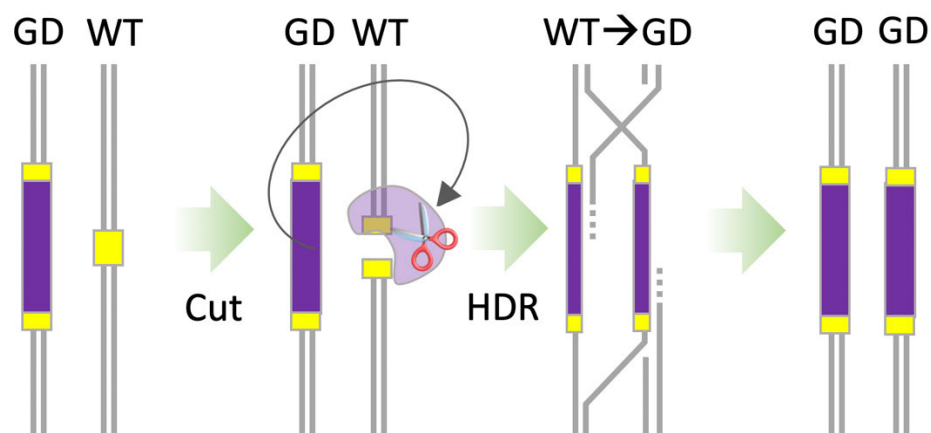


Figure 1.9 CRISPR-based homing gene drive (GD). The GD cassette (purple) expresses Cas and one or more guide RNAs that cut the sister allele on the homologous wild-type (WT) chromosome. The cell then repairs the cut by homology-directed repair (HDR), using the GD chromosome as the repair template. This process copies the GD cassette and any genetic payload (trait) onto the WT chromosome and ensures that most of mature gametes in the animal carry the GD cassette, passing on to most or all of the offspring. With this strategy, the guide RNAs must not cut the GD cassette.

Grunwald et al. (2019) successfully demonstrated a homing-based gene drive in mice, capable of driving a mCherry fluorescent genetic element embedded in the mouse tyrosinase (*Tyr*) gene, through the female germline. *Tyr* affects mouse coat colour, which enabled the frequency of the gene drive to be tracked over generations by monitoring coat colour and confirmed by PCR-based genotyping. Their initial design ubiquitously expressed the SpCas9 component of the gene drive using a synthetic CAG promoter along with a ubiquitously expressed gRNA targeting the *Tyr* gene. Homologous repair was not observed in 180 pups born across 10 families with ubiquitous SpCas9 expression.

Their successful gene drive design expressed SpCas9 with germline specificity, although the mechanism of restricting SpCas9 to the germline was crude. By crossing a Cre-inducible SpCas9 expressing mouse (CAG-LSL-SpCas9) with a germline VASA-Cre expressing mouse they were able to restrict SpCas9 expression to the germline. With this design, they achieved up to 72% homing in the female germline, but only NHEJ was observed in males harbouring the same genetic construct. The reason for the sex-specific differences in homing efficiency requires further investigation but it is likely due to timing or pattern of SpCas9 expression, which may be a result of the Cre-inducible system or a phenomenon specific to the *Tyr* locus. Importantly, Grunwald and colleagues (2019) demonstrated that CRISPR-based homing gene drive could be applied in mammals.

For population suppression, a homing gene drive can be targeted to a haplosufficient female-fertility gene (HFFG) to disrupt the coding sequence of the gene, rendering homozygous female offspring infertile, while males and heterozygous females will retain normal fertility. By targeting a homing gene drive to insert within a gene essential for female-fertility this disrupts the gene's coding sequence and ensures that most female offspring are infertile; whilst males will retain normal fertility and continue transmitting the drive. With every generation, the sex-ratio of the breeding population will become more biased towards males, eventually resulting in a population crash (Hammond et al., 2016; Prowse et al., 2017). Hammond et al. (2016) developed this system in mosquito and achieved transmission rates of 91.4 to 99.6% in caged populations.

In theory, homing gene drives targeting female reproduction could be adapted to control most vertebrate pests, including mice and rats; however, it remains unknown if such high transmission rates can also be achieved in vertebrates. Further investigation on the impact of gene drive construct size on homing efficiency is also needed. Simoni et al. (2014) observed no correlation between gene drive construct size and homing activity when testing constructs varying in size by 1.5 to 2 kb in *Drosophila*. Although as construct size in standard gene targeting experiments typically reduces HDR efficiency, it is

conceivable that future work will uncover a construct size threshold that lowers the homing activity of homing-based gene drives.

1.9.2 X-Shredder

An alternative to homing-based gene drive is an X-shredder (XS), which is a synthetic meiotic drive. In XY heterogametic species, an XS is a type of sex-ratio distorting gene drive that cuts the X-chromosome at multiple sites during spermatogenesis, thus shredding the X-chromosome beyond repair (Champer et al., 2016). To engineer a CRISPR-based XS, an XS cassette is inserted within a neutral intergenic region of the Y-chromosome. The cassette encodes a Cas nuclease, which is expressed under the control of a spermatogenesis-specific promoter, and one or more gRNAs that target conserved repetitive sequences unique to the X-chromosome (Figure 1.10). Given that X-chromosomes will be destroyed during spermatogenesis, most sperm that mature and reach the oocyte are Y-bearing, resulting in a biased sex ratio in favour of males. By placing the XS cassette on the Y-chromosome, all male offspring will inherit the cassette and continue transmitting the XS to subsequent generations (McFarlane et al., 2018).

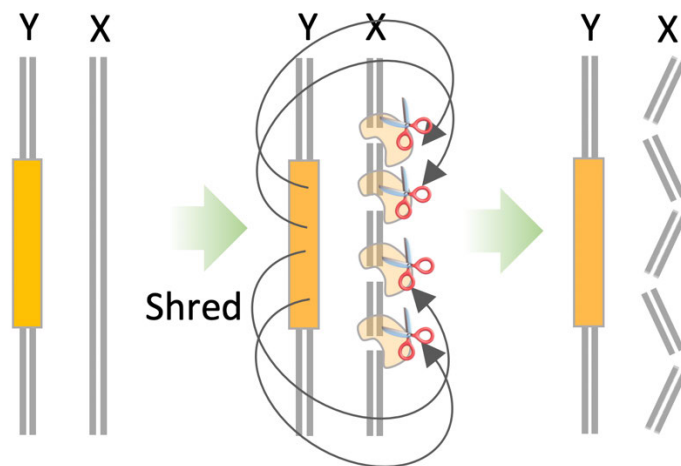


Figure 1.10 X-Shredder (XS). During spermatogenesis, Cas and guide RNA(s) are expressed from the XS cassette (orange) located on the Y-chromosome (Y) and shred the X-chromosome (X) beyond repair. Most sperm that mature and reach the oocyte are Y-bearing, resulting in most offspring being XS males.

A CRISPR-based XS has been engineered in mosquitoes, although the system expressed the XS cassette from an autosome instead of the Y-chromosome. With this

approach, Galizi et al. (2016) achieved male bias among progeny ranging from 86.1 to 94.8% in laboratory-contained mosquito populations. Although successful in mosquito, technical challenges facing the adaption of an XS into vertebrates include identifying appropriate spermatogenesis-specific promoters in target species and the transcriptional silencing of mammalian sex chromosomes during meiosis. The latter may hinder expression of Cas from the Y-chromosome, as well as the accessibility of the endonuclease to shred the X-chromosome (McFarlane et al., 2018).

1.10 Containing the spread

The two forms of sex-ratio distorting gene drives described above may only require the release of a small number of animals to spread through an entire population. The duration and extent of spread would be limited by naturally arising resistant alleles that prevent CRISPR-mediated cleavage. Resistant alleles could exist in the population before release or originate from indels generated when CRISPR-mediated cleavage is repaired by the error-prone NHEJ pathway and alters the gRNA recognition sequence (Bull, 2015; Champer et al., 2016). The rate of NHEJ-mediated repair will be dependent on the species, target site, and the stage of development at which DNA cleavage occurs. Given that natural selection tends to favour equal sex ratios, resistant alleles that restore function would spread rapidly through the population (Hammond et al., 2017; McFarlane et al., 2018).

Compared to a homing gene drive, XS should be less prone to inactivation by resistant alleles because it targets multiple sites and, therefore, would require an animal to simultaneously acquire multiple resistant alleles to incapacitate the drive (McFarlane et al., 2018). Following a similar approach, it has been suggested that the evolutionary stability of homing gene drives could be improved by using multiple gRNAs closely spaced along the target region (Esvelt et al., 2014). Several *in silico* and *in vivo* studies have now demonstrated that multiple gRNAs can assist homing gene drives to evade drive resistance and successfully spread through pest populations (Champer et al., 2018; Champer, Yang, et al., 2019; Champer et al., 2020; Prowse et al., 2017). Even if drive resistance were to prevail, it would be possible to release a second gene drive targeting a different sequence to continue suppressing the population.

Conversely, if a gene drive were not limited by naturally arising resistance, it would have the potential to spread indefinitely through a species. Therefore, it is essential to have strategies in place that could deliberately inactivate a gene drive that escaped containment or was causing unforeseen impacts. Fortunately, both homing gene drive

and XS systems can be inactivated by the release of animals bearing engineered functionally resistant alleles or a reversal gene drive that immunises the animal against the original drive (McFarlane et al., 2018). However, these reactive stop buttons are not ideal as timely deployment requires that appropriate animals are on hand, in sufficient numbers and at the required location.

More proactive approaches are needed, and scientists are working to develop gene drives that are inherently self-limiting – that is, they stay in a population transiently and could therefore be localised to targeted populations. Several self-limiting gene drive concepts have been proposed and it is hoped these systems will soon be realised (Min et al., 2017a, 2017b; Noble et al., 2016; Oberhofer, Ivy, & Hay, 2019). Of these, daisy drives have been the focus of several *in silico* population modelling studies and well publicised; daisy drives include: ‘daisy-chain’ and daisy-field’ drives. These systems are complex and have yet to be engineered in species, however, there is value in describing them here as future development would put the power in the hands of communities to deploy a gene drive locally, avoiding the need for highly complex geopolitical agreements that will be required for self-perpetuating gene drive systems.

1.10.1 Daisy-chain system

In daisy-chain drive systems, the CRISPR components are split up and scattered throughout the genome so that none of them can be copied on its own. Though physically separated, they are functionally arranged in a daisy-chain: element C causes element B to drive, and element B causes element A to drive. Element C doesn't drive, so its abundance is limited by the number of daisy drive organisms released. That means B will initially increase in abundance, then decline and vanish. In turn, A will increase even more rapidly, but eventually will run out of B and disappear (Figure 1.11a; Esvelt, 2019; Noble et al. 2016).

In other words, the elements of a daisy-chain system are similar to booster stages of a genetic rocket: those at the base of the daisy-chain help lift the payload until they run out of fuel and are successively lost (Figure 1.11b). Adding more links to the daisy chain will spread the payload to more organisms. Releasing a daisy-chain drive organism with a five-element chain (E→D→C→B→A) is hundreds of times more effective than releasing one with only element A (Esvelt, 2019; Noble et al., 2016). In theory, daisy-chain systems can do anything a standard homing gene-drive system can achieve, although the complexity of engineering such a system and having it stably inherited with little fitness

cost poses a significant hurdle. Daisy-chain drives were theorised in 2016; at present, a daisy chain of two or more homing elements has not been reported.

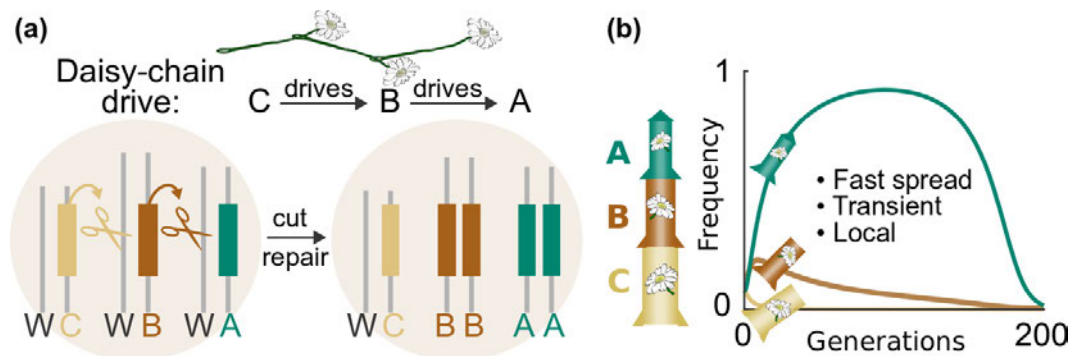


Figure 1.11 Daisy chains drive system. In a daisy chain drive system, the CRISPR components are separated and arranged so that each daisy element drives the next in the chain. The element at the end, in this case C, is not copied and is lost in half of offspring. In these organisms, B is no longer copied and is lost in turn; this process continues until the drive system stops. (b) The loss of nondriving elements to natural selection is analogous to gravity on a rocket. Adding more elements to the daisy chain allows the system to spread further before it runs out of genetic fuel and halts. Image from Esvelt and Gemmell (2017).

1.10.2 Daisy-field system

Another option is to use a form of daisy drive that doesn't require each of the booster elements to home via HDR. This significantly reduces the engineering complexity. Instead of a daisy-chain of linked elements (C→B→A), a daisy-field system scatters B elements (gRNA expression cassettes), throughout neutral loci in the genome (Figure 1.12a). As long as at least one of these is present, element A (Cas nuclease) will be copied, propagating the trait of interest. Depending on the location of the Cas nuclease, this could knock-out a gene of interest by disrupting the coding sequence or propagate a genetic cargo that contains a trait of interest. With every generation of mating with wild-type it halves the number of daisy B elements (gRNAs) eventually burning out as the generational clock winds down (Figure 1.12b; Esvelt, 2019; Min et al. 2017b).

Compared to a daisy-chain drives, a daisy-field system should be much simpler to engineer. For example, the element A (Cas nuclease) could be targeted to the site of interest using CRISPR-assisted genome engineering and the B elements (gRNA expression cassettes) scattered through the genome using lentiviral integration to

generate a daisy field animal. Daisy field drives consisting of one B element have been successfully constructed as a 'split drive' confinement strategies (section 3.3.4.2; DiCarlo et al. 2015), and the engineering technicalities of increasing the number B elements should not be inhibitory.

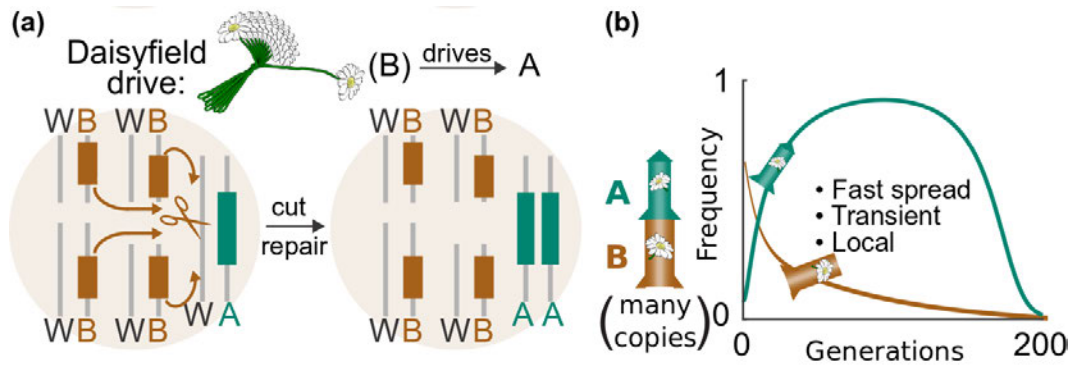


Figure 1.12 Daisy-field system. Daisy-field drive systems employ multiple daisy elements encoding the same guide RNAs. (a) A simple four-element daisy-field has four elements that target the wild-type locus harbouring the payload and CRISPR nuclease. Cutting and subsequent homology-directed repair (HDR) copies the payload and nuclease. (b) daisy-field drives are tantamount to using multiple parallel boosters, half of which run out of fuel and are lost in each generation of mating to wild-type organisms. Image from Esvelt (2017).

1.11 Mice as a model organism

Mice are the ideal species for the development and testing of population suppression gene drive technology in vertebrates. In addition to being well known invasive pests, at the population, genetic and ecological level we have more information on mice than for most other potential targets of genetic pest management (Berry & Scriven, 2005; Pocock, Hauffe, & Searle, 2005). Importantly, we also have highly developed tools for manipulating the genomes of mice. Scientists can either manipulate the zygote by directly introducing gene editing reagents or use mouse embryonic stem cell (ES) cell technology (Gurumurthy & Lloyd, 2019).

Direct zygote manipulation is highly efficient for generating simple knock-out and small knock-in alleles, however, the generation of larger more complexed knock-in alleles (>2 kb) requires HDR with a donor DNA plasmid, which is difficult to produce by direct zygote manipulation (Skarnes, 2015). In a study by Gu, Posfai, and Rossant (2018), they

reported that only 1 to 6.5% of zygotes co-injected with CRISPR-SpCas9 and plasmid repair template carried the desired modification. As such, the 'gold standard' for generating large-scale knock-in mouse models remains the well-established ES cell technology. With this approach, CRISPR-Cas and donor plasmid DNA are transfected into ES cells. Donor plasmids typically consist of homology arms (each >500 bp), the cassette of interest and selection markers, such as genes encoding antibiotic resistance or fluorescent proteins to facilitate identification of successfully transfected cells. Following selection, researchers can screen and expand correctly modified clones for injection into wild-type blastocysts to generate chimeric offspring for germline transmission (Gurumurthy & Lloyd, 2019; Singh, Schimenti, & Bolcun-Filas, 2015). Although more laborious and costly than direct zygote injection, the proven ability of ES cell technology to handle complex knock-in alleles, such as gene drive elements, is yet to be rivalled by direct zygote manipulation (Gurumurthy & Lloyd, 2019; Skarnes, 2015).

ES cell technology makes the testing of gene drive elements much easier than for other species. However, it should be noted that for the vast majority of potential vertebrate applications of gene drive, ES cell technology does not exist as embryonic stem cells have not been isolated from the species. Therefore, engineering of gene drives in these target species will likely require genome engineering via a zygotic route.

1.12 Public perception

Genetically engineered animals normally come with few ecological risks. Most engineered traits are for human benefit and will not be favoured by natural selection. By contrast, gene drives can spread through populations even if they reduce the fitness of each carrier animal. This gives gene drives more scope to escape the target population and unintentionally affect extraneous ecosystems if adequate control and containment measures are not in place (Esvelt et al., 2014; McFarlane et al., 2018). Although technical and regulatory hurdles exist for the practical use of gene drive-based control of invasive species, perhaps the greatest hurdle to be overcome will be public acceptance of the technology.

The prospects for the development and deployment of gene drive will likely hinge on public perception of whether the use of such new technologies is sufficiently warranted to solve the problems being addressed. A recent Pew Research Center study by Funk and Hefferon (2018) indicates public attitudes toward the use of genetic engineering on animals, at least in the US, tend to be supportive if the technology is being applied to a major human health issue (e.g. preventing disease transmitted by mosquitoes). The US

public was less supportive of other uses involving the environment (e.g. recovering extinct species as a means of restoring biodiversity).

Whether the general public would consider the eradication of invasive species a problem that warrants the use of genetic engineering is yet to be determined. A landscape analysis by Delborne et al. (2019) on the use of gene drives in mice showed that the research community was concerned that gene drives would only receive public support if they could effectively eradicate the species, with a general concern that if implementation of gene drive were only partially successful, the public support for the strategy and even research on gene drive would be greatly diminished.

This public perception is complicated by the fact that the science involved in invasive species eradication is often complex and is not currently well understood by or communicated to the general public. In a survey of the public on gene drive in agriculture, more than 85% of respondents had never heard of gene drive technology prior to receiving the survey (Jones et al., 2019).

Although there are risks attached to gene drive technology, the potential benefits of gene drives are equally as impactful as the risks. Gene drives could revolutionize public health, agriculture, and as discussed here, could be applied for pest control, restoring ecosystems and securing livelihoods. Public attitudes will likely be determined, not by scientific debate regarding the risks of gene drives and mitigation strategies used to manage risk, but rather on the outcomes of initial trials and how these results are communicated.

1.13 Project aims

Although the relative simplicity of the X-shredder design is an attractive prospect, there is substantial groundwork needed to adapt this system to mammals, with the most significant of these hurdles being the identification of a promoter that restricts Cas-nuclease expression to spermatogenesis. In light of this hurdle, the primary object of my PhD studentship has been to engineer a CRISPR-based homing gene drive targeting female reproduction that could be tested in a laboratory-contained population of mice.

The development of this CRISPR-based homing gene drive system was separated into three sections: (1) modelling, safeguarding and strategy design; (2) reagent development and validation; and (3) genome engineering in mouse embryonic stem cells and genotype confirmation. These sections form natural demarcation for the chapters of my thesis. Chapter 3 focuses on identifying an appropriate gene drive strategy through *in silico*

modelling, target gene identification and integration of safeguard strategies. In addition, a system for *in vitro* testing gene drives is described. Chapter 4 describes the substantial reagent design, development and validation required for this project. Chapter 5 focuses on engineering of the gene drive system in mouse embryonic stem cells and the genotyping pipeline for the confirmation of engineered cell lines.

This body of work describes all points of the gene drive development process up until the injection of genome engineered mouse embryonic stem cells into recipient blastocysts to generate gene drive transgenic mice. It is hoped that the insights gained from this work will contribute to the development of humane, safe, cost-effective and species-specific gene-drives for managing invasive populations of damaging vertebrate pests.

2 Materials and methods

All chemicals were purchased from Sigma-Aldrich unless otherwise stated.

2.1 General molecular biology

2.1.1 Genomic DNA extraction

Mouse ear clippings or cultured cells (other those cells in 96 well plate format) were collected in a 1.5 ml microfuge tube and genomic DNA (gDNA) extracted using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's guidelines.

To extract gDNA from cells cultured in 96 well plate (WP) format, cells were lysed with 100 μ l lysis buffer (section 2.7.4) per well and incubated at 37 °C for 16 hours. Lithium chloride and isopropanol DNA purification was then performed on the lysate. To do this, 10 μ l of 8 M lithium chloride and 100 μ l isopropanol were added to each well and placed on a rocking platform for 20 minutes. The plate was then centrifuged at 16,000 x g for 20 minutes and the supernatant discarded. Each well was washed twice with 150 μ l per well of 70% ethanol and the resulting DNA pellet was resuspended in 40 μ l of 0.1% Tris-EDTA (TE) buffer per well.

2.1.2 Polymerase chain reaction (PCR)

Unless otherwise stated, PCRs were performed using 100 ng of gDNA extracted as described in section 2.1.1 as template for each 50 μ l of PCR master mix. A list of primer pairs used, and their annealing temperatures are in Appendix A.2 and thermocycling conditions used for each polymerase listed below can be found the Appendix A.3.

2.1.2.1 Phusion High-Fidelity Polymerase

Phusion polymerase by NEB was used for all PCRs for T7 endonuclease I assays (T7EI) and Inference of CRISPR Edits (ICE) analyses. 25 μ l of 2X NEB Phusion PCR Master Mix with HF Buffer, 1 μ l of 10 μ M forward primer, 1 μ l of 10 μ M reverse primer, 100 ng of

gDNA, 10 µl of Fi-RED (section 2.7.4) were made up to a final volume of 50 µl with nuclease-free water.

2.1.2.2 Q5 High-Fidelity Polymerase

Q5 polymerase by NEB was used for all PCR amplified DNA fragments for molecular cloning. NEB reports an ultra-low error rate and very high-fidelity for this polymerase. 25 µl of 2X NEB Q5 PCR Master Mix, 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, 100ng of gDNA, 10 µl of 5X Fi-RED (section 2.7.4) were made up to a final volume of 50 µl with nuclease-free water.

2.1.2.3 OneTaq Polymerase

OneTaq polymerase by NEB was used for routine screening for targeted genomic integration of DNA constructs in genome engineered cell lines. 25 µl of 2X NEB OneTaq PCR Master Mix with GC Buffer, 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, 100 ng of gDNA, 10 µl of 5X Fi-RED (section 2.7.4) were made up to a final volume of 50 µl with nuclease-free water.

2.1.3 Agarose gel electrophoresis

Agarose gels were made by dissolving UltraPure agarose (Life Technologies) in 1X Tris-acetate-EDTA (TAE) buffer (section 2.7.4). PCR products <1 kb were run in 2% agarose gel, while products >1 kb were run in 1% agarose gel. For example, to make a 1% agarose gel, 1 g of agarose was added to 100 ml of 1X TAE and the mixture was heated in a microwave to completely dissolve the agarose solution. Once the gel mixture had cooled, 8 µl of 10,000X SYBR Safe DNA stain (Thermo Fisher) was added before pouring the gel. The DNA samples were loaded alongside 5 µl of GeneRuler DNA Ladder Mix (Thermo Fisher) or GeneRuler 1 kb Ladder (Thermo Fisher) for DNA length and mass approximations. These were separated by running gels at a voltage between 1–10 V/cm, depending on the requirement, and bands were visualised and imaged on a UV Transilluminator Gel Dock.

2.1.4 Purification of DNA from agarose gels and PCR reactions

To extract DNA fragments from agarose gels, the desired bands were visualised on a blue light box and excised from the gel with a scalpel. DNA from the gel slices or PCR reactions were purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences). All centrifugation steps were carried out at 16,000 x g for 1 minute. Purification was carried out according to the manufacturer's guidelines. To purify

PCRs, up to 50 μ l of PCR reaction was mixed thoroughly with 500 μ l of Capture Buffer 3 and then loaded directly into a GFX column. Gel slices were incubated with 500 μ l of Capture Buffer 3 at 60 °C until the agarose was completely dissolved before loading into the GFX column. After loading, GFX columns were centrifuged, washed once with 500 μ l of Wash 1, and then incubated with 25 μ l of Elution Buffer 4 for 1 minute before centrifugation to elute the DNA from the column into a new DNase free 1.5 ml microfuge tube.

2.1.5 Nucleic acid quantification

The concentration of DNA samples was routinely estimated using a Nanodrop 1000 spectrophotometer with the ND-1000 software (Thermo Fisher). Prior to measuring the DNA samples, the Nanodrop was initialised with 1 μ l of deionized water and then a blank measurement was taken using 1 μ l of the suspension solution without DNA. After initialising and blanking, 1 μ l of the DNA solution was pipetted onto the receiving fibre sensor on the lowest pedestal before lowering the upper pedestal arm and taking a spectral measurement. The ND-1000 software calculates the concentration of the DNA displayed in ng/ μ l (one OD260 unit = 50 ng/ μ l of dsDNA). It also displays the 260/280 and 260/230 absorbance ratios for each sample, which was used to assess the purity of DNA. A 260/280 ratio of ~1.8 is considered clean for DNA and an acceptable range for the 260/230 ratio was between 1.8–2.2.

2.1.6 DNA sequencing

All DNA sequencing in this project was by Sanger sequencing through Source BioScience based in Nottingham, UK. To prepare samples, 5 μ l of each sample at a concentration of 100 ng/ μ l for plasmid DNA and 10 ng/ μ l for PCR samples were sent by post with 5 μ l of accompanying primer in a separate tube at 3.2 μ M. A list of primers used for Sanger sequencing can be found in Appendix A.4. The data from each Sanger sequencing reaction was returned in both .ab1 and .seq formats. The chromatograms were viewed using SnapGene software.

2.1.7 Droplet digital PCR

Droplet digital PCR (ddPCR) was used to determine the copy number of the transgenes in each engineered ES cell line. ddPCR reactions were prepared at the Roslin Institute and taken on ice to the MRC Institute of Genetics and Molecular Medicine (IGMM) in Edinburgh. Primers used for ddPCR can be found in Appendix A.1 and probes are in Appendix A.5. To prepare samples all components were thawed to room temperature

(excluding the restriction enzyme). Tubes were mixed thoroughly by vortexing and centrifuged briefly to collect contents at bottom. Each 24 μ l PCR reaction contained: 12 μ l of 2X Bio-Rad ddPCR Supermix for Probes, 0.6 μ l oGM243 (36 μ M), 0.6 μ l oGM248 (36 μ M), 0.6 μ l oGM249 (36 μ M), 0.6 μ l oGM253 (36 μ M), 0.6 μ l oGM255 (10 μ M), 0.6 μ l oGM256 (10 μ M), 3 units of AluI restriction enzyme, 48 ng gDNA and made up to 24 μ l with double-distilled nuclease-free water. Samples were prepared in a 96 WP PCR plate (Eppendorf twin.tec) and immediately taken to IGMM on ice for processing. At IGMM, each 96 WP PCR plate was allowed to come to room temperature and droplets were generated using a Bio-Rad QX200 Auto DG ddPCR system. Once droplets were generated the PCR plate was foil sealed using a Bio-Rad PX1 PCR plate sealer at 180 $^{\circ}$ C for 5 seconds. The plate was then transferred to begin PCR amplification on Bio-Rad C1000 Touch Thermo Cycler using the following thermocycling conditions:

Step	Temp. (C)	Time	Cycles
Enzyme activation	95	10 minutes	1
Denaturation	94	30 seconds	40
Annealing/extension	60	1 minute	
Enzyme deactivation	98	10 minutes	1
Hold	4	Infinite	1

Following PCR amplification, the plate was transferred to the QX200 droplet reader and the resulting data analysed using Bio-Rad's QuantaSoft Pro software.

2.2 Molecular cloning

2.2.1 Restriction enzyme digestion

Restriction enzyme digestion was required for many of the molecular cloning strategies used in the project. Digests were performed following the enzyme manufacturer's guidelines. A typical restriction enzyme digest included DNA, restriction enzyme(s), buffer and distilled water. For example, 1 μ g of plasmid DNA, 10 units of each restriction enzyme, 5 μ l of the appropriate buffer and made up to 50 μ l with distilled water. This reaction was gently mixed by pipetting and incubated at the recommended temperature. DNA fragments were resolved by gel electrophoresis (section 2.1.3) and purified using a silica-based column (section 2.1.4)

2.2.2 Annealing DNA oligos

Some cloning strategies required single stranded DNA oligos (ssODN) to be annealed. A list of oligos used for annealing can be found in Appendix A.6. To do this ssODN were

resuspended in annealing buffer (10mM Tris, pH 7.5–8.0, 50 mM NaCl, 1 mM EDTA) and mixed in equimolar concentrations in a 1.5 ml microfuge tube. The tube was placed in a heating block at 95 °C for 5 minutes before turning off the heat block and allowing it to slowly cool to room temperature (~45 minutes).

2.2.3 Restriction cloning

Restriction enzyme cloning of plasmids was performed with NEB Quick Ligation Kit following the manufacturer's instructions. All restriction cloning strategies were designed and simulated using SnapGene software. Each reaction was set up in a 200 µl microfuge tube on ice, typically with a molar ratio of 1:3 vector to insert. An example of the components of a reaction are: 10 µl 2X Quick Ligase Reaction Buffer, 50 ng of 3 kb vector DNA, 37.5 ng of 1 kb insert DNA and made up to 20 µl with nuclease free water. Finally, 1 µl of Quick Ligase was added to the 20 µl mix and gently pipetting up and down. Reactions were incubated at room temperature for 5 minutes and then chilled on ice. 3 µl of the ligation reaction was transformed into 50 µl Top10 competent cells following the protocol in section 2.3.1.

2.2.4 Gibson assembly

Gibson assembly of plasmids were performed using NEB Gibson Assembly Kit in accordance with the manufacturer's guidelines. All Gibson assembly cloning strategies were designed and simulated using SnapGene software. Some Gibson assembly cloning reactions used synthesised gene fragments (gBlocks). A list of gBlocks can be found in Appendix A.6. All overlapping DNA portions were 20 to 50 bp in length. Typical Gibson assembly reactions contained a total of 0.02–0.5 pmols of DNA fragments when 1 or 2 fragments are being assembled into a vector and 0.2–1.0 pmols of DNA fragments when 4–6 fragments were being assembled. In a 200 µl microfuge tube on ice, the total DNA fragments was made up to 10 µl with deionised water before adding 10 µl of 2X NEB Gibson assembly master mix. The reaction was then incubated in a thermocycler at 50°C for 60 minutes before chilling on ice. 3 µl of the reaction was then added to 50 µl of NEB stable cells following the protocol in section 2.3.1.

2.2.5 Topoisomerase (TOPO) cloning

On occasion, PCR products with blunt ends were cloned into TOPO vectors for Sanger sequencing using Thermo Fisher Zero Blunt PCR Cloning Kit. For this, 0.5 µl of PCR product was used in the reaction if the insert was <1 kb or 2 µl if the yield was low or the product was >1 kb. The cloning reaction also contained 0.5 µl of salt solution, 0.5 µl of

pCRII-Blunt-TOPO vector and distilled water to make it up to a final volume of 3 μ l. The pCRII-Blunt-TOPO vector was added last and the solution gently mixed using the pipette tip. The mix was incubated for 5 minutes at room temperature and the 3 μ l was transformed into Top10 cells as described in section 2.3.1 followed by selection with kanamycin at 50 μ g/ml.

2.3 Bacterial molecular biology

2.3.1 Transformation protocol

50 μ l aliquots of chemically competent *E. coli* (Top10 or NEB stable cells – section 2.6) were thawed on ice for 5 minutes prior to use. Once thawed, cells were gently mixed with 3 μ l (0.001–10 ng) of plasmid DNA, and incubated on ice for 30 minutes, followed by heat shocking the mixture at 42 °C for 45 seconds, then immediately returning to ice for 5 minutes. 125 μ l of SOC medium (section 2.7.1) was then added to each transformation before incubating at 37 °C in shaking incubator for 45 minutes. 50 μ l of each culture was pipetted onto Luria-Bertani (LB) agar plates (section 2.7.1) containing carbenicillin at 50 μ g/ml and spread using a disposable plate spreader. The plates were then incubated overnight (12–16 hours) at 37 °C and analysed or stored at 4 °C for a maximum of 1 week until analysis was performed.

2.3.2 Small scale preparation of plasmid DNA

Qiagen Plasmid Spin Miniprep Kit was used to extract up to 20 μ g of plasmid DNA, following the manufacturers guidelines. All centrifugation steps were carried out at 16,000 x g unless stated otherwise.

Single bacterial colonies were inoculated in 5 ml of LB broth (section 2.7.1) supplemented with appropriate antibiotics and incubated at 37 °C with shaking overnight (12–16 hours). 1.5 ml of the overnight culture was transferred to 1.5 ml microfuge tube and spun at 6000 x g for 3 minutes to pellet bacteria. After the supernatant was discarded, the bacterial pellet was resuspended in 250 μ l of Buffer P1 and lysed in 250 μ l of Buffer P2 for no longer than 5 minutes. The cellular debris and gDNA was precipitated with the additional of 350 μ l of Buffer N3 and then pelleted by centrifugation for 10 minutes. The supernatant was collected and transferred into QIAprep spin column and centrifuged for 30 seconds. Flow through was discarded and the column was washed with 750 μ l of Buffer PE, with the flow-through again being discarded. The centrifugation step was then repeated for 1 minute to remove residual wash buffer. Each column was then transferred into a new 1.5 ml microfuge tube. 30–50 μ l of EB Buffer was added to each column and incubated at

room temperature for 1 minute. Plasmid was eluted with a final centrifugation step for 1 minute. The plasmid DNA was quantified using Nanodrop 1000 spectrophotometer (section 2.1.5) and stored at -20 °C for future use.

2.3.3 Large scale preparation of plasmid DNA

Qiagen EndoFree Plasmid Maxi Kit was used to extract up to 500 µg of plasmid DNA for large scale, endotoxin free preparations for plasmids following the manufacturers guidelines. All incubation steps were carried out at 37 °C with shaking unless otherwise stated.

Single colonies were inoculated into 5 ml of LB broth (section 2.7.1) supplemented with appropriate antibiotics and incubated for 6–8 hours. After the initial incubation, the 5 ml culture was transferred into a 500 ml conical flask containing 100 ml of LB broth supplemented with appropriate antibiotics and incubated overnight (12–16 hours). To harvest bacteria, the culture was centrifuged at 6000 x g for 15 minutes at 4 °C. The supernatant was discarded, and the cell pellet left to air-dry by inverting the open tube until most of the medium had drained. The bacterial pellet was resuspended thoroughly in 10 ml of Buffer P1 by vortexing. The cells were lysed with the addition of 10 ml of Buffer P2 and gently mixed by inverting the tube 4–6 times followed by an incubation at room temperature for no longer than 5 minutes. The lysis step was terminated with the addition of 10 ml of chilled Buffer P3 and immediate mixing by inverting the tube 4–6 times. The lysate was transferred into the QIAfilter Cartridge and incubated at room temperature for 10 minutes after which it was passed through the QIAfilter cartridge using the supplied plunger. To remove endotoxins, 2.5 ml of Buffer ER was added to the filtered lysate before inverting the tube 10 times and incubating on ice for 30 minutes. The lysate was loaded into an equilibrated QIAGEN-tip 500 column (column equilibrated by flowing 10 ml of Buffer QBT through the column) and allowed to flow through. The column was washed three times with 30 ml of Buffer QC before the plasmid DNA was eluted in 15 ml of Buffer QN into an endotoxin free 15 ml falcon tube. The DNA was precipitated with the addition of 10.5 ml of isopropanol and pelleted at 5000 x g for 1 hour at 4 °C. The supernatant was discarded, and the DNA pellet was subsequently washed with 5 ml of endotoxin-free 70% ethanol and centrifugation step was repeated at 5000 x g for 1 hour at 4 °C. The pellet was air dried for 5–10 minutes at room temperature and resuspended in 200 µl of endotoxin free Buffer TE. Plasmids were quantified using the Nanodrop 1000 spectrophotometer (section 2.1.5) and stored at -20 °C.

2.4 CRISPR reagent development

All plasmid and protein SpCas9 used in this project was delivered as a variant known as SpCas9-HF1. Herein all mention of SpCas9 is in reference to the SpCas9-HF1 (N497A/R661A/Q695A/Q926A). SpCas9-HF1 has four residue substitutions and shows at least 70% of the activity observed with native SpCas9 but exhibits considerably higher genome-wide specificity, reducing nearly all genome-wide off-target effects to nearly undetectable levels (Cebrian-Serrano & Davies, 2017; Kleinstiver et al., 2016).

2.4.1 gRNA design

SpCas9 and AsCas12a guide RNAs (gRNAs) were designed using the CRISPOR online tool (version 4.97; <http://crispor.tefor.net>) by Haeussler et al. (2016) by entering the target region of less than 2 kb, selecting the reference genome “Mus musculus – Mouse (reference) – UCSC Dec. 2011 (mm10=C57BL/6J) + SNPS” and the PAM as either “20bp-NGG – SpCas9, SpCas9-HF1, eSpCas9 1.1” for SpCas9 or “TTT(A/C/G)-21bp – Cas12a (Cpf1) – 21bp guides recommended by IDT” or “TTT(A/C/G)-23bp – Cas12a (Cpf1) recommended, 23bp guides” for AsCas12a. An image of the CRISPOR input screen is shown below in Figure 2.1. SpCas9 gRNAs were selected based on MIT specificity score (Hsu et al., 2013), predicted efficiency score (Doench et al., 2016), out-of-frame score (Bae, Park, & Kim, 2014) and potential off target mismatches in the reference genome. Although other predicted parameters are provided by CRISPOR for SpCas9 gRNAs and these can be seen in the CRISPOR output window in Figure 2.2, they were not used for gRNA selection in this project. CRISPOR provides fewer parameters for AsCas12a gRNAs and these were selected based on predicted efficiency (Kim et al., 2018), out-of-frame score (Bae et al., 2014) and off-target mismatches in the reference genome.

Once the guide RNA sequences were chosen, SpCas9 gRNA sequences for cloning in plasmid pSL70 (section 2.7.3) were screened to ensure the nucleotide at the 5' terminus was a guanine (G), if this nucleotide was not a G, an additional G was added to the 5' terminus. The human U6 RNA polymerase III promoter (hU6) within the pSL70 used for CRISPR-SpCas9 plasmid expression throughout this work requires a purine (usually a G) nucleotide to initiate gRNA transcription. The pY094 vector (section 2.7.3) for CRISPR-AsCas12a plasmid expression already has this initiating G for the hU6 to initiate gRNA transcription and therefore this modification was not required for AsCas12a gRNAs. gRNA DNA oligos (Appendix A.7) were ordered as standard, 25 nM DNA oligos from IDT with BbsI overhangs for SpCas9 gRNA oligos and BsmBI overhangs for AsCas12a gRNA oligos to clone annealed oligos to be cloned into the corresponding plasmid in the correct orientation for gRNA transcription.

Step 1

Planning a lentiviral gene knockout screen? Use [CRISPOR Batch](#)

Sequence name (optional):

Enter a single genomic sequence, < 2000 bp, typically an exon

[Clear Box - Reset to default](#)

```
GGGCAGGGCCTAGAATGTGACAGTCTGAAGACCTGGTAGAGGACCTTTGTTAGTGTGGCAGACCCAG
CCTACCCACTTGC GCGTGC GCGCCTTGTGGAGGGGGGGGGGGCGCCGGCGTGGCATAGAAGGGG
CGGGTCGGGGTGGGCGAGCTTTGAAAAGCGAGACAGACCGGCTGCCAAGCAGCGCCCGTCTCTCTG
CTTGCAGTATGGCATCCACACGGCTGATGCTGACGCCAAGCCAGACTCTGACTCTCAGAAGCTGCTT
AACGTCCTGCCTGTGCTCCCTGAGACTTCGCCTCGGCCCTGGTGGTTCCCAATTCAGGAAGTCAGTAA
TCCTCTGGTGCTTACATGGAAGCCTGGTGGCAGAAAGGTCATAGGTGAGCTCACCCAGGGGGGT
CTGGTTAGCTCTCTGTATCCCTTTAGGGGCAGCAAAGTTTGGGTTAGCTCTCTGTACTCTGTCTAGG
GGCAGCAAGTCCATGCCCATCCCTCAGTCTAATCTGCCCTTCAA
```

Text case is preserved, e.g. you can mark ATGs with lowercase.
Instead of a sequence, you can paste a chromosome range, e.g. chr1:11,130,540-11,130,751

Step 2

Select a genome

Mus musculus - Mouse (reference) - UCSC Dec. 2011 (mm10=C57BL/6J) + SNPs: C57BL/10J, C57BR/cdJ,...

Note: pre-calculated exonic guides for this species are on the [UCSC Genome Browser](#).
We have 538 genomes, but not yours? Search [NCBI assembly](#) and send a GCF_/GCA_ID to [CRISPOR support](#).

Step 3

Select a Protospacer Adjacent Motif (PAM)

20bp-NGG - Sp Cas9, SpCas9-HF1, eSpCas9 1.1

SUBMIT

33

Figure 2.1 Example of the CRISPOR input screen for SpCas9 guide RNA (gRNA) design in the mouse reference genome.

Position/ Strand	Guide Sequence + PAM + Restriction Enzymes + Variants <input type="checkbox"/> Only G- <input type="checkbox"/> Only GG- <input type="checkbox"/> Only A-	MIT Specificity Score	CFD Spec. score	Predicted Efficiency Show all scores Doench '16 Mor.-Mateos		Outcome Out-of-Frame Lindel		Off-targets for 0-1-2-3-4 mismatches + next to PAM	Genome Browser links to matches sorted by CFD off-target score <input type="checkbox"/> exons only <input type="checkbox"/> chr9 only
73 / rev	AAGGCGCGCACCGCGCAAGT GGG Cloning / PCR primers	99	99	53	68	74	81	0-0-0-0-6 0-0-0-0-0 6 off-targets	4:exon:Gm15830/Hopx 4:intron:Tenm2 4:exon:4930502E09Rik show all...

Figure 2.2 Example of the CRISPOR output screen for SpCas9 guide RNA (gRNA) design in the mouse reference genome.

2.4.2 Preparation and cloning of gRNAs in plasmids

Lyophilised gRNA DNA oligos (Appendix A.7) from IDT were resuspended to 100 μ M in nuclease-free distilled water. To prepare gRNA DNA oligos for cloning, corresponding oligos were annealed in the following reaction; 1 μ l of each oligo (100 μ M), 1 μ l of NEBuffer 2 and 7 μ l of nuclease-free distilled water were incubated at 37 °C for 30 minutes, 95 °C for 5 minutes and then ramped down to 25 °C at a rate of 5 °C per minutes. The annealed oligos were then diluted 1:100 by adding 1 μ l of annealed oligo mix to 100 μ l of nuclease-free distilled water.

To clone the annealed oligos into their respective vectors a ligation reaction for each was prepared; 0.5 μ l of vector (pSL70 or pY094 at 100 ng/ μ l), 2 μ l of diluted annealed gRNA DNA oligos, 2 μ l of T4 ligase buffer (NEB), 0.2 μ l T4 ligase (NEB), 1 μ l of BbsI or BsmBI (NEB) and made up to 20 μ l with nuclease-free distilled water. The ligation reactions were incubated in a thermocycler for 37 °C for 5 minutes followed by 21 °C for 5 minutes for six cycles. The ligation reaction was then treated with plasmid safe nuclease to digest any remaining linearized DNA. For this, the following components were added to 10 μ l of ligation reaction; 2 μ l of 10X plasmid safe buffer (Cambio), 2 μ l of ATP (10 mM), 1 μ l of plasmid safe nuclease (Cambio) and nuclease-free distilled water to give a final volume of 20 μ l. The plasmid safe reaction was incubated at 37 °C for 30 minutes followed by 70 °C for 30 minutes. 2 μ l of plasmid safe treated plasmid ligation was transformed into Top10 competent cells following the protocol in section 2.3.1. Colonies were picked and small-scale plasmid preparations performed (section 2.3.2). Cloning of gRNAs in respective vectors was confirmed by Sanger sequencing plasmids (section 2.1.6) with primer oSL35 (Appendix A.5).

2.4.3 Preparation of SpCas9 ribonucleoprotein complex

Within this project, SpCas9 protein was used for some genome engineering application to limit random integration of plasmid DNA into the host genome. For this recombinant SpCas9 protein, purified from an *E. coli* strain expressing the nuclease was purchased from IDT. The protein contains a nuclear localisation sequence (NLS) and C-terminal 6-His tag. It was provided in solution at 10 μ g/ μ l. 36 nt crRNAs (Appendix A.8) and a proprietary universal 67 nt tracrRNA were also purchased from IDT.

The crRNA and tracrRNA were duplexed. Working under RNase free conditions, crRNA and tracrRNA were resuspended in nuclease free duplex buffer (IDT) to 100 μ M. A

duplexing reaction was then prepared; 1 μ l of crRNA (100 μ M), 1 μ l of tracrRNA (100 μ M) and 98 μ l of duplex buffer added to 1.5 ml microfuge tube. The reaction was added to a heat block at 95 $^{\circ}$ C for 5 minutes. The heat block was then turned off and allowed to slowly cool to room temperature (~45 minutes). Duplexed crRNA:tracrRNA (1 μ M) was then stored at -80 $^{\circ}$ C for no more than 1 month before using.

To form ribonucleoprotein (RNP) complex, SpCas9 was thoroughly mixed by inverting the tube several times, and briefly centrifuged before diluting to 1 μ M in Opti-MEM (Thermo Fisher). For each 100 μ l of RNP solution, 8 μ l of crRNA:tracrRNA (1 μ M), 8 μ l of SpCas9 (1 μ M) and 84 μ l of Opti-MEM were mixed in a 1.5 ml microfuge tube, vortexed briefly and incubated at room temperature for 5 minutes to assemble RNP. RNP was either used immediately or stored at -80 $^{\circ}$ C for up to 6 months.

2.4.4 Screening gRNAs

2.4.4.1 T7 endonuclease I (T7EI) assay

T7EI assay was used to screen gRNA cutting efficiency due to the enzymes ability to detect and cleave miss-matched DNA. After extracting of gDNA from CRISPR-Cas transfected cells (section 2.1.1), a PCR was performed to amplify across the gRNA target site with NEB Phusion High-Fidelity polymerase (section 2.1.2.1). 2 μ l from each PCR was run on a 1% agarose gel to estimate DNA concentrations. 200 ng of PCR product from each PCR reaction was then denatured and reannealed in the thermocycler using the cycling conditions listed below:

Step	Temperature	Ramp Rate	Time
Denaturation	95 C		5 minutes
Annealing	95-85 C	-2 C per second	
	85-25 C	-0.1 C per second	
Hold	4 C		Infinite

After reannealing, 1 μ l (10 units) of T7EI was added into the reannealed PCR amplicons, mixed and incubated at 37 $^{\circ}$ C for 15 minutes. Immediately after incubation, the digestion was resolved on a 2% agarose gel. The resulting cleaved and full-length PCR products were visualised and imaged using a UV transilluminator Gel Dock.

2.4.4.2 Inference of CRISPR Edits (ICE) analysis

ICE analysis was performed by extracting gDNA from CRISPR-Cas transfected cells and a negative wild type control and PCR amplifying across the gRNA target site using NEB Phusion High-Fidelity polymerase (section 2.1.2.1). The PCR product was then purified (section 2.1.4) and sent for Sanger sequencing (section 2.1.6). ICE analysis uses the

quantitative trace data from the two returned Sanger sequencing reactions to determine the differences between the negative wild type control and the CRISPR-Cas edited cells (section 4.1.4.2). Synthego's online tool (<https://ice.synthego.com/>) was used for ICE analysis with default settings. This online tool was developed by Hsiao et al. (2019). The outputs include indel % (relative estimate of total cutting efficiency) and knock-out score, which were used to select optimal gRNAs for both SpCas9 and AsCas12a.

2.5 Mammalian cell culture

All cells were cultured at 37 °C with 5% CO₂. Culture media was refreshed as appropriate. Media and other mammalian cell culture reagent formulations can be found in section 2.7.2. All work was performed with Class II biosafety cabinets. All media, phosphate buffered saline (PBS) and trypsin were prewarmed to 37 °C before adding to cells. When required cells were counted using a haemocytometer.

2.5.1 Mouse embryonic stem (ES) cells

E14 mouse embryonic stem (ES) cells were thawed and expanded in ES cell medium on 0.1% gelatine coated flask and expanded up to T75 culture flask in feeder independent conditions. Passaging of ES cells involved aspirating off growth medium and washing twice with PBS before adding Trypsin 2X+G (section 2.7.2) and incubating for up to 8 minutes at 37 °C with 5% CO₂. Trypsin was then neutralised with an appropriate amount of cell culture medium.

2.5.1.1 Lipofection of ES cells

DNA plasmid transfection

In preparation for plasmid transfection in a 24 WP, a DNA mix of 5 µl of endotoxin free plasmid DNA (100 ng/µl) and 45 µl of Opti-MEM (Life Technologies) was prepared in 1.5 ml microfuge tube. A lipofectamine mix of 2.5 µl of Lipofectamine 2000 (Thermo Fisher) and 47.5 µl of Opti-MEM was prepared in a separate 1.5 ml microfuge tube. When two plasmids were co-transfected, ratios were equimolar.

For plasmid transfection in a 6 WP, a DNA mix of 2 µl of plasmid DNA (1 µg/µl) and 198 µl of Opti-MEM was prepared in 1.5 ml microfuge tube. A lipofectamine mix of 9 µl of Lipofectamine 2000 and 191 µl of Opti-MEM was prepared in a separate 1.5 ml microfuge tube. When two plasmids were co-transfected, ratios were equimolar.

Each mix was then vortexed briefly and incubate at room temperature for 5 minutes. For both 24 and 6 WP transfections, the DNA mix was then added to the lipofectamine mix and the DNA:lipofectamine mix was incubated at room temperature for a further 20-30 minutes. Whilst the mix was incubating, the cells were prepared by passaging and seeding at 2×10^5 cells per well in 24 WP or 8×10^5 cells per well in a 6 WP. The DNA:lipofectamine mix (100 μ l for 24 WP or 400 μ l for 6 WP) was then dripped onto cells in each well and swirled to distribute. Cells were incubated overnight 37 °C in 5% CO₂ before refreshing the medium the morning of the following day.

SpCas9 ribonucleoprotein transfection

For SpCas9 ribonucleoprotein (RNP) transfections in a 24 WP, an RNP mix of 50 μ l of SpCas9 RNP complex (section 2.4.3) was prepared in 1.5 ml microfuge tube. A lipofectamine mix of 2.5 μ l of Lipofectamine 2000 and 47.5 μ l of Opti-MEM was prepared in a separate 1.5 microfuge tube.

Each lipofectamine mix was then vortexed briefly and incubate at room temperature for 5 minutes. The RNP mix was then added to the lipofectamine mix and the RNP:lipofectamine mix was incubated at room temperature for a further up to 20-30 minutes. Whilst the mix was incubating, the cells were prepared by passaging and seeding 2×10^5 cells per well in 24 WP. The 100 μ l of RNP:lipofectamine mix was then dripped onto cells in each well and swirled to distribute. Cells were incubated overnight 37 °C in 5% CO₂ before refreshing the medium the morning of the following day.

Co-transfection of SpCas9 ribonucleoprotein and DNA plasmid

When SpCas9 RNP and DNA plasmid were co-transfected in 24 WP, plasmid was transfected into cells following the 'DNA plasmid transfection' protocol above followed by a separate transfection of SpCas9 RNP following the 'SpCas9 ribonucleoprotein transfection' protocol above 1 hour after the plasmid transfection. This separate transfection protocol was used as when an eGFP expressing plasmid and SpCas9 RNP were simultaneously transfected, no eGFP expression was observed, suggesting the plasmid was not efficiently transfecting the ES cells. Transfections separated by 1 hour overcame this issue.

Guide RNA transfection

crRNA:tracrRNA were duplex to form gRNA (section 2.4.3). For gRNA transfections in a 24 WP, an RNA mix of 2 μ l gRNA (12.5 μ M) and 48 μ l of Opti-MEM was prepared in 1.5

ml microfuge tube. A lipofectamine mix of 2.5 μ l of Lipofectamine 2000 and 47.5 μ l of Opti-MEM was prepared in a separate 1.5 ml microfuge tube.

Each mix was then vortexed briefly and incubate at room temperature for 5 minutes. The RNA mix was then added to the lipofectamine mix and the RNA:lipofectamine mix was incubated at room temperature for a further up to 20-30 minutes. Whilst the mix was incubating, the cells were prepared by passaging and seeding 2×10^5 cells per well in 24 WP. The 100 μ l of RNA:lipofectamine mix was then dripped onto cells in each well and swirled to distribute. Cells were incubated overnight 37 °C in 5% CO₂ before refreshing the medium the morning of the following day.

2.5.1.2 Fluorescence-activated cell sorting

Some cells in 6 WP format underwent fluorescence-activated cell sorting (FACS), 24 hours post-transfection, once strong eGFP expression was observed. Cells were passaged and resuspended in 400 μ l of FACS medium (section 2.7.2) and transferred into a 5 ml Falcon Test Tube with Cell Strainer Snap Cap by pipetting through the strainer cap before being sorted for eGFP expression. FACS was performed by Mr Bob Fleming or Mr Graeme Robertson at the Roslin Institute Bioimaging and Flow cytometry facility. Prior to each FACS, a negative and positive control for each fluorescent marker were used to gate the cells using a BD FACS Aria III machine. The eGFP positive cells were collected in 15 ml falcon tube with 10 ml ES cell medium with pen/strep. After sorting, cells were collected by centrifuging at 200 x g and resuspended in 500 μ l before returning to a 0.1% gelatine coated 24 WP with ES cell medium with pen/strep. The plate was incubated overnight at 37 °C in 5% CO₂ Cells were checked the following day for fluorescence.

2.5.1.3 Clonal isolation of ES cells

ES cell clonal isolation in this project was enhanced with the use of puromycin selection. After 72 hours of selection in ES cell medium with puromycin (section 2.7.2), single isolated ES cells colonies were picked. The term 'picked' refers to removing a single colony under a phase microscope using 20 μ l filter pipette tip set to 10 μ l and transferred into a round bottom 96 WP holding 30 μ l of trypsin in each well. After picking, the 96 WP was incubated for 5 minutes at 37°C before neutralising the trypsin by adding 70 μ l of ES cells medium into each well. Picked colonies in the 96 WP were then resuspended by pipetting up and down and transferred to a 0.1% gelatine coated flat bottom 96 WP and an additional 100 μ l of ES cells culture medium was added to each well. The medium was refreshed the following day once cells had attached.

2.5.1.4 Splitting a 96 well plate 1:3

Once ES cell clones became confluent, the 96 WP was split into three 96 WPs. Medium was removed from the original 96 WP and washed twice with PBS. 30 µl of trypsin 2X+G was added to each well and incubated for 8 minutes before neutralizing with 150 µl of ES cell culture medium. Three 0.1% gelatine coated 96 WPs were then prepared to collect the cells with 150 µl of ES cell culture medium. Once the collection plates were prepared, 60 µl of the 180µl of cell suspension was split into each well of the three prepared 96 WPs. The medium was refreshed the following day.

2.5.1.5 Freezing a 96 well plate

To freeze a 96 WP, the medium was aspirated off and wells washed twice with 150 µl of PBS. 30 µl of trypsin 2X+G was then added to each well and incubated for 8 minutes. The trypsin 2X+G was then neutralised with 70 µl of cell culture medium. 100 µl of Freeze mix (section 2.7.2) was then added to each well. Each well was covered with 50 µl of mineral oil suitable for embryo cell culture before and the plate was sealed with parafilm and placed into a Styrofoam box layered with wet ice. The box was then stored at -80 °C. Once frozen, plates were kept in the -80 °C for short term storage and transferred to the -150 °C for long term storage.

2.5.1.6 Freezing cells in cryovials

Media was aspirated off cells in the culture flask and cells were washed twice with a suitable amount of PBS. Trypsin 2X+G was added and incubated for 8 minutes at 37°C before neutralizing with ES cell culture medium. Cells were then resuspended before centrifuging for 5 minutes at 200 x g. The supernatant was aspirated off and the cell pellet was resuspended in ES cell culture medium (250 µl per cryovial). To each cryovial, 250 µl of freeze mix and 250µl of the cell suspension was added. Cryovials were placed into 'Mr Frosty' freezing vessel and transferred to -80 °C. Cryovials were stored at -80 °C for short term storage or transferred to the -150 °C for long term storage.

2.5.1.7 Thawing positive clones from 96 well plate

A 0.1% gelatine coated 24 WP was prepared with 1 ml of prewarmed ES cell culture medium with 2i. The 96 WP was thawed inside a waterproof bag in a 37 °C water bath (~3 minutes). Once thawed, a 200 µl pipette, set to 180µl was plunged into the well holding the clone to be thawed. With the tip of the pipette below the mineral oil, the cells were extracted. The entire 180µl was transferred into the prepared 24 WP and cultured in ES

cell culture medium with 2i until ready to passage. 2i supplement was used on thawing cells to help maintain pluripotency of the ES cells.

2.5.1.8 Thawing positive clones from a cryovial

The cryovial was thawed in a 37°C water bath (~1 minutes). Once a small piece of ice was visible, 0.5 ml of prewarmed ES cell culture medium was added to complete the thaw. The entire contents of the thawed cryovial was added to a 15 ml falcon tube, along with an additional 4 ml of ES cell culture medium. The falcon tube was centrifuged at 200 x g for 5 minutes. The supernatant was discarded, and cell pellet was resuspended in a suitable volume of ES cell culture medium with 2i for the intended purpose. Cells were culture in ES cell culture medium with 2i until ready to passage.

2.5.1.9 Differentiating ES cells to cardiomyocytes

Firstly, embryoid bodies (EBs) were generated by passaging and directly adding cells into bacterial-grade Petri dishes at 10×10^5 cells per 10 mm dish with ES cell medium without mouse leukaemia inhibitory factor (LIF) and incubating for 48 hours at 37 °C with 5% CO₂. After 48 hours, EBs had formed (Figure 2.3a), approximately six EBs were placed into each well of a 0.1% gelatine coated 6 WP with 2 ml of ES cell medium without LIF (Figure 2.3b). Plated EBs were incubated for 12 days at 37 °C with 5% CO₂, with medium changed every third day. Beating cardiomyocytes were identified under brightfield microscope and imaged on day 12.

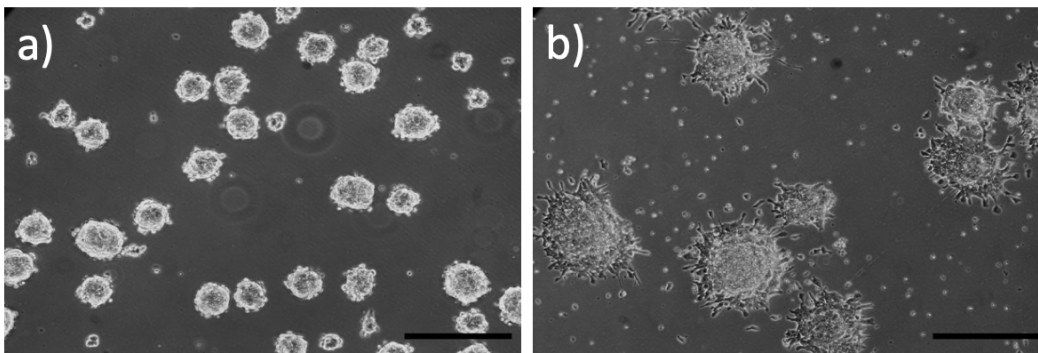


Figure 2.3 Embryoid body (EB) formation and plated EBs. a) Formation of EBs after 48 hours in suspension in ES cell medium without mouse LIF. b) EBs plated onto gelatine after 24 hours. Scale bar = 500 nm.

2.5.1.10 Karyotyping of ES cells

To karyotype ES cells, old medium was removed from ES cells in a T75 flask at 60% confluency and refreshed with 10 ml of ES cell medium with 0.1 µg/ml KaryoMAX colcemid solution in HBSS (Thermo Fisher). The plate was then incubated at 37 °C with 5% CO₂ for 3 hours before removing the medium containing the colcemid, washing twice with PBS and harvesting cells by trypsinisation. The cell pellet was resuspended in 7 ml of ES cell medium. The resuspended pellet was centrifuged at 200 x g for 5 minutes before removing the supernatant and resuspending the pellet in 300 µl ES cell medium. 5 ml of 0.4% KCl was gently added to the 300 µl cell suspension and incubated in a 37 °C circulating water bath for 10 minutes. After 10 minutes, 100 µl of freshly prepared ice-cold fixative (3:1 methanol to acetic acid) was added and gently mixed by inverting the tube 4 times. The cell suspension was then centrifuged at 300 x g for 5 minutes before removing the supernatant. 5 ml of ice-cold fixative was added to the cell pellet and resuspended gently. The cell suspension was incubated on ice for 20 minutes before centrifuging at 200 x g for 5 minutes. The supernatant was removed, and the cell pellet was resuspended in 200 µl fixative. 10 µl drops of cell suspension was placed onto Super Frost Plus microscopy slides (Thermo Fisher) from approximately 30cm above the slide. The slides were then allowed to dry overnight before staining with KaryoMAX Giemse Stain (Thermo Fisher). Images of metaphase chromosome spreads were taken with a phase contrast microscope at 60x and counted using Fiji image software.

2.5.2 Human embryonic kidney 293T (HEK) cells

HEK cells were cultured in HEK medium (section 2.7.2) at 37 °C in 5% CO₂. Cells were thawed and passaged at least once up to a T75 flask before transfecting. Passaging cells involved washing once with PBS before adding trypsin (TrypLE Express from Thermo Fisher) and incubating for 4 minutes at 37 °C in 5% CO₂. Trypsin was then neutralised with an appropriate amount of cell culture medium.

2.5.2.1 Lipofection of HEK cells with plasmid DNA

For lipofection, HEK cells were seeded into a 24 WP at 5×10^5 cells per well. After 48 hours, the cells were 80% confluent cells and ready for lipofection. Culture medium was refreshed 4 hours prior to lipofection. 1 µg of total plasmid was mixed with 100 µl of Opti-MEM medium and 4 µl of Lipofectamine 2000 in 1.5 ml microcentrifuge tube. When two plasmids were co-transfected, ratios were equimolar. The DNA-lipid complex was formed by incubating at room temperature for 5 minutes. 50 µl of DNA-lipid complex was then added to one of the wells in a 24 WP with 80% confluent HEK cells and gently swirled.

Cells were then incubated for 24 hours before refreshing medium. At 48 hours post transfection cells were viewed and imaged under an epifluorescence microscope with a FITC filter for eGFP expression or TxRed filter for dsRED expression.

2.6 Bacteria, cells and mice

Bacteria

Cells	Source	Utility
Top10 <i>E. coli</i>	Dr Spring Tan	Routine plasmid amplification
Stable competent <i>E. coli</i>	NEB	Amplifying complex plasmids

Mammalian cells

Cell line	Source	Utility
HEK 293T	Dr Dennis Headon	Testing plasmid functionality
E14 mouse ES cells (P24)	Dr Derya Ozdemir	Genome engineering

Mice

Strain	Source	Utility
C57BL/6NCrl	Roslin Institute	Ear clipping for gDNA

2.7 Reagents and buffers

2.7.1 Bacterial culture

Antibiotics

Kanamycin and carbenicillin were made up to a concentration of 50 mg/ml in distilled water and stored at -20 °C in 1 ml aliquots. The appropriate volumes of each antibiotic to give a final working concentration of 50 µg/ml was added to LB broth or agar prior to inoculating with bacteria. The antibiotic used was dependent on the purpose of the bacterial cultures.

Luria-Bertani (LB) Broth

1% (w/v) Bacto Tryptone, 0.5% (w/v) Bacto Yeast Extract and 125 mM NaCl made up with distilled water, this solution was then autoclaved on liquid cycle for 20 minutes at 15 PSI before use. An appropriate volume of antibiotics to give a final working concentration of 50 µg/ml was used as required.

Luria-Bertani (LB) Agar

1% (w/v) Bacto Tryptone, 0.5% (w/v) Bacto Yeast Extract, 125 mM NaCl and 1.5% sugar plus 15 g/L of agar made up with distilled water, this solution was then autoclaved under the same conditions as LB Broth. An appropriate volume of antibiotics to give a final working concentration of 50 ug/ml were pipetted into the melted agar once it reached approximately 50 °C. It was thoroughly mixed into the agar by gently swirling the flask before immediately pouring into 90 mm single vent petri dishes.

Super Optimal Broth (SOB) medium

2% (w/v) Bacto Tryptone, 0.5% (w/v) Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 5 mM MgSO₄ made up in distilled water, the solution was then autoclaved under the same conditions as LB Broth.

Super Optimal Borth with Catabolite repression (SOC) medium

2% (w/v) Bacto Tryptone, 0.5% (w/v) Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 5 mM MgSO₄, 20 mM glucose was made up in distilled water, the solution was autoclaved before the addition of glucose. The glucose was filter sterilized by passing a 1 M solution through a 0.2 uM filter.

2.7.2 Cell culture media and reagents

All culture media and reagent formulations (excluding Freezing medium) in this section were sterilise filtered through a Stericup-GP Sterile Vacuum 0.22 µm Filtration System (Merck).

ES cell media

Components	Amount	Company
KnockOut DMEM (1X)	500 ml	Thermo Fisher, US
Foetal calf serum (FCS)	50 ml	Life technologies, US
Non-essential Amino acids (0.1 mM)	5 ml	Sigma-Aldrich, US
L-glutamine (2mM)	5 ml	
2-mercaptoethanol (0.1 mM)	550 µl	Sigma-Aldrich, US
Mouse LIF (0.5 x 10 ⁶ U/ml)	500 µl	Millipore, UK
With pen/strep:		
Penicillin/streptomycin (10,000 U/ml)	5 ml	Gibco, US
With 2i:		
MEK/GSK3 Inhibitor Supplement	500 µl	Millipore, UK
With puromycin selection:		
Puromycin (1 mg/µl)	300 µl	Sigma-Aldrich, US

HEK medium

Components	Amount	Company
DMEM (1x) + GlutaMAX	50ml	Gibco, US
FCS	5ml	Gibco, US
Penicillin/streptomycin (10,000 U/ml)	0.6ml	Gibco, US

Trypsin 2X+G

Components	Amount	Company
EDTA (0.5M)	0.1g	Sigma-Aldrich, US
D-Glucose	0.5g	Sigma-Aldrich, US
Chicken Serum	5ml	Life Technologies, US
Trypsin (2.5%)	20ml	Thermo Fisher, US
PBS	500ml	Life technologies

Freezing medium

Components	Percentage (%)	Company
Cell culture media	40	Roslin, UK
FCS	40	Thermo Fisher, US
DMSO	20	Sigma-Aldrich, US

FACS medium

Components	Percentage (%)	Company
PBS	10 ml	Life technologies, US
FCS	1 ml	Life technologies, US

2.7.3 Plasmids

pSL70

pSL70 was initially acquired from the Addgene repository, where its official name is pSpCas9(BB)-2A-eGFP (PX458; Plasmid #48138). It expresses huSpCas9-T2A-eGFP and has a cloning backbone for a SpCas9 single gRNA (see Figure 2.4a). The plasmid was a gift to the Addgene repository from Prof Feng Zhang. After acquiring the plasmid from Addgene, Dr Simon Lillico of The Roslin Institute modified the plasmid by introducing 5 bp substitutions within the coding sequence of SpCas9 to convert the nuclease to SpCas9-HF1, as well as introducing 2 bp substitutions and two 5 bp additions to the gRNA scaffold to improve transcription and stability of the gRNAs. Subsequently, the plasmid was renamed pSL70.

pY094

pY094 was acquired from the Addgene repository (Plasmid #84743). The plasmid expresses huAsCas12a-T2A-eGFP and has a cloning backbone for an AsCas12a gRNA (see Figure 2.4b). It was a gift to the repository from Prof Feng Zhang.

pMTL23

pMTL23 is a plasmid backbone with a multiple cloning site (see Figure 2.4c). It was used as the backbone for gene drive plasmid constructs and was kindly supplied by Prof Marshall Stark from the University of Glasgow.

Cre-Reporter

Cre-Reporter was purchased from the Addgene repository (Plasmid #62732). The plasmid was a gift to the repository from Dr Niels Geijsen (see Figure 2.4d). Cre-Reporter provided the conditional *LoxP*-dsRED-*LoxP* element for the development of all Cre-inducible gene drive construct, OOEP_IVT in section 4.3.5.

pVASA-Cre

pVASA-Cre was purchased from the Addgene repository (Plasmid #15885). The plasmid was a gift to the repository from Dr Diego Castrillon (Figure 2.4e). This plasmid provides the germline specific mouse *VASA* promoter (cloned from FVB strain) for constructing OOEP_VASA plasmid in section 4.3.5.

pDonor-ROSA26

pDonor-ROSA26 was purchased from the Addgene repository (Plasmid #62732). The plasmid's official name in the repository is pDonor-MCS-ROSA26 and was a gift from Prof Charles Gersbach. It is a mouse *ROSA26* targeting vector with a multiple cloning site situated between two homology arms (See Figure 2.4f). It was used for developing ROSA26_gRNAs plasmid in section 4.3.5.

Cre-Puro

Cre-Puro plasmid was kindly supplied by Dr Stephen Meek from The Roslin Institute. The exact map of this plasmid is unknown, but an approximation is shown in Figure 2.4g. It was used from Cre-recombinase expression.

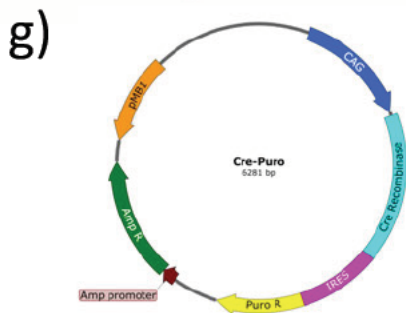
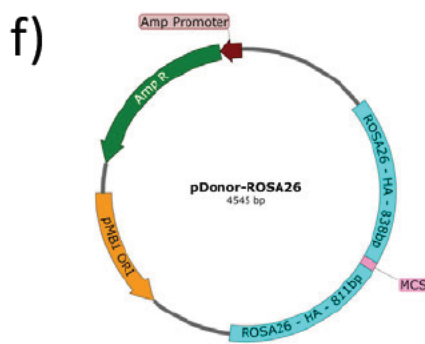
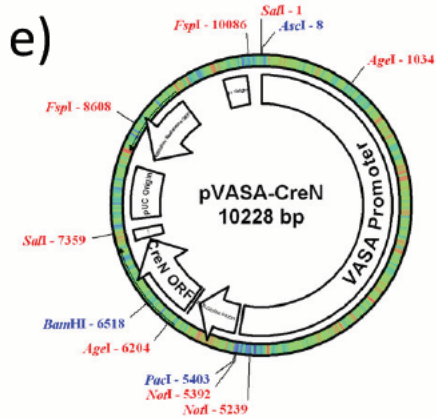
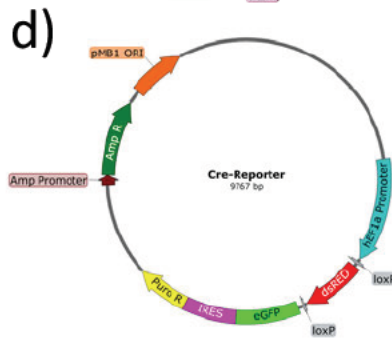
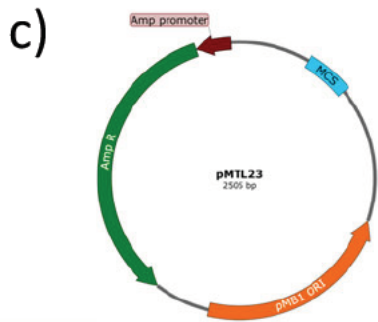
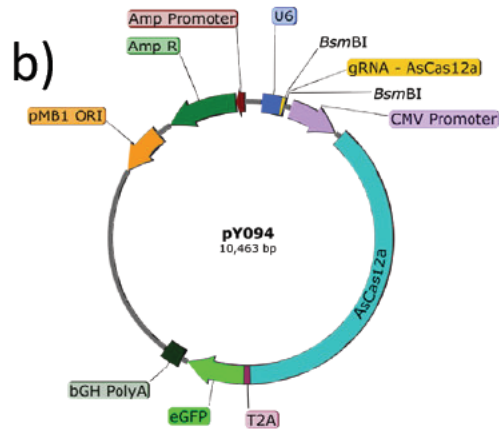
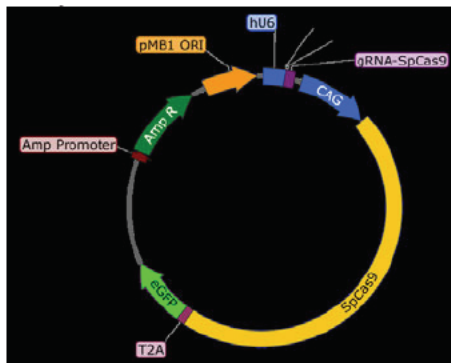


Figure 2.4 Plasmid maps. (a) pSL70, (b) pY094, (c) pMTL23, (d) Cre-Reporter, (e) pVASA-Cre, (f) pDonor-ROSA26 and (g) Cre-Puro.

2.7.4 Other buffers

50X Tris-acetate-EDTA (TAE) buffer

For 1 l of 50X TAE stock solution, 242 g of Tris, 57.1 ml glacial acetic acid, and 100 ml of 0.5 M EDTA (pH 8.0) solution was made up to final volume of 1 l with deionised water. This stock solution was diluted 50:1 with deionised water to make a 1X working solution (40mM Tris, 20mM acetic acid, and 1 mM EDTA).

5X PCR compatible loading buffer (5X Fi-RED)

10% (v/v) Ficoll-400 and a trace of cresol red dye was made up in nuclease free water in sterile conditions, the solution was filtered through 0.22 µm filter before use.

Cell lysis buffer for 96 well plate DNA extraction

For every 1 ml of 0.1% SDS, 20 µl of Proteinase K (20 mg/µl) and 4 µl of RNase A (10 mg/µl) was add. Lysis buffer was prepared fresh prior to use.

2.8 Computational tools

Bioinformatic tools

Tool	Organisation	Utility
CRISPOR (v4.97)	Tefor Infrastructure	gRNA design
Ensembl	EMBL-EBI	Genome browser
ICE CRISPR tool	Synthego	gRNA screening
PrimerBLAST	NCBI	PCR primer design
Primer3Plus	Whitehead Institute	PCR primer design
SnapGene	GSL Biotech LLC	DNA visualisation and cloning
QuantaSoft Pro	Bio-Rad	ddPCR data analysis

Other software

Software	Organisation	Utility
BioRender	BioRender	Image development
Fiji	Fiji contributors	Image analysis
GraphPad	GraphPad	Statistics and graphs
Office	Microsoft	Text, statistics and images
ZEN 2	Carl Zeiss AG	Microscopy image analysis
R packages	R core team	Population simulations

3 Gene drive population modelling, strategy formulation and experimental design

3.1 Introduction

Gene drives propagate through sexually reproducing organisms, which limits their study to *in vivo* systems (Burt & Trivers, 2006). *In vivo* genome engineering studies in mammals are typically costly and labour intensive. Furthermore, gene drive research can pose a greater risk to the environment than conventional genome engineering projects. For these reasons, it is important that substantial background work is done before considering development and testing of a gene drive system in a laboratory-contained animal population. Thoughtful strategy design and experimental planning helps safeguard wild populations and ecosystems from unintentional release, as well as ensuring value will be gained from undertaking animal research.

Important background work before embarking on the development of a population suppressing homing gene drive, include: (1) selecting candidate gene drive designs and identifying their mechanisms of inheritance; (2) *in silico* modelling these approaches in the targeted population; (3) identifying and applying safeguards and contingency strategies; (4) designing DNA constructs and gaining knowledge from *in vitro* systems; and (5) design of animal experiments, including risk assessment and regulatory approval.

3.1.1 Homing-based gene drive designs

There are three key parameters to consider when designing a CRISPR-based homing gene drive targeting female reproduction. These are the promoter used to express the Cas nuclease, the number of gRNAs, and the biology of the target gene. Although other parameters are also important, including the type of Cas nuclease used and the efficiency, specificity and characteristics of the gRNA recognition sequences, these factors are covered Chapter 4.

3.1.1.1 Promoter selection

The timing of expression of the Cas nuclease is an important parameter for the successful propagation of a gene drive as it determines where homing will take place (Grunwald et al., 2019; Li et al., 2020; Prowse et al., 2017). This can be adjusted by using promoters that drive expression at different stages for development. It is desirable to restrict expression and, therefore, homing to either germ cells or the zygote as the DNA repair machinery in somatic cells, outside these niches, typically favour NHEJ over HDR, which could reduce homing and slow propagation of the drive through the population (Gantz et al., 2015; Wang et al., 2013; Zheng et al., 2018).

As described earlier in section 1.9.1, Grunwald et al. (2019) were only able to achieve homing in female mice when SpCas9 expression was restricted to the germ line. They did not observe any HDR homing events when SpCas9 was expressed from a ubiquitous promoter but indels induced by NHEJ were present on the homologous chromosome. With ubiquitously expressed SpCas9, Grunwald and colleagues were hoping to achieve homing within the zygote. The first possibility for the lack of homing events with ubiquitous SpCas9 expression is that the homologous chromosomes are not aligned in the early zygote for inter-homologue HDR to repair DSBs. The second possibility is expression of SpCas9 is delayed in the early zygote leading to mosaic animals that went undetected in the study. The final possibility is that the outcomes were unique to the target site and constructs they used. In all cases, further investigation is needed to better understand if zygotic homing can occur in mammals with ubiquitous Cas expression.

In addition to not observing homing with ubiquitous expression, Grunwald and colleagues only confirmed homing events in the female germline and no homing events in males. There are two potential explanations here. First, despite equivalent genotypes in the male and female, Cre, SpCas9 and/or gRNA may not be well-expressed in the male germline. Second, spermatogonia continually undergo mitosis and produce new primary spermatocytes throughout the life of male mammals. By contrast, oogonia directly enlarge without further mitosis to form all of the primary oocytes during development in the embryo. The difference in the observed efficiency of inter-homologue HDR between females and males at this locus may therefore reflect a requirement for the precise timing of CRISPR–Cas activity to coincide with meiosis, although further data on this is also required.

Importantly, Grunwald et al. (2019) did demonstrate the significance of timing and pattern of expression of the Cas nuclease in a CRISPR-based homing gene drive.

3.1.1.2 gRNA array

Resistant alleles that prevent CRISPR-Cas recognising its gRNA target site could exist in the population before release or originate from indels generated when CRISPR-mediated cleavage is repaired by the error-prone NHEJ or MMEJ pathways (Champer et al., 2016; McFarlane et al., 2018). In laboratory studies, the former should be identified in the target population prior testing, the latter represents an important risk to preventing super-mendelian inheritance. A high occurrence of indel mutations in the target species would impair the spread of a suppression gene drive.

To overcome the impact of resistant alleles, it may be beneficial to utilise multiple gRNAs in a homing gene drive. The emergence of resistance within a few generations is currently one of the main causes of failure in experimental evolution gene drive studies (Hammond et al., 2017; KaramiNejadRanjbar et al., 2018). Using several gRNAs that target multiple sites decreases the rate of emergence of resistance alleles (Champer et al., 2020; Prowse et al., 2017). This strategy is similar to multi-drug therapy, whereby targeting multiple sites makes the evolution of resistance simultaneously at all sites less likely (Perron et al., 2012). Several experimental studies found that targeting multiple sites decreases the appearance of alleles resistant to cleavage (Champer et al., 2018; Champer, Yang, et al., 2019; Champer et al., 2020; Prowse et al., 2017).

Although multiplexing gRNAs is proposed as a critical component of any successful gene drive, having excessive numbers can lead to Cas activity saturation and reduced repair fidelity when homology ends for HDR around the DSB fail to line up perfectly. Using computational and experimental studies from *Drosophila*, S. Champer et al. 2020 suggested that the total number of gRNAs should be kept relatively low to achieve maximum effectiveness of multiple-gRNA drives: at least two, but well under a dozen, with the exact number depending on the type of drive and other performance characteristics.

3.1.1.3 Target genes

As described in section 1.9, the development of a homing-based sex-ratio distorting gene drive for population suppression is most efficient when targeting an essential female-fertility gene, as female reproductive performance typically controls population size. Important questions to ask when selecting candidate target genes are:

Does disruption of the fertility gene have any unintended phenotypic impacts or fitness costs on either sex?

Where does the gene play a biological role in female reproduction? (i.e. germline or soma)

Is the gene haplosufficient or haploinsufficient?

Are there confirmed knock-out models and what has been reported?

These are critical questions to be resolved before developing a gene drive and if good data is not available in the literature, scientific investigation may be required before proceeding. As demonstrated in section 3.3 of this chapter, the biology of the target gene has a significant impact on the effectiveness of a gene drive to propagate.

3.1.2 *In silico* population modelling

Just as pharmacokinetic modelling is an essential component of developing and testing a new drug, population modelling has an important role to play when developing gene drives for pest management (Sinkins & Gould, 2006; Wedell, Price, & Lindholm, 2019). *In silico* modelling provides valuable insight into the ability of a gene drive to spread and suppress. It is feasible for the human mind to elucidate how a gene drive will be inherited over several generations but at a population scale over multiple generations this becomes incomprehensible. Therefore, modelling is a valuable tool for evaluating the feasibility of gene drive strategies, and for advancing fundamental knowledge on the potential of gene drive technology (Beaghton et al., 2019; Wedell et al., 2019).

Gene drive models are only as good as the data they are based upon. Often these modelling approaches can assess key thresholds, such as how many individuals must be released for the gene drive strategy to succeed if propagating at or above a threshold. Using models, it may be impossible to absolutely know whether these simulated events will occur until the release actually takes place and the system is closely monitored. However, models are useful in gauging the probabilities of success and possible outcomes. For this reason, it is prudent for research on gene drive technology to include simulations that help to refine strategy design and identify risks (Beaghton et al., 2019; National Academies of Sciences, 2016).

3.1.3 Safeguarding gene drive research

The unintentional release or escape of gene drive animals, in this case mice, represents a risk to the environment. Best practice guidelines have been proposed by various groups of experts (Akbari et al., 2015; Lunshof & Birnbaum, 2017; National Academies of Sciences, 2016) and it is recommended that gene drive animals should be managed using

an appropriate combination of confinement strategies to mitigate these risks. Akbari et al. (2015), which was authored by a collection of 27 leading gene drive researchers from around the globe, suggest that at least two of the below confinement strategies should be implemented for gene drive research.

Ecological confinement: by conducting gene drive research in countries where the target species is not present or would have difficulty establishing in the wild.

Physical containment: by using physical barriers (e.g. secured rodent facility or nets).

Reproductive confinement: by using lab strains that cannot reproduce with wild individuals (e.g. *Drosophila* strain with chromosomal rearrangements; Akbari et al. 2015).

Molecular confinement: by molecularly restricting the gene drive to experimental population (eg. gene drive targeting an artificial sequence or by using split gene drives with the Cas gene and gRNA/s on different chromosomes; DiCarlo et al. 2015; J. Champer et al. 2019);

Although there are guidelines for the transportation of genetically altered animals, there are currently no specific guidelines for the transport of gene drive animals, and some researchers have suggested that they should not be transported or distributed to other laboratories (Akbari et al., 2015). This project plans to adhere to this recommendation.

3.1.4 Construct design and *in vitro* gene drive systems

A critical aspect of gene drive development is the design of the targeting DNA construct (Prowse et al., 2017). As gene drive constructs are typically large, including a promoter, Cas nuclease and gRNA array, plasmid DNA target constructs are usually required. Strategies for developing these constructs are highlighted in section 4.1.6. As gene drives propagate through sexually-reproduction, the animal work needed to study gene drives in vertebrates is both costly and time consuming. It would be beneficial to gain as much information as possible about a putative gene drive strategy *in vitro*, prior to commencing animal work. It may be possible to test the functionality and quantify homing efficiency of prospective gene drives using inducible gene drive systems – systems that when switched on are designed to change a cell from a heterozygous to homozygous state through HDR inter-homology repair. At the very least, the data gained from such an *in vitro* system could ensure constructs are functional prior to testing in animal studies. A key challenge of

developing such a system is finding an approach that can reliably induce expression, with no leaky gene drive expression prior to induction.

3.1.5 Experimental design for animal research

Where there is no alternative to the use of animals in research, as is the case for gene drive studies, it is important that experiments are well designed and correctly analysed in order to minimise the number of animals and suffering, while maximising the chance of generating scientifically valid results. Experiments that use too few animals may fail to pick up important effects, while those who use them in excess may waste animals or subject the animals to unnecessary suffering (Kilkenny et al., 2009).

In the UK, the laws on research using animals are set out in the Animals (Scientific Procedures) Act 1986, or 'ASPA'. The Home Office enforces the laws, including regulations on housing, environment, welfare, care, and health. Permission to work with animals is granted by the Home Office by licence only under very specific conditions. The Home Office has an inspection system to ensure compliance with laws and regulations (Wells, 2011).

Before consideration by the Home Office, each project is evaluated by one or several expert committees within their host institute. One key criterion for animal research work to go ahead is that the research cannot be carried out using non-animal methods, which is the case for gene drive. Once the experimental design has been internally approved, it is then passed on for assessment by the Home Office. This can be a lengthy process and the project cannot commence before Home Office approval is granted.

Throughout the design and assessment process there is an emphasis on the importance of the Three Rs (3Rs). These are guiding principles for the ethical use of animals in research. They were first described by Russell and Burch (1959). The 3Rs are:

Replacement: methods which avoid or replace the use of animals in research

Reduction: use of methods that enable researchers to obtain comparable levels of information from fewer animals, or to obtain more information from the same number of animals.

Refinement: use of methods that alleviate or minimise potential pain, suffering or distress, and enhance animal welfare for the animals used.

The 3Rs have a broader scope than simply encouraging alternatives to animal testing but aim to improve animal welfare and scientific quality where the use of animals cannot be avoided.

3.1.6 Aims

The aim of the work described in this chapter was to identify candidate gene drive designs, establish their mechanisms of inheritance and *in silico* model the potential impact of these approaches on a wild mouse population. Once a strategy was selected for further development, safeguards to prevent escape were included in the design and rudimentary DNA targeting constructs are proposed. Based on this planned gene drive system, animal experiments were then designed, risk assessments undertaken and regulatory approval for the work was obtained. As the final animal experiments are still ongoing and do not form part of this thesis, an overview of their design has been included within the results of this chapter. The findings in this chapter provide essential background information and justifies the work undertaken in proceeding two Chapters, 4 and 5.

3.2 Additional materials and Methods

Gene drive simulations in a wild mouse population were run in an individual based model that was adapted from a previously published model by Dr Gregor Gorjanc and Nicky Faber of the Roslin Institute. These simulations were based on code Prowse et al. (2017), which is freely available in R code from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.t78gv>. Gene drive strategies, mechanisms of inheritance and model parameters were developed by Gus McFarlane. Dr Gregor Gorjanc and Nicky Faber adjusted the Prowse et al. (2017) code to include the following parameters: simultaneous gRNA expression; CRISPR-Cas cutting efficiency of 95% (based on non-vertebrate data from Hammond et al. 2016); NHEJ repair frequencies of 0, 2 and 10%. The demographic parameters within the modelled wild mouse population were: inter-breeding interval of 0.1 years (Caughley et al., 1998); litter size of 6 (Caughley et al., 1998); survival rate at carrying capacity of 0.25; maximum survival rate of 0.54 (Prowse et al., 2017) and maximum annual population growth rate of 7.76 (Caughley et al., 1998). Simulations were performed on populations of 50,000 mice of equal sex-ratio (1:1 male to female) and run for 10 to 20 years. This population is intended to represent an isolated island population of invasive mice. Initial inoculation of all populations was with 100 male gene drive mice. Following the initial inoculation some simulations included six-monthly supplementation of male gene drive mice at rates ranging from 0 to 50% of the remaining population size. Using this model, consideration was given to the potential for evolution

resistance to develop through the formation and spread of resistance alleles and explore the conditions under which the strategies can suppress the population and assess the risk of each scenario based on invasiveness of the strategies. The code used to run these population simulations can be found at GitHub: https://git.ecdf.ed.ac.uk/HighlanderLab_public/nfaber_mouse_gd

3.3 Results

3.3.1 Candidate gene drive strategies

Three homing-based gene drive strategies for spreading female infertility were considered for development. All three strategies targeted an exonic region of an essential female-fertility gene in order to generate loss-of function mutations and subsequently spread female infertility through a population. Figure 3.1 shows the schematic construct designs of the three homing gene drive strategies that were considered. The mechanism of inheritance for each strategy is identified below.

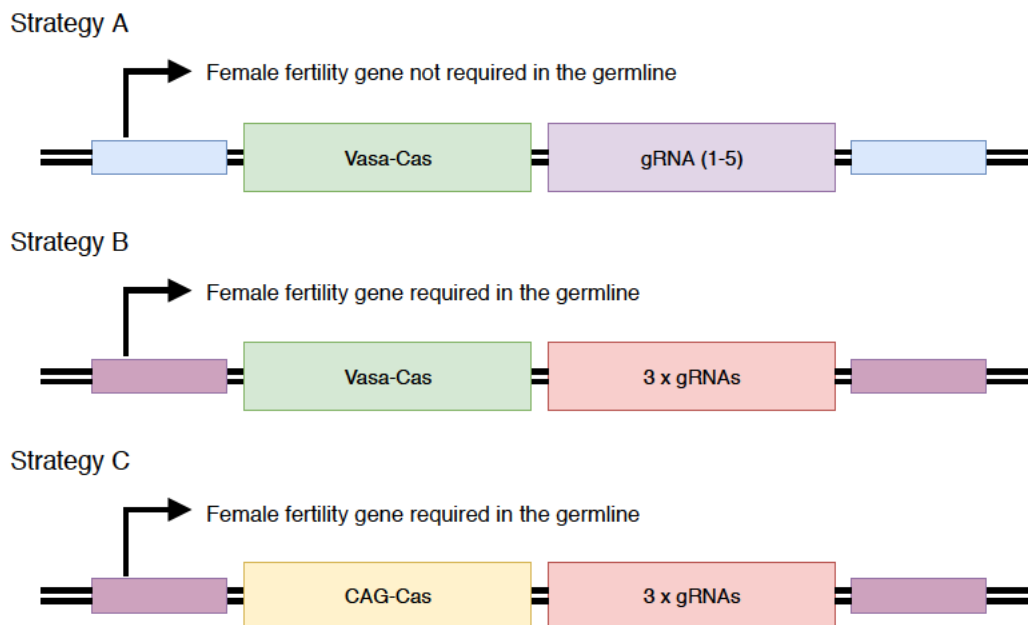


Figure 3.1 Schematic diagrams of the three gene drive constructs considered for development. Strategy A is a germline expressed gene drive targeting a haplosufficient somatic female-fertility gene. The blue boxes (Strategy A) indicate an endogenous haplosufficient female-fertility gene not required in the germline but required in the soma for female-fertility. Strategy B is a germline expressed gene drive targeting an haplosufficient female-fertility gene required in the germline. Strategy C is a constitutively expressed gene drive targeting a female fertility gene required in the germline. The purple boxes (Strategy B and C) indicate an endogenous haplosufficient female-fertility gene required in the germline for female-fertility. Arrows indicate translation start sites.

3.3.1.1 Strategy A

Strategy A targets a haplosufficient female-fertility gene that is not required in the germ cells for fertility but is required in the soma for female mice to be fertile. Therefore, disruption of both copies of the gene within the germline (induced by germline homing or NHEJ) would not compromise fertility. One candidate that has been previously suggested for this approach is the progesterone receptor gene (*PR*), mutation of which has been shown to cause recessive female infertility in mice. The anticipated mechanism of inheritance for Strategy A is shown in Figure 3.2.

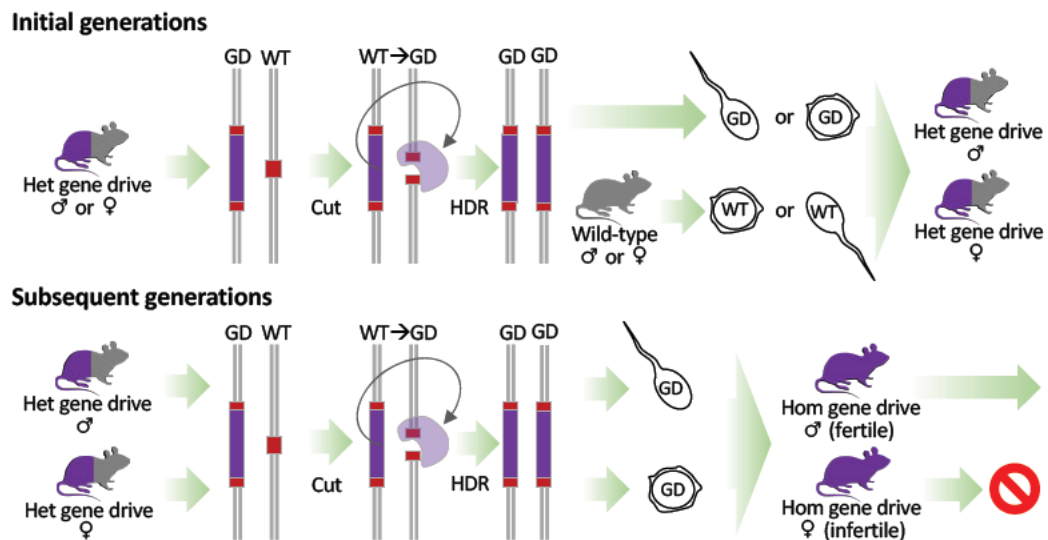


Figure 3.2 Mechanism of inheritance and suppression for Strategy A. In the initial generations, heterozygous (Het) gene drive (GD) animals are released carrying a GD cassette (purple), which disrupts the coding sequence of a haplosufficient female-fertility gene not required in the germline (HFFG; red). Within the germ cells, the GD cassette expresses the Cas nuclease and 1-5 guide RNAs (gRNAs) that cut the HFFG on the wild-type (WT) chromosome. The germ cells then repair the cut by homology-directed repair (HDR), using the GD chromosome as the repair template. This process copies the GD cassette onto the WT chromosome and ensures that most sperm or oocytes carry the GD cassette. Matings between Het GD animals and WT animals gives rise to an increasing number Het GD animals. In subsequent generations, as mating between Het GD animals becomes increasingly likely, the population declines through infertility of homozygous (Hom) GD female offspring, which are homozygous null for the HFFG. Image adapted from McFarlane et al. (2018)

3.3.1.2 Strategy B

Strategy B is similar to Strategy A; however, this approach targets a gene that is required in the germ cells for female-fertility. Therefore, disruption of both copies of the gene within the germline (induced by germline homing or NHEJ) would result in infertility of gene drive females. Unlike Strategy A this strategy does not spread through initial generations of heterozygous animals but spreads only through the male germline. Males are unaffected by loss of function of the female-fertility gene (Figure 3.3). This strategy is referred to as a 'genetically daughterless' gene drive approach, as most or all female offspring will be infertile.

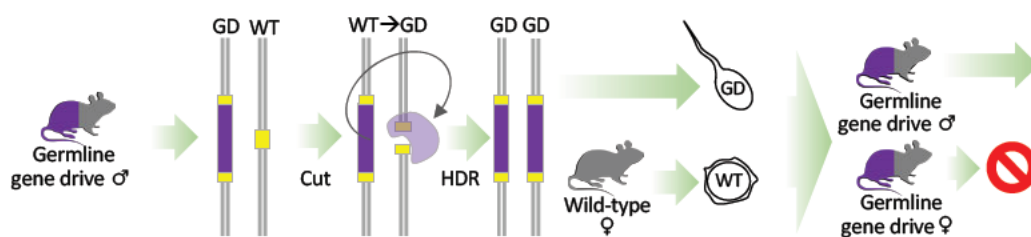


Figure 3.3 Mechanism of inheritance and suppression for Strategy B. Within the germ cells of the male, the gene drive (GD) cassette (purple) which disrupts an essential female-fertility gene that is required in the germline (EFFG; yellow), is expressed. The GD cassette encodes a Cas nuclease and 3 gRNAs which are expressed in the germ cells and cut the EFFG on the wild-type (WT) chromosome. The cut in the DNA is then repaired by homology direct repair (HDR) and copies the GD cassette onto the WT chromosome. When a GD male mates with a WT female, it passes the GD cassette onto all or most of their offspring. Germline homing via HDR will occur in offspring, which leads to homozygous germline knock-out for the EFFG. Female offspring will be infertile, whilst male offspring will retain normal fertility and pass the drive to future generations.

3.3.1.3 Strategy C

This strategy is similar to Strategy B; however, the Cas nuclease is from expressed from a constitutive promoter instead of a germline specific promoter. In this way, homing of the gene drive cassette could occur upon fertilisation in the zygote or early embryo and lead to infertile females that would be homozygous knock-out for the female-fertility gene (induced by germline homing or NHEJ). Male offspring would be unaffected and continue spreading the drive (Figure 3.4). Like strategy B, this is referred to as a 'genetically daughterless' gene drive strategy, as most or all female offspring will be infertile.

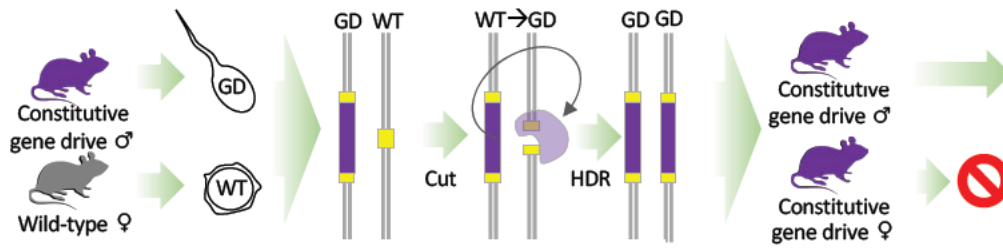


Figure 3.4 Mechanism of inheritance and suppression for Strategy C. Upon fertilisation, the sperm contributes the chromosome with the gene drive (GD) cassette (purple) to the zygote which disrupts an essential female-fertility gene that is required in the germline (EFGG; yellow). The GD cassette encodes 3 guide RNAs (gRNAs) and a Cas nuclease that is expressed by a constitutive promoter. The Cas and gRNA complex in the zygote and cut the EFGG on the maternal wild-type (WT) chromosome. The embryo then repairs the cut in the DNA by homology direct repair (HDR) and copies the GD cassette onto the maternal chromosome. This process ensures that all or most offspring will be homozygous knock-out for the EFGG. Female offspring will be infertile, whilst males will retain normal fertility and pass the drive to future generations.

3.3.2 *In silico* population simulations of proposed gene drive strategies

All three strategies were simulated using the individual-based model and parameters described in additional materials and methods section above. Strategy A has been previously modelled by Prowse et al. (2017) for a wild mouse population, however the simulations performed in this study builds on previous findings by modelling the strategy with 1–5 gRNAs and at NHEJ rates 0, 2 and 10% (Figure 3.5). Using multiple gRNAs in a gene drive is less likely to give rise to functionally resistant alleles as the targeted allele is more likely to be either: (1) converted to a gene drive allele via HDR or (2) disrupted through the introduction of multiple indels or a large deletion and become a non-functional gene. Using only one gRNA there is a greater probability that a single indel confers resistance to the gene drive whilst not disrupting the functionality of the target gene. The plots below are based on 100 gene drive animals released into a population of 50,000 wild mice, representing an isolated island population. With 0% NHEJ, Strategy A eradicated the population with any number of gRNAs. With 2% NHEJ, only gene drives with 2 or more gRNAs eradicate the population. While with 10% NHEJ, only gene drives with at least 4 gRNAs eradicated the population in most cases.

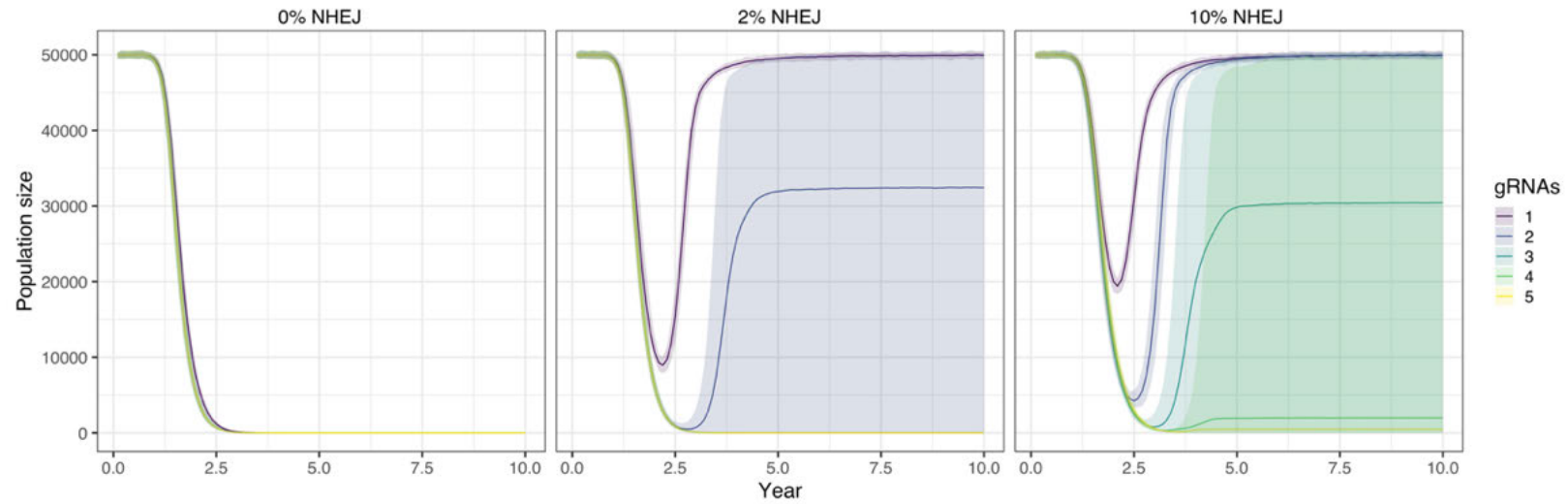


Figure 3.5 *In silico* simulation of Strategy A in a population of 50,000 wild mice. Population size over time after the introduction of 100 Strategy A gene drive male mice. Lines represent the average population size over 100 model iterations, opaque ribbons represent the 95% confidence interval. The model was run with 1 to 5 gRNAs over 3 different rates of NHEJ (0, 2 and 10%).

Strategy B and C were also modelled in the same isolated population of 50,000 wild mice. Although the molecular mechanism of gene drive propagation is different between Strategy B and C, their fundamental inheritance pattern (propagation through male germline) and underlying 'genetically daughterless' suppression approach is the same. Therefore, simulations of Strategies B and C are represented in the same plots (Figure 3.6). Although, in practice, as homing either occurs in the germline (Strategy B) or in the zygote (Strategy C) the rates of NHEJ will likely lead to differences in efficiency between these strategies.

Population simulations for Strategies B and C consisted of an initial release of 100 gene drive male mice harbouring 3 gRNAs followed by 6-monthly supplementation of gene drive males at rates ranging from 0 to 50% of the remaining population size. 3 gRNAs were selected as the fixed gRNA number in simulations of both these strategies. It was reasoned that 3 gRNAs would help combat the development of drive resistant alleles whilst being a gRNA array that remained feasible to engineer in subsequent DNA constructs developed for the *in vivo* portion of this project.

Within the simulations in Figure 3.6, following the initial release of 100 gene drive males and no further supplementation of animals (purple line), there is negligible short term and no long-term impact on mouse population size with either Strategies B and C. These strategies required substantial numbers of animals and 0% NHEJ to have significant long-term population suppression or lead to eradication of the mouse population. Under ideal conditions of 0% NHEJ it required 6-monthly population supplementation of at least 10% of the remaining population with gene drive male mice to achieve eradication within the 20-year time frame simulated here.

Although strategies B and C do not have the efficiency to act as stand-alone pest management strategies. These approaches could find use as complementary pest management tools after deploying more traditional pest control methods or to prevent that advancement of invasive fronts. In the scope of this project, strategies B and C both offer a safe approach to study gene drives in the laboratory due to their low invasive potential. The remainder of this project focused on the development of Strategy B and C for testing in a laboratory contained mouse population.

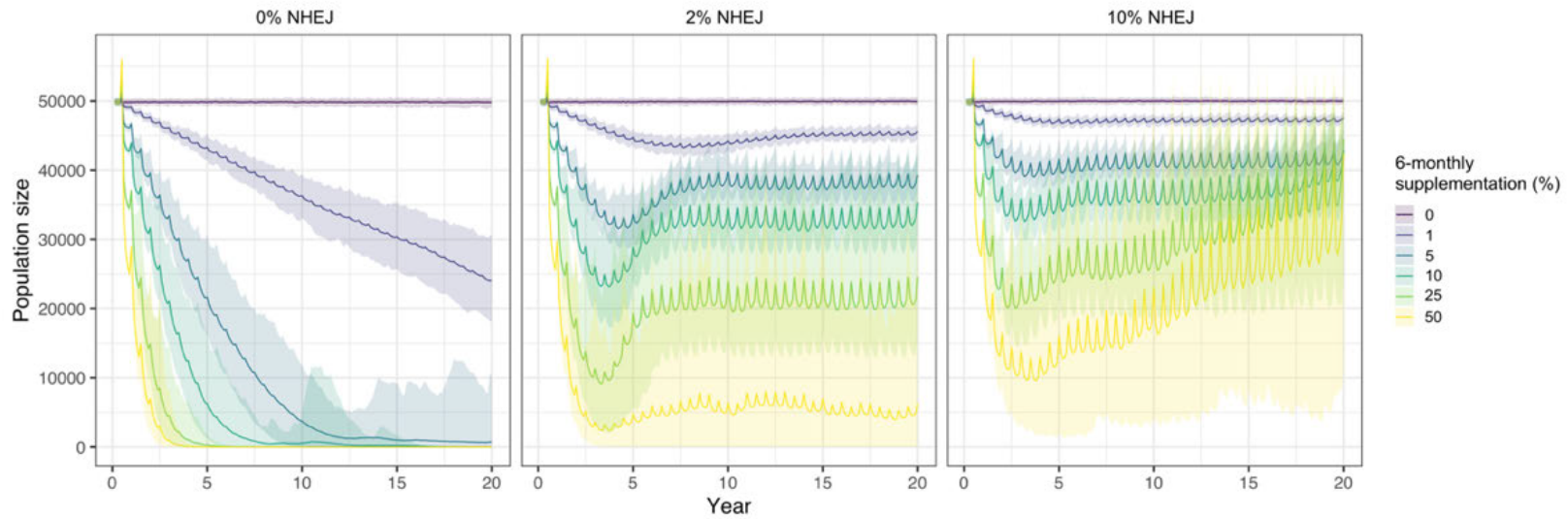


Figure 3.6 *In silico* population simulation of Strategy B and C with 3 gRNAs in a population of 50,000 wild mice. Population size over time after the initial introduction of 100 Strategy B or C gene drive male mice. Lines represent the average population size over 100 model iterations, opaque ribbons represent the 95% confidence interval. The model was run with 3 different rates of NHEJ (0, 2 and 10%) and 6-monthly supplementation ranging from 0 to 50% of the remaining population with either Strategy B or C gene drive male mice. Minor oscillations in the plot lines are due to the supplementation of gene drive mice at 6-month intervals.

3.3.3 Candidate female-fertility genes

To develop Strategy B and C, two candidate haplosufficient female-fertility genes that are required in the germline for female-fertility were identified within the scientific literature. These were *OOEP* and *NLRP5*. *OOEP* and *NLRP5* are both members of the subcortical maternal complex (SMC), which consists of at least four proteins encoded by maternal effect genes (*NLRP5*, *KHDC3*, *OOEP* and *TLE6*). These four proteins of SMC are essential for embryo cleavage beyond the two-cell stage. The SMC is believed to play an essential role in actin cytoskeleton formation within the developing embryo but its specific functions are yet to be completely clarified. (Bebbere et al., 2016; Li, Baibakov, & Dean, 2008). Previous studies have generated mouse knock-out models of *NLRP5* and *OOEP*, with both knock-out models phenocopying each other (Tashiro et al., 2010; Tong et al., 2000). The phenotype of both knock-out models is: females show normal ovulation and fertilisation, however, embryos spontaneously abort at the two-cell stage leading to infertility. In male knock-out mouse models, both *NLRP5* and *OOEP* males retain normal fertility. Importantly, knock-out mice of both sexes display no other abnormal phenotypes.

Although no undesirable phenotypes were reported in these studies, there is an increased probability of pleiotropic effects when using different strains of mice. This PhD project intends to use C57Bl6NCrl mice, while the *OOEP* knockout model reported by Tashiro et al. (2010) was in a C57BL/6J background and the mouse strain for *NLRP5* knockout model by Tong et al. (2000) was not reported. Schematic gene maps of both *NLRP5* and *OOEP* are shown below in Figure 3.7.

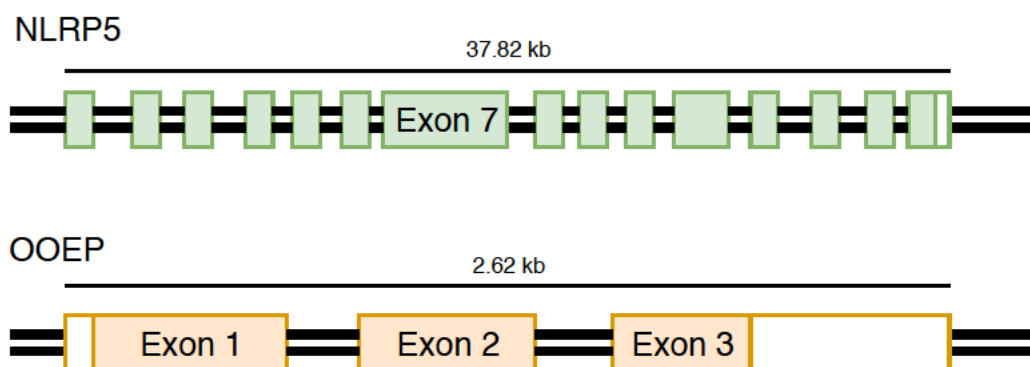


Figure 3.7 *NLRP5* and *OOEP* candidate female-fertility genes from the mouse reference genome (GRCm38.p6). The mouse *NLRP5* gene consists of 15 exons encoding the 1085 amino acid protein (transcript: ENSMUST00000108441.7). Exon 7, the major exon of the *NLRP5* was identified as the exonic target for disruption using a gene drive cassette. The mouse *OOEP* gene is substantially smaller, consisting of 3 exons encoding the 164 amino acid protein (transcript: ENSMUST00000034900.7). Exons 1 and 2 were proposed as the exonic gene drive targets within *OOEP*.

3.3.4 Safeguarding against spread

As our population simulations indicate that Strategies B and C do not operate as 'true' self-propagating gene drives, they pose less risk to the environment than Strategy A. However, in line with recommendations made by Akbari et al. (2015) additional safeguard strategies were included to further reduce the risk of an escaped animal spreading a gene drive element through a wild population. At the outset of this project no data was available on the efficiency of CRISPR-based gene drives in vertebrates. To ensure the gene drive systems developed in this project could not spread in wild mouse population, an additional two-tiers of confinement were included; these were barrier and molecular confinement.

3.3.4.1 Barrier confinement

Gene drive mice developed in this project will be housed at the Roslin Institute in a containment level (CL) 2 rodent facility to provide physical containment and inhibit mating with wild animals. The CL 2 facility for housing of animals has doubled doors, which are locked with restricted key access and have animal barriers at all exit points. The animal holding rooms within the facility have sealed pipework and removable rodent barriers that are in place during animal handling. The rodents are housed within an individually ventilated cage system. Only experienced handlers are allowed to open the cage and only one cage is opened at any one time. In addition, stock management procedures are in place to ensure all mice are identified and records kept for the Institute, University and Home Office.

Although barrier containment measures are extensive, this form of containment is known to be vulnerable to human error (Manuel, 2008), therefore, a molecular confinement strategy will also be used.

3.3.4.2 Molecular confinement

The molecular confinement strategy to be employed in this study is known as a split gene drive (SGD). A SGD system works by 'splitting up' the Cas-nuclease and gRNA array onto separate chromosomes (Del Amo et al., 2019). A Cas nuclease-only element expressed from either *NLRP5* or *OOEP* loci could spread readily in engineered mice that already express the corresponding gRNAs from an alternate chromosome but cannot bias inheritance without the gRNA (Figure 3.8). Therefore, if a SGD mouse escaped the CL 2 facility, and in the unlikely event it survived and mated with a wild mouse, the population frequency of the gRNA array will be determined by normal Mendelian dynamics and consequently limit the spread of the Cas-nuclease cassette at the *OOEP* or *NLRP5* locus.

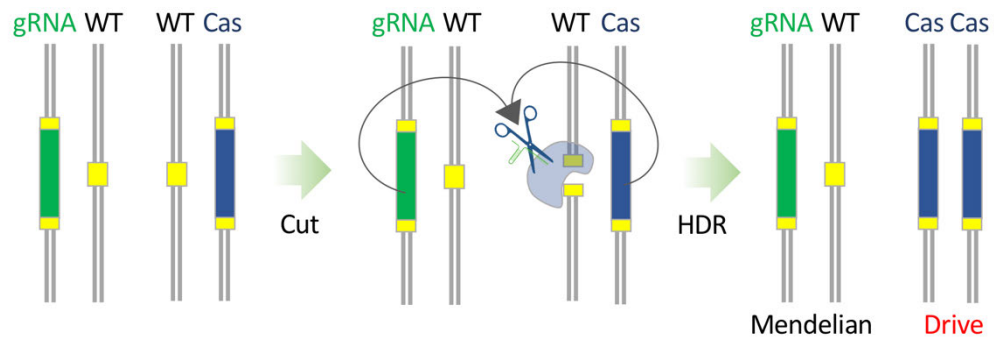


Figure 3.8 A split gene drive (SGD) separates the gRNA/s and Cas onto separate chromosomes. Within the cell, the gRNA/s and Cas are expressed, complex and cleave the target site, being the *OOEP* or *NLRP5* gene (yellow) on the homologous wild-type (WT) chromosome. Once the cut is made, homology-direct repair (HDR) copies the Cas into the cut site. Cas will only spread (drive) in the presence of the gRNA/s. As wild-type mice cannot provide the requisite gRNA/s, the SGD could not spread at a rate greater than Mendelian within a wild population.

The obvious candidate locus for expressing the gRNA array in a model mouse population is the *ROSA26*. The *ROSA26* locus encodes long non-coding RNAs that are ubiquitously expressed (Figure 3.9). The locus is considered a safe harbour and is commonly used for constitutive transgene expression in knock-in mouse models (Casola, 2010; Chen et al., 2017). Importantly, for a SGD configuration, the *ROSA26* is located on chromosome 6, while *NLRP5* and *OOEP* are on chromosome 7 and 9, respectively (Figure 3.10). Over 130 knock-in lines have been created by integrating constructs into the *ROSA26* locus. In line with the reports within the scientific literature, we anticipate no phenotypic effects as a result of construct integration and disruption of the *ROSA26* target site (Casola, 2010; Friedrich & Soriano, 1991; Yang et al., 2016).

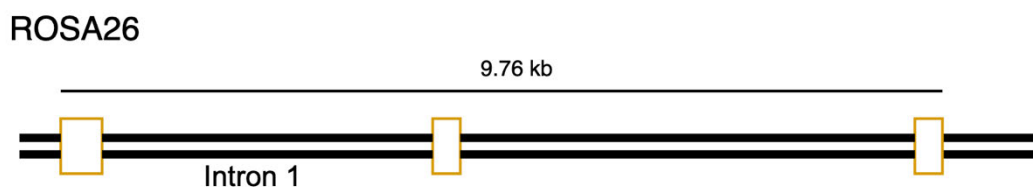


Figure 3.9 *ROSA26* from the mouse reference genome (GRCm38.p6). The mouse *ROSA26* encodes a long non-coding RNA consisting of 3 exons making up the 551 nt (transcript ENSMUST00000167415.7). Intron 1 is commonly used a safe harbour for knock-in mouse models.

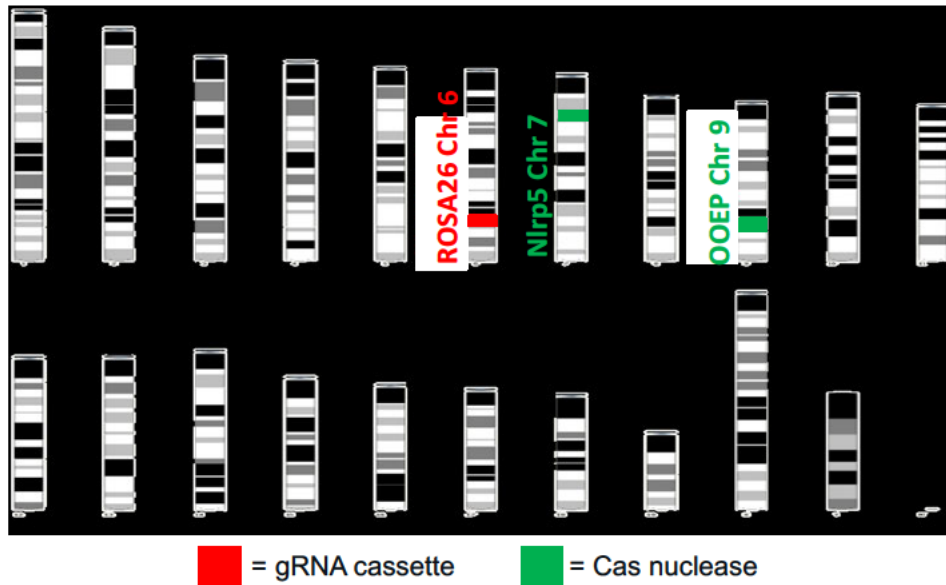


Figure 3.10 Location of the gene drive constructs in mouse genome reference (GRCm38.p6). Guide RNAs (gRNAs) will be expressed from *ROSA26* on Chromosome (Chr) 6 and Cas nuclease from either *NLRP5* on Chr 7 or *OOEP* on Chr 9.

3.3.5 Gene drive constructs

Based on *in silico* population simulations, candidate female-fertility target genes and the SGD safeguard strategy, basic construct designs were envisioned (Figure 3.11). The function of each element in the constructs is listed in Table 4.3.5. eGFP will be expressed from the same promoter as the Cas nuclease to identify expression pattern of the nuclease *in vivo*. This would be achieved with either a 2A or IRES bicistronic element. The finalised construct designs require the selection of the haplosufficient female-fertility gene to be targeted and the gRNA recognition sequences, as well as the Cas nuclease variant to be used. Selection of these construct elements is described in section 4.3. These constructs will be used to generate genome engineered ES cells for production of mice harbouring SGD systems via blastocyst injection of the engineered ES cells.

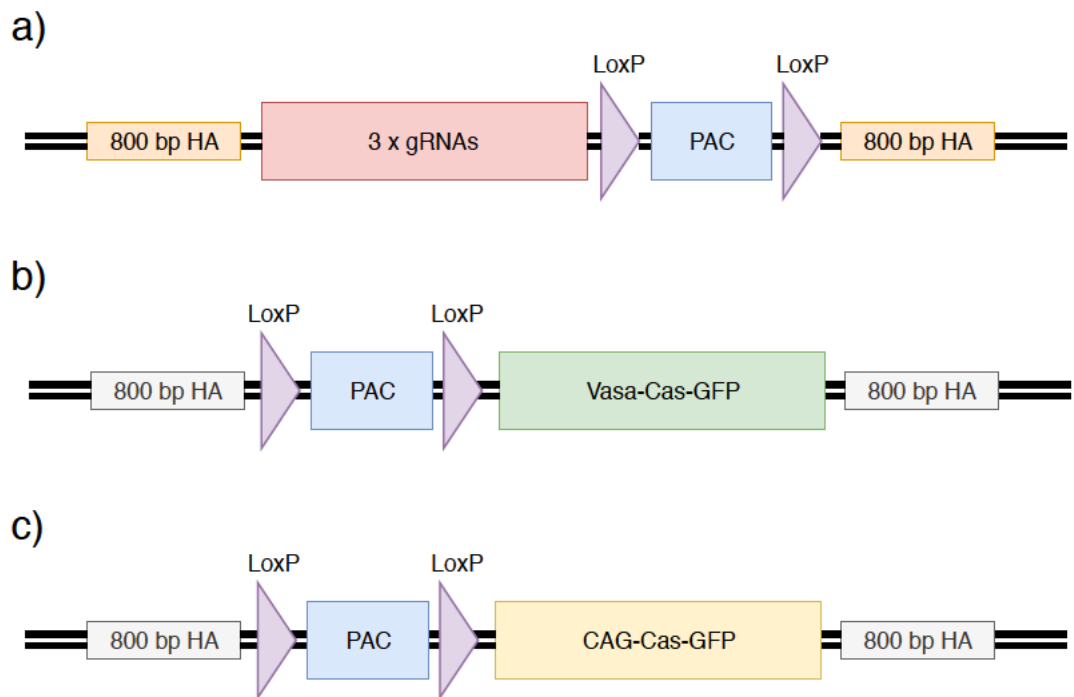


Figure 3.11 Schematic designs of the split gene drive (SGD) constructs. All constructs harbour a floxed puromycin resistance gene (PAC) for selection of genome engineered ES cells. (a) expression of 3 x gRNAs with 800 bp homology arms (HAs) for site-specific integration in the *ROSA26* locus. (b) germline expression of a Cas nuclease and GFP with 800 bp HAs for site-specific integration in *NLRP5* or *OOEP* locus. (c) constitutive expression of Cas nuclease and GFP with 800 bp HAs for site-specific integration in *NLRP5* or *OOEP* locus.

3.3.6 *In vitro* system for gene drive testing

In addition to testing homing gene drives *in vivo*, a strategy to functionally test and quantify the HDR-based homing efficiency of gene drive constructs *in vitro* was proposed. This system utilises a Cre-inducible Cas element (Figure 3.12a). To test gene drive constructs *in vitro*, heterozygous ES cells harbouring the Cre-inducible Cas gene drive construct at either candidate target gene (*OOEP* or *NLRP5*) would be generated. Clonal populations of cells would be screened to ensure that the homologous non-targeted chromosome has intact gRNA target sequences. Following the workflow shown in Figure 3.12a, heterozygote gene drive cells would be transfected with Cre mRNA to excise the floxed RFP–STOP cassette and thereby activate the drive by initiating expression of Cas-GFP. Next, Cre-treated cells would be sorted by fluorescence-activated cell sorting (FACS) to isolate GFP expressing cells, which have an active gene drive element. The HDR-homing efficiency of the gene drive could then be quantified based on the percentage of gene

drive alleles to a reference gene within the population of GFP positive cells. This could be done by assessing the transgene copy number against a reference gene by ddPCR in pooled GFP positive cells. To develop this system, the construct in Figure 3.12a was developed and tested in section 4.3.5.

3.3.7 Experimental design of animal experiments

Following the *in silico* identification of a suitable strategy, experimental design of the prospective animal studies was undertaken so the lengthy process towards obtaining University and regulatory approval through the UK Home Office for these experiments could commence. This would allow for both efficient time management of the project and prevent waste of resources if approval of animal studies was not granted.

Animal experiments will result in the generation of three lines of transgenic mice: (1) ROSA26_gRNAs mice, (2) CAG_Cas mice, and (3) VASA_Cas mice. SGD mice will be generated by separately crossing ROSA26_gRNAs mice with CAG-Cas mice and VASA_Cas. Doubly heterozygous male mice harbouring a SGD will be bred with wild-type females to produce up to 8 litters of pups. The transmission frequency of both SGDs will be quantified by genotyping all offspring born from a SGD male or female (if females are fertile) at the *OOEP* or *NLRP5* locus and *ROSA26* locus. Genotyping will be performed by a combination of PCR, Sanger sequencing and ddPCR. ddPCR will be used to determine the copy number of the *OOEP* or *NLRP5* transgene in mice genotyped as homozygous for the transgene by PCR and Sanger sequencing. The use of ddPCR is to ensure animals are not falsely genotyped as homozygous but have a large CRISPR induced deletion on the homologous chromosome. If the system 'drives' it should be transmitted at a frequency much higher than 50%. The rate of transmission between the two-SGD systems, which differ by the promoter used to express Cas, will be compared.

We will also assess the phenotypic impact of the SGD in both mouse lines, confirming if inheritance confers female infertility and assessing if any unanticipated phenotypes are present. Once SGD female pups reach sexual maturity (6 weeks), a small cohort will be mated to establish their fertility status. We expect SGD females to be infertile. We will test this by mating females up to four times (with vaginal plugs) with wild-type stud males rented from the Roslin Institute Transgenic Core that have a record of successful matings. Four plugged matings should give us insight into level of fertility reduction or infertility status.

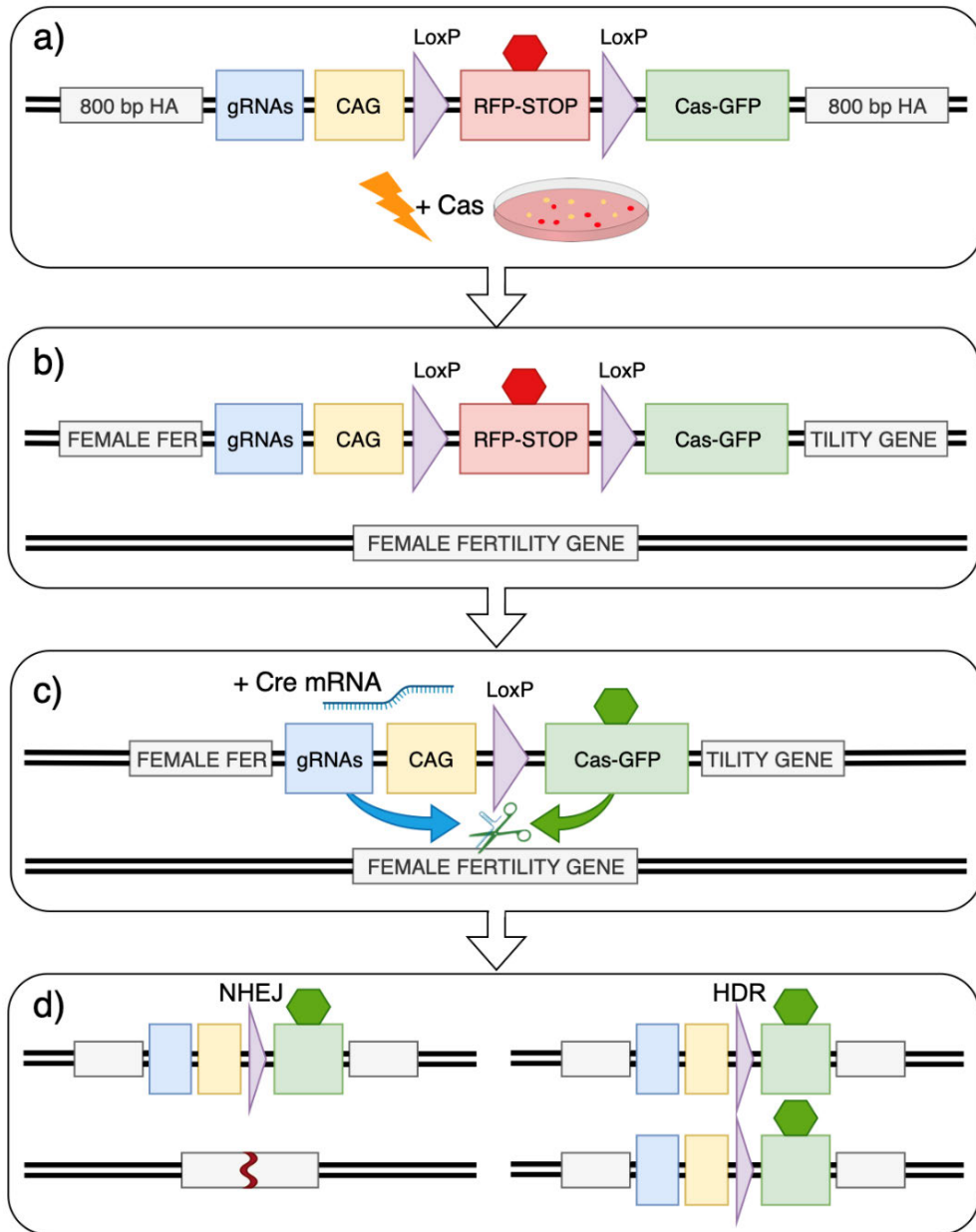


Figure 3.12 Workflow for *in vitro* testing gene drive functionality and homing efficiency in ES cells. a) Cre-inducible gene drive construct and Cas RNP would be transfected into ES cells to integrate the construct within the targeted female-fertility gene (*NLRP5* or *OOEP*). b) Fluorescence-activated cell sorting (FACS) would isolate RFP-positive clones and clones would be screened to identify heterozygotes with a WT female-fertility gene. c) Heterozygous cells would be treated with Cre-mRNA to activate the gene drive. Cells with an active gene drive element would be collected using FACS. d) The functionality and homing efficiency of the gene drive could then be quantified by using droplet digital PCR (ddPCR) to determine transgene copy number in the pooled population of GFP-positive cells.

3.4 Discussion

3.4.1 Strategy A is highly invasive

Three homing-based gene drive strategies were computationally modelled in wild mouse populations and considered for development. Of the three approaches, Strategy A was the most efficient at population control, achieving eradication in some scenarios with only 100 gene-drive carriers being introduced into a population of 50,000 wild mice. As can be seen in Figure 3.5, within the simulations, an increasing rate of NHEJ requires an increasing number of gRNAs to help combat development of resistant alleles.

The results for Strategy A are consistent with both Eckhoff et al. (2016) and Prowse et al. (2017), who modelled gene drives in mosquito mouse and respectively, and found that if functional resistant alleles that restored fertility were formed then it is likely that a gene drive will not succeed in eradicating the population. They reasoned that multiplexing gRNAs could help overcome this problem. As long as at least one site remains wild-type and, thus, cleavable, HDR can still mediate homing. The Prowse et al. (2017) model, which these simulations are based on, does not consider MMEJ-based repair outcomes but does take into account the likelihood of deletions of intervening sequences between any two of the multiplexed gRNAs. Deletion of intervening sequence between two gRNA recognition sites is assumed to result in loss of function of the female-fertility gene, regardless of the intervening sequence distance. In practice, deletion may not always lead to gene disruption.

Although this model provides insight into the potential of multiplexing gRNAs within homing gene drive designs, it is important to note that there may be additional factors to consider that are not included within this simulation. These include the potential for reduction in HDR efficiency with increasing number of gRNAs due to the imperfect homology from an increasing spacer region containing the gRNA target sites. Another factor that should also be considered is the potential of Cas nuclease activity saturation due to multiple expressed gRNAs.

A key technical hurdle with Strategy A is identifying a female-fertility gene that is not required in the germ cells for fertility but is required in the soma for female mice to be fertile. The progesterone receptor gene (*PR*) was suggested as an “attractive candidate” by Prowse et al. (2017), however, they did not mention that female mice lacking a functional *PR* gene have reproductive abnormalities, inflammation and severely limited mammary gland development (Chappell et al., 1997; Lydon et al., 1995). These phenotypes pose a significant welfare concern for the infertile gene drive females. To the

best of my knowledge no other suitable gene in mammals has been identified and this would require a separate in-depth scientific investigation to identify candidate genes.

Strategy A was eliminated as a candidate strategy for development due to its potentially highly invasive nature illustrated in the simulations and the challenges of finding a suitable target gene, which requires a separate investigation. At the outset of this project, CRISPR-based gene drive technology had not been tested in vertebrates and there remained numerous unknowns regarding the application of the technology in mice. In light of these concerns, a conservative and safeguarded strategy was considered the best path forwards.

3.4.2 Strategy B and C have low invasive potential

Fortunately, compared to Strategy A, Strategies B and C demonstrated low invasiveness within the simulations and only consistently lead to eradication of the targeted population at substantial rates of animal supplementation and in the absence of NHEJ (Figure 3.6). Unlike Strategy A, Strategies B and C do not spread through the population in somatically heterozygous individuals and therefore results in a threshold dependent spread and a requirement for constant ongoing supplementation to achieve significant suppression.

As a pest control approach, Strategies B and C may only be practical on small island populations or as a complementary tool to other traditional pest control approaches. Because of the pest status of mice and rats, a number of conventional approaches are available and can reduce densities by over 90% (Brown, 2006; Brown et al., 2007). If mice containing constructs consistent with strategies B or C were introduced following the use of conventional control, they could likely eradicate the population using a relatively smaller number of individuals to saturate the remaining population. Future modelling work should investigate this scenario.

In the scope of this project, the limited invasive potential of Strategies B and C pose a low-risk, proof-of-concept gene drive strategy to test in laboratory-contained mice. Furthermore, there are suitable target genes within the scientific literature for these approaches. Two promising candidate genes that were identified in this project are *OOEP* and *NLRP5*. However, one principle downfall of these strategies is that the homing capacity in the female germline will not be able to be qualified, as heterozygote females will likely be infertile and not give rise to offspring for genotyping.

3.4.3 Limitations of the model

Although these simulations helped guide the strategy selection, it should be highlighted that this model is simplistic, taking into account only a small number of parameters. Although much is known about wild mouse populations, there is still a need for more empirical data for increased accuracy of models. Some of the parameters, such as the 95% cutting efficiency at the target site has been taken from insect studies and it remains unknown if vertebrate systems can achieve similar efficiency (Hammond et al., 2016). Another important shortcoming of this model is that density and spatial parameters are excluded, along with the potential for behaviour resistance to gene drive animals and structured mating systems in the wild. These components are only now being implemented into the latest invertebrate gene drive models and should be considered in future vertebrate models (Champer, Kim, et al., 2019; North, Burt, & Godfray, 2019; Rode et al., 2019). Despite the limitation of the simulations presented here, they have been valuable in gaining insight and guiding decisions on strategy development.

3.4.4 Expression and design

The population simulations of Strategies B and C illustrate these approaches have the same ability to suppress a population when the rate of NHEJ is equal. However, in practice the rate of NHEJ is unlikely to be equal for both these systems. This study will test two different promoters to express the Cas nuclease, one using a synthetic CAG promoter to constitutively express the Cas nuclease, which aims to drive homing within the zygote or within the early developing embryo. The other system will use a germline specific mouse *VASA* (as known as *DDX4*) promoter to drive expression and homing in the germ cells. As Grunwald et al. (2019) report, it is likely that the timing of Cas expression will influence HDR efficiency and the resulting genome editing outcomes (Figure 3.13).

3.4.5 Safeguard strategies

Although our modelling indicates that both Strategies B and C require considerable numbers and ongoing release of gene drive animals to have an impact on population size, two additional safeguard strategies were incorporated in the experimental design: barrier confinement and molecular confinement. Barrier confinement involves housing animals in a CL 2 facility, while the molecular confinement strategy is a SGD, which works by separating the gRNA and Cas onto separate chromosomes so one allele will drive in the presence of the other but not the other way around. A SGD drive poses no greater risk to the environment than a traditional genetically engineered mouse, and the transgenes employed in this study are not predicted to confer a selective benefit. *ROSA26* is the

obvious choice for gRNA expression in a laboratory mouse project, with history of successful gene targeting events in mice (Casola, 2010).

3.4.6 *In vitro* gene drive strategy

In addition to design and modelling, I also layout the design of a strategy for *in vitro* testing homing gene drives. The approach would allow constructs to be functionally tested and HDR interhomolog repair to be quantified. One advantage of this *in vitro* system is that it could be used to assess how factors, such as RAD-51, Mek1 or Spo11, can impact HDR-mediated interhomolog repair. If successful, factors that increase interhomolog repair could be co-expressed in future gene drive designs. However, it is important to note that genome editing efficiencies and outcomes are cell type dependent (Komor, Badran, & Liu, 2017). The *in vitro* system to be developed in this project is intended for testing in ES cells, while cleavage and homing of *in vivo* gene drives is targeted to zygotes and germ cells.

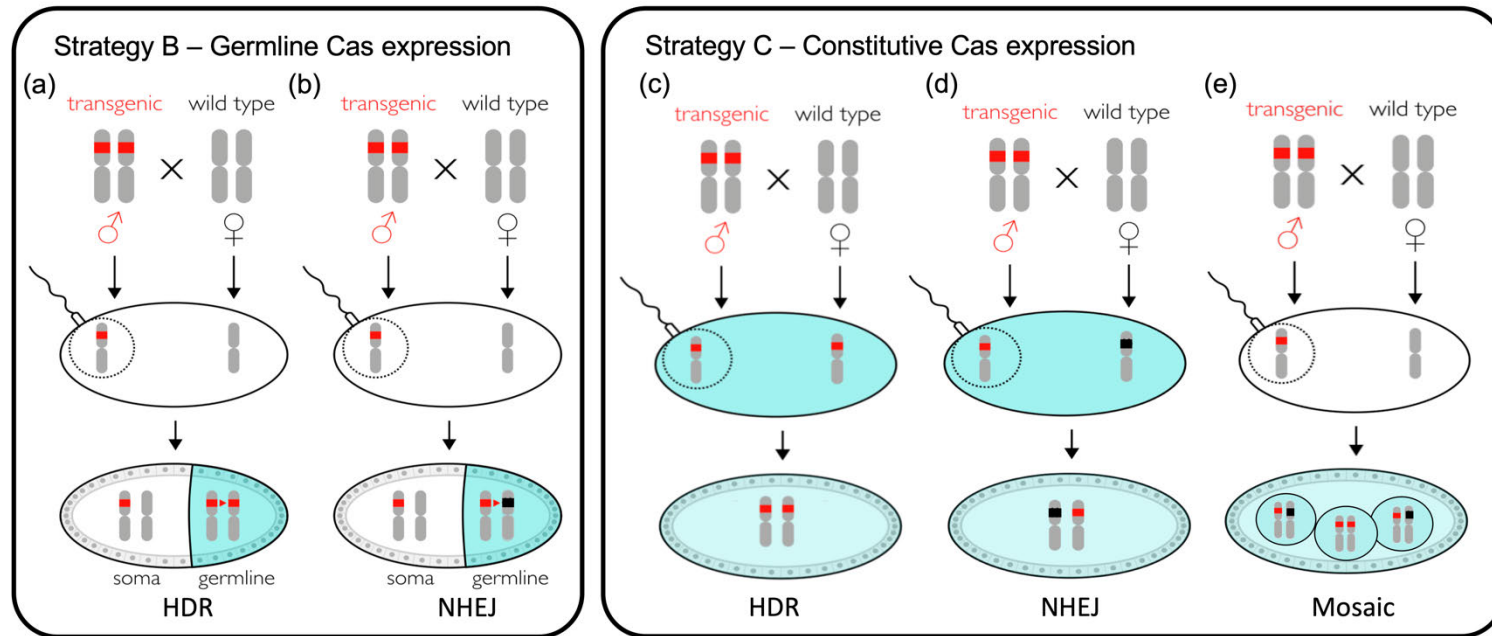


Figure 3.13 Possible genome editing outcomes from germline Cas expression (Strategy B) and constitutive Cas expression (Strategy C). Red band represents gene drive cassette. Turquoise shading represents a region of Cas expression. Black bands indicate an insertion or deletion of base pairs (indel). For Strategy B, it is expected that upon fertilisation Cas is not expressed within zygote, embryo or soma. Cas is expressed in the germline of mature males and females, which leads to cleavage of the maternal wild-type chromosome. The DSB is repaired by either (a) HDR, copying the gene drive cassette into the cut site leading to fertile male and infertile female offspring or (b) non-homologous end joining (NHEJ), leading to fertile male and infertile or fertile female offspring. For Strategy C, upon fertilisation cleavage of the maternal wild-type chromosome will either be repaired by (c) HDR, leading to fertile male and infertile female offspring; (d) NHEJ, leading to fertile male and infertile or fertile female offspring; (e) delayed Cas expression leading to a combination of both HDR and NHEJ, giving rise to mosaic offspring, which are fertile male and infertile or fertile female offspring.

3.4.7 Experimental design and regulatory approval

As may be expected, gene drive research receives greater oversight and scrutiny than conventional production of transgenic research animals. Gene drives have the potential to impact the shared environment. This project, being the first example of a vertebrate gene drive in the UK, was highly scrutinised. Ultimately, much of my time on this project was spent writing a total of nine risks assessments for genetically modified cell lines and animals, attending biosafety committee meetings, training courses, presenting to various safety committees within the University and finally writing a Project License (PPL) for the UK Home Office. After the lengthy journey navigating the regulatory landscape, a PPL for this project was granted by the Home Office until the 14th of January 2020. The cover page of the PPL and a non-technical summary can be found in Appendix A.9. The administrative burden required for progression of this gene drive project was a significant hurdle and as an emerging researcher I would consider this when weighing up future gene drive research opportunities.

3.4.8 Conclusions

Modelling can provide insight into the potential of gene drive technology at the population level. It is a valuable tool for gene drive research and helped guide the decision-making process in this project. A weak performing gene drive was purposefully selected, and multiple safeguards were incorporated into the design. It is hoped that the data generated from this project will allow for improved accuracy of subsequent gene drive models and permit the safe testing of gene drives with greater potential for real-world application. Following the experimental design of animal studies and significant university and regulatory scrutiny, the animal work received the 'green light' as safe and valuable research endeavour.

4 Reagent development and validation

4.1 Introduction

After selecting two split gene drive (SGD) strategies to develop in Chapter 3, the reagents needed to engineer these approaches in ES cells were designed, developed and tested. This included designing and screening gRNAs, Cas nuclease selection and determining the best female-fertility gene of two candidates based on cutting efficiency, as well as designing, cloning and validating targeting constructs. The majority of these reagents were validated in ES cells. Background on mouse ES cell culture can be found in section 5.1.

4.1.1 Cell transfection methods

To test genome engineering reagents *in vitro*, scientists must introduce DNA, RNA or protein into cells. In eukaryotic cells this is called transfection and can be divided into transient and stable approaches. In transient transfection, the foreign material is expressed for a limited amount of time, and any transfected DNA does not integrate into the genome, which is typically the case when screening gRNA efficiency or validating plasmid functionality. Stable transfection results in the integration of foreign DNA into the cell's genome, which is the case when engineering a gene drive in ES cells (Behringer et al., 2014; Kim & Eberwine, 2010). Many factors affect the efficiency of transfection, including the quality of the cells being cultured, optimal medium and growth conditions, and cell characteristic, as well as the quality and quantity of the transfected material (Behringer et al., 2014).

The most widely used non-viral methods for the delivery of genome editing reagents into ES cells are electroporation and lipofection (Behringer et al., 2014). Electroporation involves applying an electrical field to cells in order to increase the permeability of the cell membrane, allowing the genome editing reagents to enter into the cell (Capecchi, 2005; Kotnik et al., 2019). Lipofection achieves the same results but using synthetic cationic lipids that interact with nucleic acids or proteins to form lipid complexes, which fuse with the cell membrane to be delivered into the cell (Ma et al., 2004). As there was a robust

ES cell lipofection protocol established at The Roslin Institute with proven ability to deliver successful germline transmission, lipofection was the transfection method used throughout this project.

4.1.2 SpCas9 vs AsCas12a

This project set out to test two Cas nucleases – SpCas9 and AsCas12a – for use in the proposed SGD strategies (section 3.3.5). A biological and structural description of both SpCas9 and AsCas12a can be found in section 1.8.1 and 1.8.2, respectively. As gene drive nucleases, both SpCas9 and AsCas12a have potential advantages. SpCas9 is currently the best-characterised Cas nuclease for genome editing (Christie et al., 2020). It has a record of success in genome engineering projects in diverse species, include mice, and is currently considered the most robust Cas nuclease for genome engineering applications (Lee, Yoon, & Kim, 2020). Furthermore, SpCas9 has been successfully applied to engineer gene drives in yeast, *D. melanogaster*, mosquito and mice (DiCarlo et al., 2013; Gantz & Bier, 2015; Gantz et al., 2015; Grunwald et al., 2019).

An AsCas12a gene drive has not yet been reported. As a gene drive nuclease, AsCas12a may have the benefit of limiting the development of drive resistant alleles generated by the error-prone NHEJ repair pathway. Unlike SpCas9, AsCas12a cleaves relatively far away from the PAM (nucleotides 19 and 23; see Figure 1.8). AsCas12a-mediated NHEJ events are likely to retain the PAM sequence. Therefore, if HDR does not initially occur after AsCas12a-mediated cleavage, the continued presence of the PAM may give AsCas12a the ability to cleave again and possibly mediate HDR (Naduthodi et al., 2019). This ‘second chance’ mechanism could improve the homing frequency of a gene drive and limit the development of gene drive resistant alleles. Furthermore, *AsCas12a* (3950 bp) is smaller than *SpCas9* (4250 bp) and has the ability process its own gRNAs from a single transcript, which is beneficial when multiplexing gRNAs (Zetsche et al., 2017).

For these reasons, both AsCas12a and SpCas9 with corresponding gRNAs were screened in this project.

4.1.3 *In silico* gRNA design

Two major challenges hinder the development and application of the CRISPR-Cas systems: potential off-target effects (specificity) and on-target efficiency of the gRNAs (Liu, Zhang, & Zhang, 2019). Careful gRNA design can help resolve these issues and powerful computational approaches facilitate *in silico* design of specific gRNAs with efficiency predictions. There are a number of gRNA design tools now available, including CHOPCHOP, E-CRISP, CCTop, EuPaGDT and CRISPOR.

Within this project, the CRISPOR online tool was used as it provides an ensemble of tools in one package. At present, CRISPOR (version 4.97) contains 521 genomes and 32 gRNA types, including both SpCas9 and AsCas12a guides. In a recent review by Liu, Zhang, et al. (2019) were they evaluated 15 gRNA design tools; they state “CRISPOR may be the best tool for designing gRNAs”. CRISPOR takes genome coordinates and sequences of less than 2 kb as inputs and provides comprehensive information as outputs, which include specificity, efficiency and out-of-frame scores, and likely mismatches in the selected genome are listed for each gRNA. A detailed step-by-step methodology and screen shots of the CRISPOR interface can be found in section 2.4.1.

4.1.4 gRNA screening

Although *in silico* tools provide valuable insight, *in vitro* screening in the cell type of interest is needed to rank gRNAs. When *in vitro* screening gRNAs, it is beneficial to screen multiple gRNAs to increase the prospect of identifying guides with suitable characteristics (Kabadi & Gersbach, 2014). Unlike gRNAs for most genome engineering projects, guides for gene drives are required to cut the target site in every generation. In addition to having highly optimised gRNAs, sequence location is also critical for gene drive gRNAs to bias MMEJ events in favour of out-of-frame mutations rather than cutting at sites where micro-homology around the cut site is likely to lead to a small in-frame mutation that can create a functional but resistant allele. Within this project the SpCas9 and AsCas12a gRNAs that were designed using CRISPOR were *in vitro* screened via two methods: (1) T7 endonuclease I assay (T7EI) and (2) Sanger sequencing followed by Inference of CRISPR Edits (ICE) analysis.

Both these approaches are based on PCR amplification of relatively small regions spanning the gRNA target site. SpCas9 complexed with a single gRNA has been shown to induce large deletions, at least kilobases in size and complex lesions, such as translocations, large insertion and non-contiguous lesions at significant frequencies (Kosicki, Tomberg, & Bradley, 2018). Therefore, screening gRNAs with T7EI and ICE assays will not provide insight into the frequency of these possible outcomes.

4.1.4.1 T7 endonuclease I (T7EI) assay

In the T7EI assay, Cas nuclease and gRNA targeting the region of interest are transfected into cells, which results in cutting and repair, predominantly by NHEJ or MMEJ at the targeted site. gDNA from the pooled population is purified and the targeted region is PCR amplified. Following PCR amplification, the PCR product is denatured and re-annealed through a series of heating and cooling steps. The re-complexed PCR product should

result in heteroduplexed DNA owing to the variation of indels in the pooled population generated from the repair of CRISPR-mediated cleavage by NHEJ or MMEJ pathways. The assay takes advantage of T7 endonuclease I, which preferentially cleaves mismatched DNA, down to a 1 bp resolution. T7EI recognises and cleaves DNA heteroduplexes that contain mismatches, generating smaller-sized DNA fragments that can be resolved by agarose gel electrophoresis. This method can be used to identify the presence of CRISPR-generated mutations based on indel formation. The frequency of indels is used to make an inference of gRNA cutting efficiency (Kim et al., 2009; Kim, Kim, Kim, et al., 2014). Researchers can use densitometry to calculate the ratio of the smaller nuclease-specific band to the uncut band in the agarose gel, this helps compare gRNA efficiencies (Sentmanat et al., 2018); however, T7EI is not a quantitative method and does not provide information on the different types of indels generated. The method is commonly used to short-list gRNAs before screening by quantitative methods, such as ICE analysis.

4.1.4.2 Inference of CRISPR edits (ICE) analysis

In an ICE analysis, like the T7EI assay, the targeted region containing CRISPR-mediated indels is PCR amplified from a pooled population of cells that have been transfected with the Cas nuclease and gRNA, however, the PCR product is Sanger sequenced along with a wild-type control sample. The Sanger sequencing trace data is used to determine the relative abundance and levels of indels due to CRISPR-mediated cleavage. ICE software aligns the unedited wild-type control sequence to the edited targeted sample sequence to determine the differences between samples. The ICE software then calculates gRNA cutting efficiency based on indel frequency and provides detailed information on the different types and distributions of indels generated as a knock-out score, this is helpful in determining the likelihood of in-frame indels that could create functional resistant gene drive alleles (Hsiau et al., 2019; Sentmanat et al., 2018).

4.1.5 Construct design

As in any CRISPR-based genome engineering project, once the Cas nuclease and gRNAs have been selected, the targeting constructs can be designed. Due to the large size of gene drive constructs, plasmid DNA vectors are typically used for delivering constructs into cells for homologous recombination. Important genetic elements for plasmid generation include a plasmid backbone with a bacterial origin of replication and an antibiotic-resistance gene to allow for the propagation of the plasmid within bacteria, while selecting against any bacteria not carrying the plasmid (Howe, 2007). Genetic elements used in the gene drive systems designed in this project, including promoters,

fluorescent proteins, bicistronic elements, inducible systems, selectable markers and homology arms. These elements are discussed further below.

4.1.5.1 Promoters

Although the role of promoters is generally understood, the important function they play in specifying Cas nuclease expression in gene drives justifies a short explanation. Promoters were originally defined as a site on the DNA where the RNA polymerase recognises a specific signal that allows it to bind tightly as a prerequisite for transcriptional initiation. Today the definition is broader and also includes cis-regulatory elements responsible for controlling transcriptional machinery and determining the level and specificity of transcription (Behringer et al., 2014). Promoters can be endogenous or exogenous, where they are built into the foreign vector and introduced into the cell, as is the case of the promoters driving expression within the constructs used in this project. For reliable expression in ES cells, the *PGK-1* promoter is commonly used (Wang et al., 2008). Where the goal is high transgene expression, the CMV early enhancer-chicken B-actin promoter (CAG) is widely used (Alexopoulou, Couchman, & Whiteford, 2008). *VASA* has been described as a germline-specific promoter and is used for germline expression of Cre recombinase in the widely used *VASA-Cre* mouse line (Gallardo et al., 2007). For expressing small RNAs, such as gRNAs, expression is routinely driven by an RNA polymerase III promoter, most often by the human U6 promoter in mammalian systems (Good et al., 1997).

4.1.5.2 Fluorescent proteins

Fluorescent proteins can be used as reporter molecules and have had a major impact on our understanding of biology. The expression of these proteins has been shown to be compatible with *in vitro* and *in vivo* applications. Visualisation requires an excitation light and a compatible filter. Each fluorescent protein has its own excitation spectrum with one or two peak maxima. In response to the excitation light, each protein emits a different-wavelength light. By 'tagging' genes with a fluorescent protein by terminal fusion or using a bicistronic element (section 4.1.5.3), it is possible to monitor protein expression and trafficking (Behringer et al., 2014).

The parental proteins of the two major classes of fluorescent proteins are the green fluorescent protein (GFP), isolated from the jellyfish *Aequorea victoria*, and the red fluorescent protein dsRED, isolated from mushroom coral (*Discosoma sp.*). Many mutants of these parental proteins with spectral and intensity variations in light emissions have now been isolated (Behringer et al., 2014; Rizzo, Davidson, & Piston, 2009). This project

used the original dsRED and the enhanced GFP (eGFP), which contains two-point mutations (F64L and S65T) relative to the original GFP. These point mutations increase fluorescence and photostability of GFP, allowing for better use in mammalian cells at 37 °C (Arpino, Rizkallah, & Jones, 2012; Yang, Cheng, & Kain, 1996).

4.1.5.3 Bicistronic elements

Co-expression of multiple genes is needed for many experiments. There are a multitude of techniques to achieve this, including the use of multiple or bidirectional promoters, or the creation of bicistronic or multicistronic elements. Unlike promoters which will create unique mRNA transcripts for each gene, multicistronic vectors simultaneously express two or more separate proteins from the same mRNA (Fan et al., 2014). The two strategies most widely incorporated into plasmids for research purposes are IRES and 2A peptide sequences (Shaimardanova et al., 2019).

IRES sequences allow for the internal initiation of translation in a cap-independent manner (Hellen & Sarnow, 2001). This type of translation initiation was first discovered on picornavirus RNAs (Jang et al., 1989). The IRES sequence, which comprises ~500 bp, allows the construction of artificial bicistronic genes containing two or more coding regions. IRES elements are very useful and commonly found in bicistronic vectors, however, they do have some disadvantages. These elements are quite large (~500 bp) and it may not be feasible to express more than two genes at a time using IRES elements (Fan et al., 2014). Scientists have also reported lower expression of the downstream cistron due to factors such as the experimental cell type and the specific genes cloned into the vector (Mizuguchi et al., 2000).

Some viruses have evolved another strategy to induce the expression of multiple proteins from a single transcript. In this approach, the coding regions of these proteins are linked by a distinct sequence that encodes for an 18 to 22 amino acid peptide (2A). During translation, the 2A sequence leads to a cleavage event before its carboxyl terminus proline, releasing the protein produced that far and initiating the translation of the downstream protein at the proline. This re-initiation creates a second protein (Behringer et al., 2014). There are four main 2A peptides used in multicistronic constructs – P2A, T2A, E2A and F2A. 2A cleavage is universal in eukaryotic cells, and, although some scientists report close to 100% cleavage with some of these 2A peptides, no consensus has been reached on which peptide works best (Fan et al., 2014; Liu et al., 2017). Within this project, T2A from *Thosea asigna virus* was used when bicistronic expression was required.

4.1.5.4 Inducible systems

In many genome engineering projects, inducible expression of genetic elements is required, as was the case for the *in vitro* gene drive system outlined in section 3.3.6. Several inducible gene expression systems have been developed, these include small molecule activations systems, such as tetracycline-inducible gene expression and site-specific recombinase systems, such as Cre-*LoxP* or Flp-*FRT* (Behringer et al., 2014). These additions to the molecular toolbox have made it possible to 'switch on' gene expression when desired. Within this project, the Cre-*LoxP* system was used.

The Cre-*LoxP* system, derived from P1 bacteriophage comprises of a Cre recombinase that recognizes a 34 bp consensus sequence consisting of two 13 bp palindromic sequences (Cre recognition sites) and an 8 bp non-palindromic core that is responsible for the directional nature of the recognition site (Figure 4.1a). The 34 bp *LoxP* recognition site is small enough to be considered a 'neutral' sequence when integrated into the mouse genome and occurs at very low frequency (Kwan, 2002). Semprini et al. (2007) report that the overall frequency of primary cryptic *LoxP* sites in the mouse genome is 1.2 per megabase. Within Figure 4.1b, it can be seen that two *LoxP* sites in the same direction will result in excision of the intervening DNA by Cre-induced recombination, leaving one intact *LoxP* site. This approach can be used to excise a stop codon to initiate expression of a downstream gene of interest. In contrast, where two *LoxP* sites are in the opposing directions (Figure 4.1c), the Cre recombinase will invert the intervening piece of DNA to align or detach a promoter from a gene of interest.

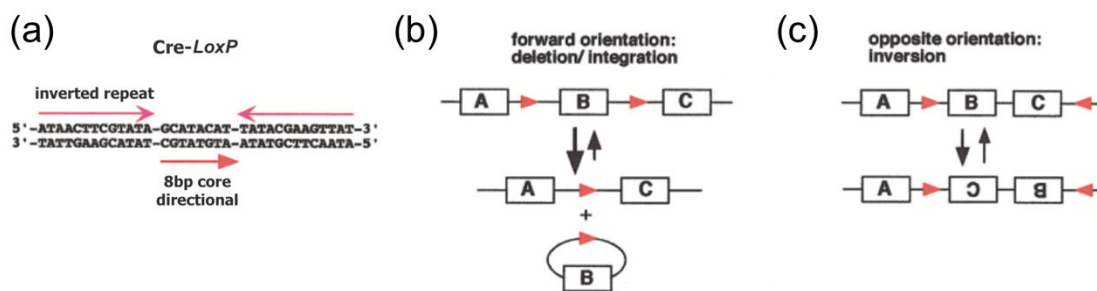


Figure 4.1 Cre-*LoxP* system. Structure and sequence of the *LoxP* recombinase recognition sites. Each site consists of two 13 bp inverted repeats flanking an asymmetric 8 bp core region that defines the directionality of the recognition site. (b) Diagram of a deletion by Cre-*LoxP* mediated recombination. (c) Diagram of an inversion by Cre-*LoxP* mediated recombination. Image was adapted from Kwan (2002).

4.1.5.5 Selectable markers

Genome engineering in ES cells allows for the use of selectable markers. Selectable markers can either be used to identify those cells where integration has occurred (positive selection) or in which the result of genome alteration was loss of a selectable marker (negative selection). This project used positive selection for genome engineering events. There are three positive selection antibiotics commonly used in ES cell-based applications. These are neomycin, hygromycin and puromycin (Behringer et al., 2014). Of these antibiotics, puromycin selection is the fastest, with sensitive cells dying within 2 days when used at 1 µg/ml concentration (Tate & Skarnes, 2011; Te Riele, Maandag, & Berns, 1992). Puromycin is an amino nucleoside antibiotic produced by *Streptomyces alboniger*. It specifically inhibits peptide transfer on ribosomes. In this way, puro impedes the growth of various insects and animals, including ES cells. The expression the *PAC* gene (puromycin-N-acetyl-transferase) confers puromycin resistance (Behringer et al., 2014). Within this project the *PAC* gene was included in all constructs for puromycin selection of genome engineering events.

4.1.5.6 Homology arms

Homology arms on targeting vectors are placed flanking the construct and positive selection markers and have the same nucleotide sequence as the two segments of the gene of interest with equivalent orientation. Relative to the orientation of the gene of interest, there are two distinguished arms – the 5' arm and the 3' arm. The sequence between these two segments should integrate by HDR (Behringer et al., 2014). Although early genetic engineering experiments typically utilised arms with at least 6 to 8 kb of total homology (Thomas & Capecchi, 1987), the double strand break induced by programmable nuclease stimulates repair at the target site, thereby reducing the size requirement of homology arms. The Zhang lab typically uses arms of 800 bp when the insert is >100 bp (Cong, 2013; Yang et al., 2013). In line with the recommendation of the Zhang lab, who pioneered CRISPR-SpCas9 gene editing in mammalian cells, this project utilised 800bp homology arms.

4.1.6 Molecular cloning techniques

After designing constructs, molecular cloning strategies are used to generate the targeting vectors. There are a broad range of cloning strategies available, including type II restriction enzyme cloning, Gibson assembly, gateway cloning with site-specific recombinases, topoisomerase cloning, Golden gate cloning with type IIs restriction endonucleases and ligation independent cloning (Casini et al., 2015; Fan et al., 2014).

This project used both restriction enzyme cloning and Gibson assembly to develop targeting vectors in the form of DNA plasmids. Constructs were generated using these cloning strategies and were transformed into bacteria for amplification, before purifying and validation of the resulting plasmid. Some background on these processes is provided below.

4.1.6.1 Restriction enzyme cloning

This cloning approach uses restriction endonucleases to create double strand DNA breaks at specific recognition sequences generating either blunt ends, where both strands of the target DNA are cut at the same spot creating blunt ends, or sticky ends, where the endonuclease cut both strands of the target DNA at different spots creating overhangs. To be able to clone a DNA insert into a cloning or expression vector using restriction enzyme cloning, both the insert and vector have to be treated with restriction enzymes that create compatible ends. If orientation of the cloning operation is important, at least one of the restriction enzymes used should be a sticky end cutter to ensure that the insert is incorporated in the correct orientation (Fan et al., 2014; Lessard, 2013).

In practice, restriction enzyme cloning typically consists of restriction digests for your insert and plasmid vector, followed by isolation and purification of the appropriate linearized DNA fragments before ligating the insert and the vector together. Ligation is accomplished by covalently connecting the sugar backbone of the two DNA fragments. This reaction is commonly performed by the T4 DNA ligase. The DNA ligase catalyses the formation of covalent phosphodiester linkages, which permanently joins the nucleotides together. After ligation, the insert is physically integrated into the vector and the complete plasmid can be transformed into bacterial cells for amplification (Allison, 2007; Lessard, 2013).

4.1.6.2 Gibson assembly

In 2009, Dr Daniel Gibson and colleagues developed a method for the easy assembly of multiple linear DNA fragments, now called Gibson assembly (Gibson et al., 2009). Regardless of fragment length or end compatibility, multiple overlapping DNA fragments (15 to 50 bp overlap) can be joined in a single isothermal reaction (Fan et al., 2014). The Gibson assembly process can be used to assemble up to six fragments in one step, resulting in scar-free assembly that does not require the presence of specific restriction sites. The reaction exploits three different enzymes: (1) T5 exonuclease, which chews back the 5' ends of the fragment, generating long overhangs which allows the single stranded regions with homology to anneal; (2) a DNA polymerase, to fill in the gaps; (3)

Taq DNA ligase, covalently joins the annealed complementary DNA fragments, removing any nicks and creating a contiguous DNA fragment. This mix of enzymes has been optimised to work at the same temperature, so the entire reaction takes an hour or less to complete at 50 °C. The sample can then be immediately transformed into bacterial cells for amplification (Fan et al., 2014; Gibson, 2011).

4.1.6.3 Transformation and amplification

After cloning a plasmid, it is typically transformed into bacteria to be amplified. Bacterial transformation is a process of horizontal gene transfer by which some bacteria take up foreign genetic material from the environment. Bacteria that have the ability to take up free extracellular DNA are termed competent cells (Howe, 2007). The majority of commercial competent cells are lab strains of *E. coli* are derived from two individual isolates, the K-12 strain and the B strain. K-12 has led to the common lab strains MG1655 and its derivatives DH5alpha and DH10b (also known as TOP10) among others, while the B strain gave rise to BL21 and its derivatives (Fan et al., 2014; Samuelson, 2011).

In the laboratory, bacterial cells can be made competent and DNA subsequently introduced by a procedure called the heat shock method. Heat shock transformation uses a calcium rich environment provided by calcium chloride to counteract the electrostatic repulsion between the plasmid DNA and bacterial cellular membrane (Dagert & Ehrlich, 1979; Koontz, 2013). A sudden increase in temperature creates pores in the plasma membrane of the bacteria and allows for plasmid DNA to enter the bacterial cell. Once inside a bacterium, the plasmid encoded origin of replication facilitates multiplication of the plasmid, which is further amplified during consecutive rounds of cell division (Howe, 2007).

The transformed plasmid usually contains an antibiotic resistance gene. For example, a plasmid carrying the gene beta lactamase (Amp R) leads to ampicillin-resistance, allowing only transformed bacteria with the plasmid to readily grow in medium containing ampicillin. Because bacteria are easy to grow in a lab, divide relatively quickly, and exhibit exponential growth rates, plasmids can be replicated easily and efficiently in a laboratory setting using this approach (Howe, 2007).

4.1.6.4 Validation and purification of plasmid

After cloning and amplifying a plasmid, it is critical that the cloning events are verified. This is important to catch any errors in cloning, sample handling or sequence assembly steps. Initially this can be done by restriction enzyme (RE) analysis, which involves

digesting the plasmid with one or more restriction enzymes and resolving digested products on an agarose gel. Putative positive plasmids with expected band patterns should then be verified by Sanger sequencing (Fan et al., 2014; Lessard, 2013). Sequencing of the entire plasmid can be performed but it is more cost effective to sequence the “insert” (the gene or genetic element cloned into the plasmid), and any other important features that differentiate this plasmid from its predecessor. Important features may include cloning junctions (where most sequence assembly errors occur), in-frame protein fusions and PCR amplified fragments.

Following Sanger sequence verification of a plasmid, it is critical that any plasmid to be introduced into ES cells is high-quality DNA, free of endotoxins, proteins, RNA and contaminating chemicals such as oxidised phenol. The DNA should be purified either with a QIAGEN EndoFree Maxi Kit or similar purification kit, or by CsCl centrifugation (Behringer et al., 2014). Once a purified plasmid preparation is obtained, functional validation of plasmids should be performed where possible, such as expression of fluorescent reporter proteins.

4.1.7 Aims

The aim of this chapter was to generate and validate the reagents for engineering two SGD strategies in mouse ES cells. This includes designing and screening SpCas9 and AsCas12a gRNAs, primer design and optimisation, and the design, cloning and validation of plasmid constructs. Furthermore, a construct for a Cre-inducible gene drive system was also developed and tested.

4.2 Additional methods

4.2.1 gRNA screening workflow

gRNAs for SpCas9 and AsCas12a were screened in E14 mouse ES cells. A workflow for screening gRNAs is provide in Figure 4.2.

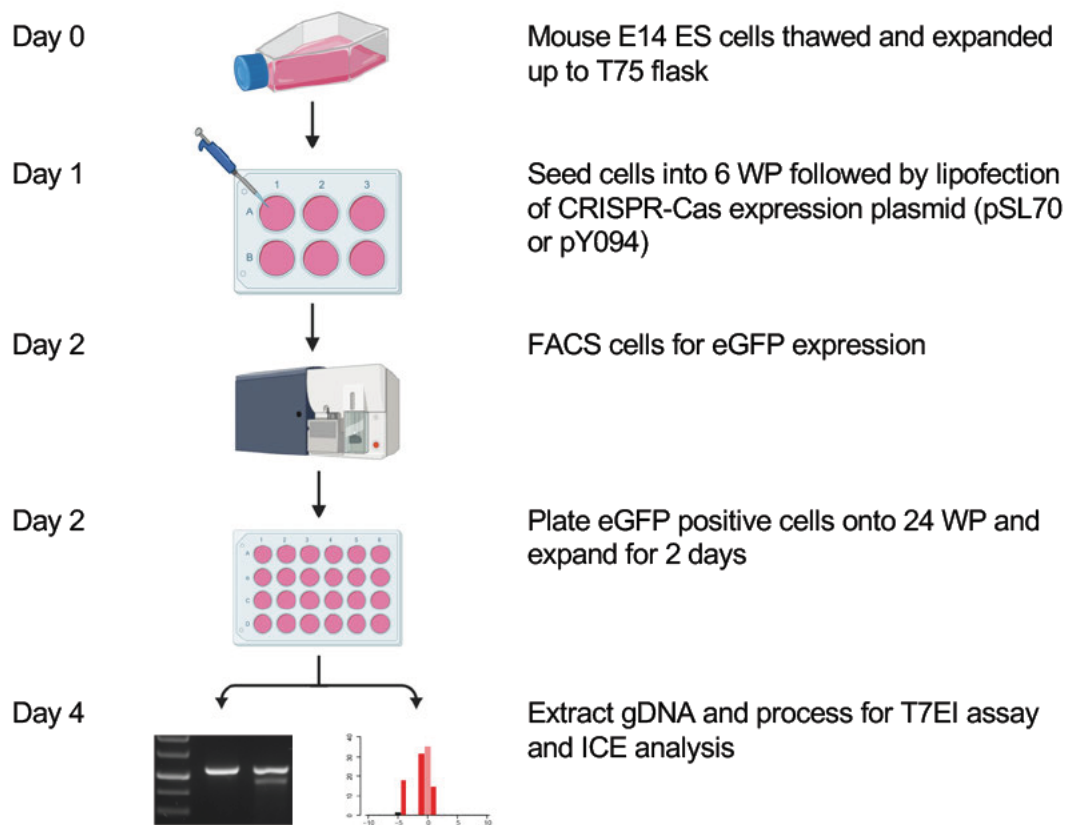


Figure 4.2 Workflow for screening plasmid expressed CRISPR-Cas in mouse E14 embryonic stem (ES) cells. Once an adequate number of ES cells were growing in log phase, ES cells were transfected following the protocol in section 2.5.1.1. Cells were then sorted for eGFP expression using fluorescence activated cell sorting (FACS; section 2.5.1.2). eGFP positive cells were plated onto a 24 WP and cultured for 2 days before extracting gDNA for T7 endonuclease I (T7EI) assay and Inference of CRISPR edited (ICE) analysis.

4.3 Results

4.3.1 *In silico* gRNA design

Using the CRISPOR tool four SpCas9 and four AsCas12a gRNAs were designed targeting both *NLRP5* and *OOEP*. The *NLRP5* gRNAs were targeted to exon 7, the largest exon of the 15 exon gene (Figure 4.3). For *OOEP*, four SpCas9 gRNAs were designed targeting exon 1 and four AsCas12a gRNAs were designed to target sites in both exons 1 and 2. *OOEP* AsCas12a gRNAs were spread across exon 1 and 2 due to the limited number of AsCas12a PAM sites (5'-TTTV) within exon 1. A further two SpCas9 and two AsCas12a gRNAs were designed targeting the *ROSA26* safe-harbour locus, located within intron 1 of the gene. This safe-harbour site corresponded to the genomic region between homology arms of pDonor-ROSA26 plasmid. To multiplex gRNAs within the SGD design, gRNAs were restricted to a short genomic region for *OOEP* and *NLRP5* to allow for efficient HDR with multiplexed gRNAs. *OOEP* SpCas9 and AsCas12a gRNAs were restricted to a 70 and 395 bp region, respectively, while *NLRP5* SpCas9 and AsCas12a gRNAs were within a 143 and 196 bp windows.

CRISPR gRNAs were selected from the CRISPOR online tool based on predictive gRNA outputs the software provides. Examples of the CRISPOR input and output screens can be found in Figures 2.1 and 2.2, respectively. gRNAs selected for this project were required to have an MIT specificity score scores greater than 50, with no identical gRNA target sites in the mouse reference genome. gRNAs were also required to have an out-of-frame score greater than 50. In addition to specificity and out of frame score, predicted efficiency was also taken into account but with less reverence. The sequences of the gRNAs and the CRISPOR data outputs can be seen in Appendix A.10. The location of the gRNAs at *OOEP*, *NLRP5* and *ROSA26* can be seen in Figure 4.3, 4.4. and 4.5, respectively.

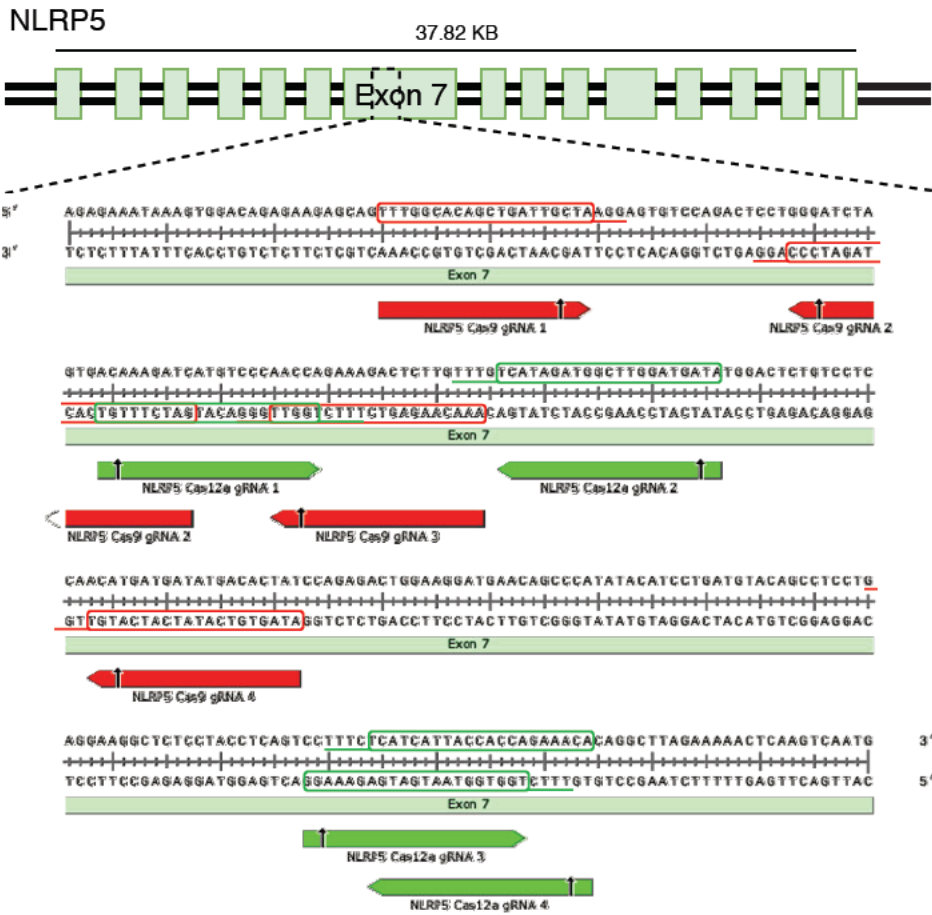


Figure 4.4 SpCas9 (red) and AsCas12a (green) gRNAs targeting exon 7 of *NLRP5*. gRNAs designed using CRISPOR online tool.

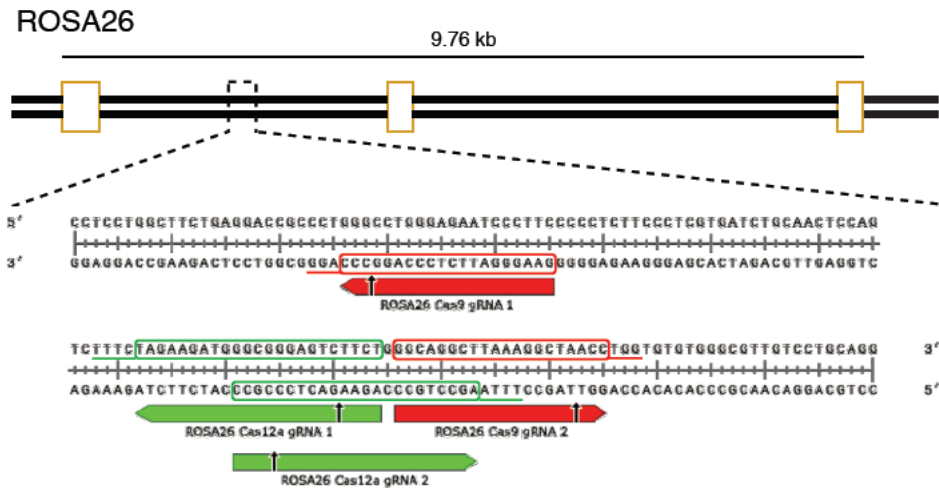


Figure 4.5 SpCas9 (red) and AsCas12a (green) gRNAs targeting intron 1 of the ROSA26 locus. gRNAs designed using CRISPOR online tool.

4.3.2 Primer design and optimisation

In order to *in vitro* screen gRNA cutting efficiency at each of the target sites, primers were designed using Primer3Plus and PrimerBLAST to amplify across the genomic region of interest (Figure 4.6). PCR protocols are in section 2.1.2 and primer sequences can be found in Appendix A.2. Primers were optimised for OneTaq polymerase using gradient PCR to determine the best annealing temperature (Figure 4.7). Robust and specific amplification was seen for all primer pairs at 61.8 and 63.2 °C. For all primer pairs, an annealing temperature of 62°C was selected for further use.

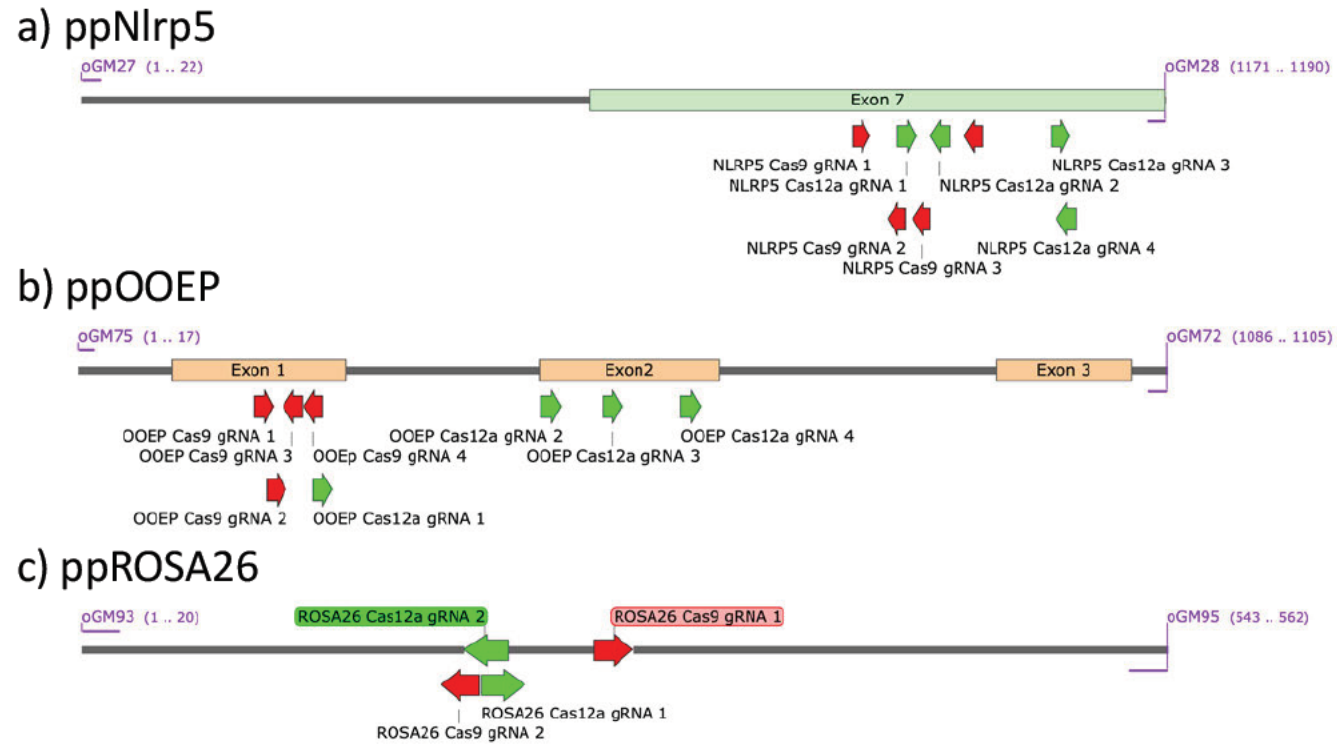


Figure 4.6 Primer pair (pp) locations. All primer pairs were designed using both Primer3Plus and Primer-Blast. (a) ppNLRP5 (oGM27 and oGM28) at *NLRP5*, (b) ppOOEP (oGM75 and oGM72) at *OOEP* and (c) ppROSA26 (oGM93 and oGM95) at *ROSA26* expected amplicon size of 562 bp. PCR primer sequences can be found in Appendix A.2. These primers were used for screening gRNAs via T7EI assay and ICE analysis.

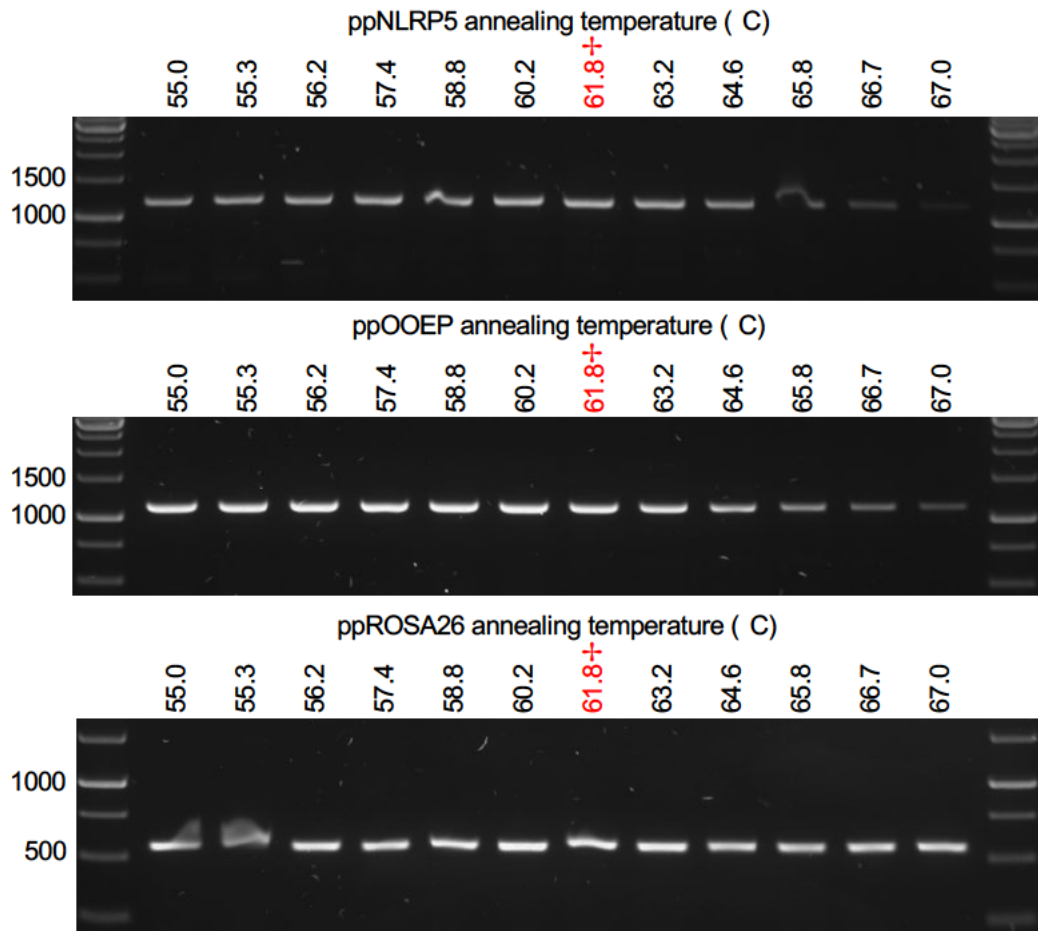


Figure 4.7 Gradient PCR of primer pair (pp) ROSA26, ppOOEP and ppNLRP5. An optimal annealing temperature of 62 °C was selected for all primers.

4.3.3 gRNA screening

The CRISPOR designed gRNAs were ordered as DNA oligos (Appendix A.7), annealed and cloned into CRISPR expression vectors following the protocol in section 2.4.2. pSL70 was used for expressing SpCas9–T2A–eGFP and corresponding gRNA, while pY094 was used for expressing AsCas12–T2A–eGFP and corresponding gRNA (see section 2.7.3 for plasmid maps). gRNAs were screened for efficiency by T7EI assay and ICE analyses in ES cells following the workflow in section 4.2.1.

4.3.3.1 T7 endonuclease I (T7EI) assays

T7EI assays were performed, and gel electrophoresis were visually assessed based on the fraction of nuclease-specific cleavage products. The substrate for each T7EI assay was a PCR product spanning the gRNA target region (see section 4.1.4.1 for further details on T7EI assay). As gRNAs within a gene drive system are required to cut within every generation, it is desirable to have highly efficient gRNAs.

For those gRNAs targeting one of the essential female-fertility genes, either *OOEP* or *NLRP5*, SpCas9 gRNAs cutting at exon 1 of *OOEP* were the most efficient by T7EI assay (Figure 4.8 and 4.9). *NLRP5* SpCas9 gRNA 4 in Figure 4.9, also cut well, however, this project aimed to multiplex gRNAs within the SGD construct and no other *NLRP5* SpCas9 gRNAs demonstrated cutting at acceptable efficiency at the *NLRP5* locus. There was evidence of AsCas12a gRNAs cutting at both the *OOEP* and *NLRP5* locus but in all cases the AsCas12a gRNAs had less nuclease-specific cleavage product than the SpCas9 counterparts. From the T7EI assay, *OOEP* SpCas9 gRNAs were short listed as potential candidates and an ICE analyses were performed to gain further information and help select the best performing three gRNAs at the *OOEP* locus for the SGD gRNA array.

All of the *ROSA26* gRNAs cut with acceptable efficiency. In order to generate a split drive system and knock-in an array of *OOEP* SpCas9 gRNAs into the *ROSA26* locus, whilst preserving the integrity of *OOEP* target site, *ROSA26* AsCas12a gRNAs were selected for further analysis. *ROSA26* AsCas12a gRNAs were also assessed by ICE analysis.

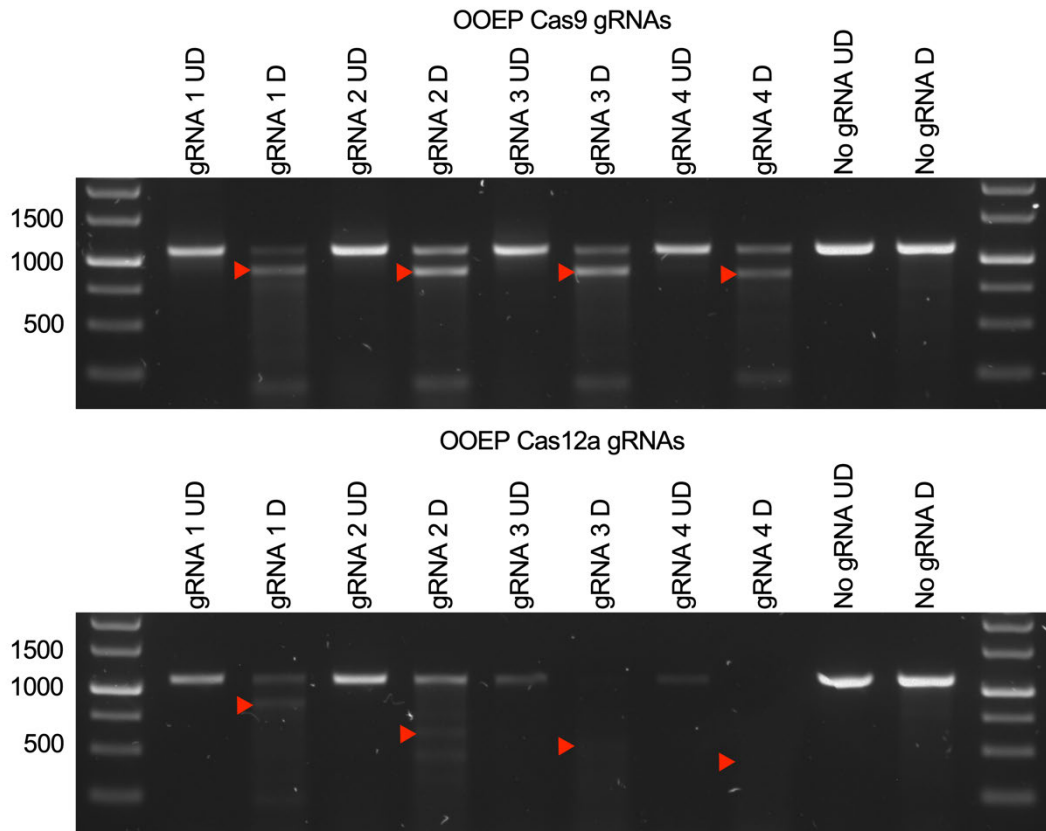


Figure 4.8 T7 endonuclease I (T7EI) assays of SpCas9 (top) and AsCas12a (bottom) gRNAs at the *OOEP* locus. D = T7EI digested PCR product. UD = Undigested PCR product. Red arrows indicate the expected size of nuclease-specific second band.

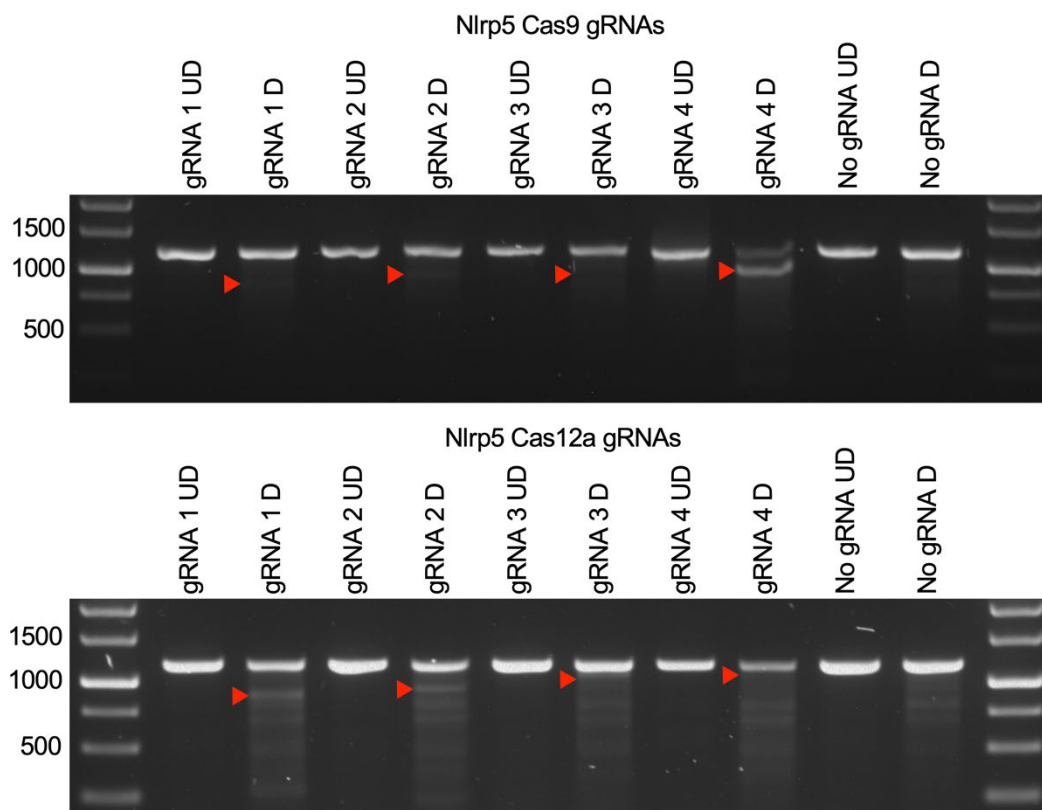


Figure 4.9 T7 endonuclease I (T7EI) assays of SpCas9 (top) and AsCas12a (bottom) at the *NLRP5* locus. D = T7EI digested PCR product. UD = Undigested PCR product. Red arrows indicate the expected size of nuclease-specific second band.

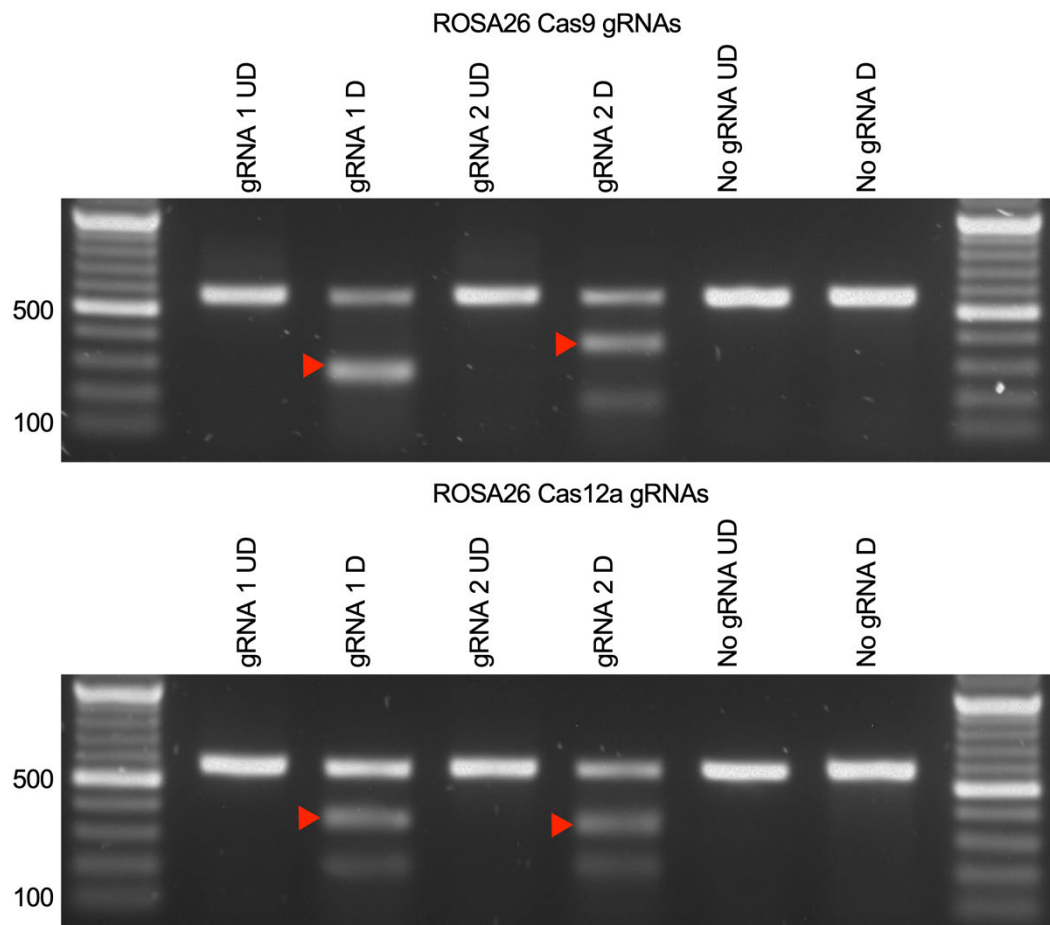


Figure 4.10 T7 endonuclease I (T7EI) assays of SpCas9 (top) and AsCas12a (bottom) at the *ROSA26* locus. D = T7EI digested PCR product. UD = Undigested PCR product. Red arrows indicate the expected size of nuclease-specific second band.

4.3.3.2 Inference of CRISPR edits (ICE) analyses

To perform the ICE analyses, Sanger sequencing of the targeted region was performed on the PCR products of gDNA extracted from the pooled population of CRISPR-Cas transfected cells. As a quality control, the sequence data was checked to verify overlapping of the trace data occurred proximal to the expected cut site, indicating indel mutations at the site within the pooled population (Figures 4.11 and 4.12). A wild-type unedited control sample was also sequenced for each target region.

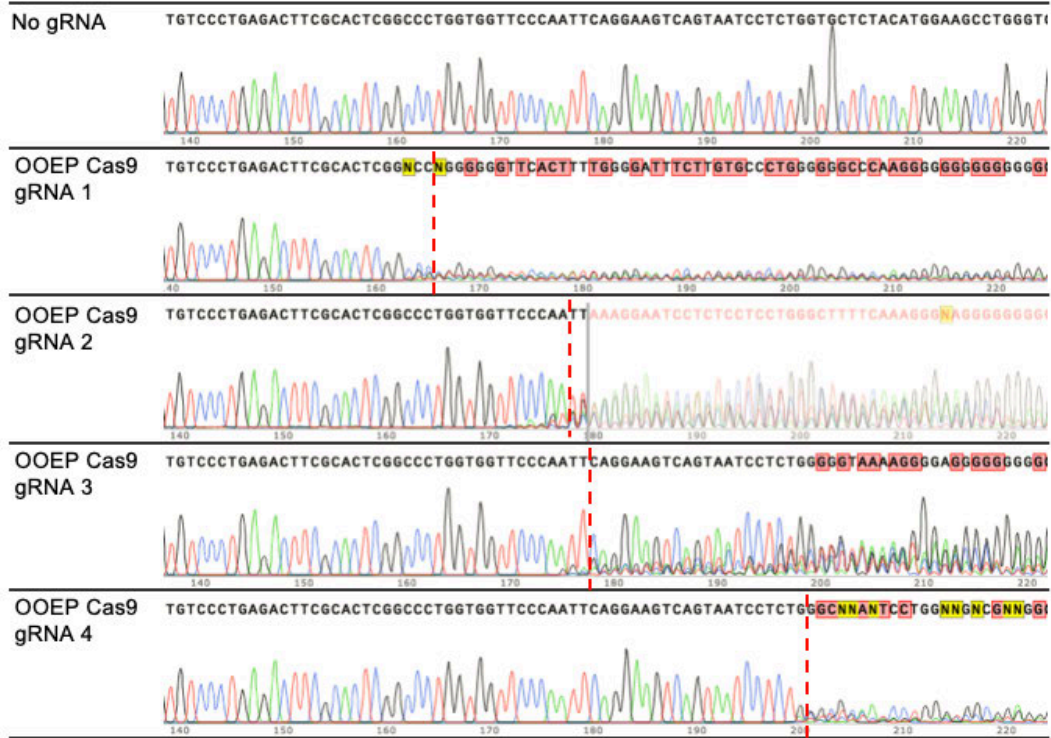


Figure 4.11 Sanger sequence chromatograms from the *OOEP* target region. Regions were PCR amplified from pooled genomic DNA using primer pair ppOOEP. PCR products were sequenced with oGM75. A clear overlapping of multiple traces can be seen proximal to the expected cleavage site (red dotted line). No gRNA is a wild-type unedited control sample.

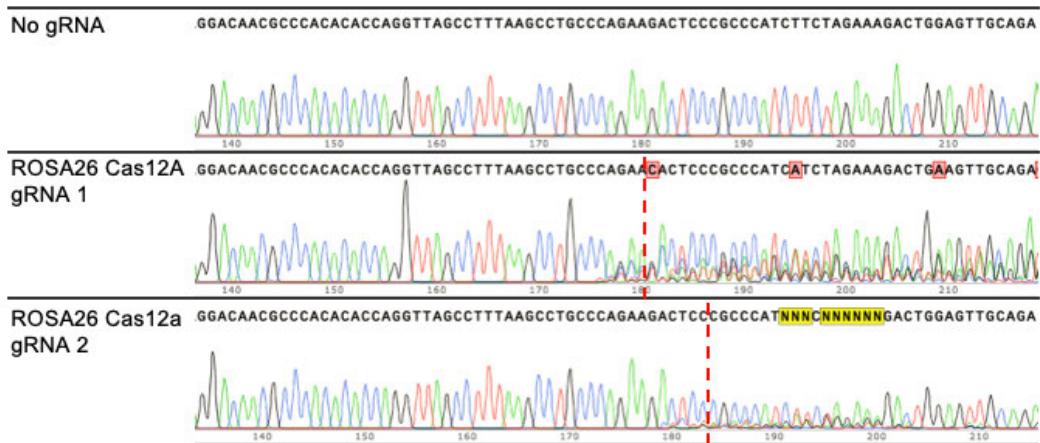


Figure 4.12 Sanger sequence chromatograms from the *ROSA26* target region. Regions were PCR amplified from pooled genomic DNA with primer pair ppROSA26. PCR products were sequenced with oGM95. A clear overlapping of multiple traces can be seen proximal to the expected cleavage site (red dotted line). No gRNA is a wild-type unedited control sample.

Within the ICE software, the Sanger sequencing files of wild-type unedited control and targeted samples were uploaded, along with the gRNA sequences. The ICE software calculates efficiency based on indel frequency and determines the profiles of all the different types of edits that are present and their relative abundances. The profile of edits is derived into a knockout-score, based on the number of edits that lead to an out-of-frame mutation. All four *OOEP* SpCas9 gRNAs has indel efficiency greater than 60% and knock-out scores greater than 50 (Figure 4.13). All *OOEP* SpCas9 gRNAs were acceptable candidates for use in SGD constructs, however due to their location, *OOEP* SpCas9 gRNAs 1, 3 and 4 were selected as the three gRNAs for inclusion in the final SGD constructs. The selection of these gRNAs prevented overlapping gRNA recognition sequences at the *OOEP* target locus. Of the two *AsCas12a* gRNAs targeting the *ROSA26* locus, *ROSA26* *AsCas12a* gRNA 1 had better indel efficiency at 55%, compared gRNA 2 at 27% (Figure 1.4), and was selected as the gRNA for genome engineering the *ROSA26* locus.

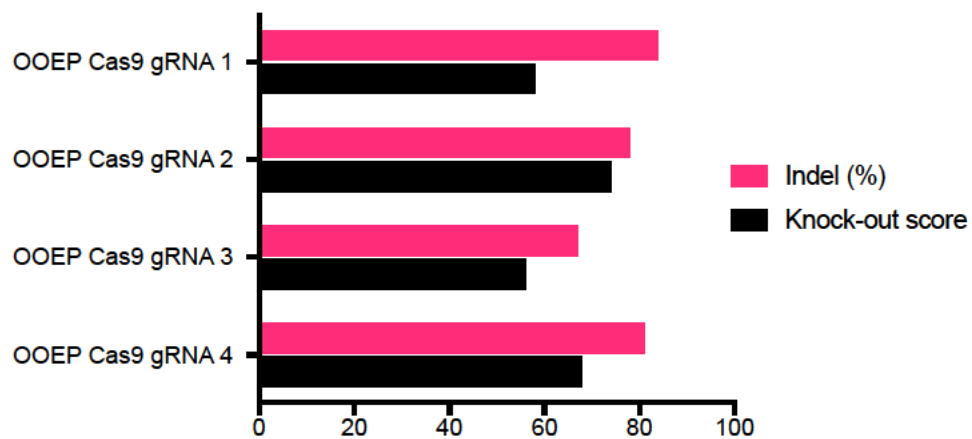


Figure 4.13 Inference of CRISPR Edits (ICE) analysis results of SpCas9 gRNAs targeting *OOEP*.

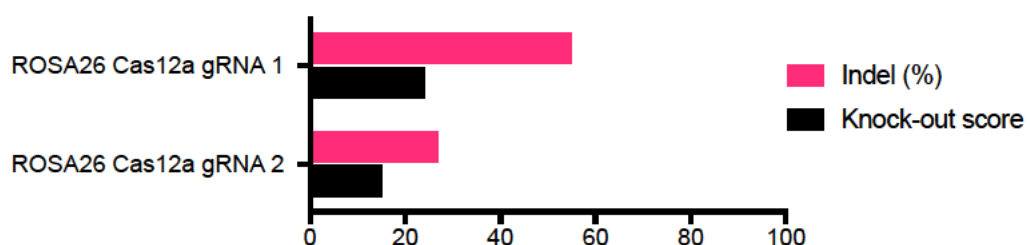


Figure 4.14 Inference of CRISPR Edits (ICE) analysis results of AsCas12a gRNAs targeting *ROSA26*.

In this project, all gRNAs were screened in E14 ES cells, which are derived from an Ola/129 strain mouse (Wakayama et al., 1999). The ultimate goal of this work is to generate SGD mice and it is the ambition to breed these SGD mice with C57Bl6NCrI (Bl6) mice. To ensure the three *OOEP* SpCas9 gRNAs selected have recognition sequences in Bl6 mice, gDNA from this strain was extracted from standard ear clippings. The *OOEP* target site was PCR amplified and sequenced. All three gRNAs have conserved recognition sequences in the Bl6 strain (Figure 4.15), confirming that the gene drive should be compatible in this strain.

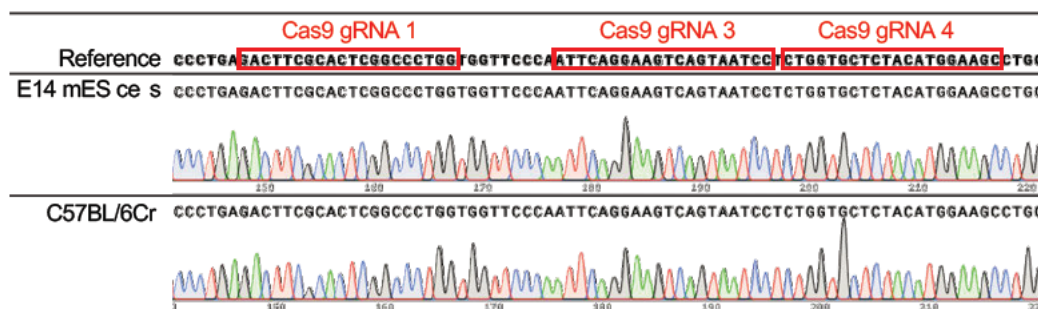


Figure 4.15 Sequence comparison of *OOEP* target site in E14 mouse embryonic stem (ES) cells and C57BL/6NCrI (Bl6) mice. The recognition site for all three of the selected SpCas9 gRNAs is conserved in Bl6 mice.

4.3.4 SpCas9 ribonucleoprotein screening

OOEP SpCas9 gRNAs 1 and 4 were also screened in ribonucleoprotein (RNP) format in mouse ES cells. RNP complex was formed following the protocol in section 2.4.3. Sequences of crRNAs can be found in Appendix A.8. These two peripheral gRNAs would be used in an RNP complex to integrate the SGD constructs into the *OOEP* locus. RNP would be used, instead of plasmid delivered CRISPR-SpCas9, to limit the likelihood of random DNA plasmid integration when genome engineering ES cells. *OOEP* SpCas9 RNP with gRNA 1 and gRNA 4 were transfected separately and also co-transfected before screening RNP cutting efficiency in transfected ES cells by T7EI assay.

When screening gRNAs in ES cells with plasmid-based CRISPR-Cas expression systems, cells underwent FACS for eGFP expression to isolate transfected cells prior to performing the T7EI assay. With no fluorophore to sort for following RNP transfection, it was not possible to FACS transfected cells prior to performing a T7EI assay. Therefore,

no inferences on the comparative efficiency of plasmid vs RNP delivered CRISPR-SpCas9 can be made here. RNP versions of *OOEP* SpCas9 gRNA 1 and 4 when transfected separately and co-transfected can mediate cleavage.

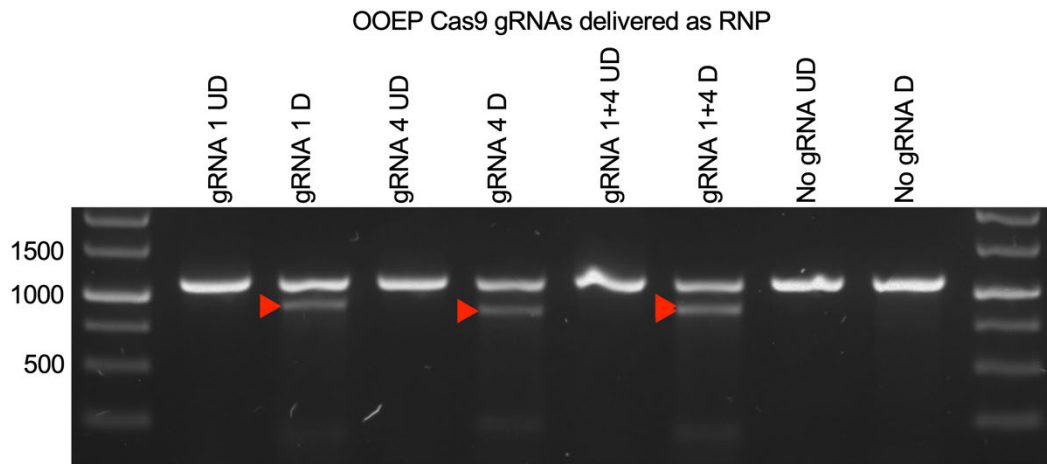


Figure 4.16 T7 endonuclease I (T7EI) assays of *OOEP* SpCas9 gRNAs 1 and 4 in ribonucleoprotein (RNP) form. D = T7EI digested PCR product. UD = Undigested PCR product. Red arrows indicate the expected size of nuclease-specific second band.

4.3.5 Construct development

After identifying target genes and gRNAs, plasmid SGD constructs were developed using the molecular cloning protocols described in section 2.2. Four plasmids were constructed in the following order: (1) *OOEP_IVT*; (2) *OOEP_CAG*; (3) *OOEP_VASA* and (4) *ROSA26_gRNAs*. Detailed workflows for construction of each plasmid vector can be seen in expandable construction histories stored on the Edinburgh data share at <https://doi.org/10.7488/ds/2744>. A list of primers, gene fragments (gBlocks) and DNA oligos used for cloning these vectors can be found in Appendices A.2 and A.6. A description of the elements contained in all plasmid vectors generated can be found in Table 4.1. A map and description of each vector, along with the validation data is provided on the following pages.

OOEP_IVT was developed to *in vitro* test interhomolog repair efficiency of gene drives (section 3.3.6). The construct (Figure 4.17a) has a Cre inducible SpCas9-T2A-eGFP. In the absence of Cre, SpCas9-T2A-eGFP is inhibited by a floxed dsRED-STOP element upstream. dsRED is expressed by the CAG promoter. With the addition of Cre, the floxed

dsRED–STOP is excised, and the CAG promoter expresses the SpCas9–T2A–eGFP. OOEP_IVT was developed by Gibson assembly of six DNA fragments in a single reaction. A subsequent Gibson assembly reaction was used to add a bGHpA regulatory element at the AgeI restriction site.

OOEP_IVT was validated by RE analysis with EcoRI+AhdI, XbaI+BspEI and ApaLI (Figure 4.17b). The plasmid was then sequenced (excluding backbone) and three SNPs identified. Two SNPs in the 3' end of the CAG promoter (Figure 4.17c). This region is very GC rich and the identified SNPs could be a result of sequencing error or replication error. *In vitro* testing confirmed the CAG promoter was functional (Figure 4.18).

The OOEP_IVT construct was developed and validated; however, it was not used for its intended purpose within this project due to unforeseen time constraints.

Table 4.1 List of genetic elements, their origin and function used in all targeting constructs developed in this project.

Element	Origin	Function
Amp promoter	pBR322 plasmid	Natural promoter (P3) for Amp R (beta-lactamase gene).
Amp R	pBR322 plasmid	Beta-lactamase gene confers ampicillin resistance to <i>E. coli</i> during plasmid amplification.
bGH PolyA	Bovine	Enhances gene expression by increasing polyadenylation and termination.
CAG	CAG promoter consists of three elements: (i) Cytomegalovirus early enhancer; (ii) chicken beta-actin promoter; and (iii) rabbit beta-globin splice acceptor.	Strong synthetic promoter commonly used to drive constitutive gene expression in mammalian cells.
dsRED	<i>Discosoma sp.</i>	Red fluorescent protein.
eGFP	<i>Aequorea Victoria</i>	Synthetically enhanced version of green fluorescent protein.
gRNA Scaffold	<i>Streptococcus pyogenes</i>	Provides the trans-activating CRISPR RNA molecule (tracrRNA) of the SpCas9 gRNA.
HA (5' or 3')	Mouse	PCR amplified homology arms for HDR at genomic target sites.
hU6	Human	Human RNA polymerase III promoter for gRNA expression.
LoxP	Bacteriophage P1	Recognition sequence for Cre recombinase.
MCS	Synthetic	Multiple cloning site.
Origin	pBR322 p asm d	Origin for plasmid replication in <i>E. coli</i> during amplification.
PAC (PuroR)	<i>Streptomyces alboniger</i>	Encodes PAC gene, which confers puromycin resistance. Used for positive selection.
PGK promoter	Mouse	Phosphoglycerate kinase promoter drives constitutive expression of PAC.
SpCas9	<i>Streptococcus pyogenes</i>	Cas9 (CRISPR associated protein 9) is an RNA-guided DNA endonuclease.
SpCas9 gRNA	<i>Streptococcus pyogenes</i>	A single gRNA that combines crRNA and tracrRNA to guide the SpCas9 endonuclease to the target site.
T2A	<i>Thosea asigna virus</i>	Bicistronic element that facilitates ribosomal skipping, leading to two proteins from one mRNA transcript.
Vasa promoter	Mouse (FVB strain)	Germline-specific expression of SpCas9 and eGFP.

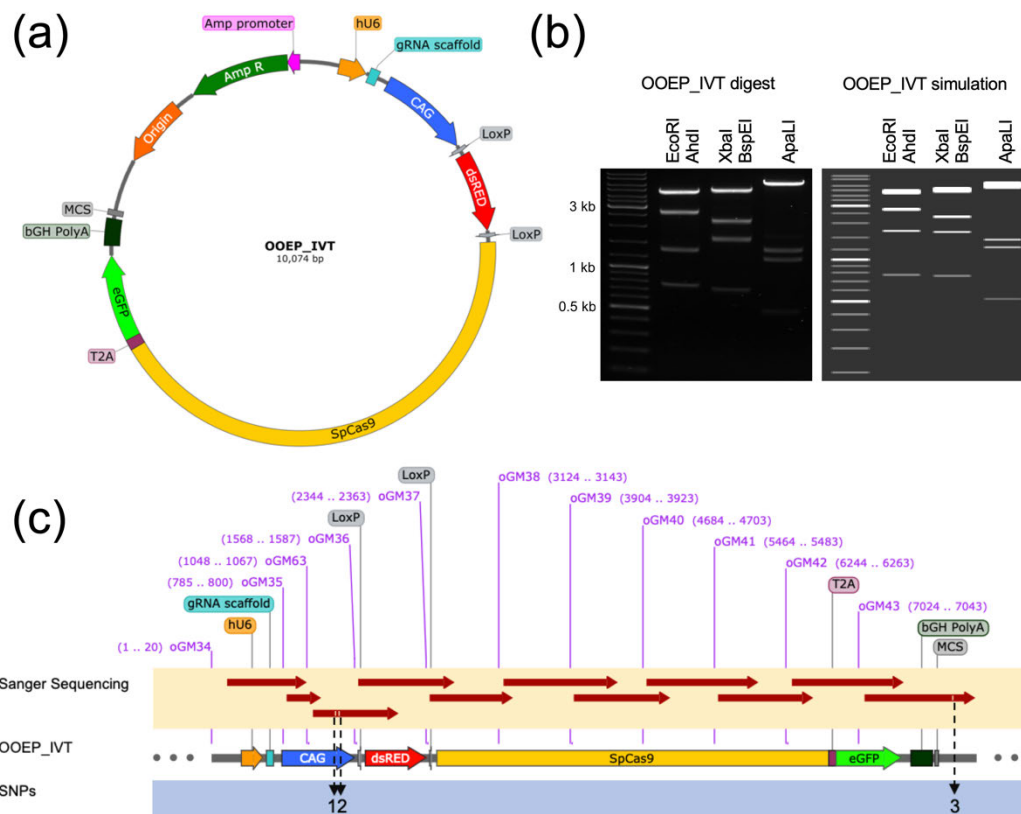


Figure 4.17 Development of OOEP_IVT plasmid. (a) Map of the 10,074 bp plasmid. See <https://doi.org/10.7488/ds/2744> for detailed construct development. (b) Restriction enzyme (RE) analysis of OOEP_IVT. Left gel is plasmid digest. Right is *in silico* simulation of anticipated digest outcome. (c) OOEP_IVT Sanger sequencing traces are presented as red arrows. Primers used for sequencing are shown in purple. The entire insert was sequenced, and three SNPs were identified.

The functionality of the floxed dsRED–stop cassette was tested in HEK cells (following the culturing and transfection protocol in section 2.5.2). This was performed by transfection of OOEP_IVT or OOEP_IVT + Cre-Puro, which is a Cre recombinase expressing plasmid (seen section 2.7.3 for Cre-Puro plasmid map). As can be seen in Figure 4.18, there is a clear transition from dsRED to eGFP expression in the presence of Cre. This confirmed that the floxed allele was working as expected, although some eGFP leakage was evident even when OOEP_IVT was transfected in the absence of Cre.

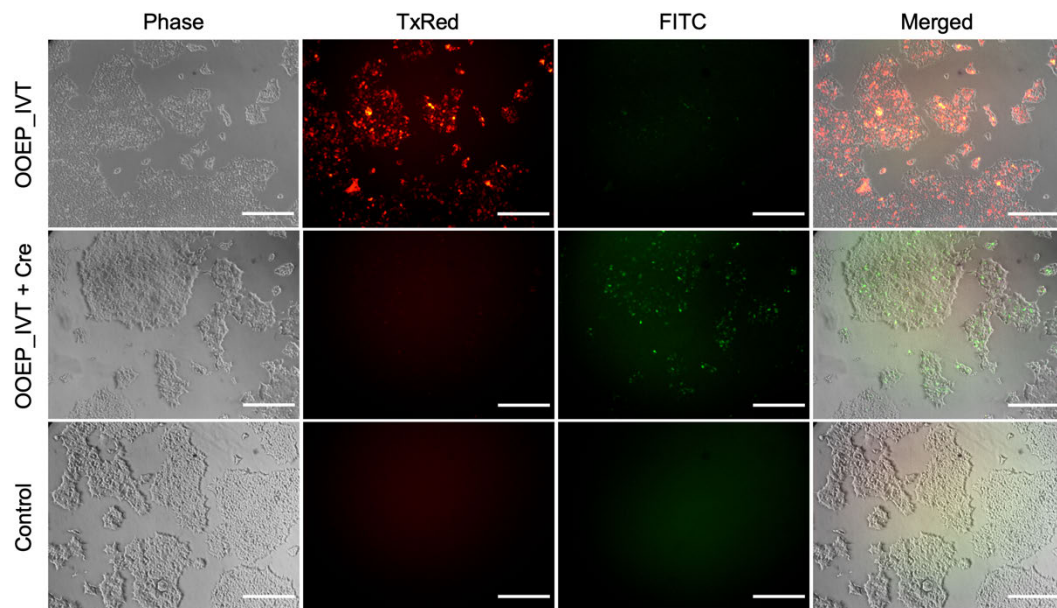


Figure 4.18 Functional testing of OOEP_IVT in HEK cells. Column 1: cells under phase. Column 2: cells under Texas Red (TxRed) filter show dsRed expression. Column 3: cells under fluorescein isothiocyanate (FITC) filter show eGFP expression. Column 4: merged images. OOEP_IVT predominantly express dsRED, although some leakage of eGFP was evident. When OOEP_IVT is co-transfected with a Cre-recombinase expressing Cre-Puro plasmid the floxed dsRED–STOP cassette is excised and cells transition to eGFP expression. Cells imaged at 24 hours post transfection. Scale bar = 500 μ m.

OOEP_CAG was developed by manipulating OOEP_IVT through four separate cloning steps. This plasmid has 800 bp OOEP homology arms flanking a CAG promoter for constitutive expression of SpCas9–T2A–eGFP (Figure 4.19a). OOEP_IVT also contains a floxed PGK–PAC immediately upstream of the CAG promoter. To construct OOEP_CAG, firstly, the KpnI and FseI sites were used to exchange the CAG–LoxP–dsRED–LoxP–SpCas9 element in OOEP_IVT for the CAG–SpCas9 element from pSL70 (plasmid map in section 2.7.3). 800 bp 3' and 5' homology arms were PCR amplified from wild-type E14 ES cell gDNA and inserted by sequential Gibson assembly reactions into BamHI and NheI restriction sites, respectively. The alignment of the homology arms relative to the OOEP SpCas9 gRNAs can be seen in Figure 4.20. Finally, a floxed PGK–PAC cassette was added by Gibson assembly into the KpnI site. OOEP_CAG was validated by RE analysis with SacI, MspA1I and AatII (Figure 4.19b). Cloning junctions and PCR amplified DNA fragments were Sanger sequenced, and no SNPs were identified (Figure 4.19c). The plasmid was then transfected into ES cells and eGFP expression was verified (Figure 4.22a).

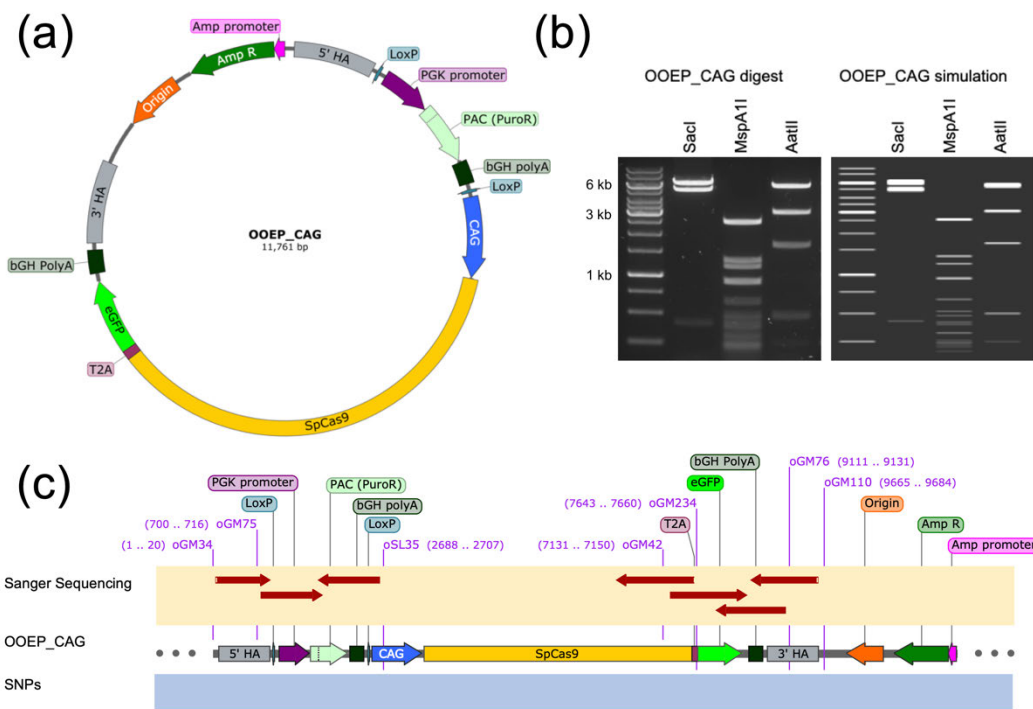


Figure 4.19 Development of OOEP_CAG plasmid. (a) Map of the 11,761 bp plasmid. See <https://doi.org/10.7488/ds/2744> for detailed construct development. (b) Restriction enzyme (RE) analysis of OOEP_CAG. Left gel is plasmid digest. Right is *in silico* simulation of anticipated digest outcome. (c) OOEP_CAG Sanger sequencing traces are presented as red arrows. Primers used for sequencing are shown in purple. Sequencing across all new inserts and junctions was performed and no SNPs were identified.

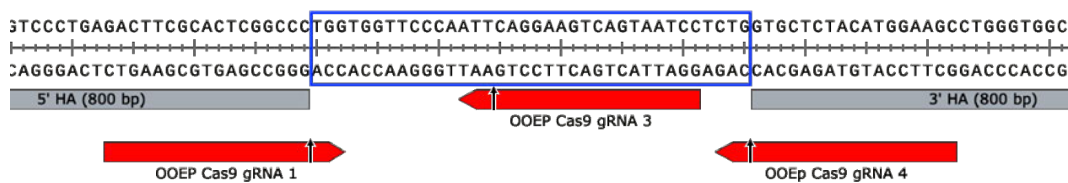


Figure 4.20 Alignment of the homology arms and gRNAs within OOEP_CAG and OOEP_VASA plasmids. The alignment is relative to the three selected OOEP SpCas9 gRNA recognition sites for use in the split gene drive strategies. The blue outlined region represents the section of DNA that will be replaced with the split gene drive cassette.

OOEP_VASA was assembled by manipulation of OOEP_CAG (see Figure 4.21a for plasmid map). Firstly, oGM237 and oGM238 were annealed and cloned into XbaI and NcoI restriction sites to replace the CAG promoter and introduce Ascl and PacI restriction sites. The VASA promoter from pVASA-Cre was then added into the Ascl and PacI

restriction sites (see section 2.7.3 for pVASA-Cre plasmid map). OOEP_VASA was validated by RE analysis with MfeI, NspI and Apal+XbaI (Figure 4.21b). Cloning junctions in OOEP_VASA were sequenced, and one SNP was identified. This was not unexpected as the VASA promoter sequence provided by AddGene for the pVASA-Cre plasmid is based upon a prediction and not actual sequence data generated from the plasmid. OOEP_Vasa was then transfected into ES cells and eGFP expression was observed (Figure 4.22b).

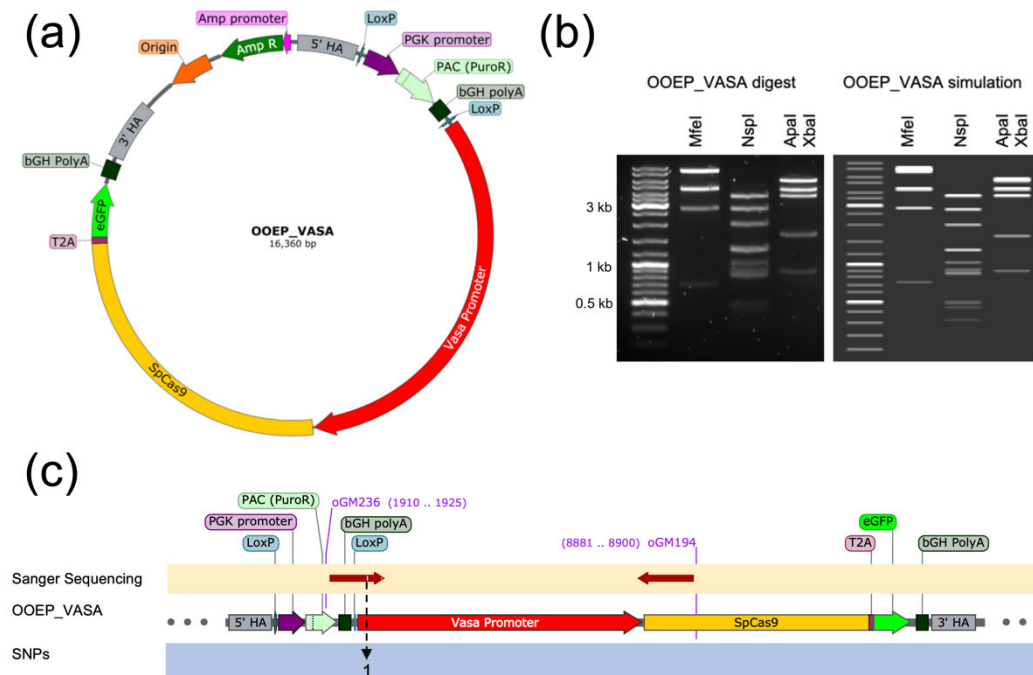


Figure 4.21 Development of OOEP_VASA plasmid. (a) Map of the 16,360 bp plasmid. See <https://doi.org/10.7488/ds/2744> for detailed construct development. (b) Restriction enzyme (RE) analysis of OOEP_VASA. Left gel is plasmid digest. Right is *in silico* simulation of anticipated digest outcome. (c) OOEP_VASA Sanger sequencing traces are presented as red arrows. Primers used for sequencing are shown in purple. Sequencing across new junctions was performed and 1 SNP was identified.

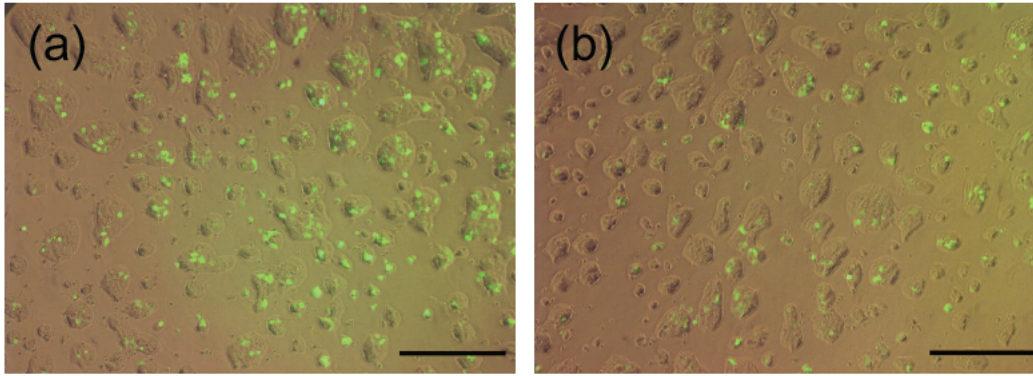


Figure 4.22 Functional testing of OOEP_CAG and OOEP_VASA plasmids in E14 mouse embryonic stem (ES) cells. Images are merged from phase contrast and fluorescein isothiocyanate (FITC) filter images taken at 24 hours post transfection. (a) Cells transfected with OOEP_CAG. (b) Cells transfected with OOEP_VASA. Scale bar = 500um.

Before commencing cloning of *ROSA26* targeting vectors, the homology arms in pDonor-ROSA26 were compared to the wild-type sequence of E14 ES cells to ensure the homology around the cut site is correct. Sequence alignment of the PCR region with the plasmid reveals there is at least 328 bp of 5' homology and 216 bp of 3' homology exists.

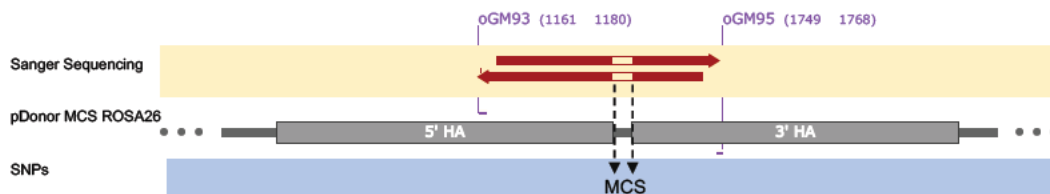


Figure 4.23 Sanger sequencing traces from E14 mouse embryonic stem (ES) cells (red arrows) aligned to the homology arms in pDonor-ROSA26. Primer pair (pp)ROSA26 were used to PCR and Sanger sequence the *ROSA26* region in ES cells. No unexpected SNPs were identified within the aligned region.

ROSA26_gRNAs plasmid was cloned through six sequential cloning steps (see Figure 4.24a for plasmid map). Firstly, pDonor-ROSA26 was digested with *Xho*I and a PCR product using oGM64 and oGM65 primers to amplify the U6-gRNA scaffold from gBlock 1 (Appendix A.6) was added into this site by Gibson assembly. OOEP SpCas9 gRNA 1 was formed by annealing oGM19 and oGM20 and cloning into the *Bsm*BI restriction sites.

U6-OOEP SpCas9 gRNA 3 was PCR amplified from pSL70 with gRNA 3 using primers oGM121 and oGM122 and added via Gibson assembly into the PmeI site. U6-OOEP SpCas9 gRNA 4 was then PCR amplified from pSL70 with gRNA4 using primers oGM127 and oGM128 and added via Gibson assembly into the BamHI site. The 3' homology arm was then adjusted to ensure the ROSA26 AsCas12a gRNA 1 would not cut this arm. Homology arm adjustment was achieved by annealing oGM147 and oGM148 and cloning into the SbfI and Sall restriction sites. The adjusted homology relative to gRNA recognition sequences can be seen in Figure 4.25. Finally, a floxed PGK-PAC was added by Gibson assembly into the NheI site. ROSA26_gRNAs was validated by RE analysis with XbaI+DraIII, NdeI and BsiWI+NheI (Figure 4.24b), before Sanger sequencing the entire insert (Figure 4.24c). One SNP in an intergenic region was identified and no further action was taken.

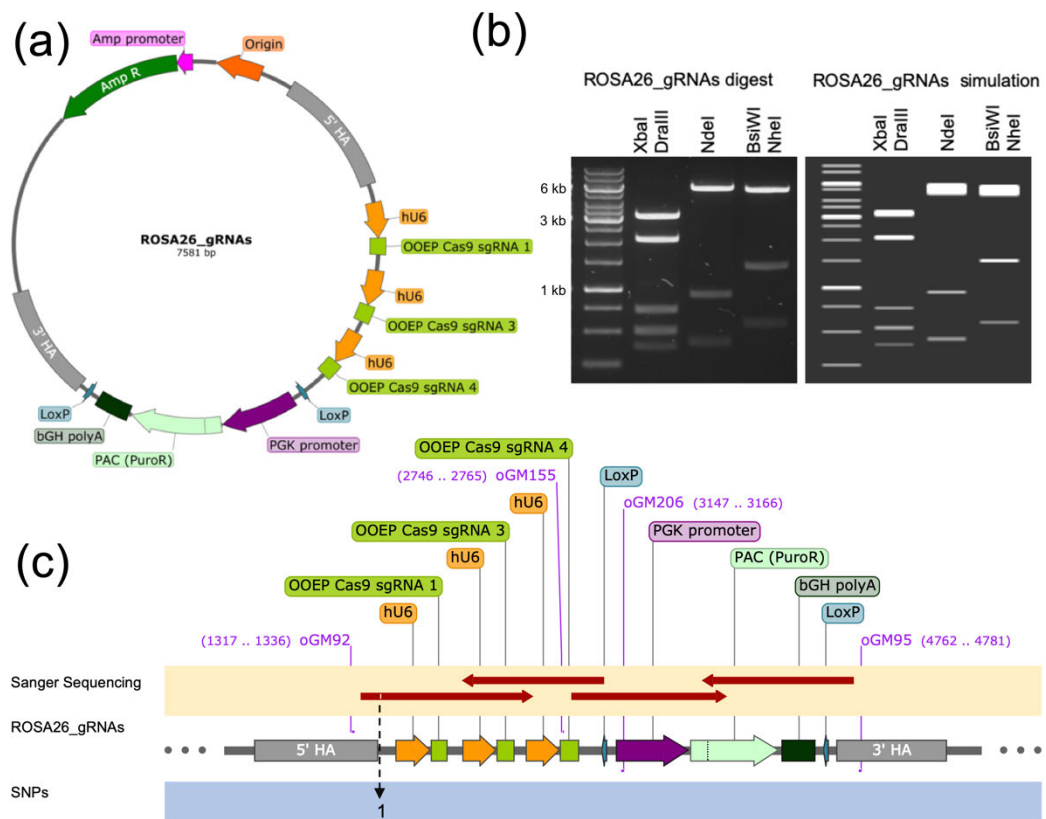


Figure 4.24 Development of ROSA26_gRNAs plasmid. (a) Map of the 7,581 bp plasmid. See <https://doi.org/10.7488/ds/2744> for detailed construct development. (b) Restriction enzyme (RE) analysis of ROSA26_gRNAs. Left gel is plasmid digest. Right is *in silico* simulation of anticipated digest outcome. (c) ROSA26_gRNAs Sanger sequencing traces are presented as red arrows. Primers used for sequencing are shown in purple. Sequencing the entire insert identified 1 intergenic SNP.

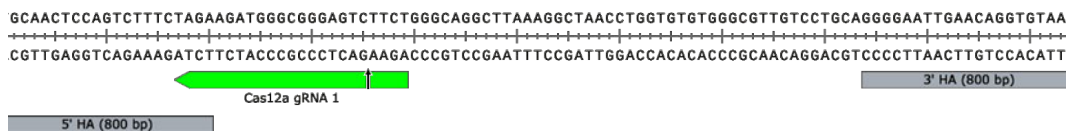


Figure 4.25 Alignment of the gRNA and adjusted homology arms within ROSA26_gRNAs plasmid. The alignment is relative to the selected ROSA26 AsCas12a gRNA 1 recognition site for integrating the gRNA array for the split gene drive systems.

4.4 Discussion

This chapter developed and validated the reagents to engineer the two SGD strategies identified in Chapter 3.

4.4.1 Designing gRNAs

A significant portion of this project was dedicated to gRNA selection. gRNAs for both SpCas9 and AsCas12a were designed against *OOEP*, *NLRP5* and *ROSA26* in the mouse reference genome (GRCm38/mm10) using CRISPOR *in silico* tool.

When designing SpCas9 gRNAs using CRISPOR the software provides a MIT specificity score, predicted efficiency score, out-of-frame score and off-target mismatches in the mouse reference genome. For the MIT specificity score (Hsu et al., 2013), scores range from 0–100 with the higher the score, the lower the probability that off-target cleavage will occur genome. All SpCas9 gRNAs were above the recommended score of 50, with the lowest scoring gRNA being *NLRP5* SpCas9 gRNA 3, at 55. The threshold of 50 or above for specificity score is based on the data from whole-genome off-target assays published by Haeussler et al. (2016). There is currently no similar specificity score available for AsCas12a but CRISPOR does provide a list of likely off-targets in the genome for both SpCas9 and AsCas12a.

CRISPOR also identifies possible off-targets for both SpCas9 and AsCas12a gRNAs in the mouse genome and ranks these by number of mismatches. This CRISPOR output is a summary of a whole-genome search for sequences similar to the crRNA sequence. Mismatched off-targets are considered if they are flanked by an NGG, NAG or NGA motif. When selecting gRNAs in this study, we consider mismatches of 0, 1 or 2 nucleotides in the recognition sequence. Or in other words, the number of locations in the genome with

no mismatches in the crRNA sequence, number of locations in the genome with 1 mismatch, and the number of locations with 2 mismatches. The threshold applied in this study required that guides have no identical recognition sequences (other than the intended target site) and a maximum of 1 location in the genome with 1 nt mismatch, providing the potential mismatch was within 12 nt “seed” region closest to the PAM. It has been demonstrated using *in vitro* cleavage assays that off-targets with a mismatch in the seed region are very inefficiently (Cho et al., 2014).

CRISPOR also provides a predicted efficiency for each SpCas9 and AsCas12a gRNAs. SpCas9 gRNA efficiency is based on the Doench et al. (2016) algorithm, which was developed from a database of 881 gRNAs in MOLM13/NB4/TF1 cells. While the AsCas12a efficiency score is based upon work by Kim et al. (2018), which analysed the indel frequencies for 15,000 target sequences delivered by lentivirus to HEK293T cells. Considering the large size of the training population this prediction efficiency is likely to be more accurate than that of the efficiency predicted with the SpCas9 algorithm. These scores give a prediction of how well a given target may be cut by the designed gRNA complexed with Cas. Scores range from 0–100 with 100 being the best. The most efficient AsCas12a gRNA had a predicted efficiency of just 24, while the least efficient *OOEP* gRNA had a score of 25.

The final predictive parameter for gRNA selection in CRISPOR is the out-of-frame score, which is a prediction of the DNA sequence after DSB repair. This also ranges from 0–100 and predicts the percentage of clones that will carry out-of-frame deletions, based on the micro-homology in the sequence flanking the expected cut site. The higher the out-of-frame score, the higher the probability that MMEJ-mediated deletions will have a length that is not a multiple of three. This score is based on an algorithm by Bae et al. (2014), which was developed for SpCas9. The AsCas12a out-of-frame predictions are also based on the Bae et al. 2014 algorithm and it remains unknown if this is useful as SpCas9 makes a blunt end cut, and therefore the out-of-frame scores may not be accurate for the staggered cut of AsCas12a. There is need for increased data to develop better tools for designing AsCas12a gRNAs.

The out-of-frame scores CRISPOR provides are valuable when aiming to generate a gene knock-out with a single gRNA but when multiplexing gRNAs, as is the case in this project, this is likely to have little predictive power on the final outcome. While assessing as much *in silico* information is valuable before *in vitro* testing, the quality of the data from these algorithms remains questionable. There are obvious benefits in checking specificity score or mismatch scores of candidate gRNAs against reference genomes; however, certainly

within this study the efficiency scores and out-of-frame scores from CRISPOR were not in accordance with the results of the T7EI assay and the ICE analyses.

4.4.2 Screening gRNAs

To *in vitro* screen the CRISPOR designed gRNAs at their respective target sites, primers were designed using Primer3Plus and PrimerBLAST to PCR amplify the targeted regions. Collectively, Primer3Plus and PrimerBLAST offer a powerful tool for designing robust primers. Primer3Plus provides an intuitive interface for Primer3, which takes into account oligonucleotide melting temperature, size, GC content, and primer-dimer possibilities, PCR product size and positional constraints within the template sequence when selecting primers (Untergasser et al., 2007). While PrimerBLAST, performs a specificity check against the reference genome when a target template and both primers are provided (Ye et al., 2012). Used together, good primers can be identified, as was the case in this project. It can be seen in Figure 4.7, that there is no evidence non-specific priming for *OOEP*, *NLRP5* or *ROSA26* in the gradient PCRs between 55 to 67 °C, although the amount of amplicon does vary with differing annealing temperatures. An integrated software package that combines Primer3Plus with PrimerBLAST or a similar tool for checking primer specificity against reference genomes would make a powerful and streamlined primer design tool.

Having designed the gRNAs and primers, guides were screened via two methods; T7EI assay and ICE analysis. The T7EI assay, offers a relatively cheap method for quickly screening gRNAs. It is not quantitative, although a rough estimate of cutting efficiency can be calculated based on the intensity of the cut band to uncut band. This was not performed in this project as the value of this calculation is questionable and of little use if an ICE analysis is later undertaken to quantify indel efficiency. Another issue with assessing cutting efficiency by T7EI is that it is dependent on the diversity of indels within the targeted population of cells. A high diversity of indels will lead to more heteroduplex formation upon denaturing and reannealing and therefore a more intense second band than would be predicted for low indel diversity, where more homoduplex formation would result in a less intense second band. This dependence on indel diversity could mean that T7EI may miss an efficient cutting gRNA that has flanking microhomologies that promotes MMEJ-mediated repair to the same indel in the majority of targeted cells. T7EI finds its value in quickly shortlisting putative gRNAs for further downstream analysis.

As an alternative to T7EI, *in vitro* Cas-RNP digestion of PCR amplified target regions has recently become popular as a tool for screening gRNAs (Mehravar, Shirazi, Mehrazar, et

al., 2019). While providing a very clear result, this approach fails to account for the impact of chromatin and nuclease accessibility to cut the target DNA in the cell type of interest. *In vitro* Cas-RNP digestion of PCR amplified target regions can be used to quickly exclude gRNAs that offer little or no Cas-mediated targeted DNA cleavage. This may be useful if the cell type used for *in vitro* gRNA screening has little biological relevance to the final targeted cell type. For example, screening gRNAs in an immortalised cancer cell line may not correlate with cutting efficiency of the same gRNAs in zygotes.

Of those gRNAs targeting essential female-fertility genes in this project, *OOEP* SpCas9 gRNAs were more efficient than all others when assessed by T7EI assay. All four *OOEP* SpCas9 gRNAs generated indels with good efficiency, displaying a strong nuclease-specific cleavage product. There was also evidence of indels generated by *OOEP* AsCas12a gRNA 1 and 2, although the signal was not as strong as that of SpCas9 gRNAs targeting the same gene. Unfortunately, *OOEP* AsCas12a gRNAs 3 and 4 needed more PCR product to get an accurate T7EI read out but as *OOEP* SpCas9 gRNAs clearly outperformed both *OOEP* AsCas12a gRNA 1 and 2, the poor quality T7EI assays of *OOEP* AsCas12a gRNAs 3 and 4 were not repeated.

NLRP5 proved less amenable to CRISPR-cleavage. Of the *NLRP5* SpCas9 gRNAs, gRNA 4 was the only efficient cutter. As this project was seeking to multiplex gRNAs, a single *NLRP5* gRNA was not sufficient. All AsCas12a gRNAs at *NLRP5* locus had mild evidence of indel formation but this was minor when compared that that of *OOEP* SpCas9 gRNAs. Based on T7EI assay results it was decided that the *OOEP* SpCas9 gRNAs would move on for further validation by ICE analysis.

The T7EI assay was used to select between SpCas9 and AsCas12a gRNAs in this project, however, direct cutting efficiency comparisons between these nucleases may not be valid with this assay. SpCas9 cleavage generates blunt ends, which in the absence of a DNA repair template is frequently repaired by the error-prone NHEJ pathway; while AsCas12a generates sticky-end overhangs, which may be less prone to the formation of indels as the break could be repaired by the higher fidelity single strand annealing (SSA) pathway. If this were the case, T7EI assays would underestimate the inferred cutting efficiency of AsCas12a gRNAs when compared to SpCas9 gRNAs. The same underestimation would also be seen in ICE analysis. Further investigation is needed but if the 5 bp overhangs from AsCas12a cleavage is sufficient to initiate Rad51-independent SSA, this could be compensated for when screening AsCas12a gRNAs by indel-based approaches to more accurately assess the cutting efficiency of AsCas12a gRNAs.

In parallel to screening the gRNAs targeting *OOEP* and *NLRP5*, SpCas9 and AsCas12a gRNAs were screened against the *ROSA26* locus, intended as the safe harbour for expressing the gRNA array for the SGD strategies. Both AsCas12a and SpCas9 gRNAs cut *ROSA26* with good efficiency. It is worth noting that CRISPOR offers two gRNA lengths for AsCas12a, 21 nt or 23 nt. IDT recommends 21 nt based on their inhouse testing of AsCas12a. Within this study, *OOEP* and *NLRP5* AsCas12a gRNAs were designed to be 21 nt, while the two *ROSA26* AsCas12a gRNAs were designed at 23 nt. Although no conclusions can be made here due to the potential for loci-dependent differences and the insufficient sample size, the 23nt gRNAs all cut with good efficiency, while no 21 nt AsCas12a gRNA targeting *OOEP* or *NLRP5* cut with an acceptable level of efficiency.

From the T7EI assays performed at *ROSA26*, AsCas12a gRNAs were selected for further screening. The selection of AsCas12a gRNAs was based on the need to preserve the *OOEP* locus when knocking-in an array of *OOEP* targeting SpCas9 gRNAs into the *ROSA26* locus.

ICE analyses were performed on the four *OOEP* SpCas9 gRNAs and the two *ROSA26* AsCas12a gRNAs. These assays gave further insight into the cutting efficiency. All four *OOEP* SpCas9 gRNAs had indel efficiencies greater than 60% and knock-out scores greater than 50. Based on the location of the gRNAs and to avoid overlap of the recognition sequences within the genome, *OOEP* SpCas9 gRNAs 1, 3 and 4 were selected for the SGD design. These gRNA recognition sites were also present in the B16 mouse genome and, therefore, the SGD would be compatible with this strain for laboratory contained breeding. Of the two AsCas12a *ROSA26* gRNAs assessed by ICE analysis, gRNA 1 was the most efficient cutter and would used for integrating the *OOEP* SpCas9 gRNA array into this safe harbour locus.

At the completion of gRNA screening, it was determined that the SGD would target the *OOEP* essential female-fertility gene, using SpCas9 with gRNAs 1, 3 and 4. *ROSA26* would be engineered with the *ROSA26* AsCas12a gRNA 1.

T7EI and ICE analyses were able to effectively screen gRNAs in this project, however, the gold standard for screening gRNAs involves targeted next-generation sequencing (NGS) to perform deep sequencing of the targeted region. Targeted NGS is an extremely sensitive method of detecting editing outcomes, and the high-throughput sequence-based data provides a comprehensive view of the indels generated. This level of detailed data analysis does not come cheap. At present, NGS is costly and time consuming to analyse

(Güell, Yang, & Church, 2014; Zischewski, Fischer, & Bortesi, 2017). In the future, it is likely that the cost will be reduced and simplified analysis pipelines with graphical interfaces will be developed, making it a more accessible method for genome engineers. Furthermore, an abundance of NGS data of CRISPR edits would have the added benefit of providing more deep sequencing data to improve the accuracy of gRNA design tools, such as CRISPOR.

4.4.3 Benefits of Cas ribonucleoprotein

gRNAs in this project were screened by lipofection of plasmids expressing both the Cas nuclease and corresponding gRNA into ES cells. Following transfection, integration of the dsDNA molecules into the host genome is always a possibility, and this increases in the presence of DSBs. This issue was recently highlighted when the US Food and Drug Administration found additional plasmid sequence proximal to the target site in genome edited hornless bulls (Norris et al., 2020). It is therefore desirable to limit the amount of dsDNA entering cells when undertaking genome engineering projects. When utilising CRISPR-Cas a common method of reducing the amount of additional dsDNA delivered into a cell is to use Cas nuclease in the form of mRNA or protein along with the gRNA, instead of expressing from a plasmid. With mRNA, Cas expression decreases as RNA is degraded within the cell. With Cas protein, a complex between the nuclease and gRNA is formed, called ribonucleoprotein (RNP) which can then be introduced into the cell. RNP has instant activity, with the protein degrading relatively shortly after transfection. Reports in human cells indicate SpCas9 RNP is substantially degraded within 24 hours following transfection (Dodsworth et al., 2020).

Within this project, it was decided that SpCas9 RNP would be used for CRISPR-assisted knock-in of both constructs targeting the *OOEP* locus, limiting the amount of dsDNA entering engineered cells to just the targeting construct (*OOEP_CAG* or *OOEP_VASA*), reducing the likelihood of undesirable random plasmid integration. For genome engineering the *OOEP* locus, the two peripheral gRNAs (*OOEP* SpCas9 gRNA 1 and 4) were screened by T7EI assay to confirm CRISPR-mediated cleavage could be achieved using these gRNAs in RNP format. *OOEP* SpCas9 gRNAs 1 and 4 had sufficient cutting activity as RNP. Due to cost, AsCas12a was not available in RNP format for this project and would be delivered as a traditional plasmid expression system when engineering *ROSA26*.

4.4.4 Cloning

Whilst screening gRNAs and determining the exact target region, OOEP_IVT was cloned and functionally tested in HEK cells. This plasmid was developed to *in vitro* test interhomolog repair efficiency of gene drive systems. The construct in OOEP_IVT utilises a ‘traffic light’ Cre-inducible system. It was cloned by stitching together six fragments in a single Gibson Assembly reaction, followed by an additional Gibson Assembly reaction to add a polyadenylation signal. Gibson Assembly is undoubtedly a powerful tool for plasmid cloning and this is clearly demonstrated here by its ability to yield a 9789 bp plasmid from six separate fragments in a single reaction. Gibson assembly, in conjunction with the reduced cost and improvements in speed of commercial DNA synthesis now makes it possible to generate almost any conceivable plasmid design in a standard molecular biology laboratory.

The OOEP_IVT construct was transfected into HEK cells and it proved capable switching from dsRED to SpCas9–T2A–eGFP expression in the presence of Cre recombinase. However, there was evidence of leaky transient expression of Cas9–T2A–eGFP in the absence of Cre. The functionality of the plasmid was not tested in ES cells. One potential solution to prevent leaking of SpCas9–eGFP in the absence of Cre is to introduce a polyadenylation sequence, such as bGHpA or SV40 immediately downstream of the dsRED within the floxed cassette. Although this system shows promise, it was decided to not develop this OOEP_IVT construct further and concentrate solely on the development of the *in vivo* based SGD strategies. There are three main reasons for this decision: (1) the potential data gained may not be biologically comparable to a gene drive expressed in germ cells or zygotes; (2) it is likely to be difficult and time consuming to isolate a heterozygous cell line with an intact wild-type allele on the homologous chromosome; (3) leaky expression of the SpCas9–T2A–eGFP is likely to compromise the results, and would therefore need to be rectified before continuing. Although this *in vitro* system was not further developed in this project, it has promise as a tool for quantifying interhomolog repair efficiency, and if developed further could find application for evaluating HDR enhancing and NHEJ inhibiting molecules.

At the completion of gRNA screening, cloning of OOEP_CAG commenced by modifying OOEP_IVT through a series of four Gibson assembly cloning reactions, which included adding 800 bp homology arms to the construct. Next, OOEP_VASA was generated by modifying OOEP_CAG, through two restriction enzyme cloning steps.

To develop the array of three *OOEP* gRNAs for integration into *ROSA26*, the *ROSA26_gRNAs* plasmid was cloned using the pDonor-Rosa26 plasmid from Prof Charles Gersbach as the base plasmid. A complicated five step cloning series, involving both Gibson assembly and restriction enzyme cloning was performed. This included removing 63 bp of the 3' homology arm to accommodate the *ROSA26* AsCas12a gRNA 1. This adjustment meant that the homology arms were not immediately flanking the expected cut site but instead the 5' homology arm started 15 bp downstream and the 3' arm was 48 bp from upstream the cut site. This spacing between the cut and initiation of homology arms is likely to reduce HDR efficiency but given that the *ROSA26_gRNAs* construct has a puromycin selection cassette this reduced efficiency should be tolerable.

When designing the *ROSA26_gRNAs* construct to express three SpCas9 gRNAs, consideration was given to using a polycistronic gRNA-tRNA array. This approach uses a single RNA polymerase III promoter to transcribe all gRNAs in one single transcript, with the tRNA sequences being removed by the endogenous tRNA-processing system to generate separate gRNAs (Dong et al., 2017). Using this approach, Dong et al. (2017) were able to express 2 to 6 gRNAs for precise genome editing of one to three histone deacetylase (HDAC) genes in human cells. This approach would limit the likelihood of recombination of the repetitive hU6 promoter sequences. For gene drive designs for real world application, such a system should be considered for improved evolutionary stability. Although the concept of the polycistronic gRNA-tRNA shows promise, to best of my knowledge there is currently no data on its effectiveness *in vivo*, and as robust and reliable expression of each guide was desired, it was decided that the *ROSA26_gRNAs* construct would use a separate hU6 polymerase III promoter for each of the three *OOEP* gRNAs.

All plasmids were validated by RE analysis and Sanger sequencing. For *ROSA26_gRNAs* and the parental *OOEP_IVT* plasmid, the entire construct was Sanger sequenced. When sequencing *OOEP_CAG* and *OOEP_VASA*, which were based on *OOEP_IVT*, only the cloning junctions or PCR amplified fragments were sequenced. Plasmids were also functionally tested when a fluorescent protein was expressed in the construct. All plasmids expressed fluorescent proteins upon transfection. *OOEP_VASA* used the mouse *VASA* promoter to drive expression of SpCas9-T2A-eGFP, which should primarily restrict expression to germ line, however, low levels of expression are reported in mouse ES cells (Gordeeva, Lifantseva, & Khaidukov, 2011). On transient transfection of the *OOEP_VASA* plasmid in mouse ES cells, expression of eGFP was observed. The clear observation of eGFP expression in Figure 4.22 may be explained by a high copy number of the transiently transfected plasmid.

Plasmids are the bread and butter of molecular biology labs, but their development and validation can be laborious and costly (Casini et al., 2015). A number of commercial vector building companies now offer cost effective and efficient plasmid development and validation services. These companies usually employ a library of common genetic elements in conjunction with DNA synthesis technology, so users can design constructs through online tools and get immediate quotes and expected delivery times. These vector building companies are now applying high-throughput NGS technologies for rapid plasmid validation. The high costs of sample preparation for NGS sequencing have prevented most laboratories from using these methods for routine plasmid validation at small-scale, however, it becomes viable when done at scale by commercial plasmid construction services. NGS can sequence the entire plasmids providing scientists with even more information to aid in the reproducibility of scientific research.

4.4.5 Conclusions

This chapter designed, developed and validated the reagents needed to genome engineer two SGD strategies in mouse ES cells. During the process, SpCas9 was selected as the nuclease to target the *OOEP* female-fertility gene in the two SDGs to be developed. Three *OOEP* SpCas9 gRNAs were identified for expression from the *ROSA26* locus by integrating a gRNA array into this locus using a AsCas12a-assisted genome engineering approach. With validated reagents at the ready, it was now possible to engineer these SGD systems in ES cells.

5 Genome engineering split gene drives in mouse embryonic stem cells

5.1 Introduction

There are a number of approaches used to generate genetically modified mice, including ES cell technology (Robertson et al., 1986), ex vivo direct injection of zygotes (Thomas & Capecchi, 1987), electroporation of genome editing reagents into zygotes (Chen et al., 2016) and, most recently, *in situ* electroporation of genome editing reagents into zygotes within the oviduct of female mice (Ohtsuka et al., 2018). Within this project, ES cell technology has been selected as the desired route. ES cell technology remains the 'gold standard' for generating mouse models with complex or large genetic alterations, such as the constructs of a CRISPR-based gene drive (Chen et al., 2016). With this approach genome engineered ES cells harbouring the split gene drive (SGD) constructs will be injected into blastocysts to generate chimeras. Some background on the culturing and maintenance of ES cells is provide below, along with a description of gene targeting approaches in ES cells, and genotyping and screening techniques for assessing genome engineered ES cell.

5.1.1 ES cell technology

The isolation of ES cells has revolutionised biomedical research, becoming an important tool for the study of early development (Gross & Kioussi, 2014). In the mouse, these cells provide the basis for ES cell technology for the generation of genetically altered mice. ES cells are characterised by self-renewal, pluripotency, and the ability to contribute to the tissues of chimeras generated by injection of ES cells into the early embryo (Romito & Cobellis, 2016). Injected ES cells can contribute to the germ cells of the resulting chimeras and germline competent chimeras can be bred to produce mice with the ES cell-derived genotype (Bradley et al., 1984). Mouse ES cells were first isolated from the pre-implantation epiblast of embryos in 1981 (Evans & Kaufman, 1981; Martin, 1981), and

their ability to contribute to chimeras, colonise the germline and produce healthy offspring was demonstrated in 1984 (Bradley et al., 1984). This finding provoked a major effort to introduce targeted genetic modifications into mice by implementing homologous recombination in ES cells. It was not until the 1990s that the technology became relatively routine (Capecchi, 2005).

A major reason for the time gap between demonstrating germline contribution and generating genetically modified mice with ES cells is that cell cultures were frequently found to be aneuploid, particularly following clonal selection. This problem gradually reduced as appreciation spread of the relatively fastidious demands of ES cell culture compared with other cell types (Mulas et al., 2019; Robertson et al., 1986). In particular, ES cells must be passaged frequently to avoid overgrowth, which confers advantage to genetically abnormal cells (Mulas et al., 2019). Since these early breakthroughs, ES cells can now be readily genetically modified, genotyped and clonally expanded, and advances such as CRISPR-Cas have made it possible to engineer complex genetic systems, such as floxed genes or possibly gene drives, in ES cells (Andersson-Rolf et al., 2017; Mulas et al., 2019).

The 129 strain of mice is considered the most efficient for derivation of ES cells. Until recently, the 129 strain ES cell lines were the most commonly used and remain an important resource (Behringer et al., 2014; Gardner & Brook, 2002). E14 ES cells, that are used in this project, are derived from the white-bellied, pink eyed dilute, light chinchilla substrain 129P2/Ola, homozygous for chinchilla (*Tyrc-ch*) and pink eyed dilution (*Oca2p*). E14 ES cells were first derived in 1985 by Dr Martin Hooper in Edinburgh, UK (Hooper et al., 1987; Wakayama et al., 1999). The C57BL/6 mouse, comprising of two main substrains, C57BL/6N and C57BL/6J, is now the most widely used strain in biomedical research, as it is an excellent model for studying metabolism, behaviour, immunology, and oncology (Rogers, 2018). Although C57BL/6-derived ES cells, such as the JM8 ES cell types, are now widely used, they are generally considered less stable and inefficient when compared to 129-derived ES cells (Behringer et al., 2014; Bouabe & Okkenhaug, 2013).

5.1.2 Optimal ES cell culture

Regardless of the strain of ES cells, to maintain their developmental potential, optimal culture conditions are required. The power of ES cell technology depends upon the genetic and phenotypic fidelity of ES cells during propagation. If ES cells are cultured inappropriately, they progressively acquire genetic lesions such as aneuploidy that will

compromise their developmental and germline potential. Variants that have undergone chromosomal rearrangements or mutations could be selected for in suboptimal culture conditions, outgrowing the euploid pluripotent cells. Suboptimal conditions include limited supply of some nutrients or growth factors or prolonged culture at densities which favour differentiation (Behringer et al., 2014; Mulas et al., 2019).

Although subjective, an initial choice of correctly genotyped clones with a high probability of contributing to the germline can be partially based on colony morphology and growth rate. The cells should grow in dense, three-dimensional colonies, having distinct edges with a minimal number of flattened colonies and fibroblast like outgrowths. Clones with unusually high or low growth rates should be avoided. Once genetically engineered clones have been identified it is important to reduce the time they spend in culture by freezing down clones as early as possible during the screening process and recovering these low passage number clones for generating chimeras (Czechanski et al., 2014).

5.1.3 Media formulation

High-quality medium and supplements are crucial for successful culturing of ES cells. Traditionally, ES cells are cultured in Dulbecco's modified Eagle's medium (DMEM). However, KnockOut D-MEM, which is based on DMEM but with reduced osmolarity to approximate that of mouse embryonic tissues, has become popular. This medium is optimised for the growth of undifferentiated ES cells and was used when culturing ES cells in this project. This basal medium contains vitamins, amino acids, salts, glucose, and a pH indicator (Behringer et al., 2014; Mulas et al., 2019). They contain no proteins and require supplementation (see section 2.7.2 for ES cell medium formulation). Important components that are supplemented are serum or serum replacement and differentiation inhibitors, such as leukaemia inhibitory factor (LIF) and MEK/GSK3 inhibitors (2i).

The variability in ES cell culture is due primarily to serum. Foetal calf serum (FCS) comes from the blood drawn from a bovine foetus. Serum is typically produced in large batches from many animals. However, samples may differ depending on the age, diet and antibiotic usage in pregnant animals, the country of origin, the breed and other factors creating lot-to-lot variations (Behringer et al., 2014; van der Valk et al., 2018). Some vendors test FCS lots for compatibility with ES cell culture, but this comes at significant cost. FCS is increasingly substituted with serum-free supplements for the culture of ES cells. The use of serum replacement provides a more defined culture medium and eliminates some of the variability found between serum batches (Behringer et al., 2014).

The original culture conditions for ES cells comprised co-culture with a feeder layer of mitotically arrested mouse embryonic fibroblasts and medium containing FCS (Robertson et al., 1986). This was effective, but this complex system was simplified with the discovery that a major contribution of feeders was to provide the cytokine LIF (Smith et al., 1988; Williams et al., 1988). LIF is a secreted polypeptide cytokine that inhibits the spontaneous differentiation of ES cells. Addition of LIF increases the robustness of ES cell cultures on feeders and this remains a widely used system. LIF can also support ES cell culture without the requirement of feeders, which is the case for E14 ES cells when conditions are tightly controlled (Nichols, Evans, & Smith, 1990; Ying et al., 2003). The role of LIF in the maintenance of pluripotency and self-renewal, by activation of the STAT3 pathway, was shown by Niwa et al. (1998).

In addition to LIF, the small molecule inhibitors of the MEK and GSK3 pathways (2i) have been applied to the propagation and derivation of ES cells. ES cells appear morphologically and molecularly relatively homogeneous when maintained in defined medium in which the MEK signalling pathway is blocked and GSK3 is partially inhibited (Mulas et al., 2019; Wray et al., 2011). Under this dual inhibition of these pathways, ES cells exhibit transcriptome similarity to the pre-implantation epiblast (Boroviak et al., 2014). Importantly, male ES cells maintained in 2i can retain a euploid karyotype and germline chimaera competency over multiple passages (Mulas et al., 2019).

5.1.4 Targeting strategies

There are a number of factors to be considered when designing a genome engineering strategy. Of these factors, consideration should be given to the transfection method, format of genome editing reagents and clonal isolation method employed. For ES cells, the transfection method is likely to be either lipofection or electroporation and the decision is often based on the availability of equipment and established protocols in the host laboratory. The commercial availability of SpCas9 RNP and the benefits it offers over mRNA and plasmid formats has seen a rapid increase in its application, despite it is more expensive than plasmid-based CRISPR-Cas formats (Kim, Kim, Cho, et al., 2014; Liang et al., 2015; Schumann et al., 2015). The accompanying HDR repair template is usually delivered as a plasmid or linearized dsDNA or ssDNA. For large constructs, such as CRISPR-based gene drives, plasmid HDR templates are generally used for their stability and reduced random genomic integration compared to linearized dsDNA. ssDNA templates are best for smaller knock-in alterations due to difficulties associated with synthesising ssDNA templates longer than 200 bp (Cong, 2013; Song & Stieger, 2017).

Following transfection of reagents, modified cells need to be isolated and in ES cells this is usually accomplished by antibiotic selection and picking of resistant colonies.

As integration of the targeting construct occurs in only a fraction of the ES cells transfected, the inclusion of an antibiotic-resistance cassette in the targeting construct greatly improves screening efficiency. This cassette keeps the cells alive after integration when the culture is treated with the corresponding antibiotic (Behringer et al., 2014). In this project, a *PAC* cassette was included in all constructs to confer resistance to puromycin. During puromycin treatment, all or most sensitive cells should die, and when plated at an appropriate density the resistant cells harbouring the construct within their genome grow into individual colonies to be isolated by picking (see section 2.5.1.3 for description of picking). The clonal populations are expanded, typically in 96 WP format, replica-plated, frozen and gDNA extracted from a replicate for genotyping (Behringer et al., 2014). Through the entirety of the genome engineering pipeline, it is desirable to keep cells in culture for as short a period of time as possible to limit the likelihood of differentiation or chromosomal abnormalities arising (Gaztelumendi & Nogués, 2014; Mulas et al., 2019).

5.1.5 Genotyping ES cells

One benefit of ES cell technology is that genotyping can be performed prior to generating the animals. A genotyping strategy should be established before initiating an experiment. Traditionally, ES cells are genotyped by Southern blot, and this remains the benchmark technique (Carofino & Justice, 2015; Zhou et al., 2018). There are a number of other methods for genotyping genetically altered ES cells, including PCR, quantitative PCR (qPCR), Sanger sequencing and droplet digital PCR (ddPCR). A description of these techniques is described below. Emerging NGS-based screening techniques are discussed in section 5.4.5.

5.1.5.1 Southern blot

The method of Southern blotting was first published in 1975 by Edwin Southern, whom the technique is named after (Southern, 1975). A Southern blot is a method used for detection of a specific DNA sequence in DNA samples. Southern blotting combines transfer of electrophoresis-separated DNA fragments to a filter membrane and subsequent fragment detection by probe hybridisation (Chowdhury & Dubey, 2014). Southern blots performed with restriction enzyme-digested gDNA can be used to determine the size and copy number of a genetic element in the genome of ES cell lines (Wang et al., 2015). Although a powerful technique, Southern blotting is labour intensive,

costly, requires large quantities of gDNA, and in most cases involves the use of hazardous radioactive probes (Hoebeeck, Speleman, & Vandesompele, 2007).

5.1.5.2 PCR

PCR was invented in 1983 by Kary Mullis. Since its invention it has become a fundamental technique in molecular biology. Using PCR, targeted DNA sequences can be exponentially amplified for further analysis (Bartlett & Stirling, 2003). The majority of PCR methods rely on thermal cycling, where reactants, including template DNA, primers and DNA polymerase are exposed to repeated cycles of heating and cooling to permit DNA melting and enzyme-driven DNA replication (Green et al., 2012). For screening HDR genome editing events using PCR, primers are typically designed flanking the homology arms. These primer pairs span from inside the construct to outside the homology arms on both the 5' and 3' ends. Only ES cell lines with site-specific integration will amplify a PCR product, which can be resolved by gel electrophoresis or Sanger sequenced (Behringer et al., 2014). It is important to note that PCR can amplify very small amounts of DNA and this level of sensitivity creates its own drawbacks, primarily the amplification contaminating DNA can lead to false positives (Borst, Box, & Fluit, 2004). Furthermore, conventional PCR screening approaches cannot detect concurrent random integration or quantify copy number of the genetic construct in the genome and, therefore, accurate characterisation of a clonal ES cell population requires additional techniques, such as Southern blotting, qPCR or ddPCR (Collier et al., 2017; Mancini et al., 2011; Wang et al., 2015).

5.1.5.3 Quantitative PCR

Quantitative PCR (qPCR) is a modification of standard PCR to quantify the amount of target DNA by introducing fluorescent or intercalating dyes to detect PCR product as it accumulates in real time during PCR thermocycling reaction (Green et al., 2012). qPCR can be used as a relative quantification of transgene copy number. For such an analysis, the primers and assay require optimisation to generate robust and reproducible results. To measure DNA copy number, the amplicon should be located within the unique transgene sequence and performed on cells that have undergone at least one passage prior to screening to ensure residue HDR template is sufficient degraded or diluted through cell division. This assay also uses a control gene that is known to have two copies in the genome (Li et al., 2019; Ma & Chung, 2014).

A master mix is prepared, which includes gDNA template, primers, probes or dyes and polymerase. The assay is performed on a qPCR instrument and data is collected in real

time. There are two approaches to the assay: fluorescent probes and intercalating dyes. In either approach, fluorescence should double with every cycle of PCR, and the amount of starting template can be determined from the number of cycles required to achieve a specified threshold level of fluorescence. Relative quantification of the transgene to the control gene can be used to determine copy number of the transgene in the clone of interest (Green et al., 2012; Ma & Chung, 2014).

Transgene copy number analysis by qPCR can be easily scaled to analysis large numbers of samples but this approach suffers from a number of setbacks. It can be affected by DNA quality, results may fall between integers (e.g. transgene copy number of 1.4) making interpretation difficult and, unlike droplet digital PCR (section 5.1.5.5) triplicates are usually required for each run (Taylor, Laperriere, & Germain, 2017).

5.1.5.4 Sanger sequencing

PCR products amplified from the gDNA of a targeted ES cell line can be Sanger sequenced to confirm that underlying DNA sequence of the integration is correct. The method was developed by Frederick Sanger and colleagues in the 1970s that is based on selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during *in vitro* DNA replication (Sanger, Nicklen, & Coulson, 1977). Modern Sanger sequencing typically uses fluorescently labelled dideoxynucleotides that are detected by a laser after capillary electrophoresis to generate a chromatogram with peaks corresponding to incorporation of the four different fluorescent dyes coupled to dideoxynucleotides: ddA, ddT, ddC and ddG (Smith et al., 1986).

Although one could use individual Sanger sequencing reactions to cover an entire transgene, this testing approach can be costly and labour intensive and rarely performed on large constructs, such as the gene drive constructs used in this project. Sanger sequencing is of value in assessing the DNA sequence on the homologous chromosome in the ES cells that are heterozygous. Only one allele in ES cells is usually modified by HDR and Sanger sequencing can determine whether CRISPR-induced indels by NHEJ are present on the homologous chromosomes (Jarazo, Qing, & Schwamborn, 2019).

5.1.5.5 Droplet Digital PCR

Droplet Digital PCR (ddPCR) is a recent technology that has been commercially available since 2011. ddPCR technology uses DNA polymerase in a standard PCR reaction to amplify a target DNA fragment from a sample using validated primers and fluorescently labelled probes that are positioned between the two PCR primers (Taylor et al., 2017). Because the probe is sequence specific, it will only detect the presence of a single

amplicon within the reaction. ddPCR reactions are partitioned into thousands of individual reaction droplets prior to amplification. In this project, a Bio-Rad ddPCR system was used, in which a 20 µl PCR reaction is divided into 20,000 nanolitre-sized droplets through a water-oil emulsion technique (Quan, Sauzade, & Brouzes, 2018).

After partitioning the PCR in 20,000 droplets, the emulsion undergoes multiple amplification cycles to reach the end point of the reaction, then each droplet is checked for fluorescence in a cytometer with a binary readout of “0” or “1”. The fraction of fluorescing versus non-fluorescing droplets is recorded, allowing calculation of the template DNA quantity in the starting sample Poisson Algorithm. Poisson correction compensates the presence of more than one copy of target gene in any droplet. Unlike analogue quantitative PCR (qPCR), quantification with ddPCR does not require a standard curve and PCR reactions have a greater tolerance for inhibitor substances and inefficient amplification compared to qPCR (Quan et al., 2018).

In ES cell genome engineering projects, ddPCR can be used to efficiently determine the copy number of the transgene in clonal populations. This can be done by using two sets of primers and probes, one targeting a reference gene that is known to have two copies in the genome and the other targeting a region within the construct of interest. Examples of commonly used reference genes in mice are *GAPDH*, *PGK1*, *GUSB* and *TFRC* (Boda et al., 2009). In this project, the Bio-Rad recommended *TFRC* gene was used, which encodes a cell surface receptor necessary for cellular iron uptake by the process of receptor-mediated endocytosis (Locke et al., 2015). Within the construct, the *PAC* gene was targeted, as this element was present in all three constructs. The ratio of positive PCR reactions (droplets) of the reference gene (*TFRC*) to those positive for gene of interest (*PAC*) is the basis for determine the copy number of the transgene. However, ddPCR does not provide information of the size of the integration, which can be obtained when performing Southern blotting.

5.1.6 Karyotyping

The normal mouse mitotic karyotype consists of 40 chromosomes (Behringer et al., 2014). ES cell lines should be karyotyped as subsets of aneuploid cells can be selected with continuous passages (Gaztelumendi & Nogués, 2014). Variation in frequency of germline transmission has been observed since the first applications of ES cell technology and this can be largely attributed to the chromosomal abnormalities that provide a growth advantage and presumably interfere with the ability to form functional germ cells in chimeras (Codner et al., 2016; Suzuki et al., 1997). Chromosomal abnormalities, in

particular, the trisomy of chromosome 8, occur rather frequently in ES cells during gene targeting (Kim et al., 2013). Trisomy of chromosome 8 confers as a selective growth advantage and leads to depletion and eventual loss of normal ES cells during consecutive passages. ES cells with trisomy 8 rarely contribute to the germline (Liu et al., 1997).

Suboptimal culture conditions and enzymatic dissociation of cells with trypsin may account for the chromosomal abnormalities accumulated during continuous culture (Gaztelumendi & Nogués, 2014; Rebuzzini et al., 2008). It is possible that future development of more optimal ES cell culture conditions may facilitate the maintenance of a normal karyotype, even after prolonged culture. Until then, it is essential to perform karyotype analysis of ES cells lines used for the generation of chimeras. ES cell clones with at least 50% of the chromosome spreads containing 40 chromosomes should be selected for chimera production (Codner et al., 2016; Cotton et al., 2015).

5.1.7 Functional analysis of ES cells *in vitro*

Another advantage of ES cell mediated transgenesis is that it allows for some introduced genetic elements to be tested *in vitro* before proceeding to generate animals. Examples include testing floxed alleles by transfection of Cre recombinase and checking for gene expression from the integrated constructs (Flemer & Bühler, 2015; Huang et al., 2008). This lets researchers gain valuable insight into the functionality of the introduced genetic elements, limiting the generation and unwanted waste of animals with non-functional genetic systems.

ES cells can develop into many cell types if they are placed into an appropriate differentiating environment. This can occur *in vitro* by modifying culture conditions to induce differentiation (Zakrzewski et al., 2019). The desire to differentiate cells towards a certain lineage has led to an increasing number of specialised culture conditions that can bias the differentiation of ES cells into many cell types, including pancreatic cells, haematopoietic cells, endothelial cells, and cardiomyocytes (Valamehr et al., 2011).

Most ES cell differentiation protocols start through an intermediate stage called the embryoid body (EB). EBs are round structures composed of ES cells that have undergone some of the initial stages of differentiation. EBs can be manipulated to generate more specific cell types (Kibschull, 2017). Differentiating ES cells into cardiomyocytes was performed in this project as a method for screening for basic developmental potential, in this case the capacity to contribute to the cardiovascular system.

5.1.8 Aim

This chapter aimed to genome engineer three ES cell lines: OOEP_CAG, OOEP_VASA and ROSA26_gRNAs using a CRISPR-assisted genome engineering approach. Engineered cell lines would be clonally isolated and genotyped by PCR, Sanger sequencing and ddPCR, as well as confirmation by karyotyping, functional testing of genetic constructions and *in vitro* differentiation. Validated ES cell lines could be used for blastocyst injection to generate two lines of SGD mice: OOEP_CAGxROSA26_gRNAs and OOEP_VASAxROSA26_gRNAs.

5.2 Additional methods

5.2.1 ES cell genome engineering pipeline

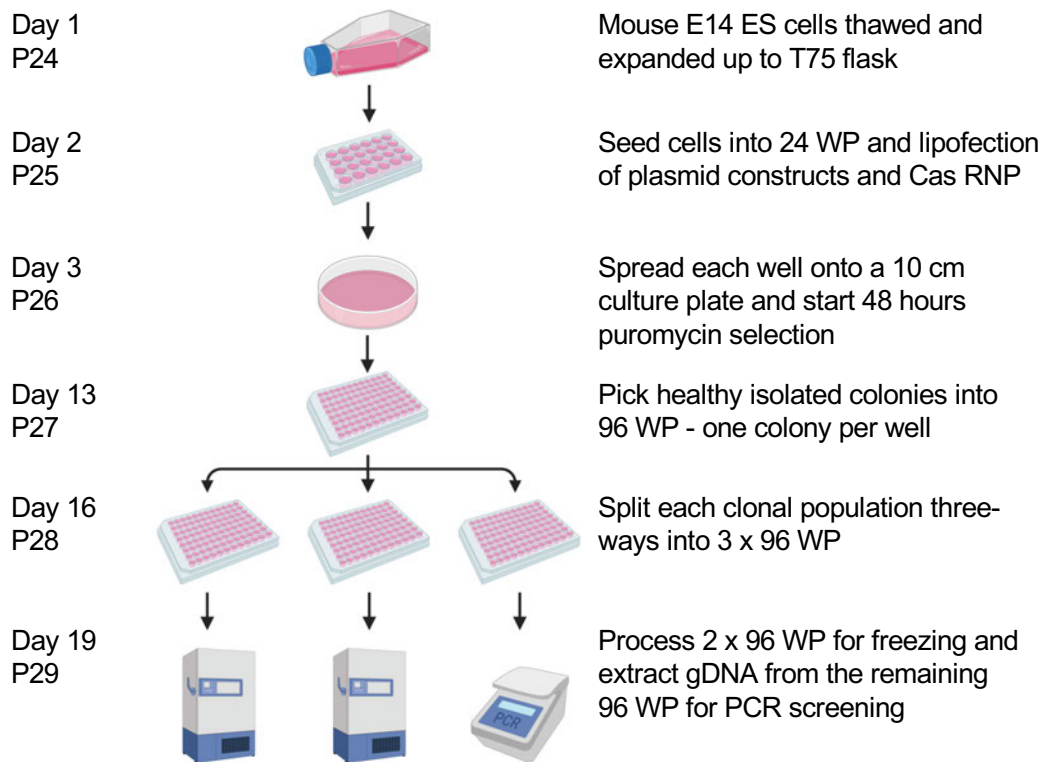


Figure 5.1 Workflow for genome engineering ES cells. Cells are seeded into a 24 well plate (WP) on day 2 at 2×10^5 cells per well. P = passage number.

5.2.2 Expansion of positive ES cell clones

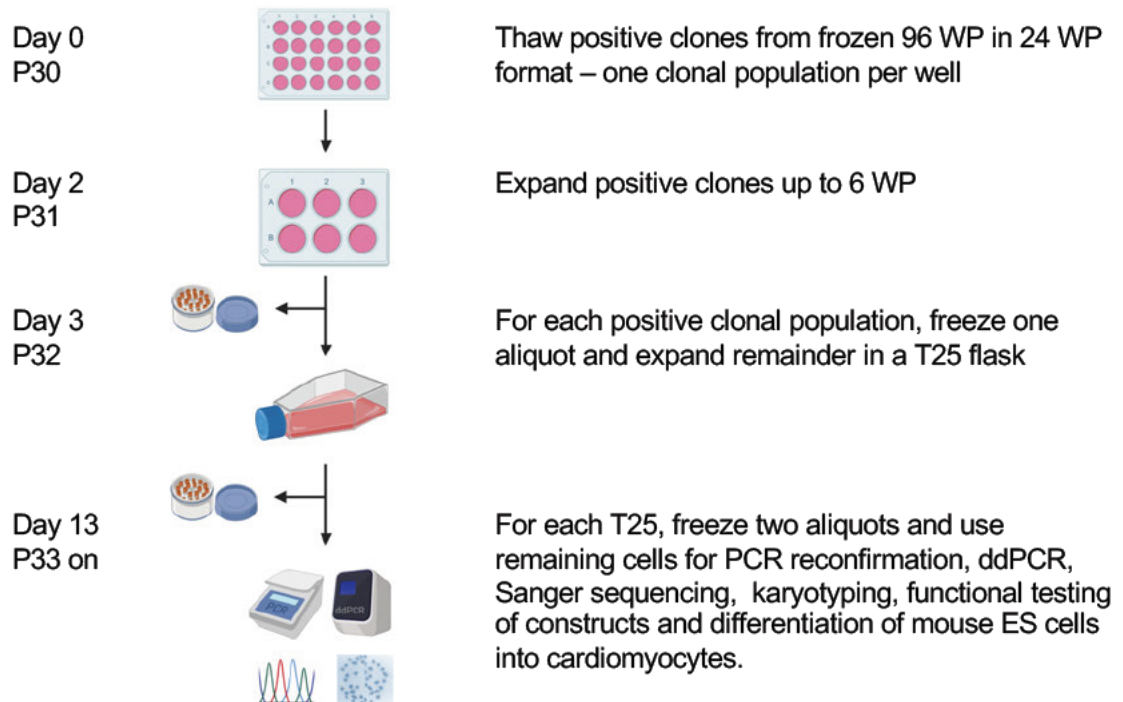


Figure 5.2 Workflow for expanding PCR positive genome engineered ES clones from frozen 96 well plate (WP). P = passage number.

5.3 Results

5.3.1 Transfections

E14 ES cells were genome engineered following the workflow outlined in Figure 5.2.1. *OOEP*-targeted cell lines were engineered by co-transfection of *OOEP_CAG* or *OOEP_VASA* plasmids with SpCas9 RNP of *OOEP* SpCas9 gRNAs 1 and 4. *ROSA26_gRNAs* cell lines were engineering by co-transfection of *ROSA26_gRNAs* plasmid and pY094 expressing *ROSA26* AsCas12a gRNA 1. A schematic diagram of the three-transfection combinations can be seen in Figure 5.3. All transfections were performed by lipofection following the protocol in section 2.5.1.1.

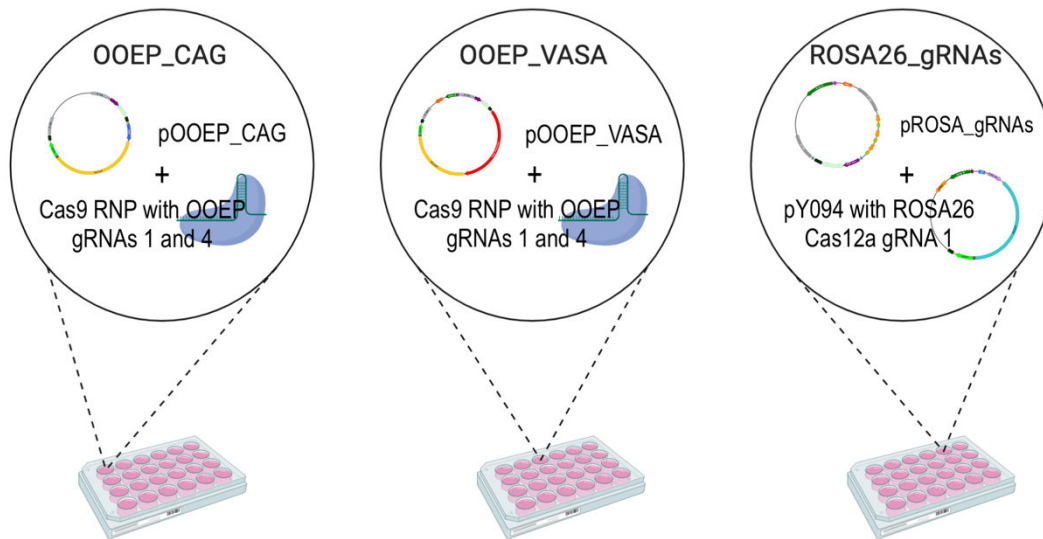


Figure 5.3 Schematics representation of the transfection combinations used for genome engineer three ES cell lines: OOEP_CAG, OOEP_VASA and ROSA26_gRNAs. Detailed plasmid (p) maps for pOOEP_CAG, pOOEP_VASA and pROSA26_gRNAs can be found in section 4.3.5. A plasmid map of pY094 is shown in section 2.7.3.

5.3.2 Primer design and PCR screening

In order to screen ES cells for CRISPR-mediated genomic integration at the target site, PCR primers flanking the homology arms of each construct were designed to confirm construct integration at the target site. To confirm the copy number of each construct in the clonally isolated populations, ddPCR was used to quantify *PAC* against a mouse reference gene, *TFRC*. Primers and probes for homology arm flanking PCRs and ddPCR screening can be seen in Figure 5.4. All primer and probe sequences can be found in Appendices A.2 and A.5, respectively.

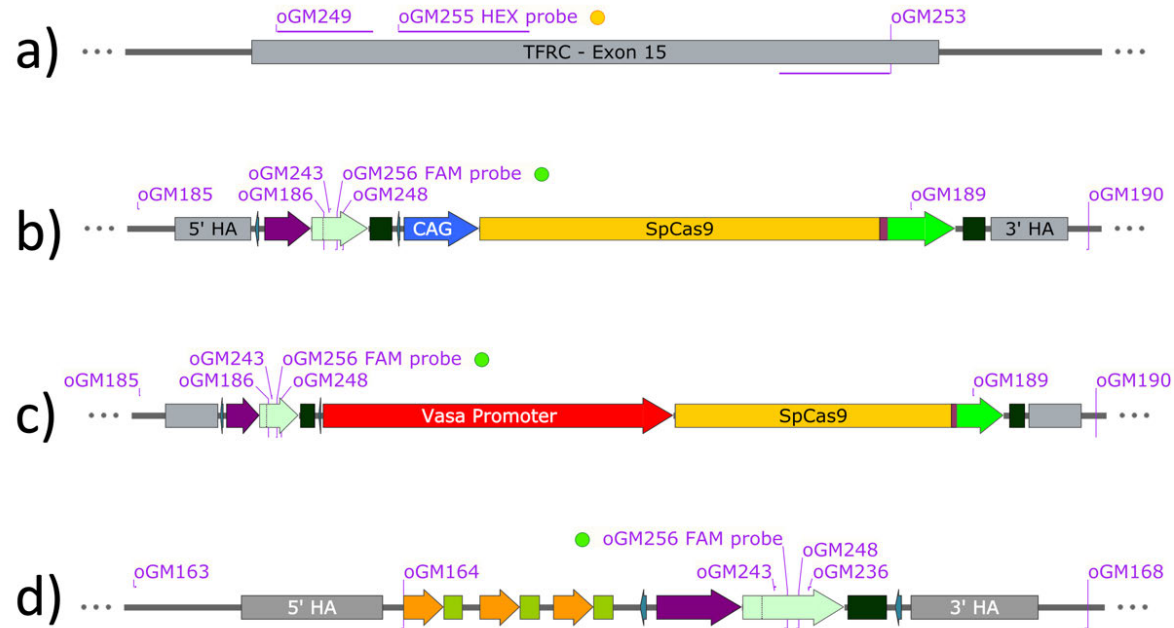


Figure 5.4 Primers and probes for screening clonal cell populations. a) oGM253 and oGM249 amplify 126 bp of the *TFRC* reference gene used in droplet digital PCR (ddPCR) with the oGM255 hexachlorofluorescein (HEX) probe. b) and c) oGM185 and oGM186 flank the 5' homology arm of *OOEP_CAG* and *OOEP_VASA* edited cells to amplify a 1981 bp product. oGM189 and oGM190 flank the 3' homology arm of *OOEP_CAG* and *OOEP_VASA* edited cells to amplify a 1888 bp product. oGM243 and oGM248 amplify 150 bp of the *PAC* gene used for ddPCR with the oGM256 fluorescein (FAM) probe. d) oGM163 and oGM164 flank the 5' homology arm of *ROSA26_gRNAs* edited cells to amplify a 1597 bp product. oGM236 and oGM168 flank the 3' homology arm of *ROSA26_gRNAs* edited cells to amplify a 1650 bp product. oGM243 and oGM248 primers, along with the oGM256 FAM probe were also used in ddPCR experiments for *ROSA26_gRNAs* edited cells.

Homology arm flanking PCRs were performed as an initial screen of targeted ES clones. To do this gDNA was extract from clonal populations in 96 WP format. 5' and 3' flanking homology arm PCRs were run using primers in Figure 5.4. Homology arm flanking primers spanned the junction between each arm of the targeting constructs and the adjacent flanking sequence of the host cell gDNA. Gel electrophoresis can be seen in Appendix A.11 and a summary of the PCR screens on each 96 WP are presented in Figure 5.5. From PCR screening, OOEP_CAG, OOEP_VASA and ROSA26_gRNAs had 27, 39 and 12 clones respectively, with anticipated PCR bands supporting HDR events at both the 5' and 3' ends of the targeted constructs.

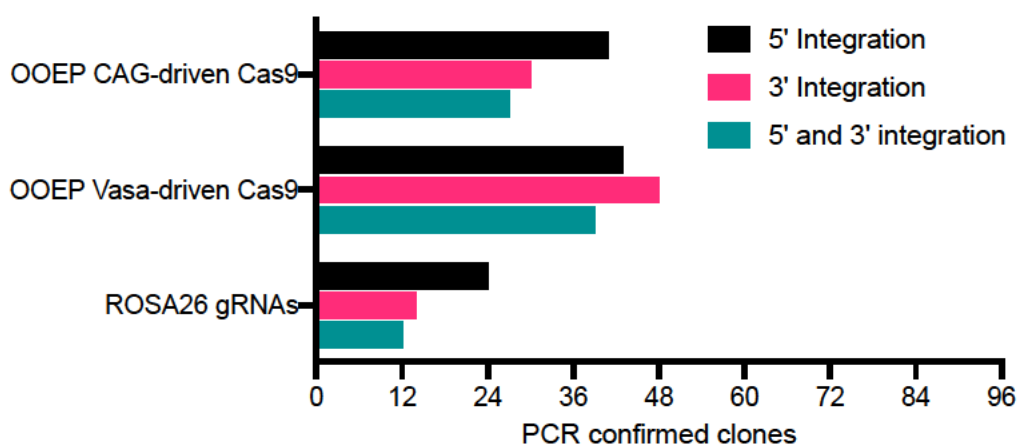


Figure 5.5 Summary of PCR analysis using 5' and 3' homology arm flanking primers in clonal populations of OOEP_CAG, OOEP_VASA and ROSA26_gRNAs edited clones in 96 well plate format. OOEP_CAG, OOEP_VASA and ROSA26_gRNAs had 27, 39 and 12 clones, respectively, with putative 5' and 3' integration.

5.3.2 ddPCR primer testing

Cells that had evidence of 5' and 3' homology arm integration by PCR were then recovered from a frozen 96 WP, expanded and aliquots frozen following the workflow in Figure 5.2. Some cells were maintained in culture and fresh gDNA was extracted from each clonal population of cells for analysis by ddPCR. ddPCR was used to determine the copy number of the *PAC* gene in each cell line. As a simple method to determine primer specificity, endpoint PCRs were run using the ddPCR polymerase (BioRad SuperMix) and the recommended thermocycling conditions. Figure 5.6 shows clean bands with the

expected product sizes of 150 bp for the *PAC* amplicon and 126 bp for the *TFRC* amplicon, indicating that these primers are suitable for ddPCR experiments.

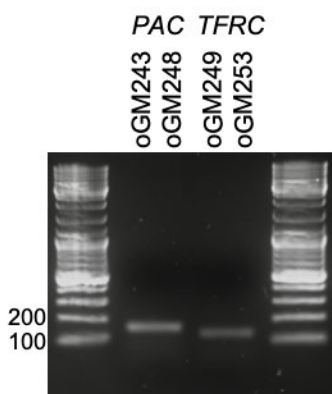


Figure 5.6 Gel electrophoresis of PCR products to check specificity of ddPCR primers. *PAC* amplicon was expected to be 150 bp with primers oGM243 and oGM248, and *TFRC* amplicon was expected to be 126 bp with primers oGM249 and oGM253. ddPCR supermix used in the reaction and run for 40 cycles.

5.3.3 ddPCR results

Tested ddPCR primer pairs were used to determine the copy number of the transgenes in each putatively PCR-positive ES clone using ddPCR protocol outlined in section 2.1.7. The data generated from the protocol was analysed using BioRad Quantasoft Pro to determine the copy number of the *PAC* gene in each of targeted ES clones. A summary of the data can be seen in Figure 5.7 and quality control graphs of each ddPCR experiment is available in Appendix A.12. Being diploid cells, the number of copies of the reference gene *TFRC* was set at two. By comparing the number of copies of the *PAC* gene to *TFRC* in each cell line, the copy number of the targeting construct in each clonal population was quantified. OOEP_CAG and OOEP_VASA had 12 and 20 clones with a single copy of *PAC*, respectively. ROSA26_gRNAs had only 2 clones with a single copy of *PAC*. Selection of clones was based on the presence of tight Poisson 95% confidence limits and a minimum of 950 fluorescent PCR droplets read, in most cases several thousand fluorescent droplets were read (see ddPCR quality control Appendix A.12). These clones were considered to be putative heterozygous for their respective constructs.

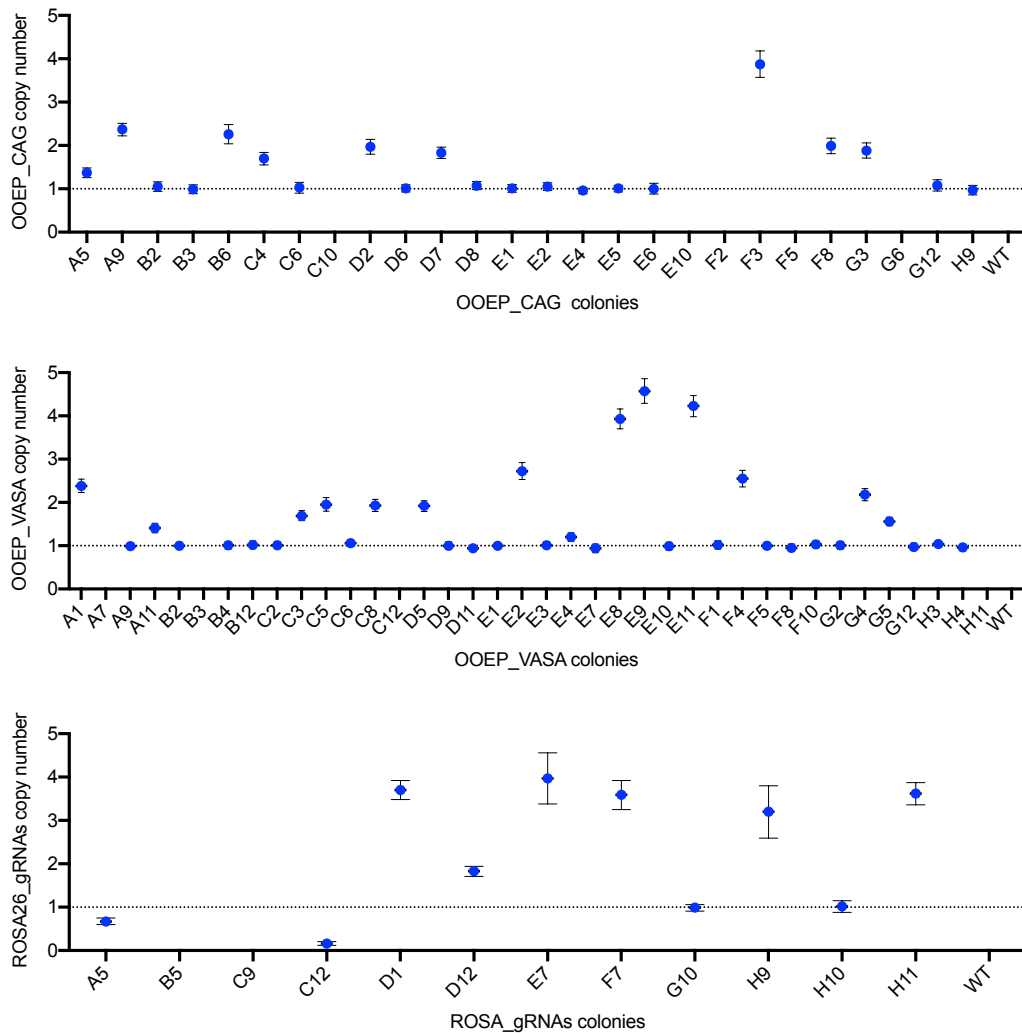


Figure 5.7 Summary of ddPCR transgene copy number results. Copy number of *PAC* gene in clonal ES cell populations against TRFC reference. OOEP_CAG, OOEP_VASA and ROSA26_gRNAs had 12, 20 and 2 clones with one copy of the *PAC* gene, respectively. The *PAC* gene was used to quantify construct copy number in each cell line. Error bars reflect Poisson 95% confidence limits.

5.3.4 Reconfirming PCRs

In parallel to ddPCR screening, gDNA from the same thawed cells was used to reconfirm 5' and 3' integration of the targeting construct by PCR. Based on the results from these reconfirmed PCRs and the ddPCR analysis, 4 heterozygous clones of both OOEP_CAG and OOEP_VASA and the 2 heterozygous clones from ROSA26_gRNAs were selected for further screening. Clones labelled in red with a cross (+) in Figures 5.8, 5.9 and 5.10 were selected for further screening.

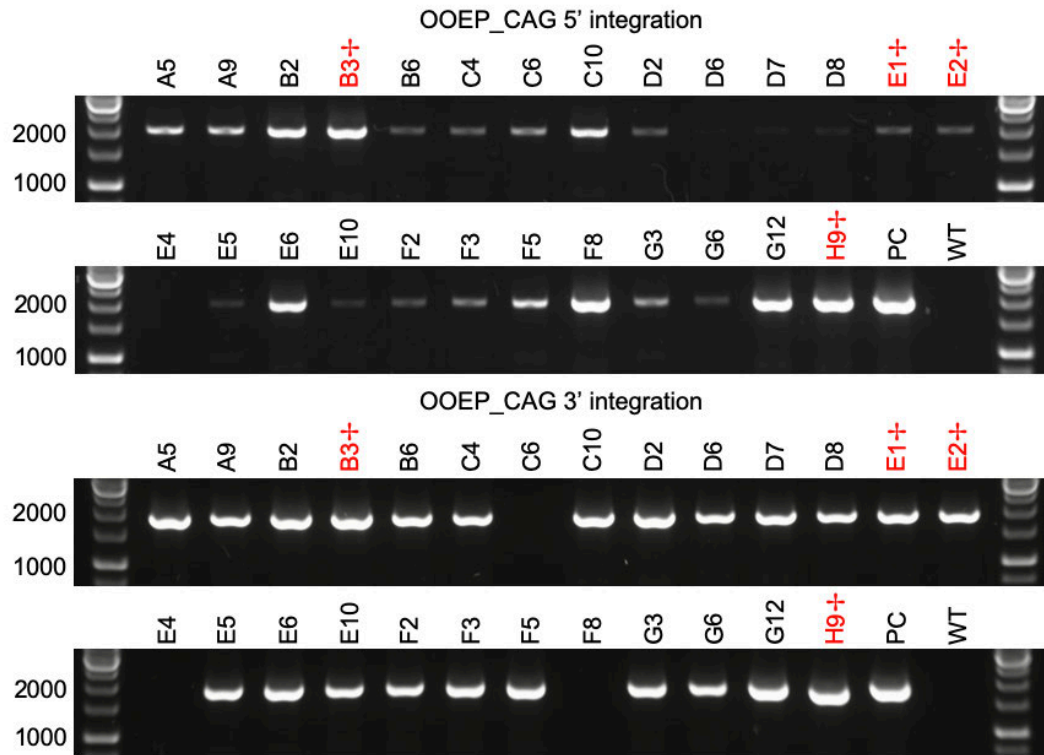


Figure 5.8 Reconfirmation of homology arm integration in thawed OOEP_CAG target cells. 5' flanking PCRs used oGM185 and oGM186 to amplify an expected product of 1981 bp. 3' flanking PCRs used oGM189 and oGM190 to amplify an expected product of 1888 bp. Red labels with + denote those clones that were selected for further screening.

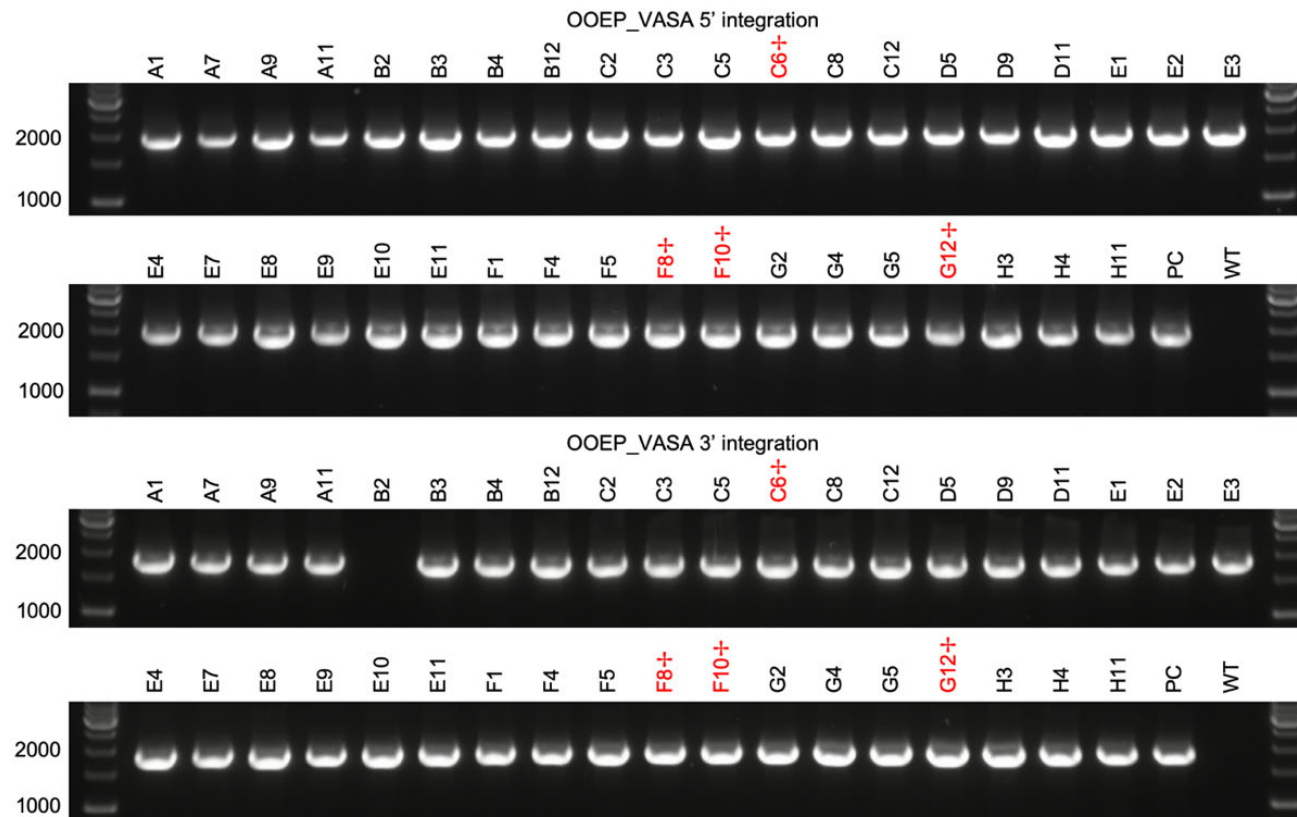


Figure 5.9 Reconfirmation of homology arm integration in thawed OOEP_VASA targeted cells. 5' flanking PCRs used oGM185 and oGM186 to amplify an expected product of 1981 bp. 3' flanking PCRs used oGM189 and oGM190 to amplify an expected product of 1888 bp. Red labels with + denote those clones that were selected for further screening.

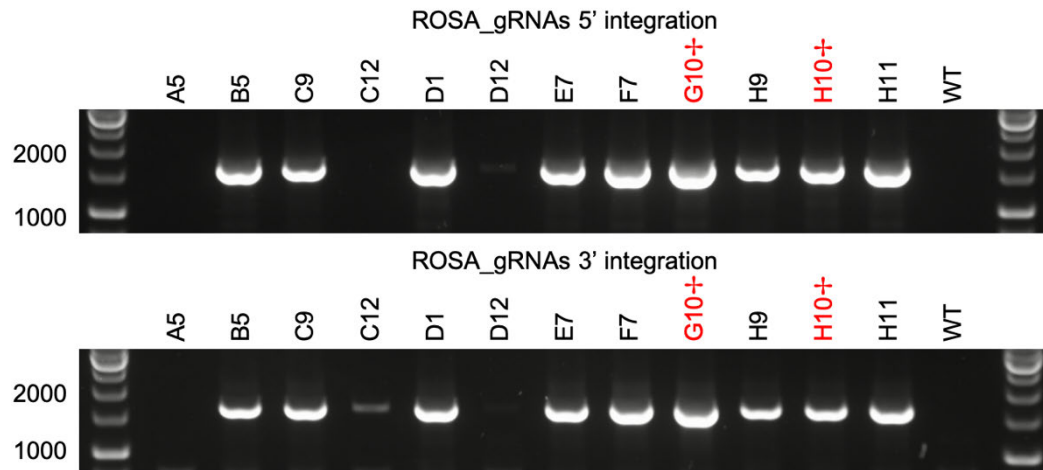


Figure 5.10 Reconfirmation of homology arm integration in thawed ROSA26_gRNAs targeted cells. 5' flanking PCRs used oGM163 and oGM164 to amplify an expected product of 1597 bp. 3' flanking PCRs used oGM189 and oGM190 to amplify an expected product of 1650 bp. Red labels with + denote those clones that were selected for further screening.

5.3.5 Wild-type allele sequencing

PCR and Sanger sequencing of the homologous chromosome (termed 'wild-type' here) was performed on each of the selected heterozygous clones from section 5.3.4. *OOEP* was amplified with primer pair ppOOEP and *ROSA26* was amplified with ppROSA26 before sending for Sanger sequencing. Of the four *OOEP_CAG* clones, B3 and E2 wild-type chromosomes had 45 and 36 bp deletions, respectively. E1 had a 11 bp insertion and 1 bp deletion, and H9 had no indel present (Figure 5.11). *OOEP_VASA* clones had deletions at the target site of all four of the wild-type chromosomes sequenced, ranging from 8 to 42 bp in size (Figure 5.12). ROSA26_gRNAs clone H10 had a 10 bp deletion in the wild-type chromosome, while G10 had multiple overlapping traces following the expected cut site, suggesting G10 was a polyclone (Figure 5.13). ROSA26_gRNAs G10 was subsequently subcloned to isolate clonal populations.

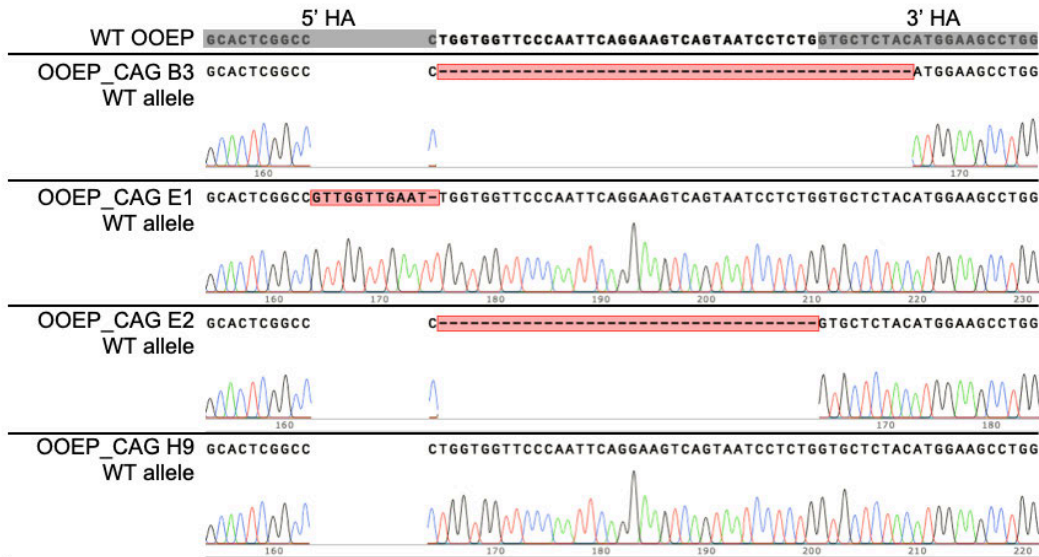


Figure 5.11 Sanger sequencing of the *OOEP* target region on the homologous wild-type (WT) chromosome in heterozygous *OOEP_CAG* targeted cells.

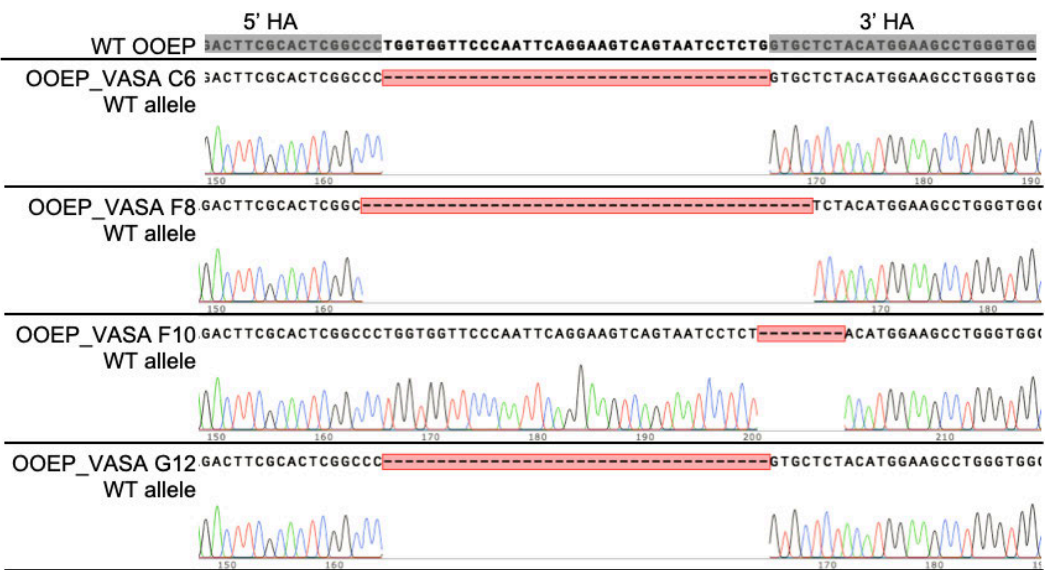


Figure 5.12 Sanger sequencing of the *OOEP* target region on the homologous wild-type (WT) chromosome in heterozygous *OOEP_VASA* targeted cells.

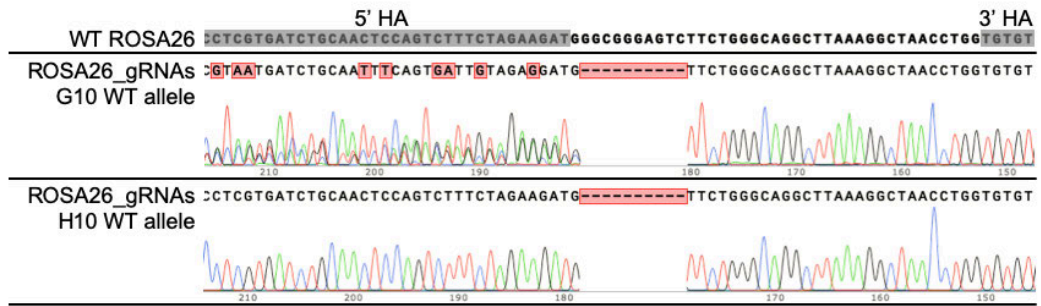


Figure 5.13 Sanger sequencing of the *ROSA26* target region on the homologous wild-type (WT) chromosome in heterozygous *ROSA26_gRNAs* targeted cells. The *ROSA26_gRNAs* G10 clone shows multiple overlapping traces proximal to the cut site, indicating a polyclone.

5.3.6 Subcloning and reconfirmation of G10 polyclonal line

ROSA26_gRNAs G10 was subcloned by spreading ES cells at low density into a 6 WP and picking healthy, isolated colonies into a 96 WP following the protocol in section 2.5.1.3. Isolated clones were then expanded up to a 6 WP format and aliquots frozen. gDNA was extracted from remaining cells for each clone and PCRs run to reconfirm 5' and 3' integration (Figure 5.14). All G10 clones were confirmed for 5' and 3' integration. A2, A5 and B1 cell lines for *ROSA26_gRNAs* G10 were then selected for Sanger sequencing of the wild-type chromosome.

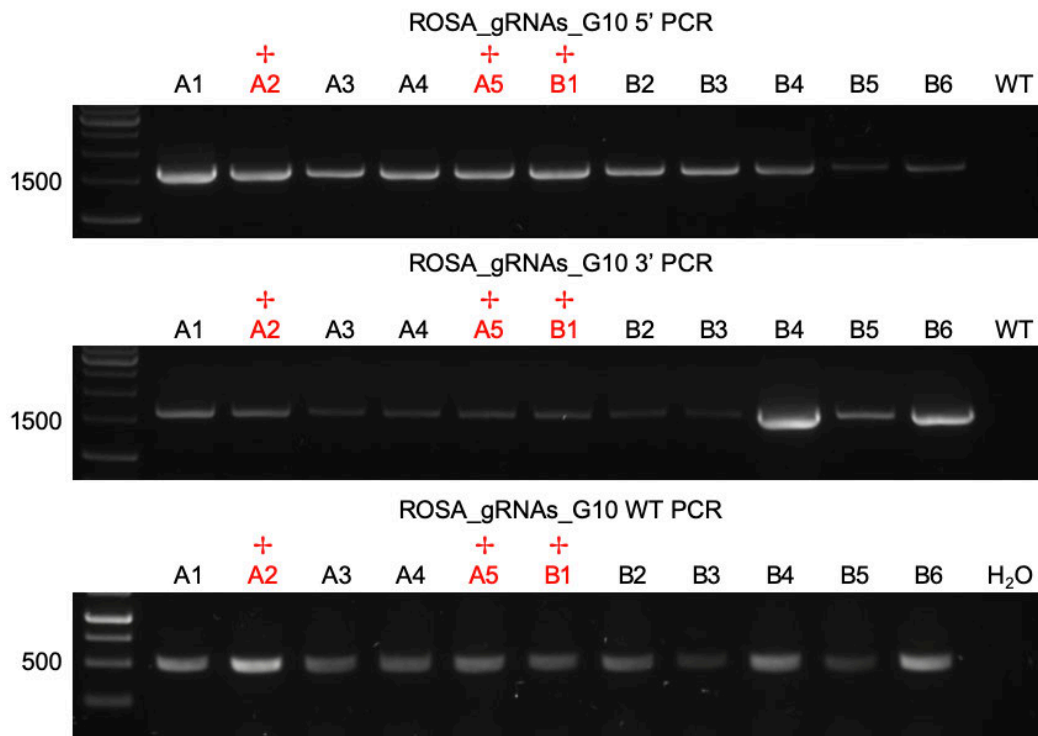


Figure 5.14 PCR analysis of the subcloned ROSA26_gRNAs G10 subclones. All clones were confirmed by 5' and 3' homology arm flanking PCRs. Expected band size of 5' homology flanking PCR (top) is 1597 bp. Expected band size of 5' homology flanking PCR (middle) is 1650 bp. Expected band size of wild-type (WT) allele is 562 bp. Red labels with + denote those clones that were selected for further screening.

The Sanger sequencing results from the three selected clones identified two allelic variants in the ROSA26_gRNAs G10 polyclone. The wild-type chromosome in ROSA26_gRNAs G10(B1) clone contained no CRISPR-mediated indel, while both G10(A2) and G10(A5) had a 10 bp deletion at the expected AsCas12a cleavage site. As the ddPCR identified that the G10 polyclone to be heterozygous, due to the sensitivity of the ddPCR, these subclones are also expected to be heterozygous for the construct. ROSA26_gRNAs G10(A2) and (B1) were selected for further screening. ROSA26_gRNAs G10(A2) and (B1) were selected as they represent two distinct clones of the G10 polyclone, while ROSA26_gRNAs G10(A2) and G10(A5) are likely to be derived from the same parental cell.

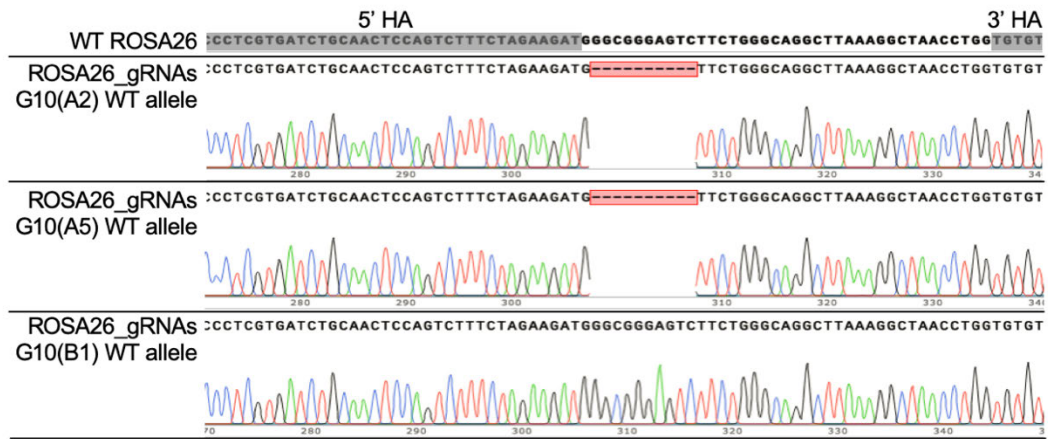


Figure 5.15 Sanger sequencing of the *ROSA26* target region on the homologous chromosome in heterozygous *ROSA26_gRNAs* G10 subclones. G10(A2) and G10(B1) clones were selected for further analysis.

5.3.7 Karyotyping

Selected ES clones from PCR, ddPCR and Sanger sequencing screening were karyotyped following the protocol in section 2.5.1.10. From published reports by Cotton et al. (2015) and Codner et al. (2016), a threshold of 50% euploidy metaphase spreads (40 chromosomes) was required from to qualify an ES clone for further screening. Codner et al. (2016) report that germline transmission efficiency of the ES cell derived genotypes drops from over 80% to just over 20% when there is less than 50% euploid metaphase chromosome spreads. Although their findings are based off counting 30 chromosome spreads per clone, within this study only 25 spreads were counted per clone. A summary of the karyotyping can be seen in Figure 5.16 and representative karyotype images can be seen in Figure 5.17. Of all the 11 clones karyotyped, only clones OOEP_CAG B3 and OOEP_CAG H9 were removed from further screening. OOEP_CAG B3 was below 50% threshold, while OOEP_CAG H9 had only 55% euploidy counts and was precautionarily removed. All OOEP_VASA and *ROSA26_gRNAs* clones had acceptable karyotypes, above the 50% threshold.

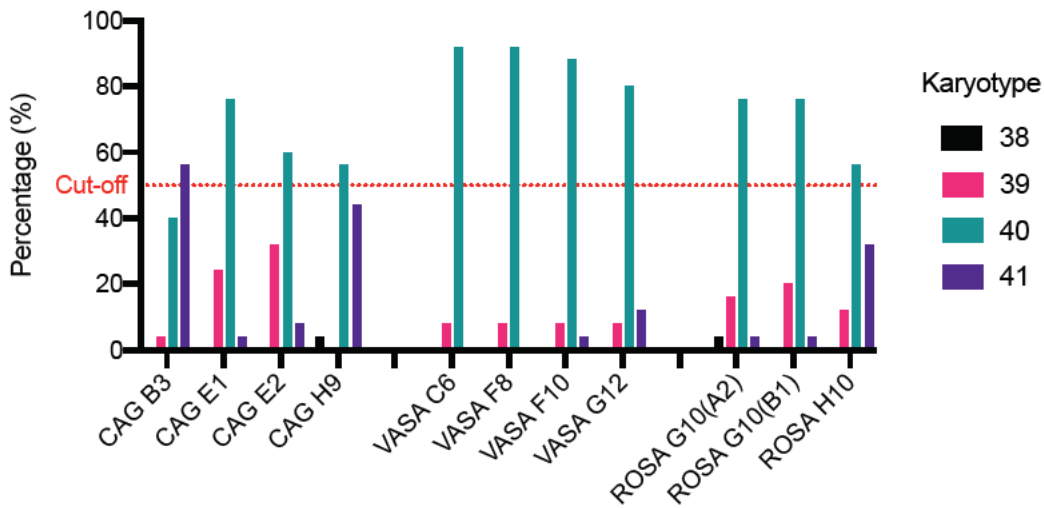


Figure 5.16 Summary of chromosome counting from metaphase spreads. Percentages are based off counts from 25 spreads for each clone. OOEP_CAG B3 and H9 were excluded from further analysed.

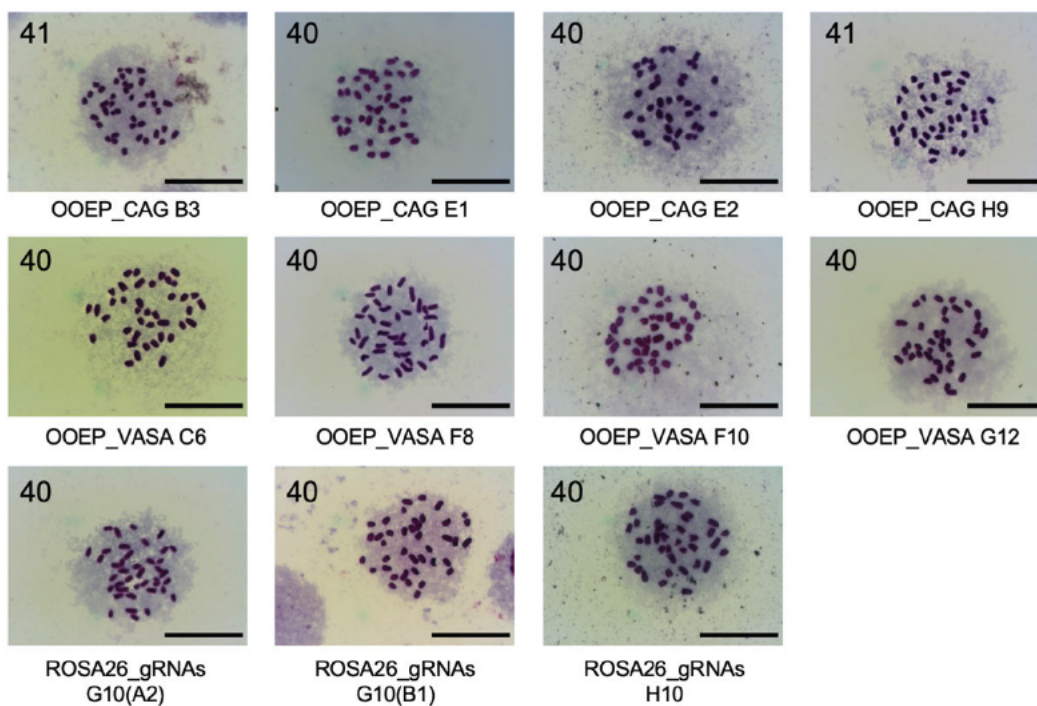


Figure 5.17 Representative images of metaphase chromosome spreads used for karyotyping of the clonal ES cell populations. Number in top left-hand corner shows chromosome count of the spread. Images taken at 60x. Scale bar is 30 μ m.

5.3.8 Morphology of selected ES clones

Selected ES cells with selected genotypes and karyotypes were thawed from passage 33 frozen aliquots following the protocol in section 2.5.1.8. Recovered cells were then imaged under a phase contrast microscope two days after thawing to compare morphology of clones (Figure 5.17). Some colonies in OOEP_VASA F8 had signs of differentiation, with cells spreading flat and forming cobblestone-like structures, while some colonies in ROSA26_gRNAs H10 were flatter than desired with rough surfaces and the ability discriminate individual cells within those flat colonies. All other thawed clones had acceptable morphology that are typically rounded or oval in shape with smooth domed tops and a phase contrast bright edge.

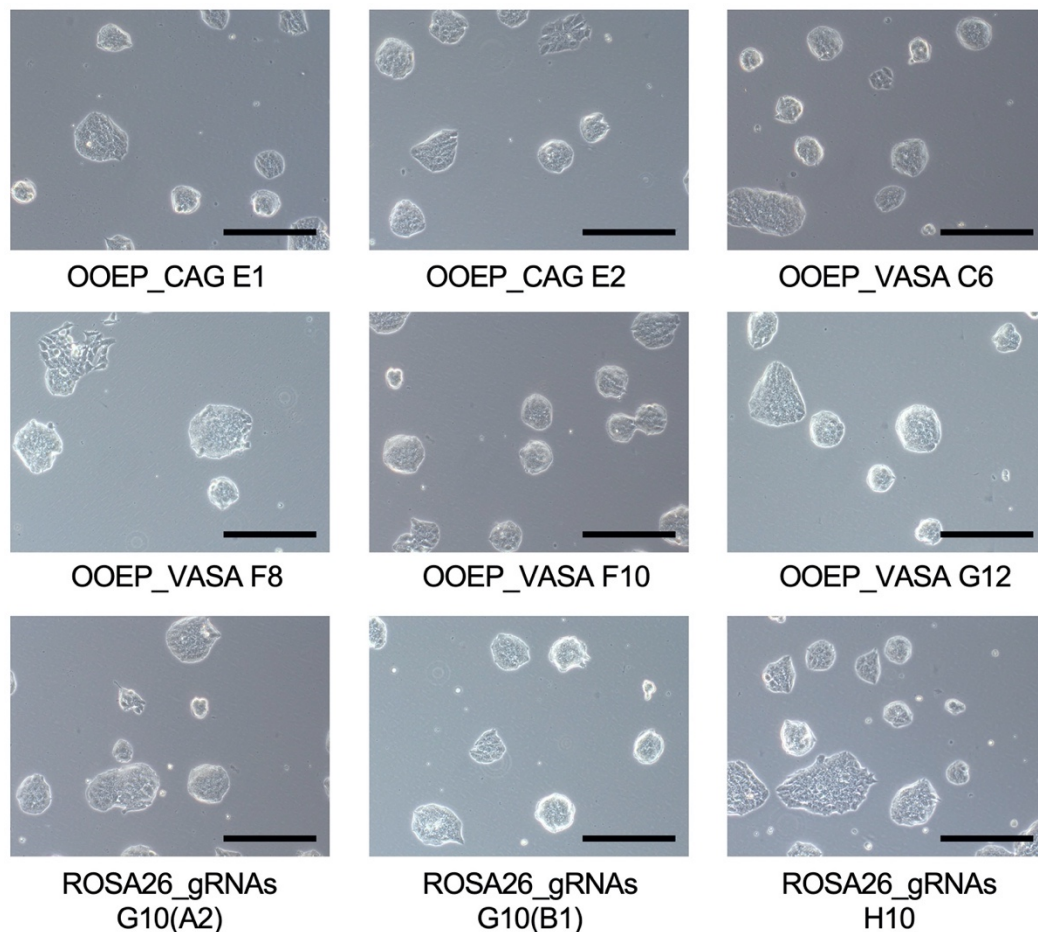


Figure 5.17 Morphology of selected ES clones plated on gelatine at low density. Images taken under phase contrast two days after thawing at passage 34. Scale bar is 200 μ m.

5.3.9 Cardiomyocyte differentiation

Following the protocol in section 2.5.1.9, each of the selected clones were differentiated into EBs then plated onto gelatine (Figure 2.3) until contracting cardiomyocytes were formed. All clones successfully differentiated into beating cardiomyocytes and this was used a simplistic method of assessing each clone's developmental capacity. Images of regions of contracting cardiomyocytes for each differentiated ES clone can be seen in Figure 5.18.

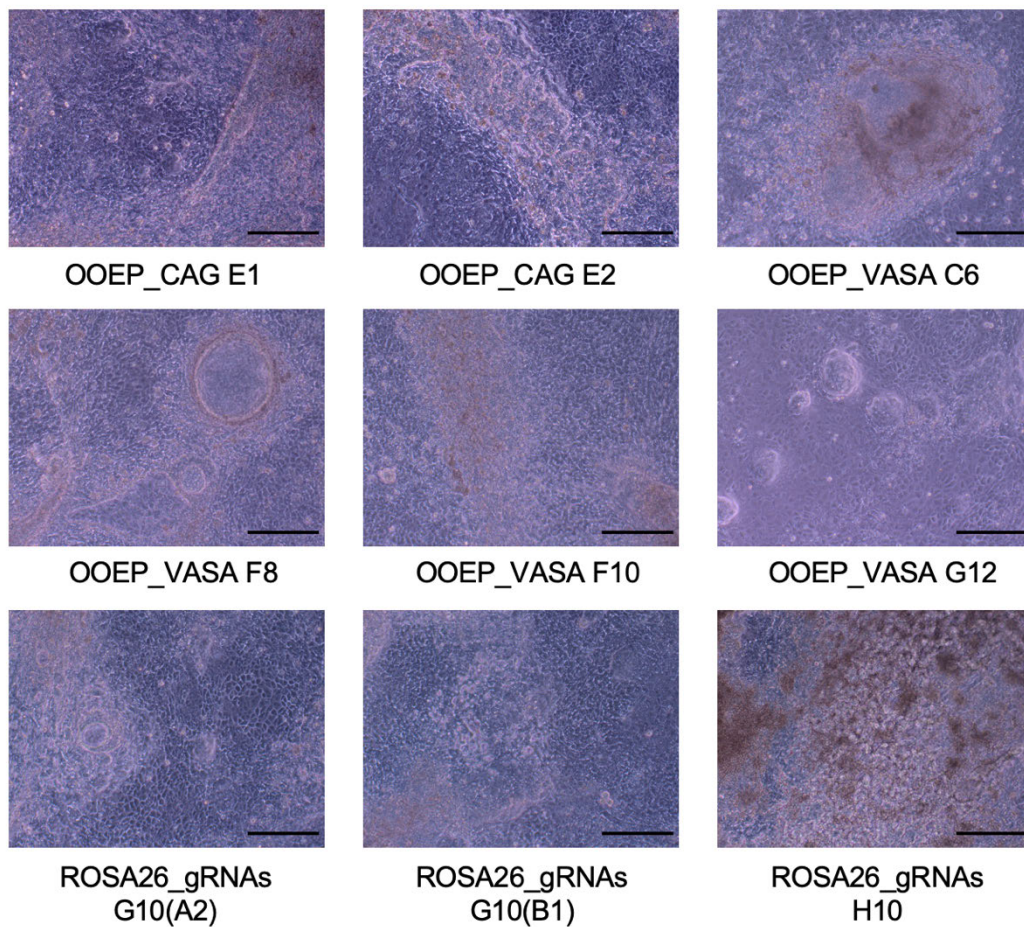


Figure 5.18 Images of regions of beating cardiomyocytes, differentiated from selected clonal ES cell lines. Images taken at day 12 after plating embryoid bodies onto gelatine. Scale bar is 200 μ m.

5.3.10 Floxed removal of *PAC* cassette

Functional testing of the constructs was then performed on clones with acceptable genotype, karyotype and differentiation capacity. This included testing the functionality of the *LoxP* sites surrounding the *PAC* cassette in each clone. To screen the functionality of the *LoxP* sites, Cre mRNA was transfected into each clone. PCR primers were designed that span the floxed cassette of each construct (Figure 5.19). In those clones with no Cre mRNA, the floxed cassette remains intact and yields a PCR amplicon of 2008 bp, 1910 bp and 1843 bp for OOEP_CAG, OOEP_VASA and ROSA26_gRNAs, respectively (Figure 5.20). In those cells treated with Cre mRNA, a high proportion of cells are transfected and the floxed cassette is excised, leading to a smaller PCR amplicon of 500 bp, 402 bp and 335 bp for OOEP_CAG, OOEP_VASA and ROSA26_gRNAs, respectively, demonstrating functionality of *LoxP* sites (Figure 5.21). All of the 9 clones screened had functional *LoxP* sites flanking the *PAC* cassette.

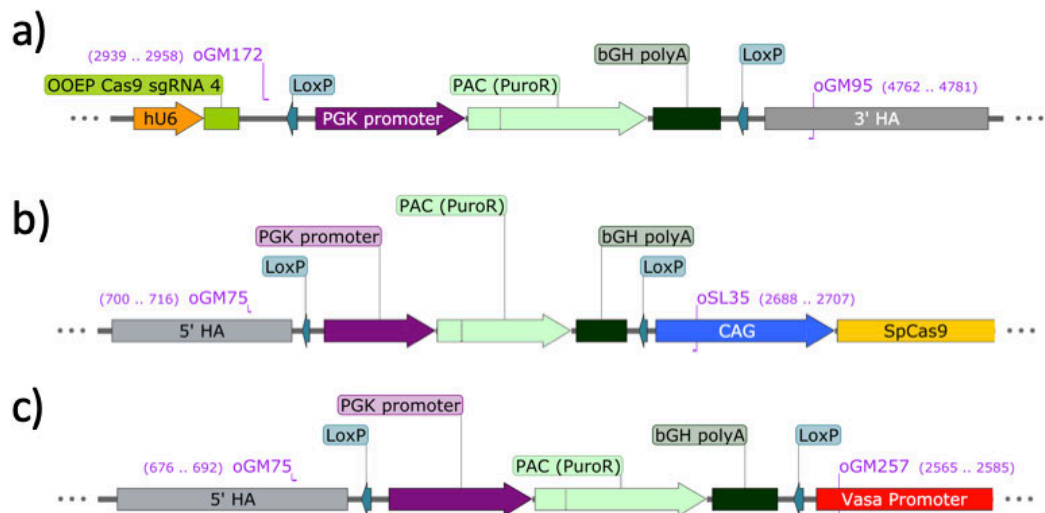


Figure 5.19 Primer design for testing the functionality of the *LoxP* sites flanking the *PAC* cassette in each cell line. a) ROSA26_gRNAs cells use oGM172 and oGM95 primers to amplify a 1843 bp product with an intact *PAC* cassette or 335 bp when the *LoxP* sites are recombined by Cre recombinase. b) OOEP_CAG cells use oGM75 and oSL35 primers to amplify a 2008 bp product or 500 bp when the *LoxP* sites are recombined. c) OOEP_VASA cells use oGM75 and gGM257 primers to amplify a 1910 bp product or 402 bp when the *LoxP* sites are recombined.

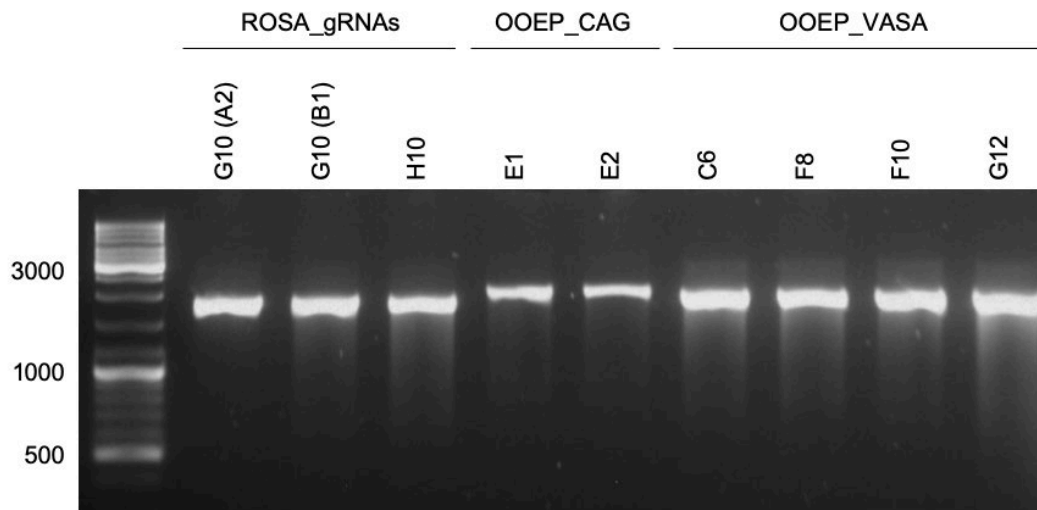


Figure 5.20 Flanking PCRs of the floxed *PAC* cassette in selected clones in the absence of Cre recombinase.

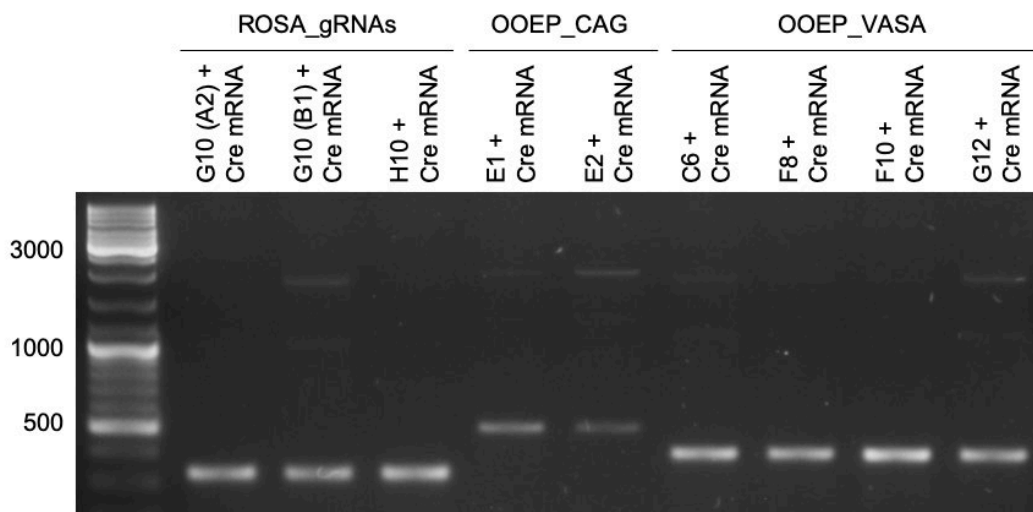


Figure 5.21 Flanking PCRs of the floxed *PAC* cassette in selected clones in the presence of Cre mRNA.

5.3.11 Functional validation of SpCas9 and gRNA expression

As no eGFP expression was observed in any of the OOEP_CAG or OOEP_VASA engineered ES clones using an epifluorescence microscope with a FITC filter (images not shown), it was necessary to undertake further screening to determine if SpCas9–T2A–eGFP was expressed in these clones. Three options were considered: (1) fluorescence

flow cytometry could be used to get better detection resolution of eGFP expression; (2) use an eGFP antibody to increase the fluorescence intensity; (3) functional testing could be undertaken by transfecting a validated gRNA (in the absence of exogenous SpCas9) into OOEP_CAG and OOEP_VASA clones. The latter option was selected.

In Section 4.3.3.1, *NLRP5* SpCas9 gRNA 4 was identified as a highly efficient guide by T7EI assay. To test if SpCas9 was expressed in the two selected OOEP_CAG clones and four selected OOEP_VASA clones, *NLRP5* SpCas9 gRNA 4 was ordered as an crRNA (Appendix A.8) and duplexed with IDT tracrRNA to form a gRNA. The *NLRP5* gRNA 4 was then transfected into OOEP_CAG and OOEP_VASA cells and a T7EI assay was performed to determine if CRISPR induced indels were present at *NLRP5*. Figure 5.22 shows that transfected *NLRP5* SpCas9 gRNA 4 complexed with endogenously expressed SpCas9 and resulted in CRISPR-mediated indels in the *NLRP5* target site in all OOEP_CAG and OOEP_VASA clones. OOEP_VASA C6 and F8, had the weakest evidence of cleavage at the *NLRP5* locus. Importantly, this assay confirmed SpCas9 was expressed in all of the selected OOEP_CAG and OOEP_VASA clones.

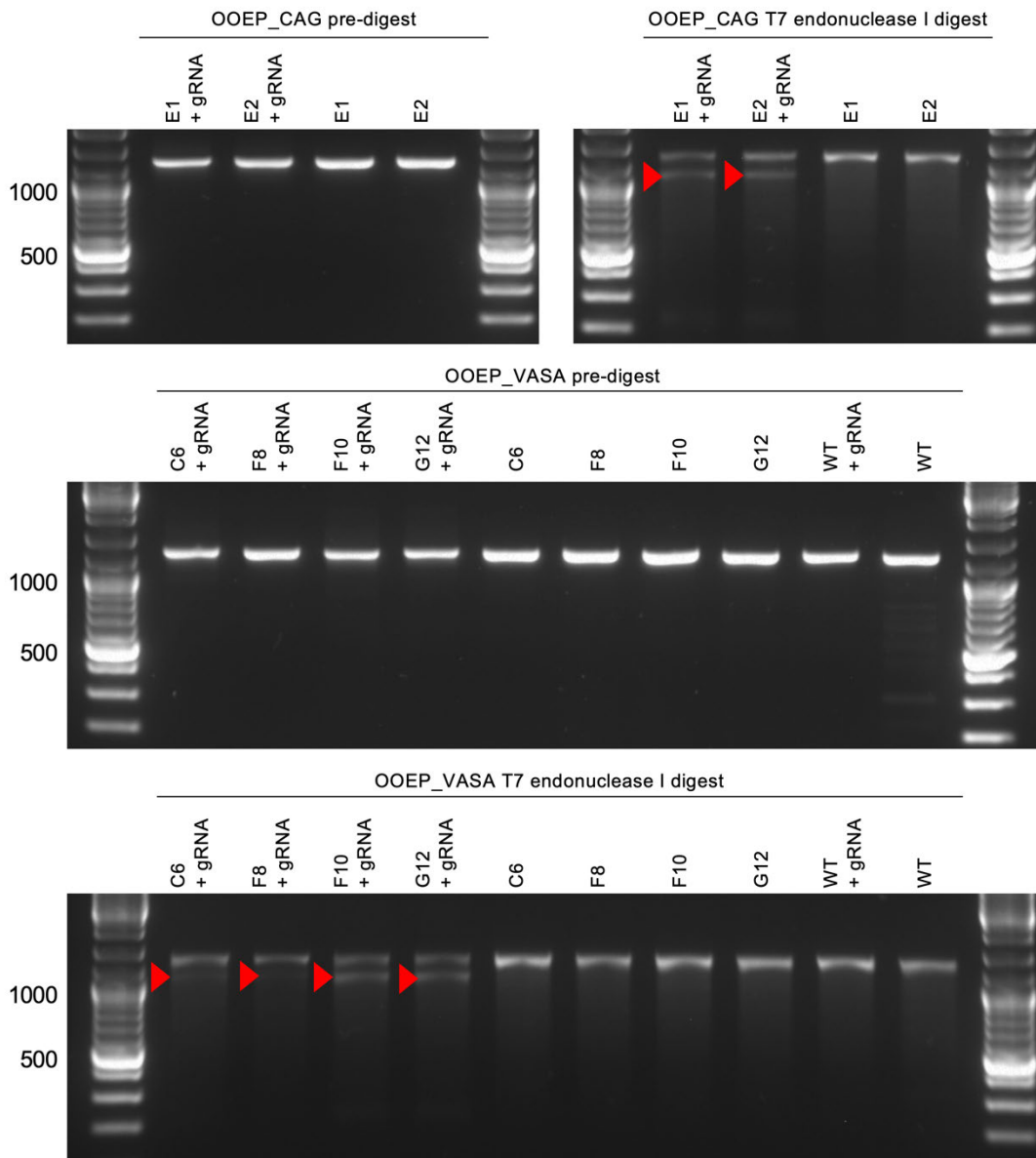


Figure 5.22 Functional analysis of SpCas9 expression using T7 endonuclease I assay on selected cell lines transfected with *NLRP5* gRNA 4 (gRNA). All cell lines transfected with the gRNA, displayed the expected second nuclease-specific band following gel electrophoresis. PCRs were performed using primer pair ppNLRP5. Red arrows indicate the expected size of nuclease-specific second band.

In a similar way, gRNA expression in ROSA26_gRNAs clones was confirmed by transfecting a SpCas9 only expressing plasmid into each of the three selected ROSA26_gRNAs clones. Expression of the *OOEP* gRNAs should result in SpCas9-mediated indels at *OOEP*. The *OOEP* target site was screened using a T7EI assay. ROSA26_gRNAs was genome engineered using an AsCas12a-assisted approach and

had not previously been exposed to SpCas9 nuclease. In Figure 5.23, it can be seen that all three ROSA_gRNAs clones showed robust *OOEP* gRNA expression, evidenced by the nuclease-specific bands in all T7EI assays.

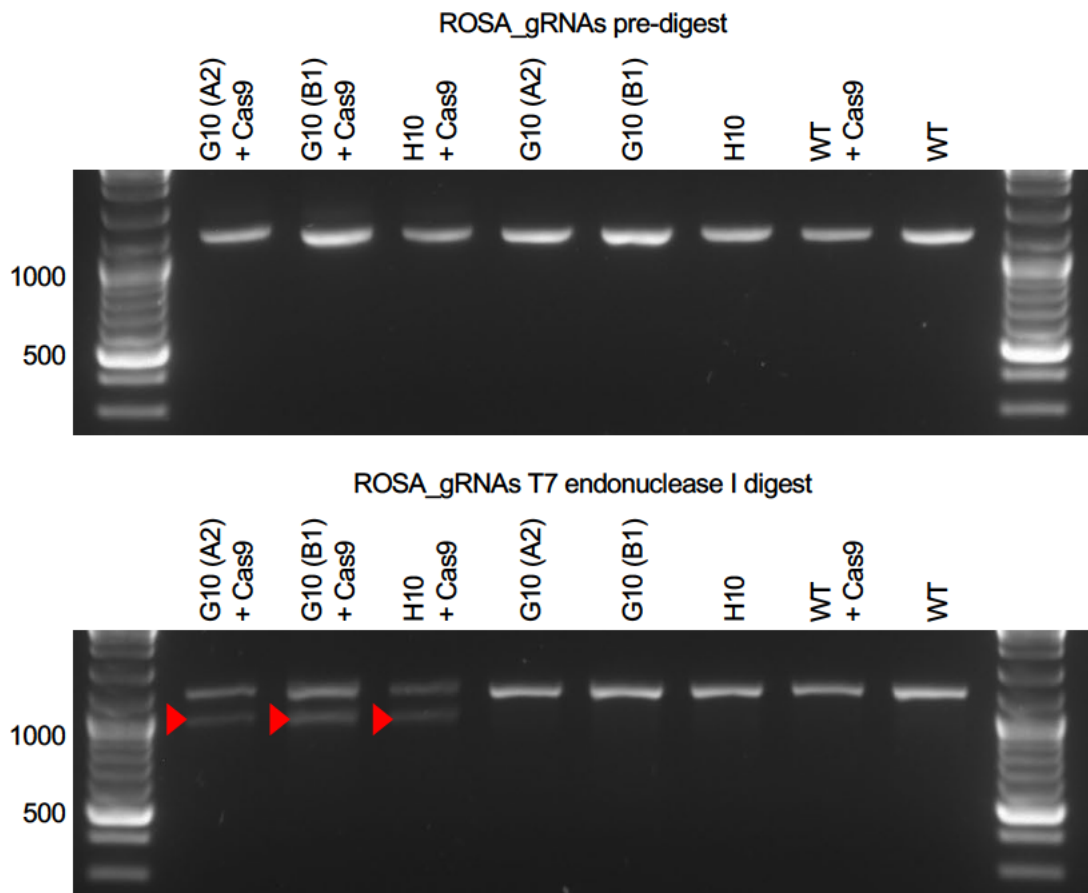


Figure 5.23 Functional analysis of SpCas9 expression using T7 endonuclease I assay on selected cell lines transfected with SpCas9 expressed from pSL70. All cell lines transfected with the SpCas9 expressing vector and no gRNA, displayed the expected second nuclease-specific band in gelatine electrophoresis. PCRs were performed using primer pair ppOOEP. Red arrows indicate approximately where the three OOEP gRNAs are expected result in nuclease-specific second band.

From this screening workflow, OOEP_CAG clones E1 and E2, OOEP_VASA clones C6, F8, F10 and G12, and ROSA26_gRNAs clones G10(A2), G10(B1) and H10 were selected as candidates for blastocyst injection to generate chimeras required for *in vivo* testing the two SGDs. Satisfied with the genotype, karyotype and functionality of constructs in each of the prospective clones, the final decision on which of these candidate clones to be used for blastocyst injection will be based on morphology and growth rate on recovery of low passage number clones.

5.4 Discussion

5.4.1 ES cell technology

ES cell technology remains the gold standard for the production of genome engineered mice that harbour large or complex constructs; however, the approach is labour intensive, time-consuming and costly (Champer, Kim, et al., 2019; Skarnes, 2015). The pipeline of transfection, isolation, screening and freezing, expanding and validating, prior to injection for the generation of chimeric offspring deters many scientists. I was asked on several occasions during this project “why don’t you try zygote injection; it will save so much time?” Excluding construct design, generation and gRNA screening, at least six months is typically required to produce and breed ES cell-derived chimeras by blastocyst injection, with no assurance that the ES cell genotype will gain germline transmission (Skarnes, 2015).

A further hindrance to ES cell technology is that not all ES cell types are robust, especially those derived from mice other than the 129 strain (Behringer et al., 2014). Often multiple independent targeted clones must be injected before successful production of chimeric pups. Genome instability of cultured ES cells is one reason for failure of germline transmission, overtly manifested by the loss or gain of whole chromosomes in culture. This highlights the requirement for karyotyping each clone prior to injection, which was performed in this project. Despite these difficulties, more than 25,000 genetically modified mouse strains that have so far been produced using ES cell technology (Skarnes, 2015).

Although direct zygotic injection is faster and cost-effective, there are benefits to ES cell technology. *In vitro* transfection protocols have the ability to engineer thousands of cells simultaneously, and the application of selectable markers such as the *PAC* cassette used in this project allows enrichment for, and subsequent characterisation of, the desired genotype, prior to commencing animal work. There have been efforts to develop well defined pipelines for generating genome engineered ES cells. The field would benefit from a stringent methodology for ES cell culture and recently Mulas et al. (2019) in Cambridge, UK detailed a standardised 2i medium composition and cell handling procedures for robust propagation and genetic manipulation of ES cells. This is a much needed contribution to the field of mouse genome engineering.

Although there are efforts to improve ES cell technology, researchers have undoubtedly embraced the relatively simplistic strategy of direct addition of CRISPR-Cas reagents to zygotes for generating rudimentary alteration. Simple indels induced by NHEJ or the integration of ssODN template by HDR for knock-ins of <100bp are now routinely carried

out in genome engineering facilities using direct injection or electroporation of zygotes. However, more complex alleles, such as gene drive constructs, that usually require HDR with a donor plasmid are more difficult to produce by zygote injection, although not impossible (Burgio, 2018; Skarnes, 2015). Typically, only a few percent of embryos co-injected with a CRISPR-Cas and a donor plasmid carry the desired targeted modification (Gu et al., 2018). Because the founder animals are often mosaic, transmission of the genetic alteration to the next generation is also not guaranteed (Mehravar, Shirazi, Nazari, et al., 2019). Thus, further improvement of CRISPR-assisted targeting in zygotes is required to rival the broad range of highly complex genetic modifications possible with ES cell technology.

5.4.2 Non-model species

Although ES cell technology is a valuable tool in mice, for nearly all non-model mammalian species, embryonic stem cells have not been derived and, therefore, gene drives for real-world application will likely require a zygotic route for genome engineering. Further emphasising the need for new or improved approaches for genome engineering large constructs into zygotes with high efficiency. The majority of emerging techniques for genome engineer are focused on creating relatively small genome alterations or corrections. Examples of these strategies are highlighted in an article I co-authored during my PhD, titled “On-Farm Livestock Genome Editing Using Cutting Edge Reproductive Technologies”. A copy of this paper can be found in Appendix A.13. Other examples of novel molecular genome editing strategies include prime editing, that reverse transcribes an RNA repair template (Anzalone et al., 2019); and SpCas9 covalently tethered to a ssODN repair template (Aird et al., 2018). The instability of RNA and ssODN templates will make it challenging to deliver the large constructs required for CRISPR-based gene drives (>5 kb) directly into zygotes using these novel techniques.

5.4.3 Genotyping

No matter the genome engineering approach used, genotyping is required and is one of the most time consuming aspects of any project. PCR screens offer a good method to rapidly determine if the cassette is present and quickly narrow down likely candidate cell lines or animals. PCR using homology arm flanking primers, spanning from inside the integrated construct to outside the homology arm, were used in this project to identify cell lines with the expected band sizes for 5' and 3' targeted integration. This approach was ideal for shortlisting candidates, but the information gained from standard PCR does not fully characterise the clones, offering no insight into the copy number of the integrated construct nor the fidelity of the integration. Southern blotting can be used to gain

information on the copy number of constructs in the genome and the expected size of the integration at the target site (Wang et al., 2015). Southern blotting was not performed in this project due to its cost and safety issues with the use of hazardous radioactive probes (see section 5.1.5.1).

5.4.4 Droplet digital PCR (ddPCR)

To overcome the drawbacks of Southern blotting, new methods are available and becoming routine in genome engineering facilities. One approach for identifying the copy number of the construct within a cell line is ddPCR and this was used in this project. Cell lines with 5' and 3' construct integration at the target site confirmed by PCR were screened by ddPCR to determine the copy number of the construct. The *PAC* gene that is present in all construct was probed for copy number. When determining copy number by ddPCR it is essential to digest the gDNA prior to analysis. When choosing a restriction enzyme, keep in mind that in most cases, the construct will integrate in a head-to-tail tandem array, known as a concatemer (Smirnov et al., 2020). If a restriction enzyme that has no sites in the construct is selected, the analysis could miss concatemers. In this project, gDNA was digested with *AluI*, which is recognised as an efficient cutter and has cleavage sites within the *PAC* transgene, allowing for identification of concatemers.

From the ddPCR analysis, clones with one copy of the *PAC* gene, and PCR confirmed 5' and 3' integration, were considered to be heterozygous with no additional copies of the construct within their genome. The wild-type homologous chromosome in the heterozygous clones was PCR amplified and Sanger sequenced within each heterozygous cell line to identify any CRISPR-induced indels present on the homologous chromosomes (Figures 5.11 to 5.13). Using this genotyping approach clones were narrowed down to four OOEP_CAG, four OOEP_VASA and three ROSA_gRNAs clones (section 5.3.5). Although one of the initial two ROSA-gRNAs clones (G10) was a polyclone and required subcloning for clonal isolation.

All samples in the ddPCR analysis had more than 10,000 droplets (individual PCR reactions) read, however, it would have been beneficial to increase the concentration of gDNA in some samples. The protocol was designed to have 0.8 copies of template per droplet. Several samples were below this calculation, for instance, ROSA26_gRNAs clones H10, one of the samples that underwent complete screening, had 13 droplets positive for *PAC* and *TRFC*, 319 droplets positive for *PAC* only, 632 droplets positive for *TRFC* only, and 11,925 with no call (Appendix A.12). Only 8% of droplets contained template. Although the 95% Poisson confidence limits for this clone are reassuringly tight, adding

more gDNA template would improve the statistics of this analysis. This variance in gDNA template is likely due to inaccurate gDNA quantification on the Nanodrop spectrophotometer. Using a DNA quantification system with improved accuracy, such as a Qubit fluorometer, would likely help in template distribution in the reaction.

This method of applying PCR and ddPCR can be used to identify heterozygous cell lines, although it does not characterise the size or sequence of the integrated construct. This approach of using PCR and ddPCR should not be used for identifying homozygous lines. It is possible that randomly integrated plasmid accompanied by a large CRISPR-induced deletion at one of the targeted alleles, which evades PCR detection, could give a false positive result. Full characterisation of homozygous cell lines requires Southern blotting or NGS technology. It is also worth noting that plasmid DNA lacking the *PAC* gene could integrate into the genome and go undetected by this screening process. Furthermore, within the ROSA26_gRNAs clones, no screening was undertaken for random integration of pY094 plasmid that provided the expression of AsCas12a and associated gRNA. A rudimentary method for screening for random integration of this plasmid would be a PCR specific to the ampicillin gene that is located on the plasmid backbone, however, this is liable to false positives due to the highly sensitive nature of PCR. If pY094 is integrated into the genome, it is possible to breed out in the resulting mice.

Overall, ddPCR provides an efficient and reliable method for high throughput screening of transgene copy number.

5.4.5 New generation screening technologies

Next generation sequencing (NGS) technologies are now being applied to genotype modified cell lines and animals. These methods can be high throughput and provide improved data outputs (Cain-Horn et al., 2017; Nicholls et al., 2019). A good example of an NGS based genotyping approach for cell line screening is Targeted Locus Amplification (TLA) technology, where gDNA is cross-linked, fragmented and ligated before sequencing using outward facing primers complementary to a sequence within the transgene construct. Tens to hundreds of kilobases of the transgene and surrounding gDNA can be sequenced with TLA. This enables robust detection of SNPs, concatemers and random integration sites in the genome (Hottentot et al., 2017). TLA is therefore a useful method in genome engineering projects when comprehensive or allele-specific genetic information is needed; for instance, gene drive animals for potential real-world application.

Another NGS technology, Oxford Nanopore Technologies (ONT) sequencing produces long reads which can easily cover the entire length of a gene drive construct in one molecule (Jain et al., 2018). Although ONT is gaining traction in many biological fields, it remains plagued by high error rates. McCabe et al. (2019) showed that the high error rates inherent to ONT sequencing can be offset by very deep sequencing coverage. They established a workflow analysis of sequencing data to identify the genomic DNA samples that include the correct allele for accurate genotyping of genetically modified mice. ONT sequencing could also be applied to ES cell clones prior to injection. Several web-based online platforms have been developed to analyse the NGS data, including CRISPR-GA, BATCH-GE, CRISPResso, Cas-analyzer and CRISPRMatch (Li et al., 2019). However, even though effective, NGS-based analysis still requires multi-step operations and are costly in both time and money.

Although not imminent, with the continuing reduction in cost of whole genome sequencing (WGS), this NGS technique may become a viable option for validating genome engineered cell lines and animals in the future. WGS is the only method to fully characterise the genome and is rarely deployed, only in high-value projects, such as genome engineered livestock and ex vivo engineered therapeutic cell lines (Ashmore-Harris & Fruhwirth, 2020; Norris et al., 2020). The main barriers of WGS remain cost of the library preparation and downstream analysis, including genome assembly. With increases in computation power and the development of user-friendly software pipelines for rapid analysis of WGS data, what is now considered 'overkill' for most projects may soon be the norm.

5.5.6 Ribonucleoprotein (RNP)

Within this study, SpCas9 delivered as RNP had significantly better knock-in efficiency than plasmid delivered AsCas12a. Although these differences in knock-in efficiency could be locus, vector or nuclease dependent, it is worthwhile highlighting some of the reported differences in RNP delivered CRISPR-Cas rather than Cas and gRNA expressed from plasmid vectors.

Plasmid delivery approaches are considerable cheaper but can be associated with higher levels of unintended edits (Lino et al., 2018). This editing is likely due to the relatively long period that plasmid persists inside the cells, continuing to express Cas and gRNA, which provides ample opportunity for off-target cleavage to occur. Several studies have demonstrated that the frequency of off-target mutations is lower when CRISPR-Cas is delivered as RNP rather than plasmid DNA (Kim, Kim, Cho, et al., 2014; Liang et al.,

2015). For instance, Liang et al. (2015) found that for the gene OT3-18, the ratio of off-target to on-target mutations was 28-fold lower when using RNPs relative to plasmid delivery of SpCas9.

Another potential problem with plasmid transfection is the random integration of all or part of the plasmid DNA into the genome of targeted cells. It is desirable to limit the amount of dsDNA entering cells to reduce the likelihood of this occurring. Kim, Kim, Cho, et al. (2014) found a prevalence of plasmid-derived insertions at off-target sites, transfection using RNPs avoids the risk of the CRISPR-Cas expression system integrating into the genome.

SpCas9 RNP also has been reported to have increased HDR efficiency, which was certainly the case in this project. Differences in efficiency is likely due to the immediate activity of the RNP. When RNP is delivered to the nucleus with a donor plasmid, RNP stimulates HDR at the target site without delay, when fresh repair template is readily available. Published findings show a consist improvement in knock-in efficiencies (Kim, Kim, Cho, et al., 2014; Liang et al., 2015; Schumann et al., 2015). These findings align with improved levels of HDR seen when using RNP in this project.

5.4.7 Techniques for increasing HDR efficiency

HDR was at an acceptable level in this project when using a puromycin selection, however, it is unlikely that selection could be implemented when engineering gene drives for real world application in wild species, due to the lack of ES cell technology available in these species. Therefore, some discussion on strategies aimed at improving HDR efficiency in the absence of antibiotic selection is provided here.

If gene drives are to be engineered directly in embryos, introduction of CRISPR reagents at the 2-cell stage of development rather than in zygotes has proven successful at increasing HDR (Gu et al., 2018). The rationale for performing the microinjections at the 2-cell stage is twofold. First, at this developmental time point, the embryo undergoes its major genome activation, which is associated with an increased open chromatin state that may improve the accessibility of the CRISPR-Cas and the HDR template to the target site. Second, HDR is typically restricted to the S and G2 phases of the cell cycle. Embryos display an exceptionally long G2 phase of the cell cycle during the 2-cell stage, which could provide an extended time for HDR repair to occur (Plaza Reyes & Lanner, 2018). Gu et al. (2018) showed that in mice the traditional approach where injection is performed in the fertilized zygote generated 1 to 6.5% knock-in efficiencies of both alleles whereas 2-cell stage injection resulted in 31.5 to 35% targeting efficiencies.

One concern related to the 2-cell stage embryo injection might be the plausible increase in mosaicism compared to zygote delivery, as there are now two independent microinjections and four potential genome editing events that need to be resolved in the 2-cell embryo (Plaza Reyes & Lanner, 2018). Furthermore, performing the microinjection at the 2-cell stage includes an increased technical difficulty, which might make its implementation in wild species more challenging when dealing with embryos where little embryology is known. Finally, it will be important to explore if this finding extends to other species or if this is a particular feature of 2-cell stage mouse embryos.

An alternate or complementary approach for boosting HDR is the use of chemical and biochemical compounds to interfere with the DNA repair machinery. These work either by blocking the NHEJ pathway, with siRNA or inhibitors like SCR7, or by stimulating the HDR pathway, with HR effectors like RAD51 or with stimulating molecules like RS-1 (Song et al., 2016; Wilde et al., 2018), or by delivering the DNA template in a vector designed to make use of alternative DNA repair pathways like microhomology-mediated end-joining (MMEJ) or single strand annealing (Nakade et al., 2014; Quadros et al., 2017; Sakuma et al., 2016).

One example of a technique that exploits the MMEJ repair pathway is Precise Integration into Target Chromosome (PITCh) by Sakuma et al. (2016) that improves efficiency of plasmid donor template integration in both ES cell and zygotes or embryos. PITCh uses MMEJ-directed repair vectors with short microhomology arms (10–40 bp) flanking the construct. Three DSBs are necessary for knock-in a PITCh vector: one on either side of the construct to excise it from the carrier plasmid and one in between the 5' and 3' microhomologies in the genomic locus. The first two breaks can be induced via a generic PITCh-gRNA; the third break requires a gRNA specific to the target site in the genome. These DSBs allow the microhomology arms (5' and 3') to anneal, knocking the construct into the locus by a MMEJ repair pathway. Although HDR efficiencies for PITCh are high, with a report of 12% knock-in efficiency for a 5 kb construct in mouse zygotes (Aida et al., 2016), one potential drawback of PITCh is the frequency to which linearized plasmid can randomly integrate into the host genome. More data is needed on this.

5.4.8 Karyotyping

One of the main concerns with ES cell technology is on injection cells fail to generate chimeras or gain germline transmission from the founding chimeras. These failures are often attributed to the chromosomal abnormalities (Suzuki et al., 1997). Classically, analysis of cell karyotype is by Giemsa stained metaphase spreads (Bakker et al., 2015).

This was employed in this project to identify the percentage of aneuploidies in each clone. To assist with spreads, the mitotic arrestant colcemid, which greatly increases the frequency of mitotic spreads was used; however, this has the concomitant and deleterious effect of contracting the chromosomes, thus making detailed karyotyping of the chromosomes challenging. Precise G-banding, in the absence of a mitotic arrestant, with the resolution to detect inter- and intra-chromosomal abnormalities requires extensive training, is instrument and labour intensive, and is therefore often outsourced to commercial laboratories (Behringer et al., 2014).

In this project, 25 metaphase chromosome spreads were counted for each clone under 60x phase microscope. A threshold of no greater than 50% aneuploidy counts in each clone was used. It would be desirable to increase the number of spreads counted to improve the statistical power. For instance, previous reports have used 30 metaphase spreads per clone, but they reported no difference in outcome when counting either 30 or 50 spreads. The 50% threshold for acceptable clones for injection was initially proposed by Cotton et al. (2015) based on 8 clones in their study with less than 50% euploid cells that did not yield germline transmission. Codner et al. (2016) reconfirmed that the threshold for efficient germline transmission is acceptable at 50%. In their study of 708 JM8 ES cells clones, those clones that had >50% euploidy had 80% chance of germline transmission, while in clones with less than 50% euploidy, germline transmission drops to under 20%. Although, 50% was the predetermined cut-off, the OOEP-CAG H9 clone was precautionarily excluded from further screening with 55% euploidy spreads and 44% with a 41 chromosome count.

An additional finding from Codner et al. (2016), which looked at 708 JM8 ES cell clones, was that over 99% of total aneuploidy events involved at least one of these four chromosomes (1, 8, 11 and Y), with trisomy of chromosome 8 being the most common. Clones OOEP_CAG B3 and H9 in this project may carry this trisomy, as 56% and 44% of their spreads, respectively, had 41 chromosomes. Codner et al. (2016) devised and validated a ddPCR protocol for evaluating chromosome copy number from ES cell-derived gDNA. In their protocol, the copy number of a single gene on each aneuploidy-prone chromosome (1, 8, 11 or Y) is assayed relative to that of a calibration gene (present on either Chr 10 or Chr 17), each shown to be generally maintained at a stable diploid state by cytogenetic karyotype analysis. Their technique enables laboratories that are non-specialist, or work with large numbers of clones, to precisely screen ES cells for the most common aneuploidies prior to microinjection. This method would have been valuable to use in this project.

5.4.9 Morphology and *in vitro* differentiation

The indefinite self-renewal ability and plasticity of ES cells allows for *in vitro* generation of an unlimited number of cell types. This ability has caught much interest recently as the generation of high-quality specialised cells through robust and reproducible directed differentiation protocols is in demand for both research and biomedical applications (Zakrzewski et al., 2019). This demand is reliant on protocols that can maintain ES cells in a pluripotent state, without unwanted differentiation. Although modern ES cell culturing protocols and media formulation substantially improve our ability to maintain pluripotency, suboptimal conditions can lead to differentiation and some differentiation can be seen in clone OOEP_VASA F8 in Figure 5.17. This could be due to a raft of factors, including lack of passaging or poor recovery on thawing. As there are alternative OOEP_VASA clones, it is unlikely that the F8 clone will be used for blastocyst injection.

In this project, this capacity of ES cells to differentiation into specific cell types was used to gain further functional insight into the developmental potential of the genome engineered ES cell clones that were generated. Each genotyped and karyotyped ES clone was differentiated into contracting cardiomyocytes as a method of gauging their developmental potential. All clones were differentiated into contracting cardiomyocytes via an EB intermediate within 14 days. An alternative method for analysis of the pluripotent potential of these ES cells would be to screen for the expression of pluripotent markers, such NANOG and OCT4, or alkaline phosphatase staining (Ghimire et al., 2018).

5.4.10 Functional validation

Finally, selected clones were screened for functionality of the constructs they harbour. It was concerning that eGFP was not visible under an epifluorescence microscope with FITC filter in any of the OOEP_CAG or OOEP_VASA clones that should express SpCas9–T2A–eGFP. Fortunately, SpCas9 expression in these clones could be tested by transfection with a validated gRNA. *NLRP5* SpCas9 gRNA 4 was transfected into two OOEP_CAG and four OOEP_VASA clones and a T7EI assay performed on the gDNA extracted from the cells. All clones cut the *NLRP5* target site, indicating that SpCas9 was expressed in the two OOEP_CAG and four OOEP_VASA clones tested. In a similar way, expression of the gRNAs in ROSA-gRNAs clones was confirmed by transfecting a SpCas9-only expressing plasmid followed by a confirmatory T7EI assay at the *OOEP* target site of the endogenously expressed gRNAs.

LoxP sites were also tested for their ability to excise the floxed *PAC* cassette. It was plausible to remove the floxed *PAC* cassette in each ES clone prior to injection by

transfection of clones with Cre mRNA and isolation of clones with excised floxed cassettes. However, to keep the passage number of all clones as low as possible, excision was not performed *in vitro* and only functional testing of *LoxP* sites were performed. After blastocyst injection of the engineered ES cells, it is possible to remove the floxed *PAC* cassette *in vivo*, by crossing the transgenic lines with VASA-Cre mice that specifically express Cre recombinase within their germline.

Undoubtedly, one of the key benefits that ES cell technology offers is the scrupulous level of genotyping and functional testing that can be undertaken before commencing *in vivo* work. Although the ES cell lines generated in this project were well characterised it does not guarantee these cells will gain germline transmission. There are a number of factors that can prevent germline transmission. The karyotyping assessment in this project was basic and the ES cells used were at a high passage at the outset (passage 24). Although previous mice have been generated at The Roslin Institute with this parental E14 ES cell line at a passage number up to 43, it would be desirable to use ES cells with lower passage number and this would certainly be a consideration if I were to utilise ES cell technology in future projects. Furthermore, the potential of the ES cells developed in this project will be greatly influenced by embryo manipulation and culture conditions during microinjection.

5.4.11 Conclusions

Although ES cell technology requires significant time and effort for generating genome engineered mice, it remains the benchmark and has played a central role in developing complex mouse. Accurate genotyping and karyotyping are key hurdles for most genome engineering projects and a range for new technologies, including ddPCR and NGS technologies are helping to increase throughput and improve data. Despite its time consuming nature, ES cell technology provides an excellent path to test and tune gene drive strategies in mice but technologies capable of delivering large constructs specifically and efficiently into target sites within zygotes will likely be need for real-world application of gene drives in wild species.

6 Discussion

This project set out to investigate the potential of gene drive technology as a pest management tool in rodents. This involved designing, modelling and developing a safeguarded gene drive strategy that could spread female infertility through a laboratory-contained mouse population. In Chapter 3, gene drive strategies were designed and modelled *in silico*. The reagents for developing the two selected split gene drive (SGD) strategies were then generated and validated in Chapter 4. Within Chapter 5, mouse ES cells were engineered to harbour the SGD elements, and cell lines are now ready to generate two SGD mice models for *in vivo* assessment.

6.1 Ambitions and realities

At the outset of this project, it was my ambition to generate a founding cohort of SGD mice for a future *in vivo* study. This ambition was not realised, although substantial progress was made towards achieving this goal in the allotted 3-year timeframe. Being the first gene drive project at the University of Edinburgh and the first vertebrate gene drive project within the UK, there were significant delays due to intensive, and justified, biosafety and ethical scrutiny. The project was evaluated by The Roslin Institute Biosafety Committee, the University of Edinburgh Biosafety Committee, the University of Edinburgh Animal Welfare and Ethical Review Body (AWERB) committee and the Home Office. This involved numerous presentations and volumes of documentation, including nine separate risk assessments. Although steps towards obtaining approval for generating SGD mice commenced in June 2017, final approval in the form of Project License (PPL) from the Home Office was not granted until the 14th of January 2020. As of August 2020, the SARS-CoV-2 outbreak had stalled impending animal experiments.

6.2 Planned *in vivo* testing

Because of these delays, there remains significant work to be undertaken in order to test the SGDs developed in this project *in vivo*. Steps that remain to be undertaken include generating the SGD mice using the three engineered ES cell lines (OOEP_CAG,

OOEP_VASA and ROSA26_gRNAs), and studying the transmission pattern and phenotype conferred by the SGD in OOEP_CAGxROSA26_gRNAs and OOEP_VASAxROSA26_gRNAs mice. These steps are not trivial, and include blastocyst injection of the ES cells, obtaining germline transmission from the F0 chimeras, genotyping F1 offspring, potentially crossing ES cell-derived mice with VASA-Cre mouse lines to remove the floxed *PAC* cassette, crossing subsequent lines to establish a cohort SGD mice, which will require ongoing genotyping and phenotyping. Figure 6.1 shows a schematic overview of the planned *in vivo* testing.

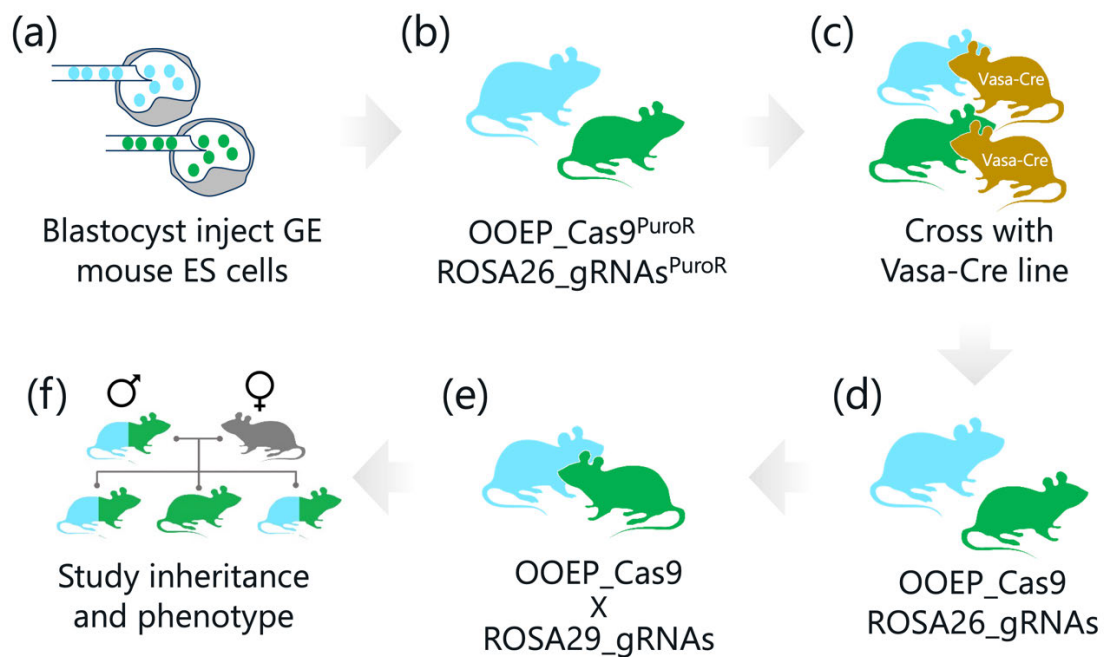


Figure 6.1 Schematic overview of planned *in vivo* work. a) Blastocyst inject confirmed OOEP_CAG, OOEP_VASA (collectively labelled OOEP_Cas9 in figure) and ROSA26_gRNAs mouse ES cells. b) Recipient females give birth to three separate heterozygous mouse lines: OOEP_CAG^{PuroR}, OOEP_VASA^{PuroR} (collectively represented as blue) and ROSA26_gRNAs^{PuroR} (green). All three lines harbour the floxed *PAC* cassette (PuroR). c) Each line will be crossed with Vasa-Cre line (brown) to excise the floxed *PAC* cassette. d) Establish small breeding colonies of each of the three OOEP_CAG, OOEP_VASA and ROSA26_gRNAs lines. e) Cross OOEP_CAG and OOEP_VASA lines with ROSA_gRNAs line to generate the two split gene drive lines: (1) OOEP_CAGxROSA_gRNAs and (2) OOEP_VASAxROSA_gRNAs. f) Breed split gene drive male mice with wild-type females (grey) to produce up to 8 litters of pups (approximately 48 pups in total for each split gene drive). The transmission frequency of the split gene-drive will be assessed by PCR-based genotyping of ear clippings. Once split gene-drive female pups reach sexual maturity (6 weeks), a small cohort of pups will be mated to establish their fertility status. Females presenting an infertility phenotype will be mated up to 4 times to confirm this infertility status.

It should be noted that the ES cell lines generated in this project, although well characterised, are by no means guaranteed of achieving chimera production or germ line transmission of the ES cell-derived genotype. This is an issue that has plagued ES cell technology since its inception. If OOEP_CAGxROSA26_gRNAs and OOEP_VASAxROSA26_gRNAs mice are successfully generated, male founders will be mated with C57BL/6NcrJ (B6) wild-type females to produce 8 litters of pups for each of the two SGD models. Anticipated *in vivo* experiments have been well thought out due to the rigorous internal and external assessment of this project and details on this can be found in section 3.3.7 and within the non-technical summary from the granted PPL in Appendix A.9. The AWERB committee approved the inclusion of this redacted non-technical summary within this thesis.

The C57BL/6NcrJ wild-type mice to be used are genetically compatible with both SGDs, having conserved DNA sequences at OOEP gRNA target sites (Section 4.3.3.2). As this is a proof-of-concept study, the breeding plan was knowingly underpowered to reduce animal numbers. For a gene drive to be successful as a pest management tool, it will be required to propagate at a rate much greater than 50%. It is expected that this effect could be seen in the selected sample size. Assuming each of the 8 litters contains 6 pups, this equates to 48 pups for each SGD strategy. If successful in this initial proof-of-concept study, there is potential to undertake an additional study to statistically validate the best performing SGD system.

The SGD approaches developed in this project feature multilayer safeguard design. After identifying the two SGD strategies and generating constructs, Grunwald et al. (2019) published the first demonstration of a CRISPR-based gene drive in mice. Of the gene drive strategies they reported, self-propagation was only observed in the female germline when expression of SpCas9 was restricted to the germ cells. In this project, female germline homing in OOEP_VASAxROSA26_gRNAs mice may not be detected as all females that inherit the SGD should be OOEP knock-out within their germ cells and infertile. As female infertility in OOEP knock-out mice results in spontaneous abortion at the 2-cell embryo stage, embryos from OOEP_VASAxROSA26_gRNAs cannot be genotyped. OOEP_VASAxROSA26_gRNAs females could be phenotypically assessed for their fertility status, but it will be unknown if infertility was due to HDR-mediated homing, or NHEJ or MMEJ-based indel mutations. This issue is not anticipated with OOEP_VASAxROSA26_gRNAs males, which should retain normal fertility, or OOEP_CAGxROSA26_gRNAs mice where homing for both sexes is anticipated to occur in the zygote or early embryo.

Once generated, SGD mice will be maintained at The Roslin Institute, which has a modern, well-resourced and environmentally stabilised rodent facility. Husbandry will be performed by experienced animal technicians. Careful selection of the targeted female-fertility gene (*OOEP*) based on previous studies that have shown *OOEP* knock-out mice are phenotypically normal other than female infertility indicates animals are unlikely to suffer in any way. Furthermore, as the embryo in *OOEP* knock-out females aborts at the 2-cell stage it is unlikely to be associated with any adverse clinical signs such as uterine infection and is unlikely to be physiologically recognised as a pregnancy or influence subsequent ovulation cycles.

6.3 Value of fundamental research

The research progress made in this in this project and its future objectives can be classified as applied science, and like all applied sciences, it relies on fundamental research. A wide variety of selfish genetic elements occur naturally in many types of organisms and, although fundamental biologists have been studying these elements since the 1920s their origins and our understanding of the molecular mechanics remains basic (Burt & Trivers, 2006; National Academies of Sciences, 2016). Selfish genetic elements are a fascinating and a particularly challenging field of genetics. Continued research into this field, along with replenishment of interest and funding due to the publicity surrounding gene drive technology will likely identify new classes of naturally occurring selfish genetic elements and improve our understanding of how these elements function. This improved understanding could guide the development of new more robust synthetic gene drive strategies.

In addition to advancing our fundamental knowledge on naturally occurring selfish genetic elements, population genetics and ecosystem dynamics are two fields of study that are essential to determining the efficacy of gene drives and their biological and ecological outcomes. There are considerable gaps in knowledge regarding the implications of gene drives for species fitness, gene flow in and among populations, and the dispersal of individuals, and how factors such as mating behaviour, population structure, and generation time could influence a gene drive's effectiveness (National Academies of Sciences, 2016). The population simulation performed in this project did not incorporate many of these important factors. Addressing these knowledge gaps to gain better understanding of the biological and ecological outcomes of gene drivers will require non-trivial interdisciplinary research.

6.4 Interdisciplinary nature of gene drive research

To emphasise the importance of interdisciplinary research when developing gene drives, simply look at the contents of this thesis, written by an emerging biotechnologist, who has the skills to design gene drives, without necessarily a deep understanding of population genetics or ecosystem dynamics. Yet, these parameters are required to accurately understand the effects of gene drives on populations and ecosystems. The power of interdisciplinary collaboration was briefly highlighted in Chapter 3 of this document. Within this Chapter, collaboration with Dr Gregor Gorjanc and Nicky Faber, both population modellers, led to valuable simulations of gene drive strategies in wild mouse populations. Fields that will be required to collaborate for gene drive technology to be realised, include: fundamental geneticist, population modellers, molecular biologist, ecologists, ethicists, animal technicians, regulators, policymakers, and communication specialists. This requirement for interdisciplinary interaction will be both rewarding and challenging.

6.5 Phased testing pathway

With the recent successes in developing gene drives in laboratory, there is increasing interest in progressing this technology to be ready for real-world applications. However, the path towards applying gene drives requires a phased testing approach, such as the one developed by the World Health Organization (WHO) for testing genetically modified mosquitoes (World Health Organization, 2014). Each step in such a pathway promotes careful study and evaluation, includes checkpoints to determine whether research should move to the next phase. A phased testing framework can also be adapted to laboratory research on gene drive. For example, in this project, experiments aimed to start with by developing a SGD in laboratory-contained inbred mouse strain, to avoid issues associated with a failure of containment, before considering the develop and testing of any self-sustaining gene drive strategies in wild mice.

In accordance with the WHO phased testing pathway for genetically modified mosquitoes, an idealised pathway to develop of gene drives for real-world applications includes five steps: Phase 0 – Research Preparation: includes identification of ideal attributes and containment strategies, establish risk assessment needs and plan regulatory requirements; Phase 1 – Laboratory-based research: includes using computational methods to model gene drives, undertaking risks assessment and obtaining regulatory approval, developing containment strategies, optimising gRNA design and enhancing HDR efficiency, evaluating effects on animal fitness and measuring gene drive stability over multiple generations. The work in this project pertained to Phases 0 and 1. The

remaining three phases of WHO testing pathway, include: Phase 2 – Field-based research, Phase 3 – Active environmental release and Phase 4 – Post-release surveillance.

6.6 Risk assessment

A common concern raised with field-trials and active environmental release of gene drives is that it will be challenging to accurately predict the consequences that could arise from gene drive animals as the environments they will be released into do not exist as ‘ecological vacuums’ (Webber, Raghu, & Edwards, 2015). Individual species are connected to other species in the community through direct trophic links (e.g. native eagle preys on introduced mice) and through indirect trophic links (e.g. introduced mice competes with native rodent for the same resources). These links create dynamic feedbacks that affect the relative abundances of different species (Wootton, 1994). These feedback loops can become highly complex and makes accurate prediction difficult (Scheffer, 2009). Although prediction is difficult, the theoretical and empirical insights that ecological risk assessments can provide will play a critical role as testing progresses beyond laboratory-contained research.

In addition to the assessing likely ecosystem changes from a gene drive release, good ecological risk assessment will also make comparisons between alternate strategies, incorporate relevant people’s concerns, and can be used to identify sources of uncertainty, giving it the ability to quantify the probability of different outcomes. Reliable data and robust models are particularly crucial for ecological risk analysis, which is challenging to obtain for gene drives. Contained laboratory studies, such as the proceeding *in vivo* study from this project, and *in silico* modelling currently represents the best approaches for providing data to reduce uncertainty.

Although a significant proportion of this project was dedicated to risk assessment, no in-depth ecological analysis was undertaken. This project aimed to develop two SGDs in an inbred laboratory mice within a secure rodent facility. The SGD was not self-sustaining and, therefore, the risk this project posed to the environment was assessed to be no greater than a conventional genetically altered laboratory mouse.

6.7 Governance and regulations

With gene drive research moving rapidly, particularly in mosquitoes focused at controlling malaria, it is likely that field trials will soon be undertaken. It is essentially that adequate regulations and governance for the environmental release of gene drive organisms are in

place before this occurs. The governance of research begins with the personal responsibility of investigators, which is formalised in professional guidelines and can be legally enforceable. Currently, institutions, funders, and professional societies work to encourage professional best practices in research. The research in this project was conducted at The Roslin Institute, that receives funding from the BBSRC and is part of the University of Edinburgh. The Roslin Institute, therefore, is subject to University of Edinburgh and BBSRC's Biosafety Policies. Beyond the professional, institutional and funding body guidelines, there remains much regulatory uncertainty regarding gene drive technology around the globe (Brossard et al., 2019).

In many countries, there are regulations in place to guard against unintended effects of releasing genetically engineered organisms into the environment. These regulations were developed for animals that have conventional genetic modifications that are typically for welfare or human benefit and reduce the fitness of the animal outside a controlled environment. Current regulations call for a step-by-step process where the first releases are in small numbers in isolated areas (Brossard et al., 2019; National Academies of Sciences, 2016). With gene drives, the intention is for the engineered animal genotype to spread, to some extent, in the environment, so there is a need to amend established regulations. At present, no nation has regulations in place specifically for gene drives and no case of release of an organism with a gene drive has been recorded (Brossard et al., 2019).

The most broad-ranging and widely accepted international governance system for genetic resources and biosafety is the United Nations Convention on Biological Diversity (UNCBD), as implemented through the Cartagena and Nagoya Protocols. Many countries are now developing regulatory systems in response to the Cartagena Protocol. Many such systems are predicated on a strong precautionary, nearly preventative approach (National Academies of Sciences, 2016). To cope with the unique aspects of gene drives, existing approaches to governance need to be adapted. Integrating new policy and law for gene drives into existing international governance poses significant challenges but it is necessary, and this need was recently highlighted in April 2020 by a UNCBD Ad Hoc Technical Expert Group (AHTEG) Report on Risk Assessment.

6.8 Engagement

Gene drives represent a case of post-normal science for which technical expertise alone is not enough to address the intricacies surrounding a scientific issue. Consideration should be given not only to technical concerns but also social and ethical implications.

Unlike normal scientific issues for which risk assessment can be primarily based on scientific inputs, post-normal science requires input from a multitude of perspectives when assessing risks and benefits (Brossard et al., 2019; Ravetz, 1999). The decision to deploy a gene drive to suppress a pest species will involve more than a technical assessment of the risks involved, and responsible decision-making regarding their use will require consensus from communities, stakeholders and publics (Figure 6.2).

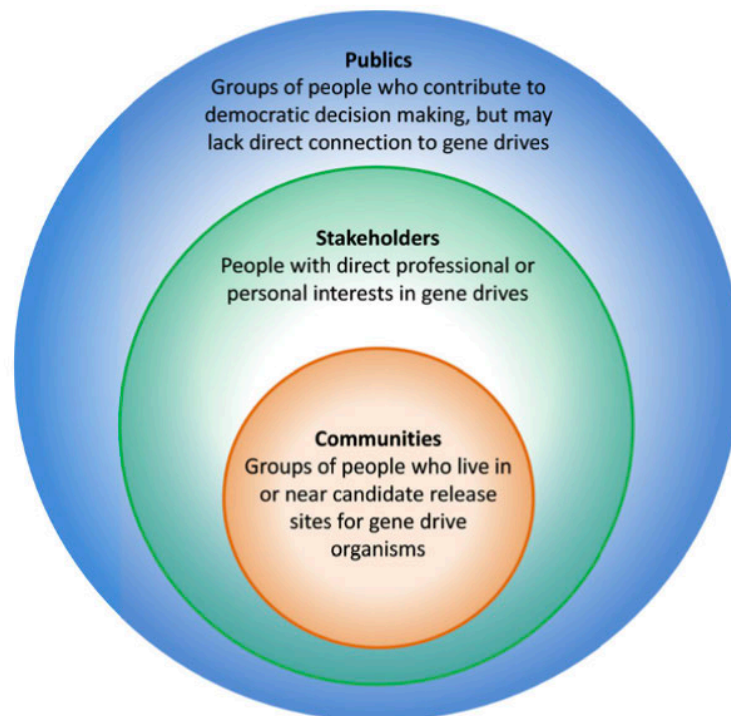


Figure 6.2 Communities, publics, and stakeholders. These audiences for engagement are based on geographic proximity and interests. Image from National Academies of Sciences (2016).

Engaging with communities, stake holders and publics should be considered an essential part of any gene drive project. The outcomes of engagement can be as crucial as the scientific outcomes for decision making. This project did include media engagement. At the start of this project, I outlined the goals and strategy of my PhD research in a press conference at the Science Media Centre at Wellcome Centre in London. This press conference was linked to an article I co-authored, titled “CRISPR-Based Gene Drives for Pest Control” (Appendix A.1), and lead to several press articles, including write ups in the Times, Guardian, Financial Times, Telegraph and The Australian, along with significant

social media attention. This helped inform publics, stake holders and communities, encouraging discussion and debate on the topic. Over the course of my PhD, I have also developed publicly available videos on genome engineering technologies and actively engaged in scientific discussion on Twitter.

In addition to engaging through media channels, I contributed to a Wellcome Trust funded research project during my PhD on “Talking about gene drive: An exploration of language to enable understanding and deliberation in Africa, Europe, North America and Australasia.” Decisions involving governance and potential future use of gene drive technology will require meaningful and relevant dialogue among and between publics, stakeholders and communities. This project looks to identify what language devices, such as storytelling or metaphors, are most useful for communicating gene drive technology across audiences.

Following on from my PhD, I am now involved a small pilot project investigating the potential of gene drive to manage invasive grey squirrel in UK. This project has been promoted through print and digital media, and although it has been largely well received it has attracted some criticism in social media channels. Very few emerging technologies are realised without first encountering criticism and caution. Caution is warranted with early applications of gene drive. In a similar way that genetically modified foods were condemned, the way gene drives are used in their first attempts could have a lasting, potentially generational impact on society’s view of this technology. All gene drive research should include a thoughtful engagement plan from the outset.

6.9 Human values

Through my own experiences of engagement, in both social and academic fora, it has become clear that not everyone will be affected by gene drive research and applications in the same way. The expectation that people should have a voice in fundamental decisions that affect their environment is important. The increased power that gene drive technology might give scientists and governments to alter, and perhaps remove invasive pests, will be intrinsically disagreeable to some people. An increased ability to conserve species and ecosystems through gene drive technology may be intrinsically attractive to others. Approaches to ensure that communities participate meaningfully in decision making about the use of gene drive will be essential for the progression of this technology. One consistently uniting force that both sides of the gene drive debate can embrace is the strong and widely shared commitment of most to protect human welfare and the environment.

6.10 Other vertebrate applications

From an ethical standpoint, some may find the use of gene drive to rescue a species from a devastating disease more acceptable than rescuing a native species through the demise of a competing invasive species. For instance, chronic wasting disease (CWD) is a form of prion disease in cervids that affects primarily deer, elk and moose. The infectious prions responsible for CWD are caused by misfolding and aggregation of a prion protein (PrP) and are spread through saliva, urine and faeces. Once an animal is infected, misfolded PrP builds up in the brain, causing neuronal loss and eventually leading to death (Benestad et al., 2016). In some states in northern US, prevalence of CWD in deer is up to 30% in free-ranging populations and 90% in captive deer (Williams, 2005; Williams & Young, 1992). Studies in rodents and large animals have shown that animals do not require PrP (Denning et al., 2001; Weissmann & Flechsig, 2003). Gene drives could be used to knock-out the *PrP* gene to rescue infected cervid populations from this highly infectious disease. However, the long generation interval in cervids would mean the gene drive spreads very slowly through the population (Whitelaw & McFarlane, 2019).

Gene drives could also be applied to rescue threatened amphibians from chytridiomycosis. Chytridiomycosis is the disease caused by the fungus *Batrachochytrium dendrobatidis*, which has been implicated in mass die-offs and a conservative estimate suggests that chytridiomycosis has caused the severe decline or extinction of over 200 species (Fisher & Garner, 2007). In a study Olson et al. (2013), they detected *B. dendrobatidis* in 516 of 1240 (42%) amphibian species sampled. As the fungus reproduces mostly asexually (Fisher, Garner, & Walker, 2009), it cannot be controlled with a gene drive itself. However, resistance to *B. dendrobatidis* infection varies both within and among amphibian species. Major histocompatibility complex (MHC) peptides play an important role in the innate immune system of vertebrates and a specific MHC allele in the lowland leopard frog (*Lithobates yavapaiensis*) has been shown to increase survival of infected individuals (Savage & Zamudio, 2011). Gene drive could be used to spread the lowland leopard frog's resistant MHC allele through threatened amphibian populations, rescuing species from this killer fungus and potentially saving them from extinction (Whitelaw & McFarlane, 2019).

For further information on other potential applications of gene drive technology in vertebrates, see Appendix A.14, which contains an article titled "Accelerating Evolution", which I co-authored during my PhD for the Royal Society of Biology magazine, The Biologist.

When looking to apply gene drive in a non-model species there are a number of factors to consider, such as the genetic feasibility of engineering the species. A lack of reproductive and embryological information on the species may limit genetic engineering capabilities. Other issues may be the identification of suitable promoters, establishing stable transgene expression, access to a genome sequence data and genetic variation in the population that would likely prevent the spread of a gene drive. Evolutionary issues should also be strictly scrutinised, such as mating structure and barriers to mating between naive and engineered animals.

6.11 Final remarks

This project set out on a pioneering path to study gene drive technology as a potential tool for rodent pest management. The work encountered regulatory hurdles that slowed progression but was ultimately able to design, model, and engineer safeguarded two SGD approaches in mouse ES cells. These cells can now be used to generate SGD mouse models for *in vivo* studies. Throughout this journey, it has become clear that technical challenges of engineering a gene drive are relatively minor in comparison to the ecological, regulatory and ethical questions that remain regarding gene drive technology. It is hoped that work undertaken in this project will contribute to answering some of these questions and aid in the future development of safe, cost-effective and species-specific technologies for humanely managing damaging invasive vertebrate pests.

The scale of opportunities for gene drive technology and depth of interdisciplinary research needed before real-world application of this technology can be justified is truly exciting and I hope to continue contributing to this field. If used wisely, I believe synthetic selfish genetic elements can help solve many issues in public health, agriculture and conservation.

References

- Adelman, Z. N., & Tu, Z. (2016). Control of Mosquito-Borne Infectious Diseases: Sex and Gene Drive. *Trends in Parasitology*, 32(3), 219-229. doi:http://dx.doi.org/10.1016/j.pt.2015.12.003
- Aida, T., Nakade, S., Sakuma, T., Izu, Y., Oishi, A., Mochida, K., . . . Tanaka, K. (2016). Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. *BMC Genomics*, 17(1), 979-979. doi:10.1186/s12864-016-3331-9
- Aird, E. J., Lovendahl, K. N., St. Martin, A., Harris, R. S., & Gordon, W. R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Communications Biology*, 1(1), 54. doi:10.1038/s42003-018-0054-2
- Akbari, O. S., Bellen, H. J., Bier, E., Bullock, S. L., Burt, A., Church, G. M., . . . Wildonger, J. (2015). Safeguarding gene drive experiments in the laboratory. *Science*, 349(6251), 927.
- Alexopoulou, A. N., Couchman, J. R., & Whiteford, J. R. (2008). The CMV early enhancer/chicken β actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors. *BMC cell biology*, 9(1), 2.
- Allison, L. A. (2007). *Fundamental molecular biology*. Blackwell Pub.
- Andersson-Rolf, A., Mustata, R. C., Merenda, A., Kim, J., Perera, S., Grego, T., . . . Koo, B. K. (2017). One-step generation of conditional and reversible gene knockouts. *Nat Methods*, 14(3), 287-289. doi:10.1038/nmeth.4156
- Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol*, 38(7), 824-844. doi:10.1038/s41587-020-0561-9
- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., . . . Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576(7785), 149-157. doi:10.1038/s41586-019-1711-4
- Ardlie, K. G. (1998). Putting the brake on drive: meiotic drive of t haplotypes in natural populations of mice. *Trends in Genetics*, 14(5), 189-193.
- Arpino, J. A. J., Rizkallah, P. J., & Jones, D. D. (2012). Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222. *PLoS One*, 7(10), e47132-e47132. doi:10.1371/journal.pone.0047132
- Ashmore-Harris, C., & Fruhwirth, G. O. (2020). The clinical potential of gene editing as a tool to engineer cell-based therapeutics. *Clinical and Translational Medicine*, 9(1), 15. doi:10.1186/s40169-020-0268-z

- Backus, G. A., & Gross, K. (2016). Genetic engineering to eradicate invasive mice on islands: modeling the efficiency and ecological impacts. *Ecosphere*, 7(12), e01589. doi:10.1002/ecs2.1589
- Bae, S., Park, J., & Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, 30. doi:10.1093/bioinformatics/btu048
- Bakker, B., van den Bos, H., Lansdorp, P. M., & Fojier, F. (2015). How to count chromosomes in a cell: An overview of current and novel technologies. *Bioessays*, 37(5), 570-577. doi:10.1002/bies.201400218
- Bartlett, J. M. S., & Stirling, D. (2003). A Short History of the Polymerase Chain Reaction. In J. M. S. Bartlett & D. Stirling (Eds.), *PCR Protocols* (10.1385/1-59259-384-4:3pp. 3-6). Totowa, NJ: Humana Press.
- Bauer, H., Schindler, S., Charron, Y., Willert, J., Kusecek, B., & Herrmann, B. G. (2012). The nucleoside diphosphate kinase gene *Nme3* acts as quantitative trait locus promoting non-Mendelian inheritance. *PLoS Genet*, 8(3), e1002567. doi:10.1371/journal.pgen.1002567
- Bauer, H., Véron, N., Willert, J., & Herrmann, B. G. (2007). The t-complex-encoded guanine nucleotide exchange factor *Fgd2* reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes Dev*, 21(2), 143-147. doi:10.1101/gad.414807
- Beaghton, A. K., Hammond, A., Nolan, T., Crisanti, A., & Burt, A. (2019). Gene drive for population genetic control: non-functional resistance and parental effects. *Proceedings of the Royal Society B: Biological Sciences*, 286(1914), 20191586. doi:doi:10.1098/rspb.2019.1586
- Bebbere, D., Masala, L., Albertini, D. F., & Ledda, S. (2016). The subcortical maternal complex: multiple functions for one biological structure? *Journal of assisted reproduction and genetics*, 33(11), 1431-1438. doi:10.1007/s10815-016-0788-z
- Behringer, R., Gertsenstein, M., Nagy, K. V., & Nagy, A. (2014). *Manipulating the mouse embryo: a laboratory manual*: Cold Spring Harbor Laboratory Press.
- Belfort, M., & Bonocora, R. P. (2014). Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Methods in molecular biology (Clifton, N.J.)*, 1123, 1-26. doi:10.1007/978-1-62703-968-0_1
- Benestad, S. L., Mitchell, G., Simmons, M., Ytrehus, B., & Vikøren, T. (2016). First case of chronic wasting disease in Europe in a Norwegian free-ranging reindeer. *Veterinary Research*, 47(1), 88. doi:10.1186/s13567-016-0375-4
- Berry, R., & Scriven, P. (2005). The house mouse: a model and motor for evolutionary understanding. *Biological Journal of the Linnean Society*, 84(3), 335-347.
- Boch, J., & Bonas, U. (2010). Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. In N. K. VanAlfen, G. Bruening, & J. E. Leach (Eds.), *Annual Review of Phytopathology, Vol 48* (Vol. 48, pp. 419-436).
- Boda, E., Pini, A., Hoxha, E., Parolisi, R., & Tempia, F. (2009). Selection of Reference Genes for Quantitative Real-time RT-PCR Studies in Mouse Brain. *Journal of Molecular Neuroscience*, 37(3), 238-253. doi:10.1007/s12031-008-9128-9

- Boroviak, T., Loos, R., Bertone, P., Smith, A., & Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nature cell biology*, 16(6), 513-525.
- Borst, A., Box, A. T. A., & Fluit, A. C. (2004). False-Positive Results and Contamination in Nucleic Acid Amplification Assays: Suggestions for a Prevent and Destroy Strategy. *European Journal of Clinical Microbiology and Infectious Diseases*, 23(4), 289-299. doi:10.1007/s10096-004-1100-1
- Bouabe, H., & Okkenhaug, K. (2013). Gene targeting in mice: a review. *Methods in molecular biology (Clifton, N.J.)*, 1064, 315-336. doi:10.1007/978-1-62703-601-6_23
- Bradley, A., Evans, M., Kaufman, M. H., & Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309(5965), 255-256. doi:10.1038/309255a0
- Brossard, D., Belluck, P., Gould, F., & Wirz, C. D. (2019). Promises and perils of gene drives: Navigating the communication of complex, post-normal science. *Proceedings of the National Academy of Sciences*, 116(16), 7692. doi:10.1073/pnas.1805874115
- Brown, P. R. (2006). Short-and long-term demographic changes in house mouse populations after control in dryland farming systems in Australia. *Wildlife Research*, 33(6), 457-466.
- Brown, P. R., Huth, N. I., Banks, P. B., & Singleton, G. R. (2007). Relationship between abundance of rodents and damage to agricultural crops. *Agriculture, ecosystems & environment*, 120(2-4), 405-415.
- Bull, J. (2015). Evolutionary decay and the prospects for long-term disease intervention using engineered insect vectors. *Evolution, medicine, and public health*, 2015(1), 152-166.
- Burgio, G. (2018). Redefining mouse transgenesis with CRISPR/Cas9 genome editing technology. *Genome Biology*, 19(1), 27. doi:10.1186/s13059-018-1409-1
- Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society B: Biological Sciences*, 270(1518), 921-928. doi:10.1098/rspb.2002.2319
- Burt, A., & Koufopanou, V. (2004). Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Current Opinion in Genetics & Development*, 14(6), 609-615. doi:10.1016/j.gde.2004.09.010
- Burt, A., & Trivers, R. (2006). *Genes in conflict: the biology of selfish genetic elements*. Massachusetts, United States: Harvard University Press.
- Cain-Hom, C., Splinter, E., van Min, M., Simonis, M., van de Heijning, M., Martinez, M., . . . Warming, S. (2017). Efficient mapping of transgene integration sites and local structural changes in Cre transgenic mice using targeted locus amplification. *Nucleic Acids Research*, 45(8), e62-e62. doi:10.1093/nar/gkw1329
- Capecchi, M. R. (1989). Altering the genome by homologous recombination. *Science*, 244(4910), 1288-1292. doi:10.1126/science.2660260
- Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature Reviews Genetics*, 6(6), 507-512. doi:10.1038/nrg1619

- Carofino, B. L., & Justice, M. J. (2015). Tissue-Specific Regulation of Oncogene Expression Using Cre-Inducible ROSA26 Knock-In Transgenic Mice. *Current protocols in mouse biology*, 5(2), 187-204. doi:10.1002/9780470942390.mo140150
- Casini, A., Storch, M., Baldwin, G. S., & Ellis, T. (2015). Bricks and blueprints: methods and standards for DNA assembly. *Nature Reviews Molecular Cell Biology*, 16(9), 568-576. doi:10.1038/nrm4014
- Casola, S. (2010). Mouse models for miRNA expression: the ROSA26 locus. *Methods in molecular biology (Clifton, N.J.)*, 667, 145-163. doi:10.1007/978-1-60761-811-9_10
- Caughley, J., Bomford, M., Parker, R., Sinclair, R., Griffiths, J., & Kelly, D. (1998). *Managing vertebrate pests: rodents*. Canberra: Bureau of Resource Sciences and Grains Research and Development Corporation.
- Cebrian-Serrano, A., & Davies, B. (2017). CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mammalian genome : official journal of the International Mammalian Genome Society*, 28(7-8), 247-261. doi:10.1007/s00335-017-9697-4
- Champer, J., Buchman, A., & Akbari, O. S. (2016). Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics*, 17(3), 146-159. doi:10.1038/nrg.2015.34
- Champer, J., Kim, I., Champer, S., Clark, A., & Messer, P. (2019). Suppression gene drive in continuous space can result in unstable persistence of both drive and wild-type alleles. *bioRxiv*, 10.1101/769810, 769810. doi:10.1101/769810
- Champer, J., Liu, J., Oh, S. Y., Reeves, R., Luthra, A., Oakes, N., . . . Messer, P. (2018). Reducing resistance allele formation in CRISPR gene drive. *Proceedings of the National Academy of Sciences*, 115(21), 5522-5527.
- Champer, J., Yang, E., Lee, Y. L., Liu, J., Clark, A., & Messer, P. (2019). Resistance is futile: a CRISPR homing gene drive targeting a haplolethal gene. *bioRxiv*. 651737.
- Champer, S., Oh, S. Y., Liu, C., Wen, Z., Clark, A., Messer, P., & Champer, J. (2020). Computational and experimental performance of CRISPR homing gene drive strategies with multiplexed gRNAs. *Science Advances*, 6(10), eaaz0525. doi:10.1126/sciadv.aaz0525
- Chappell, P. E., Lydon, J. P., Conneely, O. M., Malley, B. W. O., & Levine, J. E. (1997). Endocrine defects in mice carrying a null mutation for the progesterone receptor gene. *Endocrinology*, 138(10), 4147-4152.
- Chen, F., Ji, J., Shen, J., & Lu, X. (2017). When Long Noncoding RNAs Meet Genome Editing in Pluripotent Stem Cells. *Stem cells international*, 2017, 3250624-3250624. doi:10.1155/2017/3250624
- Chen, S., Lee, B., Lee, A. Y.-F., Modzelewski, A. J., & He, L. (2016). Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. *The Journal of biological chemistry*, 291(28), 14457-14467. doi:10.1074/jbc.M116.733154
- Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, 24(1), 132-141. doi:10.1101/gr.162339.113

- Chowdhury, M. R., & Dubey, S. (2014). Role of Cytogenetics and Molecular Genetics in Human Health and Medicine. In *Animal Biotechnology* (pp. 451-472): Elsevier.
- Christie, K. A., Robertson, L. J., Conway, C., Blighe, K., DeDionisio, L. A., Chao-Shern, C., . . . Moore, C. B. T. (2020). Mutation-Independent Allele-Specific Editing by CRISPR-Cas9, a Novel Approach to Treat Autosomal Dominant Disease. *Molecular Therapy*, doi.org/10.1016/j.ymthe.2020.05.002. doi:doi.org/10.1016/j.ymthe.2020.05.002
- Codner, G. F., Lindner, L., Caulder, A., Wattenhofer-Donz , M., Radage, A., Mertz, A., . . . Teboul, L. (2016). Aneuploidy screening of embryonic stem cell clones by metaphase karyotyping and droplet digital polymerase chain reaction. *BMC cell biology*, 17(1), 30-30. doi:10.1186/s12860-016-0108-6
- Collier, R., Dasgupta, K., Xing, Y.-P., Hernandez, B. T., Shao, M., Rohozinski, D., . . . Thilmoney, R. (2017). Accurate measurement of transgene copy number in crop plants using droplet digital PCR. *The Plant Journal*, 90(5), 1014-1025. doi:10.1111/tpj.13517
- Cong, L. (2013). CRISPR: Zhang Lab's CRISPR Frequently Asked Questions. Retrieved from <https://www.addgene.org/crispr/zhang/faq/>
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), 819-823. doi:10.1126/science.1231143
- Cotton, L. M., Meilak, M. L., Templeton, T., Gonzales, J. G., Nenci, A., Cooney, M., . . . Viney, E. (2015). Utilising the resources of the International Knockout Mouse Consortium: the Australian experience. *Mammalian Genome*, 26(3-4), 142-153.
- Craig, G. B., Hickey, W. A., & Vandehey, R. C. (1960). Inherited Male-Producing Factor in *Aedes aegypti*. *Science*, 132(3443), 1887-1889. doi:10.1126/science.132.3443.1887
- Czechanski, A., Byers, C., Greenstein, I., Schrode, N., Donahue, L. R., Hadjantonakis, A.-K., & Reinholdt, L. G. (2014). Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. *Nature protocols*, 9(3), 559-574. doi:10.1038/nprot.2014.030
- Dagert, M., & Ehrlich, S. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene*, 6(1), 23-28.
- Del Amo, V. L., Bishop, A. L., S nchez C, H. M., Bennett, J. B., Feng, X., Marshall, J. M., . . . Gantz, V. M. (2019). Split-gene drive system provides flexible application for safe laboratory investigation and potential field deployment. *bioRxiv*, 10.1101/684597, 684597. doi:10.1101/684597
- Delborne, J., Shapiro, J., Mahmuc, F., Ford, T., George, D., & Dermer, S. (2019). *Exploring Stakeholder Perspectives on the Development of a Gene Drive Mouse for Biodiversity Protection on Islands*. Retrieved from <https://go.ncsu.edu/ges-gene-drive-landscape>:
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., . . . Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602-607. doi:10.1038/nature09886
- Denning, C., Burl, S., Ainslie, A., Bracken, J., Dinnyes, A., Fletcher, J., . . . Clark, A. J. (2001). Deletion of the alpha(1,3)galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep. *Nat Biotechnol*, 19(6), 559-562. doi:10.1038/89313

- DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research*, *41*(7), 4336-4343. doi:10.1093/nar/gkt135
- Dodsworth, B. T., Hatje, K., Meyer, C. A., Flynn, R., & Cowley, S. A. (2020). Rates of homology directed repair of CRISPR-Cas9 induced double strand breaks are lower in naïve compared to primed human pluripotent stem cells. *Stem Cell Research*. 101852.
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., & Donovan, K. F. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, *34*. doi:10.1038/nbt.3437
- Dong, F., Xie, K., Chen, Y., Yang, Y., & Mao, Y. (2017). Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells. *Biochem Biophys Res Commun*, *482*(4), 889-895. doi:10.1016/j.bbrc.2016.11.129
- Early, R., Bradley, B. A., Dukes, J. S., Lawler, J. J., Olden, J. D., Blumenthal, D. M., . . . Tatem, A. J. (2016). Global threats from invasive alien species in the twenty-first century and national response capacities. *Nat Commun*, *7*, 12485. doi:10.1038/ncomms12485
- Eckhoff, P. A., Wenger, E. A., Godfray, H. C. J., & Burt, A. (2016). Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proceedings of the National Academy of Sciences*, *114*(2), E255. doi:10.1073/pnas.1611064114
- Esvelt, K. M. (2017). Daisy drive: Safe Genes research plan. *Responsive Science*. Retrieved from <https://www.responsivescience.org/pub/daisy-drive>.
- Esvelt, K. M. (2019). Daisy Drive Systems. *MIT Media Lab*. Retrieved from <http://www.sculptingevolution.org/daisydrives>
- Esvelt, K. M., & Gemmell, N. (2017). Conservation demands safe gene drive. *PLOS Biology*, *15*(11), e2003850. doi:10.1371/journal.pbio.2003850
- Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *Elife*, *3*, e03401. doi:10.7554/eLife.03401
- Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, *292*(5819), 154-156.
- Fan, M., Ferenc, M., Jucheim, A., LaManna, C., Monroe, M., Morgan, K., . . . Taylor-Parker, J. (2014). Plasmids 101: A Desktop Resource. In: United States of America: Addgene.
- Fisher, M. C., & Garner, T. W. (2007). The relationship between the emergence of *Batrachochytrium dendrobatidis*, the international trade in amphibians and introduced amphibian species. *Fungal Biology Reviews*, *21*(1), 2-9.
- Fisher, M. C., Garner, T. W., & Walker, S. F. (2009). Global emergence of *Batrachochytrium dendrobatidis* and amphibian chytridiomycosis in space, time, and host. *Annual review of microbiology*, *63*, 291-310.
- Flemr, M., & Bühler, M. (2015). Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. *Cell Reports*, *12*(4), 709-716. doi:<https://doi.org/10.1016/j.celrep.2015.06.051>

- Fraser, M. J., Jr. (2012). Insect transgenesis: current applications and future prospects. *Annu Rev Entomol*, 57, 267-289. doi:10.1146/annurev.ento.54.110807.090545
- Friedrich, G., & Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes & development*, 5(9), 1513-1523.
- Funk, C., & Hefferon, M. (2018). *Most Americans Accept Genetic Engineering of Animals That Benefits Human Health, but Many Oppose Other Uses*. Retrieved from https://www.pewinternet.org/wp-content/uploads/sites/9/2018/08/PS_2018.08.16_biotech-animals_FINAL.pdf:
- Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, 31(7), 397-405. doi:<https://doi.org/10.1016/j.tibtech.2013.04.004>
- Galizi, R., Doyle, L., Menichelli, M., Bernardini, F., Deredec, A., Burt, A., . . . Crisanti, A. (2014). A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nature Communications*, 5. doi:10.1038/ncomms4977
- Galizi, R., Hammond, A., Kyrou, K., Taxiarchi, C., Bernardini, F., O'Loughlin, S. M., . . . Crisanti, A. (2016). A CRISPR-Cas9 sex-ratio distortion system for genetic control. *Science Reports*, 6, 31139. doi:10.1038/srep31139
- Gallardo, T., Shirley, L., John, G. B., & Castrillon, D. H. (2007). Generation of a germ cell-specific mouse transgenic Cre line, Vasa-Cre. *Genesis*, 45(6), 413-417.
- Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science*, 348(6233), 442. doi:10.1126/science.aaa5945
- Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences*, 112(49), E6736-E6743. doi:10.1073/pnas.1521077112
- Gardner, R. L., & Brook, F. A. (2002). Reflections on the biology of embryonic stem (ES) cells. *International journal of developmental biology*, 41(2), 235-243.
- Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., . . . Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320), 67-71.
- Gaztelumendi, N., & Nogués, C. (2014). Chromosome Instability in mouse Embryonic Stem Cells. *Sci Rep*, 4(1), 5324. doi:10.1038/srep05324
- Ghimire, S., Van der Jeught, M., Neupane, J., Roost, M. S., Anckaert, J., Popovic, M., . . . Heindryckx, B. (2018). Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. *Sci Rep*, 8(1), 5884. doi:10.1038/s41598-018-24051-5
- Gibson, D. G. (2011). Chapter fifteen - Enzymatic Assembly of Overlapping DNA Fragments. In C. Voigt (Ed.), *Methods in Enzymology* (Vol. 498, pp. 349-361): Academic Press.
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343-345. doi:10.1038/nmeth.1318

- Good, P. D., Krikos, A. J., Li, S. X. L., Bertrand, E., Lee, N. S., Giver, L., . . . Engelke, D. R. (1997). Expression of small, therapeutic RNAs in human cell nuclei. *Gene Therapy*, 4(1), 45-54. doi:10.1038/sj.gt.3300354
- Gordeeva, O. F., Lifantseva, N. V., & Khaidukov, S. V. (2011). Expression patterns of germ line specific genes in mouse and human pluripotent stem cells are associated with regulation of ground and primed state of pluripotency. *Russian Journal of Developmental Biology*, 42(6), 355-375. doi:10.1134/S1062360411060038
- Green, M. R., Hughes, H., Sambrook, J., & MacCallum, P. (2012). *Molecular cloning: a laboratory manual*.
- Gross, A. P., & Kioussi, C. (2014). Culturing and Differentiating Mouse Embryonic Stem Cells. In C. Kioussi (Ed.), *Stem Cells and Tissue Repair: Methods and Protocols* (10.1007/978-1-4939-1435-7_1pp. 1-8). New York, NY: Springer New York.
- Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R. S., Bier, E., & Cooper, K. L. (2019). Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature*, 566(7742), 105-109. doi:10.1038/s41586-019-0875-2
- Gu, B., Posfai, E., & Rossant, J. (2018). Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat Biotechnol*, 36(7), 632-637. doi:10.1038/nbt.4166
- Güell, M., Yang, L., & Church, G. M. (2014). Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics*, 30(20), 2968-2970.
- Gupta, R. M., & Musunuru, K. (2014). Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of Clinical Investigation*, 124(10), 4154-4161. doi:10.1172/JCI72992
- Gurumurthy, C. B., & Lloyd, K. C. K. (2019). Generating mouse models for biomedical research: technological advances. *Disease models & mechanisms*, 12(1).
- Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., . . . Concordet, J.-P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, 17(1), 148. doi:10.1186/s13059-016-1012-2
- Hamilton, W. D. (1967). Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, 156(3774), 477-488. doi:10.1126/science.156.3774.477
- Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., . . . Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat Biotechnol*, 34(1), 78-83. doi:10.1038/nbt.3439
- Hammond, A., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., . . . Crisanti, A. (2017). The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *Plos Genetics*, 13(10), e1007039.
- Hellen, C. U., & Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes & development*, 15(13), 1593-1612.
- Herrmann, B. G., & Bauer, H. (2012). The mouse t—haplotype: a selfish chromosome—genetics, molecular mechanism, and evolution. *Evolution of the house mouse*, 3, 297.

- Hickey, W. A., & Craig, G. B., Jr. (1966). Genetic distortion of sex ratio in a mosquito, *Aedes aegypti*. *Genetics*, 53(6), 1177-1196.
- Hoebeeck, J., Speleman, F., & Vandesompele, J. (2007). Real-time quantitative PCR as an alternative to Southern blot or fluorescence in situ hybridization for detection of gene copy number changes. *Methods in molecular biology (Clifton, N.J.)*, 353, 205-226. doi:10.1385/1-59745-229-7:205
- Hooper, M., Hardy, K., Handyside, A., Hunter, S., & Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature*, 326(6110), 292-295. doi:10.1038/326292a0
- Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 327(5962), 167-170. doi:10.1126/science.1179555
- Hottentot, Q. P., van Min, M., Splinter, E., & White, S. J. (2017). Targeted Locus Amplification and Next-Generation Sequencing. *Methods in molecular biology (Clifton, N.J.)*, 1492, 185-196. doi:10.1007/978-1-4939-6442-0_13
- Howal, G., Donald, J., Galvan, J. P., Russel, J., Parkes, J., Samaniego, A., . . . Tershy, B. (2007). Invasive Rodent Eradication on Islands. *Conservation Biology*, 21(5), 1258-1268. doi:10.1111/j.1523-1739.2007.00755.x
- Howe, C. (2007). Simple cloning. In *Gene Cloning and Manipulation* (2 ed., DOI: 10.1017/CBO9780511807343.005pp. 52-73). Cambridge: Cambridge University Press.
- Hsiao, T., Conant, D., Maures, T., Waite, K., Yang, J., Kelso, R., . . . Stoner, R. (2019). Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*, 10.1101/251082, 251082. doi:10.1101/251082
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., & Agarwala, V. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 31. doi:10.1038/nbt.2647
- Huang, J., Deng, K., Wu, H., Liu, Z., Chen, Z., Cao, S., . . . Liu, L. (2008). Efficient Production of Mice from Embryonic Stem Cells Injected into Four- or Eight-Cell Embryos by Piezo Micromanipulation. *STEM CELLS*, 26(7), 1883-1890. doi:10.1634/stemcells.2008-0164
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, 169(12), 5429-5433. doi:10.1128/jb.169.12.5429-5433.1987
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., . . . Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*, 36(4), 338-345. doi:10.1038/nbt.4060
- Jang, S. K., Davies, M., Kaufman, R., & Wimmer, E. (1989). Initiation of protein synthesis by internal entry of ribosomes into the 5'nontranslated region of encephalomyocarditis virus RNA in vivo. *Journal of Virology*, 63(4), 1651-1660.
- Jarazo, J., Qing, X., & Schwamborn, J. C. (2019). Guidelines for Fluorescent Guided Biallelic HDR Targeting Selection With PiggyBac System Removal for Gene Editing. *Frontiers in genetics*, 10(190). doi:10.3389/fgene.2019.00190
- Jasin, M. (1996). Genetic manipulation of genomes with rare-cutting endonucleases. *Trends in Genetics*, 12(6), 224-228. doi:https://doi.org/10.1016/0168-9525(96)10019-6

- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), 816-821. doi:10.1126/science.1225829
- Jones, M. S., Delborne, J. A., Elsensohn, J., Mitchell, P. D., & Brown, Z. S. (2019). Does the U.S. public support using gene drives in agriculture? And what do they want to know? *Science Advances*, 5(9), eaau8462. doi:10.1126/sciadv.aau8462
- Kabadi, A. M., & Gersbach, C. A. (2014). Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods (San Diego, Calif.)*, 69(2), 188-197. doi:10.1016/j.ymeth.2014.06.014
- KaramiNejadRanjbar, M., Eckermann, K. N., Ahmed, H. M. M., Sánchez C, H. M., Dippel, S., Marshall, J. M., & Wimmer, E. A. (2018). Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proceedings of the National Academy of Sciences*, 115(24), 6189. doi:10.1073/pnas.1713825115
- Kibschull, M. (2017). Differentiating mouse embryonic stem cells into embryoid bodies in AggreWell plates. *Cold Spring Harbor Protocols*, 2017(6), pdb. prot094169.
- Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F. W., Cuthill, I. C., Fry, D., . . . Altman, D. G. (2009). Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals. *PLoS One*, 4(11), e7824. doi:10.1371/journal.pone.0007824
- Kim, H., & Kim, J. (2014). A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics*, 15(5), 321-334. doi:10.1038/nrg3686
- Kim, H. J., Lee, H. J., Kim, H., Cho, S. W., & Kim, J.-S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome research*, 19(7), 1279-1288.
- Kim, H. K., Min, S., Song, M., Jung, S., Choi, J. W., Kim, Y., . . . Kim, H. (2018). Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat Biotechnol*, 36(3), 239-241. doi:10.1038/nbt.4061
- Kim, J. M., Kim, D., Kim, S., & Kim, J.-S. (2014). Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nature Communications*, 5(1), 3157. doi:10.1038/ncomms4157
- Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research*, 24(6), 1012-1019. doi:10.1101/gr.171322.113
- Kim, T. K., & Eberwine, J. H. (2010). Mammalian cell transfection: the present and the future. *Analytical and Bioanalytical Chemistry*, 397(8), 3173-3178. doi:10.1007/s00216-010-3821-6
- Kim, Y. M., Lee, J.-Y., Xia, L., Mulvihill, J. J., & Li, S. (2013). Trisomy 8: a common finding in mouse embryonic stem (ES) cell lines. *Molecular cytogenetics*, 6(1), 3-3. doi:10.1186/1755-8166-6-3
- Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), 490-495. doi:10.1038/nature16526

- Komor, A. C., Badran, A. H., & Liu, D. R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*, 168(1), 20-36. doi:10.1016/j.cell.2016.10.044
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), 420-424.
- Koontz, L. (2013). Chapter Three - Explanatory Chapter: Introducing Exogenous DNA into Cells. In J. Lorsch (Ed.), *Methods in Enzymology* (Vol. 529, pp. 29-34): Academic Press.
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., & Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. *Nature*, 351(6322), 117-121.
- Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol*, 36(8), 765-771. doi:10.1038/nbt.4192
- Kotnik, T., Rems, L., Tarek, M., & Miklavčič, D. (2019). Membrane Electroporation and Electropermeabilization: Mechanisms and Models. *Annual Review of Biophysics*, 48(1), 63-91. doi:10.1146/annurev-biophys-052118-115451
- Krull, C. R., Galbraith, J. A., Glen, A. S., & Nathan, H. W. (2014). Invasive vertebrates in Australia and New Zealand. In A. Stow, G. I. Holwell, & N. Maclean (Eds.), *Austral Ark: The State of Wildlife in Australia and New Zealand* (DOI: 10.1017/CBO9781139519960.012pp. 197-226). Cambridge: Cambridge University Press.
- Kwan, K. M. (2002). Conditional alleles in mice: practical considerations for tissue-specific knockouts. *Genesis*, 32(2), 49-62. doi:10.1002/gene.10068
- Lander, E. S. (2016). The heroes of CRISPR. *Cell*, 164(1-2), 18-28.
- Lee, H., Yoon, D. E., & Kim, K. (2020). Genome editing methods in animal models. *Animal cells and systems*, 24(1), 8-16. doi:10.1080/19768354.2020.1726462
- Lessard, J. C. (2013). Chapter Seven - Molecular Cloning. In J. Lorsch (Ed.), *Methods in Enzymology* (Vol. 529, pp. 85-98): Academic Press.
- Li, B., Ren, N., Yang, L., Liu, J., & Huang, Q. (2019). A qPCR method for genome editing efficiency determination and single-cell clone screening in human cells. *Sci Rep*, 9(1), 18877. doi:10.1038/s41598-019-55463-6
- Li, L., Baibakov, B., & Dean, J. (2008). A Subcortical Maternal Complex Essential for Preimplantation Mouse Embryogenesis. *Developmental Cell*, 15(3), 416-425. doi:https://doi.org/10.1016/j.devcel.2008.07.010
- Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., . . . Akbari, O. S. (2020). Development of a confinable gene drive system in the human disease vector *Aedes aegypti*. *Elife*, 9, e51701. doi:10.7554/eLife.51701
- Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., . . . Chesnut, J. D. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology*, 208, 44-53. doi:https://doi.org/10.1016/j.jbiotec.2015.04.024

- Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: a review of the challenges and approaches. *Drug delivery*, 25(1), 1234-1257. doi:10.1080/10717544.2018.1474964
- Liu, G., Zhang, Y., & Zhang, T. (2019). Computational approaches for effective CRISPR guide RNA design and evaluation. *Computational and structural biotechnology journal*, 18, 35-44. doi:10.1016/j.csbj.2019.11.006
- Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., & Ma, W. (2019). Methodologies for Improving HDR Efficiency. *Frontiers in genetics*, 9(691). doi:10.3389/fgene.2018.00691
- Liu, X., Wu, H., Loring, J., Hormuzdi, S., Disteche, C. M., Bornstein, P., & Jaenisch, R. (1997). Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. *Developmental dynamics: an official publication of the American Association of Anatomists*, 209(1), 85-91.
- Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., . . . Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep*, 7(1), 2193-2193. doi:10.1038/s41598-017-02460-2
- Locke, M. E. O., Milojevic, M., Eitutus, S. T., Patel, N., Wishart, A. E., Daley, M., & Hill, K. A. (2015). Genomic copy number variation in *Mus musculus*. *BMC Genomics*, 16(1), 497-497. doi:10.1186/s12864-015-1713-z
- Lorvelec, O., & Pascal, M. (2005). French attempts to eradicate non-indigenous mammals and their consequences for native biota. *Biological Invasions*, 7(1), 135-140. doi:10.1007/s10530-004-9643-8
- Lunshof, J. E., & Birnbaum, A. (2017). Adaptive risk management of gene drive experiments: biosafety, biosecurity, and ethics. *Applied Biosafety*, 22(3), 97-103.
- Lydon, J. P., DeMayo, F. J., Funk, C. R., Mani, S. K., Hughes, A. R., Montgomery, C., . . . O'Malley, B. W. (1995). Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes & development*, 9(18), 2266-2278.
- Ma, H., Liu, Q., Diamond, S. L., & Pierce, E. A. (2004). Mouse embryonic stem cells efficiently lipofected with nuclear localization peptide result in a high yield of chimeric mice and retain germline transmission potency. *Methods*, 33(2), 113-120. doi:https://doi.org/10.1016/j.ymeth.2003.11.008
- Ma, L., & Chung, W. K. (2014). Quantitative analysis of copy number variants based on real-time LightCycler PCR. *Current protocols in human genetics*, 80, 7.21.21-27.21. doi:10.1002/0471142905.hg0721s80
- Mancini, C., Messana, E., Turco, E., Brussino, A., & Brusco, A. (2011). Gene-targeted embryonic stem cells: real-time PCR assay for estimation of the number of neomycin selection cassettes. *Biological procedures online*, 13, 10-10. doi:10.1186/1480-9222-13-10
- Manser, A., Cornell, S. J., Sutter, A., Blondel, D. V., Serr, M., Godwin, J., & Price, T. A. R. (2019). Controlling invasive rodents via synthetic gene drive and the role of polyandry. *Proceedings of the Royal Society B: Biological Sciences*, 286(1909), 20190852. doi:10.1098/rspb.2019.0852
- Manser, A., Lindholm, A. K., König, B., & Bagheri, H. C. (2011). Polyandry and the decrease of a selfish genetic element in a wild house mouse population. *Evolution: International Journal of Organic Evolution*, 65(9), 2435-2447.

- Manser, A., Lindholm, A. K., Simmons, L. W., & Firman, R. C. (2017). Sperm competition suppresses gene drive among experimentally evolving populations of house mice. *Molecular Ecology*, 26(20), 5784-5792.
- Manuel, J. (2008). Oversight Without Obstruction: The Challenge for High-Containment Labs. *Environmental Health Perspectives*, 116(11), A486-A489. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2592294/>
- Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., & Zhu, J.-K. (2013). Application of the CRISPR-Cas system for efficient genome engineering in plants. *Molecular plant*, 6(6), 2008-2011.
- Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*, 78(12), 7634-7638.
- McCabe, C. V., Codner, G. F., Allan, A. J., Caulder, A., Christou, S., Loeffler, J., . . . Teboul, L. (2019). Application of long-read sequencing for robust identification of correct alleles in genome edited animals. *bioRxiv*, 10.1101/838193, 838193. doi:10.1101/838193
- McDermott, S. R., & Noor, M. A. F. (2010). The role of meiotic drive in hybrid male sterility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1544), 1265-1272. doi:doi:10.1098/rstb.2009.0264
- McFarlane, G. R., Whitelaw, C. B. A., & Lillico, S. G. (2018). CRISPR-Based Gene Drives for Pest Control. *Trends Biotechnol*, 36(2), 130-133. doi:10.1016/j.tibtech.2017.10.001
- Meerburg, B. G., Brom, F. W., & Kijlstra, A. (2008). The ethics of rodent control. *Pest Manag Sci*, 64(12), 1205-1211. doi:10.1002/ps.1623
- Mehravar, M., Shirazi, A., Mehrazar, M. M., & Nazari, M. (2019). In Vitro Pre-validation of Gene Editing by CRISPR/Cas9 Ribonucleoprotein. *Avicenna journal of medical biotechnology*, 11(3), 259-263.
- Mehravar, M., Shirazi, A., Nazari, M., & Banan, M. (2019). Mosaicism in CRISPR/Cas9-mediated genome editing. *Developmental Biology*, 445(2), 156-162. doi:<https://doi.org/10.1016/j.ydbio.2018.10.008>
- Min, J., Noble, C., Najjar, D., & Esvelt, K. M. (2017a). Daisy quorum drives for the genetic restoration of wild populations. *bioRxiv*, 10.1101/115618, 115618. doi:10.1101/115618
- Min, J., Noble, C., Najjar, D., & Esvelt, K. M. (2017b). Daisyfield gene drive systems harness repeated genomic elements as a generational clock to limit spread. *bioRxiv*, 10.1101/104877, 104877. doi:10.1101/104877
- Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., & Hayakawa, T. (2000). IRES-Dependent Second Gene Expression Is Significantly Lower Than Cap-Dependent First Gene Expression in a Bicistronic Vector. *Molecular Therapy*, 1(4), 376-382. doi:10.1006/mthe.2000.0050
- Mojica, F. J. M., & Montoliu, L. (2016). On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals. *Trends in Microbiology*, 24(10), 811-820. doi:<https://doi.org/10.1016/j.tim.2016.06.005>
- Mulas, C., Kalkan, T., von Meyenn, F., Leitch, H. G., Nichols, J., & Smith, A. (2019). Defined conditions for propagation and manipulation of mouse embryonic stem cells. *Development (Cambridge, England)*, 146(6), dev173146. doi:10.1242/dev.173146

- Naduthodi, M. I. S., Mohanraju, P., Südfeld, C., D'Adamo, S., Barbosa, M. J., & van der Oost, J. (2019). CRISPR-Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1. *Biotechnology for biofuels*, *12*, 66-66. doi:10.1186/s13068-019-1401-3
- Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., . . . Suzuki, K.-i. T. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature Communications*, *5*(1), 5560. doi:10.1038/ncomms6560
- National Academies of Sciences, Engineering, and Medicine. (2016). *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values* (10.17226/23405). Washington, DC: The National Academies Press.
- Nemudryi, A. A., Valetdinova, K. R., Medvedev, S. P., & Zakian, S. M. (2014). TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta naturae*, *6*(3), 19-40.
- Nicholls, P. K., Bellott, D. W., Cho, T.-J., Pyntikova, T., & Page, D. C. (2019). Locating and Characterizing a Transgene Integration Site by Nanopore Sequencing. *G3 (Bethesda, Md.)*, *9*(5), 1481-1486. doi:10.1534/g3.119.300582
- Nichols, J., Evans, E. P., & Smith, A. G. (1990). Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development*, *110*(4), 1341-1348.
- Niwa, H., Burdon, T., Chambers, I., & Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev*, *12*(13), 2048-2060. doi:10.1101/gad.12.13.2048
- Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A. L., . . . Esvelt, K. M. (2016). Daisy-chain gene drives for the alteration of local populations. *bioRxiv*, 10.1101/057307. doi:10.1101/057307
- Norris, A. L., Lee, S. S., Greenlees, K. J., Tadesse, D. A., Miller, M. F., & Lombardi, H. A. (2020). Template plasmid integration in germline genome-edited cattle. *Nat Biotechnol*, *38*(2), 163-164.
- North, A. R., Burt, A., & Godfray, H. C. J. (2019). Modelling the potential of genetic control of malaria mosquitoes at national scale. *BMC Biol*, *17*(1), 26. doi:10.1186/s12915-019-0645-5
- Oberhofer, G., Ivy, T., & Hay, B. A. (2019). Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proceedings of the National Academy of Sciences*, *116*(13), 6250-6259. doi:10.1073/pnas.1816928116
- Ohtsuka, M., Sato, M., Miura, H., Takabayashi, S., Matsuyama, M., Koyano, T., . . . Gurumurthy, C. B. (2018). i-GONAD: a robust method for in situ germline genome engineering using CRISPR nucleases. *Genome Biology*, *19*(1), 25. doi:10.1186/s13059-018-1400-x
- Olson, D. H., Aanensen, D. M., Ronnenberg, K. L., Powell, C. I., Walker, S. F., Bielby, J., . . . Fisher, M. C. (2013). Mapping the Global Emergence of *Batrachochytrium dendrobatidis*, the Amphibian Chytrid Fungus. *PLoS One*, *8*(2), e56802. doi:10.1371/journal.pone.0056802

- Perron, G. G., Kryazhimskiy, S., Rice, D. P., & Buckling, A. (2012). Multidrug therapy and evolution of antibiotic resistance: when order matters. *Applied and environmental microbiology*, 78(17), 6137-6142. doi:10.1128/AEM.01078-12
- Pickar-Oliver, A., & Gersbach, C. A. (2019). The next generation of CRISPR–Cas technologies and applications. *Nature Reviews Molecular Cell Biology*, 20(8), 490-507. doi:10.1038/s41580-019-0131-5
- Pimentel, D. (2007, 13/06/17). *Environmental and Economic Costs of Vertebrate Species Invasions into the United States*. Paper presented at the Managing Vertebrate Invasive Species: Proceedings of an International Symposium, National Wildlife Research Center, Fort Collins.
- Plaza Reyes, A., & Lanner, F. (2018). Time Matters: Gene Editing at the Mouse 2-Cell Embryo Stage Boosts Knockin Efficiency. *Cell Stem Cell*, 23(2), 155-157. doi:https://doi.org/10.1016/j.stem.2018.07.008
- Pocock, M. J., Hauffe, H. C., & Searle, J. B. (2005). Dispersal in house mice. *Biological Journal of the Linnean Society*, 84(3), 565-583.
- Pratt, J., Venkatraman, N., Brinker, A., Xiao, Y., Blasberg, J., Thompson, D. C., & Bourner, M. (2012). Use of zinc finger nuclease technology to knock out efflux transporters in C2BBel1 cells. *Curr Protoc Toxicol*, Chapter 23, Unit 23.22. doi:10.1002/0471140856.tx2302s52
- Prowse, T. A. A., Cassey, P., Ross, J. V., Pfitzner, C., Wittmann, T. A., & Thomas, P. (2017). Dodging silver bullets: good CRISPR gene-drive design is critical for eradicating exotic vertebrates. *Proceedings of the Royal Society B: Biological Sciences*, 284(1860). doi:10.1098/rspb.2017.0799
- Quadros, R. M., Miura, H., Harms, D. W., Akatsuka, H., Sato, T., Aida, T., . . . Sakai, D. (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biology*, 18(1), 92.
- Quan, P.-L., Sauzade, M., & Brouzes, E. (2018). dPCR: A Technology Review. *Sensors (Basel, Switzerland)*, 18(4), 1271. doi:10.3390/s18041271
- Ravetz, I. (1999). What is post-normal science. *Futures-the Journal of Forecasting Planning and Policy*, 31(7), 647-654.
- Rebuzzini, P., Neri, T., Zuccotti, M., Redi, C. A., & Garagna, S. (2008). Chromosome number variation in three mouse embryonic stem cell lines during culture. *Cytotechnology*, 58(1), 17-23. doi:10.1007/s10616-008-9164-x
- Rizzo, M. A., Davidson, M. W., & Piston, D. W. (2009). Fluorescent protein tracking and detection: applications using fluorescent proteins in living cells. *Cold Spring Harbor Protocols*, 2009(12), pdb. top64.
- Robertson, E., Bradley, A., Kuehn, M., & Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature*, 323(6087), 445-448. doi:10.1038/323445a0
- Robertson, P. A., Adriaens, T., Lambin, X., Mill, A., Roy, S., Shuttleworth, C. M., & Sutton-Croft, M. (2017). The large-scale removal of mammalian invasive alien species in Northern Europe. *Pest Management Science*, 73(2), 273-279. doi:10.1002/ps.4224

- Rocha-Martins, M., Cavalheiro, G. R., Matos-Rodrigues, G. E., & Martins, R. A. (2015). From Gene Targeting to Genome Editing: Transgenic animals applications and beyond. *An Acad Bras Cienc*, 87(2 Suppl), 1323-1348. doi:10.1590/0001-3765201520140710
- Rode, N. O., Estoup, A., Bourguet, D., Courtier-Orgogozo, V., & Débarre, F. (2019). Population management using gene drive: molecular design, models of spread dynamics and assessment of ecological risks. *Conservation Genetics*, 20(4), 671-690. doi:10.1007/s10592-019-01165-5
- Rogers, A. B. (2018). Stress of Strains: Inbred Mice in Liver Research. *Gene expression*, 19(1), 61-67. doi:10.3727/105221618X15337408678723
- Romito, A., & Cobellis, G. (2016). Pluripotent Stem Cells: Current Understanding and Future Directions. *Stem cells international*, 2016, 9451492-9451492. doi:10.1155/2016/9451492
- Rouet, P., Smih, F., & Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and cellular biology*, 14(12), 8096-8106. doi:10.1128/mcb.14.12.8096
- Russell, W. M. S., & Burch, R. L. (1959). *The principles of humane experimental technique*: Methuen.
- Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.-I. T., & Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nature protocols*, 11(1), 118-133. doi:10.1038/nprot.2015.140
- Samuelson, J. C. (2011). Recent Developments in Difficult Protein Expression: A Guide to E. coli Strains, Promoters, and Relevant Host Mutations. In J. T. C. Evans & M.-Q. Xu (Eds.), *Heterologous Gene Expression in E.coli: Methods and Protocols* (10.1007/978-1-61737-967-3_11pp. 195-209). Totowa, NJ: Humana Press.
- Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, 32(4), 347-355. doi:10.1038/nbt.2842
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.
- Savage, A. E., & Zamudio, K. R. (2011). MHC genotypes associate with resistance to a frog-killing fungus. *Proceedings of the National Academy of Sciences*, 108(40), 16705-16710.
- Scheffer, M. (2009). *Critical transitions in nature and society* (Vol. 16): Princeton University Press.
- Schumann, K., Lin, S., Boyer, E., Simeonov, D. R., Subramaniam, M., Gate, R. E., . . . Doudna, J. A. (2015). Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proceedings of the National Academy of Sciences*, 112(33), 10437-10442.
- Semprini, S., Troup, T. J., Kotelevtseva, N., King, K., Davis, J. R., Mullins, L. J., . . . Mullins, J. J. (2007). Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombineering techniques. *Nucleic Acids Res*, 35(5), 1402-1410. doi:10.1093/nar/gkl1108

- Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P., & Pruett-Miller, S. M. (2018). A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep*, 8(1), 888. doi:10.1038/s41598-018-19441-8
- Shaimardanova, A. A., Chulpanova, D. S., Kitaeva, K. V., Abdrakhmanova, I. I., Chernov, V. M., Rutland, C. S., . . . Solovyeva, V. V. (2019). Production and application of multicistronic constructs for various human disease therapies. *Pharmaceutics*, 11(11), 580.
- Silver, L. M. (1993). The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends in Genetics*, 9(7), 250-254. doi:https://doi.org/10.1016/0168-9525(93)90090-5
- Simoni, A., Siniscalchi, C., Chan, Y.-S., Huen, D. S., Russell, S., Windbichler, N., & Crisanti, A. (2014). Development of synthetic selfish elements based on modular nucleases in *Drosophila melanogaster*. *Nucleic Acids Research*, 42(11), 7461-7472. doi:10.1093/nar/gku387
- Singh, P., Schimenti, J. C., & Bolcun-Filas, E. (2015). A mouse geneticist's practical guide to CRISPR applications. *Genetics*, 199(1), 1-15. doi:10.1534/genetics.114.169771
- Singleton, G. (1997). Integrated management of rodents: A Southeast Asian and Australian perspective. *BELGIAN JOURNAL OF ZOOLOGY*, 127, 157-169.
- Singleton, G. (2003). *Impacts of rodents on rice production in Asia*. Paper presented at the IRRI Discussion Paper Series, Los Baños, Philippines.
- Sinkins, S. P., & Gould, F. (2006). Gene drive systems for insect disease vectors. *Nature Reviews Genetics*, 7(6), 427-435. doi:10.1038/nrg1870
- Skarnes, W. C. (2015). Is mouse embryonic stem cell technology obsolete? *Genome Biology*, 16(1), 109. doi:10.1186/s13059-015-0673-6
- Smih, F., Rouet, P., Romanienko, P. J., & Jasin, M. (1995). Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Research*, 23(24), 5012-5019. doi:10.1093/nar/23.24.5012
- Smirnov, A., Fishman, V., Yunusova, A., Korablev, A., Serova, I., Skryabin, B. V., . . . Battulin, N. (2020). DNA barcoding reveals that injected transgenes are predominantly processed by homologous recombination in mouse zygote. *Nucleic Acids Research*, 48(2), 719-735. doi:10.1093/nar/gkz1085
- Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., & Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, 336(6200), 688-690.
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., . . . Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, 321(6071), 674-679.
- Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A., & Kucherlapati, R. S. (1985). Insertion of DNA sequences into the human chromosomal β -globin locus by homologous recombination. *Nature*, 317(6034), 230-234. doi:10.1038/317230a0
- Snow, A. A. (2019). Genetically Engineering Wild Mice to Combat Lyme Disease: An Ecological Perspective. *BioScience*, 69(9), 746-756. doi:10.1093/biosci/biz080

- Song, F., & Stieger, K. (2017). Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Molecular therapy. Nucleic acids*, 7, 53-60. doi:10.1016/j.omtn.2017.02.006
- Song, J., Yang, D., Xu, J., Zhu, T., Chen, Y. E., & Zhang, J. (2016). RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nature Communications*, 7(1), 10548. doi:10.1038/ncomms10548
- Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, 98(3), 503-517. doi:10.1016/s0022-2836(75)80083-0
- Sutter, A., & Lindholm, A. K. (2015). Detrimental effects of an autosomal selfish genetic element on sperm competitiveness in house mice. *Proceedings of the Royal Society B: Biological Sciences*, 282(1811), 20150974.
- Suzuki, H., Kamada, N., Ueda, O., Jishage, K., Kurihara, Y., Kurihara, H., . . . Toyoda, Y. (1997). Germ-line contribution of embryonic stem cells in chimeric mice: influence of karyotype and in vitro differentiation ability. *Exp Anim*, 46(1), 17-23. doi:10.1538/expanim.46.17
- Tashiro, F., Kanai-Azuma, M., Miyazaki, S., Kato, M., Tanaka, T., Toyoda, S., . . . Miyazaki, J.-I. (2010). Maternal-effect gene *Ces5/Ooep/Moep19/Floped* is essential for oocyte cytoplasmic lattice formation and embryonic development at the maternal-zygotic stage transition. *Genes to Cells*, 15(8), 813-828. doi:10.1111/j.1365-2443.2010.01420.x
- Tate, P. H., & Skarnes, W. C. (2011). Bi-allelic gene targeting in mouse embryonic stem cells. *Methods*, 53(4), 331-338. doi:https://doi.org/10.1016/j.ymeth.2010.12.025
- Taylor, S. C., Laperriere, G., & Germain, H. (2017). Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci Rep*, 7(1), 2409. doi:10.1038/s41598-017-02217-x
- Te Riele, H., Maandag, E. R., & Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proceedings of the National Academy of Sciences*, 89(11), 5128-5132.
- Thomas, K. R., & Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, 51(3), 503-512.
- Tong, Z.-B., Gold, L., Pfeifer, K. E., Dorward, H., Lee, E., Bondy, C. A., . . . Nelson, L. M. (2000). *Mater*, a maternal effect gene required for early embryonic development in mice. *Nature Genetics*, 26(3), 267-268. Retrieved from <http://dx.doi.org/10.1038/81547>
- Towns, D. R., Atkinson, I. A. E., & Daugherty, C. H. (2006). Have the Harmful Effects of Introduced Rats on Islands been Exaggerated? *Biological Invasions*, 8(4), 863-891. doi:10.1007/s10530-005-0421-z
- UNCBD Ad Hoc Technical Expert Group (AHTEG). (2020). *REPORT OF THE AD HOC TECHNICAL EXPERT GROUP ON RISK ASSESSMENT*. Retrieved from <https://www.cbd.int/doc/c/a763/e248/4fa326e03e3c126b9615e95d/cp-ra-ahteg-2020-01-05-en.pdf>
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, 35(suppl_2), W71-W74.

- Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, *11*(9), 636-646. doi:10.1038/nrg2842
- Valamehr, B., Tsutsui, H., Ho, C.-M., & Wu, H. (2011). Developing defined culture systems for human pluripotent stem cells. *Regenerative medicine*, *6*(5), 623-634. doi:10.2217/rme.11.54
- van der Valk, J., Bieback, K., Buta, C., Cochrane, B., Dirks, W., Fu, J., . . . Liebsch, M. (2018). Fetal bovine serum (FBS): past–present–future. *Altex*, *35*(1), 1-20.
- Wakayama, T., Rodriguez, I., Perry, A. C. F., Yanagimachi, R., & Mombaerts, P. (1999). Mice cloned from embryonic stem cells. *Proceedings of the National Academy of Sciences*, *96*(26), 14984. doi:10.1073/pnas.96.26.14984
- Wang, B., Li, K., Wang, A., Reiser, M., Saunders, T., Lockey, R. F., & Wang, J.-W. (2015). Highly efficient CRISPR/HDR-mediated knock-in for mouse embryonic stem cells and zygotes. *BioTechniques*, *59*(4), 201-208. doi:10.2144/000114339
- Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*, *153*(4), 910-918. doi:https://doi.org/10.1016/j.cell.2013.04.025
- Wang, R., Liang, J., Jiang, H., Qin, L.-J., & Yang, H.-T. (2008). Promoter-dependent EGFP expression during embryonic stem cell propagation and differentiation. *Stem cells and development*, *17*(2), 279-290.
- Webber, B. L., Raghu, S., & Edwards, O. R. (2015). Opinion: Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? *Proceedings of the National Academy of Sciences*, *112*(34), 10565. doi:10.1073/pnas.1514258112
- Wedell, N., Price, T. A. R., & Lindholm, A. K. (2019). Gene drive: progress and prospects. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1917), 20192709. doi:doi:10.1098/rspb.2019.2709
- Weissmann, C., & Flechsig, E. (2003). PrP knock-out and PrP transgenic mice in prion research. *British Medical Bulletin*, *66*(1), 43-60. doi:10.1093/bmb/66.1.43
- Wells, D. J. (2011). Animal welfare and the 3Rs in European biomedical research. *Annals of the New York Academy of Sciences*, *1245*(1), 14-16. doi:10.1111/j.1749-6632.2011.06335.x
- Whitelaw, B., & McFarlane, G. R. (2019). Accelerating evolution. *The Biologist (RSB)*, *66*(6), 18-21. Retrieved from <https://thebiologist.rsb.org.uk/biologist/158-biologist/features/2245-accelerating-evolution>
- Wilde, J. J., Aida, T., Wienisch, M., Zhang, Q., Qi, P., & Feng, G. (2018). Efficient Zygotic Genome Editing via RAD51-Enhanced Interhomolog Repair. *bioRxiv*, 10.1101/263699, 263699. doi:10.1101/263699
- Williams, E. S. (2005). Chronic wasting disease. *Vet Pathol*, *42*(5), 530-549. doi:10.1354/vp.42-5-530
- Williams, E. S., & Young, S. (1992). Spongiform encephalopathies in Cervidae. *Rev Sci Tech*, *11*(2), 551-567. doi:10.20506/rst.11.2.611

- Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., . . . Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*, 336(6200), 684-687.
- Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., . . . Crisanti, A. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*, 473(7346), 212-215. doi:10.1038/nature09937
- Windbichler, N., Papathanos, P. A., Catteruccia, F., Ranson, H., Burt, A., & Crisanti, A. (2007). Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Research*, 35(17), 5922-5933. doi:10.1093/nar/gkm632
- Witmer, G. W., & Fuller, P. L. (2011). Vertebrate species introductions in the United States and its territories. *Current Zoology*, 57(5), 559-567. doi:10.1093/czoolo/57.5.559
- Wootton, J. T. (1994). The nature and consequences of indirect effects in ecological communities. *Annual review of ecology and systematics*. 443-466.
- World Health Organization. (2014). *Guidance framework for testing of genetically modified mosquitoes* (Retrieved from <https://apps.who.int/iris/handle/10665/127889>). Geneva: World Health Organization.
- World Health Organization. (2019). *World malaria report 2019* (Retrieved from <https://apps.who.int/iris/handle/10665/330011>). Geneva: World Health Organization.
- Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., & Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature cell biology*, 13(7), 838-845.
- Yang, D., Song, J., Zhang, J., Xu, J., Zhu, T., Wang, Z., . . . Chen, Y. E. (2016). Identification and characterization of rabbit ROSA26 for gene knock-in and stable reporter gene expression. *Sci Rep*, 6(1), 25161. doi:10.1038/srep25161
- Yang, H., Wang, H., Shivalila, Chikdu S., Cheng, Albert W., Shi, L., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*, 154(6), 1370-1379. doi:10.1016/j.cell.2013.08.022
- Yang, T.-T., Cheng, L., & Kain, S. R. (1996). Optimized Codon Usage and Chromophore Mutations Provide Enhanced Sensitivity with the Green Fluorescent Protein. *Nucleic Acids Research*, 24(22), 4592-4593. doi:10.1093/nar/24.22.4592
- Yanik, M., Ponnampalani, S. P. G., Wimmer, T., Trimborn, L., Müller, C., Gambert, I., . . . Stieger, K. (2018). Development of a Reporter System to Explore MMEJ in the Context of Replacing Large Genomic Fragments. *Molecular Therapy - Nucleic Acids*, 11, 407-415. doi:10.1016/j.omtn.2018.03.010
- Yao, R., Liu, D., Jia, X., Zheng, Y., Liu, W., & Xiao, Y. (2018). CRISPR-Cas9/Cas12a biotechnology and application in bacteria. *Synthetic and Systems Biotechnology*, 3(3), 135-149. doi:https://doi.org/10.1016/j.synbio.2018.09.004
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13(1), 134.

- Ying, Q.-L., Nichols, J., Chambers, I., & Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell*, 115(3), 281-292.
- Zakrzewski, W., Dobrzyński, M., Szymonowicz, M., & Rybak, Z. (2019). Stem cells: past, present, and future. *Stem cell research & therapy*, 10(1), 68-68. doi:10.1186/s13287-019-1165-5
- Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., . . . Regev, A. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163(3), 759-771.
- Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., . . . Zhang, F. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol*, 35(1), 31-34. doi:10.1038/nbt.3737
- Zheng, J., Li, H., Zhang, Q., Sun, L., Liu, X., & Luo, C. (2018). Homologous recombination occurs frequently at innate GT microsatellites in normal somatic and germ cells in vivo. *BMC Genomics*, 19(1), 359. doi:10.1186/s12864-018-4758-y
- Zhou, D., Tan, L., Li, J., Liu, T., Hu, Y., Li, Y., . . . Wang, A. (2018). Identification of Homologous Recombination Events in Mouse Embryonic Stem Cells Using Southern Blotting and Polymerase Chain Reaction. *Journal of visualized experiments : JoVE*, 10.3791/58467(141). doi:10.3791/58467
- Zischewski, J., Fischer, R., & Bortesi, L. (2017). Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology Advances*, 35(1), 95-104. doi:https://doi.org/10.1016/j.biotechadv.2016.12.003

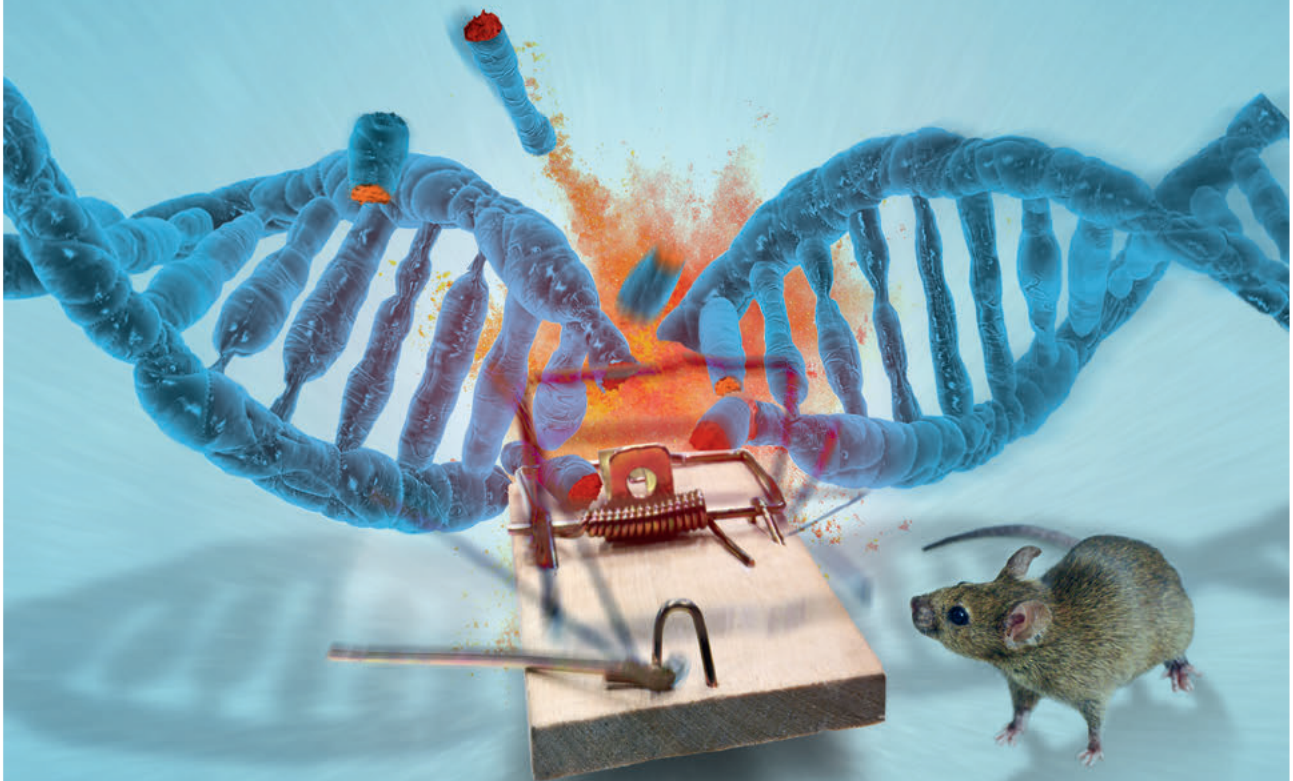
A Appendices

A.1 Publication: CRISPR-Based Gene Drives for Pest Control

The following pages contain a paper I co-authored during my PhD and has been referred to within this document (including section 1.8.4) The corresponding journal cover image is also presented, which was created in collaboration with Norrie Russell and Gordon MacPherson of The Roslin Institute.

Trends in Biotechnology

Volume 36 Number 2
February 2018
ISSN 0167-7799



Focus on Applications
of CRISPR

CellPress
REVIEWS

15. Malik, D.J. *et al.* (2017) Formulation, stabilisation and encapsulation of bacteriophage for phage therapy. *Adv. Colloid Interface Sci.* Published online May 14, 2017. <http://dx.doi.org/10.1016/j.cis.2017.05.014>

Focus on Applications of
CRISPR

Forum

CRISPR Based Gene Drives for Pest Control

Gus R. McFarlane,¹
C. Bruce A. Whitelaw,^{1,*} and
Simon G. Lillico¹

Clustered regularly interspaced short palindromic repeats (CRISPR)-based gene drives (GDs) could be used to spread desirable genetic elements through wild populations. With the imminent development of this technology in vertebrates, we believe that it is timely to highlight two forms of sex-ratio distorting GDs that show potential as pest management tools.

In agriculture and wildlife, invasive pests are well known: rabbits and cane toads in Australia, mink in the UK, and the omnipresent infestation of rodents around the globe. It is estimated that in the USA, introduced rats cost the economy more than US\$27 billion per year. [1]. To counter the impact of vertebrate pests, control measures are deployed that include shooting, poison baiting, trapping, and the release of biological agents (Figure 1A). These methods are costly and inadequate, and often lead to unwanted suffering in both target and non target species [2]. GDs may offer a more cost effective, humane, and species specific alternative than current approaches.

Gene Drives for Sex-Ratio Distortion

The GD concept has emerged from observations that naturally occurring selfish genetic elements, such as homing endonuclease genes and transposons, are preferentially inherited at frequencies greater than predicted by Mendelian inheritance. This 'super Mendelian' inheritance allows these elements to drive through a population even if they reduce the fitness of an individual organism [3] (Figure 1B). The recent discovery and repurposing of RNA guided CRISPR endonucleases into a set of gene editing tools (Box 1) allows the development of synthetic GDs in a standard molecular biology laboratory [4].

One potential application of GD is to distort the sex ratio of a population. By skewing the sex ratio away from the favored Fisherian ratio of 1:1 male to female, it is possible to manipulate the reproductive performance of a population. In most pest species, female procreative capacity is responsible for maintaining the overall population size. Therefore, an efficient means of population suppression is to bias the sex ratio in favor of males [5]. A grossly male population will result in a population decline, while an all male population will lead to eradication. Here, we describe two forms of CRISPR based sex ratio distorting GDs, homing GD and X shredder (XS), both of which have the potential to drive maleness. To date,

these GD systems have only been engineered in proof of concept studies in mosquitoes [6,7], with the focus on controlling vector borne diseases. After insects, invasive vertebrate pests are likely to be the next GD target.

Homing Gene Drive Targeting Female Fertility

A homing GD works by copying or 'homing' itself into a target site in the genome. To build a CRISPR based homing GD, an animal is engineered with a GD cassette that expresses a CRISPR endonuclease, such as Cas9, and one or more guide RNAs (gRNAs) (Box 1) from one allele that can cut at a conserved target site on the sister allele on the homologous chromosome (Figure 1C). After CRISPR mediated cleavage, homology directed repair (HDR) results in the CRISPR machinery and any additional payload included in the GD cassette being copied onto the homologous chromosome [4,8]. This process ensures homozygosity for the GD cassette.

Targeting a homing GD to a haplosufficient female fertility gene (HFFG) can be used to disrupt the coding sequence of the gene, rendering homozygous female offspring infertile, while males and heterozygous females will retain normal fertility. Importantly, for this strategy to drive through a population, the homing event should occur only in germline cells that are precursors to sperm or eggs. This can

Box 1. CRISPR Cas9

The CRISPR-Cas prokaryotic immune defence system, and its associated proteins (Cas), has been repurposed into a set of gene-editing tools. Currently, the CRISPR-Cas9 system from *Streptococcus pyogenes* is the most widely used CRISPR system for genetic manipulation. The system comprises two components: a guide RNA (gRNA) and a nonspecific endonuclease (Cas9; Figure 1). The gRNA includes an 89-nucleotide 'scaffold' sequence that Cas9 binds to and a 20-nucleotide user-defined 'targeting' sequence that delivers the endonuclease to the correct site in the genome by Watson-Crick base pairing with the target sequence. Provided that the genomic target is immediately adjacent to a protospacer adjacent motif (PAM; 5'-NGG-3' for *S. pyogenes* Cas9), the endonuclease cuts the DNA generating a double-stranded DNA break (DSB). The natural DNA repair machinery of the cell then recognizes that the DNA is damaged and repairs the DSB by one of two pathways: (i) in the presence of a DNA repair template, such as the homologous chromosome or an exogenous DNA template, the DSB is repaired through high-fidelity homology-directed repair (HDR). HDR can be used to make accurate repairs or precisely edit the DNA sequence; or (ii) in the absence of a DNA repair template, a DSB is repaired by the error-prone nonhomologous end-joining (NHEJ) pathway. Repair by NHEJ yields deletion or insertion mutations (indels) [12].

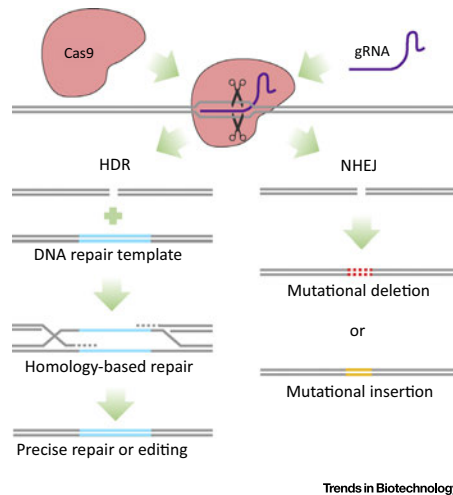


Figure 1. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Cas9 Gene Editing. The Cas9 endonuclease is guided to the target sequence in the genome by the guide RNA (gRNA). At the target site, Cas9 cleaves the DNA, creating a double-stranded DNA break (DSB). A DSB induced by Cas9 can be repaired by homology-directed repair (HDR) or nonhomologous end joining (NHEJ). HDR can precisely repair or edit the DNA sequence. NHEJ-mediated repair produces insertion or deletion mutations (indels).

be achieved by using a germline specific promoter to express Cas9. By restricting homing to the germline, this will initially allow rapid spread of the GD and accumulation of fertile heterozygous GD animals that produce mostly GD gametes [7,9] (Figure 1C). As mating between heterozygous GD carriers becomes increasingly likely, the population will decline due to infertility of the homozygous GD female offspring, which are homozygous null for the HFFG [9]. With every generation, the sex ratio will become more biased towards males, eventually resulting in a population crash. Hammond and colleagues [7] developed this system in mosquitoes and achieved transmission rates of 91.4–99.6% in caged populations. In theory, homing GDs could be adapted to control most vertebrate pests and several groups are currently undertaking pilot studies in mouse models.

X-Shredder

In XY heterogametic species, an XS is a type of sex ratio distorting GD that cuts

the X chromosome at multiple sites during spermatogenesis, thus shredding the X chromosome beyond repair [8]. To engineer a CRISPR based XS, an XS cassette is inserted within a neutral intergenic region of the Y chromosome. The cassette encodes Cas9, which is expressed under the control of a spermatogenesis specific promoter, and one or more gRNAs that target conserved repetitive sequences unique to the X chromosome. Given that most X chromosomes are destroyed during spermatogenesis, most sperm that mature and reach the oocyte are Y bearing, resulting in a biased sex ratio in favor of males. By placing the XS cassette on the Y chromosome, all male offspring will inherit the cassette and continue transmitting the XS to subsequent generations.

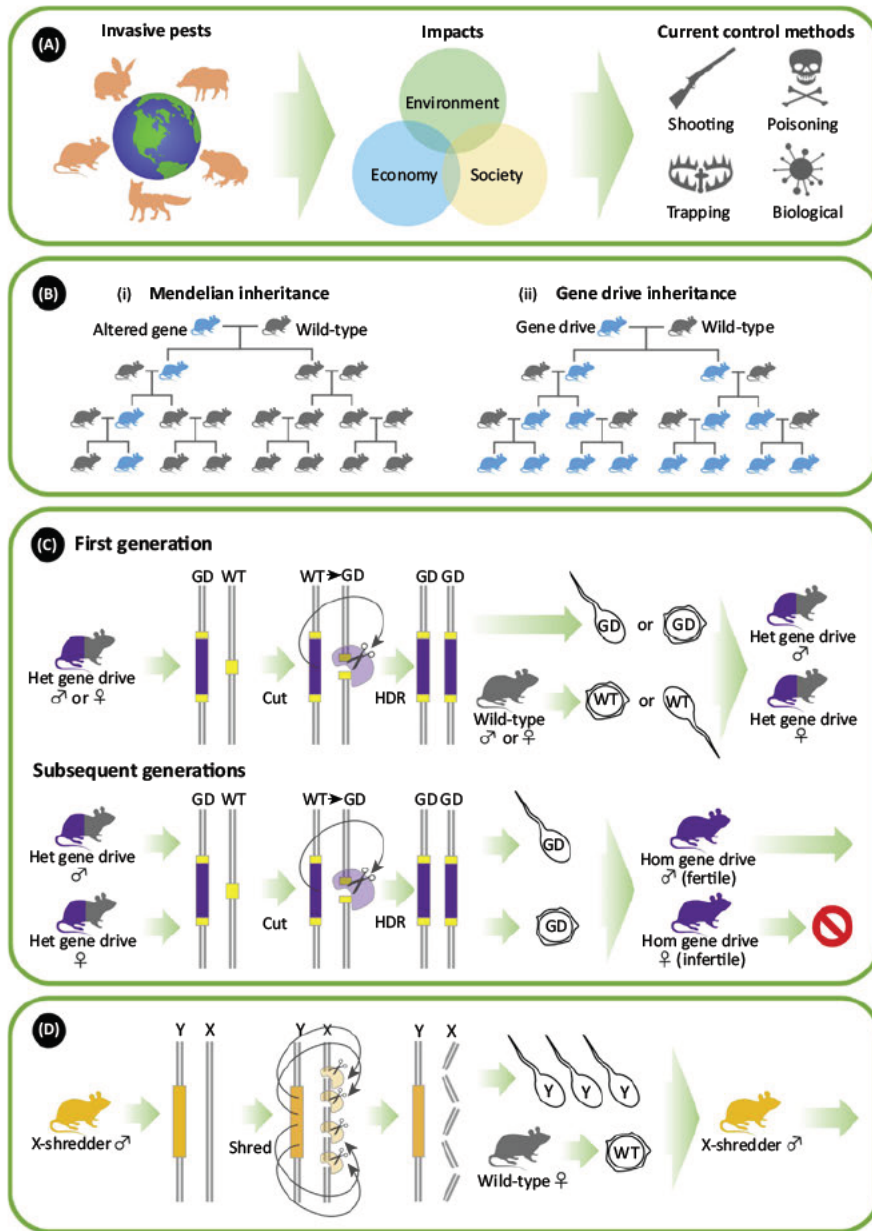
A CRISPR based XS has been engineered in mosquitoes, although the system was commendably safeguarded by expressing the XS cassette from an autosome instead of the Y chromosome. With

this approach, Galizi and colleagues [6] achieved male bias among progeny ranging from 86.1% to 94.8% in laboratory contained mosquito populations. Although successful in mosquito, technical challenges facing the adaptation of an XS into vertebrates include identifying appropriate spermatogenesis specific promoters in target species and the transcriptional silencing of mammalian sex chromosomes during meiosis. The latter may hinder expression of Cas9 from the Y chromosome, as well as the accessibility of the endonuclease to shred the X chromosome.

Drive Resistance and Inactivation Strategies

The two forms of GDs described above are self-perpetuating and, in theory, would only require the release of a small number of engineered animals to initiate drive. The duration and extent of spread would be limited by naturally arising resistant alleles that prevent CRISPR mediated cleavage. Resistant alleles could exist in the population before release or originate from indels generated when CRISPR mediated cleavage is repaired by the error-prone nonhomologous end joining (NHEJ) pathway and alters the gRNA recognition sequence [8]. The rate of NHEJ mediated repair will be dependent on the species, target site, and the stage of development that DNA cleavage occurs. Given that natural selection tends to favor equal sex ratios, resistant alleles that restore function would spread rapidly through the population [10].

Of the two strategies presented here, XS should be less prone to inactivation by resistant alleles because it targets multiple sites and, therefore, would require an animal to simultaneously acquire multiple resistant alleles to incapacitate the drive. Following a similar approach, it has been suggested that the evolutionary stability of homing GDs could be improved by using multiple gRNAs closely spaced along the target region [4]. To test this hypothesis,



Trends in Biotechnology

Figure 1. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Based Gene Drive (GD) Strategies for Controlling Vertebrate Pests. (A) Invasive pests are a global concern. Current control strategies are inhumane, costly, and often inadequate. (B) (i) Mendelian inheritance of an altered gene. (ii) (Figure legend continued on the bottom of the next page.)

Prowse and colleagues [9] used *in silico* modeling to demonstrate that multiple gRNAs are necessary for homing GDs to evade drive resistance and successfully suppress vertebrate pest populations. Even if drive resistance were to prevail, it would be possible to release a second GD targeting a different gene to continue suppressing the population.

Conversely, if a GD were not limited by naturally arising resistance, it would have the potential to spread indefinitely through a species. Therefore, it is essential to have strategies in place that could deliberately inactivate a GD that escaped containment or was causing unforeseen impacts. Fortunately, both homing GD and XS systems can be inactivated by the release of animals bearing engineered functionally resistant alleles or a reversal gene drive that immunizes the animal against the original drive [4,8]. However, it is important to recognize that, with the current technology, once either of these systems is released, complete reversion to a wild type genotype would not be possible because residual Cas9 and gRNA would still be present.

Risks and Benefits

Genetically engineered animals normally come with few ecological risks. Most engineered traits are for human benefit and will not be favored by natural

selection. By contrast, GDs can spread through populations even if they reduce the fitness of each carrier animal [4]. This gives GDs more scope to escape the target population and unintentionally affect extraneous ecosystems. However, the potential benefits of GDs are equally as impactful as the risks. GDs could revolutionize public health, agriculture, and, as discussed here, could be applied for pest control and ecological restoration. In line with the recent decision at the United Nations Convention on Biodiversity, we believe that the potential benefits of GDs warrant further investigation.

Concluding Remarks

For the first time, we have the makings of a technology that could reduce or eliminate a pest population in a humane and species specific manner. If proven effective, the decision to deploy a GD should be based on substantiated research and involve public engagement to ensure there is societal consensus. With the rapid progress in this space, the risks associated with current GD architectures are likely to be reduced with the realization of self limiting GD strategies [11].

Acknowledgement

G.R.M. gratefully acknowledges the support provided by the Commonwealth Scholarship Commission. This work was part funded by the BBSRC through ISP BB/P013759/1.

¹The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, EH25 9RG, UK

*Correspondence:
bruce.whitelaw@roslin.ed.ac.uk (C. Bruce A. Whitelaw).
<https://doi.org/10.1016/j.tibtech.2017.10.001>

References

1. Pirmentel, D. *et al.* (2007) Environmental and economic costs of vertebrate species invasions into the United States. In *Managing Vertebrate Invasive Species: Proceedings of an International Symposium* (Wtmer, G.W., ed.), pp. 2–8, USDA APHIS Wildlife Services
2. National Academies of Sciences, Engineering, and Medicine (2016) *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*, The National Academies Press
3. Burt, A. and Trivers, R. (2009) *Genes in Conflict: The Biology of Selfish Genetic Elements*, Harvard University Press
4. Esvelt, K.M. *et al.* (2014) Concerning RNA guided gene drives for the eradication of wild populations. *eLife* 3, e03401
5. Hamilton, W.D. (1967) Extraordinary sex ratios. A sex ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* 156, 477–488
6. Galizi, R. *et al.* (2016) A CRISPR Cas9 sex ratio distortion system for genetic control. *Sci. Rep.* 6, 31139
7. Hammond, A. *et al.* (2016) A CRISPR Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* 34, 78–83
8. Champer, J. *et al.* (2016) Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* 17, 146–159
9. Prowse, T.A.A. *et al.* (2017) Dodging silver bullets: good CRISPR gene drive design is critical for eradicating exotic vertebrates. *Proc. R. Soc. B Biol. Sci.* 284, 20170799
10. Hammond, A.M. *et al.* (2017) The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *bioRxiv* Published online June 12, 2017. <http://dx.doi.org/10.1101/149005>
11. Noble, C. *et al.* (2016) Daisy chain gene drives for the alteration of local populations. *bioRxiv* Published online June 7, 2017. <http://dx.doi.org/10.1101/057307>
12. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas Systems. *Science* 339, 819–823

Inheritance of a GD. (C) A homing GD targeting female fertility. In the first generation, heterozygous (Het) GD animals are released carrying a GD cassette (purple), which disrupts the coding sequence of a haplosufficient female-fertility gene (HFFG; yellow). Within the germ cells, the GD cassette expresses Cas9 and one or more guide RNAs (gRNAs) that cut the HFFG on the wild-type (WT) chromosome. The germ cells then repair the cut by homology-directed repair (HDR), using the GD chromosome as the repair template. This process copies the GD cassette onto the WT chromosome and ensures that most sperm or eggs carry the GD cassette. Matings between Het GD and WT animals will give rise to an increasing number Het GD animals. In subsequent generations, as mating between Het GD animals becomes increasingly likely, the population will decline through infertility of homozygous (Hom) GD female offspring, which are homozygous null for the HFFG. (D) X-Shredder (XS). During spermatogenesis, Cas9 and gRNA(s) are expressed from the XS cassette (orange) located on the Y-chromosome (Y) and shred the X-chromosome (X) beyond repair. Most sperm that mature and reach the oocyte are Y-bearing, resulting in most offspring being XS males.

A.2 PCR primers

Table A.2 PCR primers.

Primer pair name	Primer	Sequence	Polymerase	Annealing temp. (C)
ppNLRP5	oGM27	GGCTCAGAAGAAGCTAGACTGA	Phusion	62
	oGM28	TTCTCGAGGACCAGCTGAGA		
ppOOEP	oGM72	TCGTGAAGGCGGGTAAGAAA	Phusion	64
	oGM75	GCCGGCGTGGCATAGAA		
ppROSA26	oGM93	GCACGTTTCCGACTTGAGTT	Phusion	64
	oGM95	GGCGGATCACAAGCAATAAT		
ppGA1	oGM9	TGAAATCACTTTTTTTTTCAGGTTGGAGGCCGCCTAGTCCAGTGTGG	Q5	62
	oGM10	GTGGTCCATTATAGTCCATGGTGGCAAACCTCGAGAGATCTAACGG		
ppGA2	oGM106	CTTCCTTGACCCTGGAAGTCCCTAATAGAGGGCCCGTTTAAACCC	Q5	62
	oGM107	ACTCATACTAGTAATAGCCCACCGGCAGCTGGTTCTTTCCGCCTC		
ppGA3	oGM108	TGAGTCCCGCGGAAGGCGCTACAGGGTGCTCTACATGGAAGCCTG	Q5	64
	oGM109	GCCTTTTTACGGTTCCTGGCGGATCCGGAAAGCCAGCCAGTTTTA		
ppGA4	oGM139	GCCAGTCACTATGGCGTGCTGGTCGACCATCTCCAGAAAACAGCTCAA	Q5	65
	oGM140	TACCGTAAGTTATGTAACGGGTACCGGGCCGAGTGCGAAGTCTCA		
ppGA5	oGM208	TGAGACTTCGCACTCGGCCCGGTACATTCCTGCAGCCCAATTCCG	Q5	62
	oGM209	TACCGTAAGTTATGTAACGGGTACGCCGCTCTAGAACTAGTGGAT		
ppGA6	oGM64	TCTAGAGAATTCTCTAGAAATATTCACGTTGTAACGACGGCCA	Q5	62
	oGM65	CTGCGGCCGCGTCGACGTTTAAACCAACGGGTACCTCTAGAGCCA		
ppGA7	oGM121	CTAGAGGTACCCGTTGGTTTATTTCCCATGATTCCTTCAT	Q5	62
	oGM122	CTGCGGCCGCGTCGACGTTTGGATCCAGCCATTTGTCTGCAGAATT		
ppGA8	oGM127	CGCCAATTCTGCAGACAAATGGCTGTATTTCCCATGATTCCTTCAT	Q5	62
	oGM128	CCATCTGCGGCCGCGTCGACGTTTGTGTACAAGCCATTTGTCTGCAGAATT		
ppGA9	oGM239	ACGAAGCTGTTCTGTATTTCCAGAGATTCTGCAGCCCAATTCCG	Q5	62
	oGM240	CCCCTGCAGACAACGCCACACAGGCCGCTCTAGAACTAGTGGAT		

ppTFRC	oGM249	GGGTCTAAGTCTACAGTGGC	Bio-Rad SM	60
	oGM253	CGATCATTGATTTCCCTCATGAC		
ppPAC	oGM243	ATCGGCAAGGTGTGGGTC	Bio-Rad SM	60
	oGM248	TTCCATCTGTTGCTGCGC		
ppOOEP_5'	oGM185	ATGGGCTTAGCAAGTAAGGTACA	OneTaq	56
	oGM186	GGTCCGGATCGACGGT		
ppOOEP_3'	oGM189	CGACTTCTTCAAGTCCGCC	OneTaq	56
	oGM190	CAACCTACATACTCCCCGAACC		
ppROSA26_5'	oGM163	ACTACTGTGTTGGCGGACTG	OneTaq	56
	oGM164	GTGGCTAACGTACGGGAGAG		
ppROSA26_3'	oGM236	GCGTCTCGCCGACCA	OneTaq	56
	oGM168	CCATGAGTCAAGCCAGTCCAAG		
ppgRNAs_LoxP	oGM172	TACACAAACGTCGACGAAGC	OneTaq	52
	oGM95	GGCGGATCACAAGCAATAAT		
ppCAG_LoxP	oGM75	GCCGGCGTGGCATAGAA	OneTaq	55
	oSL35	GTCAATAGGGGGCGTACTTG		
ppVASA_LoxP	oGM75	GCCGGCGTGGCATAGAA	OneTaq	57
	oGM257	TGTGGGTCCAGAGTCTTACCA		

A.3 PCR thermocycling conditions

Table A.3 PCR thermocycling conditions for each polymerase.

	NEB Phusion High-Fidelity Polymerase			NEB Q5 High-Fidelity Polymerase			NEB OneTaq Polymerase		
	Temp. (C)	Time	Cycles	Temp. (C)	Time	Cycles	Temp. (C)	Time	Cycles
Initial denaturation	98	2 min	1	98	30 s	1	94	30 s	1
Denaturation	98	30 s	30	98	10 s	30	94	20 s	32
Annealing	See primer list	30 s		See primer list	30 s		See primer list	30 s	
Extension	72	20 s/kb		72	20 s/kb		68	1 min/kb	
Final extension	72	5 min	1	72	2 min	1	68	5 min	1

A.4 Primers used for Sanger sequencing

Table A.4 Primers used for Sanger sequencing.

Primer	Sequence	Primer	Sequence
oGM34	CACATTTCCCGAAAAGTGC	oGM75	GCCGGCGTGGCATAGAA
oGM35	TAACTTACGGTAAATGGCCC	oGM76	GCACTGGGTTGTCCAAAGATG
oGM36	TCAGGTTGGAGGCCGCCTAG	oGM92	AAGGGAGCTGCAGTGGAGTA
oGM37	CCAGTAGAGCTAGCTGAATA	oGM93	GCACGTTTCCGACTTGAGTT
oGM38	TCCAGCTGGTGCAGACCTAC	oGM95	GGCGGATCACAAGCAATAAT
oGM39	TCCGCATCCCCTACTACGTG	oGM110	GTGTGGAATTGTGAGCGGAT
oGM40	AGGATATACAAAAGGCACAG	oGM155	GGATTACTGACTTCCTGAAT
oGM41	TCCGGAAGGATTTCCAGTTT	oGM194	CAGAAACAGGTCGGCGTACT
oGM42	TGGCCCTGCCCTCCAAATAT	oGM206	TGCTAAAGCGCATGCTCCAG
oGM42	TGGCCCTGCCCTCCAAATAT	oGM234	CTCGACGTCACCGCATGT
oGM43	AGCAGCACGACTTCTTCAAG	oGM236	GCGTCTCGCCCGACCA
oGM63	CATCGCTATTACCATGGTCC	oSL35	GTCAATAGGGGGCGTACTTG

A.5 ddPCR probes

Table A.5 Probes for ddPCR.

Probe name	Sequence	Gene	Probe type
oGM255	TCCGCTCGTGGAGACTACTTCCGTGCT	<i>TFRC</i>	IDT 100 nm PrimeTime 5' HEX/ZEN/3' IBFQ
oGM256	CCGCCCCCGCTTCGACGCTC	<i>PAC</i>	IDT 100 nm PrimeTime 5' 6-FAM/ZEN/3' IBFQ

A.6 Gene fragments and DNA oligos for cloning

Table A.6 Gene fragments (gBlocks) and ssDNA oligos used for cloning plasmid DNA constructs.

Name	Sequence	Fragment type
gBlock1	ACGTTGTA AACGACGGCCAGTGAATTGCCGGCGATATCGTAAGCTTTATAAGACCTAGGTAGTGCA CAATGCATCACCTTACTCTCCCGTACGTTAGCCACTATTTCCCATGATTCCTTCATATTTGCATATACG ATACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATAC GTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAATGGACTATCA TATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAC CGGAGACGGGATACCGTCTCTGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTGTTTAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGCAGACAAATGGCTCTAGAGGTAC CCGTTACATAACTTACGGTAAATGGCCCGCCTGGC	IDT gBlock
gBlock2	TCGGCATGGACGAGCTGTACAAGGAATTCTAACTAGAGCTCGCTGATCAGCCCCCGTGCCTTCCTTG ACCCTGGAAGTCCCTAACCGGTGGGCTATTACTAGTATGAGTCCCGCGGAAGGCGCTACAGGGATC CGCCAGGAACCGTAAAAAGGCCGCGTTCCTCGCGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCT	IDT gBlock
oGM237	CTAGACCTAGGCAATTGGGCGCGCCTACTTAATTAAGCTGCAGGC	IDT DNA Oligo
oGM238	CATGGCCTGCAGGCTTAAGTTAATTAAGTAGGCGCGCCAATTGCCTAGGT	IDT DNA Oligo
oGM147	TCGACGAAGCTGTTCTGTATTTCCAGAGCTAGCTGTGTGGGCGTTGTCCTGCA	IDT DNA Oligo
oGM148	GGACAACGCCACACAGCTAGCTCTGGAATACAGAACAGCTTCG	IDT DNA Oligo

A.7 gRNA DNA oligos

Table A.7 DNA oligos for cloning crRNAs into pSL70 and pY094 plasmids.

gRNA	Annealing oligos	Sequence
<i>OOEP</i> SpCas9 gRNA 1	oGM19	CACCGACTTCGCACTCGGCCCTGG
	oGM20	AAACCCAGGGCCGAGTGCGAAGTC
<i>OOEP</i> SpCas9 gRNA 2	oGM100	CACCGCCCTGGTGGTTCCCAATTC
	oGM101	AAACGAATTGGGAACCACCAGGGC
<i>OOEP</i> SpCas9 gRNA 3	oGM98	CACCGGATTACTGACTTCCTGAAT
	oGM99	AAACATTCAGGAAGTCAGTAATCC
<i>OOEP</i> SpCas9 gRNA 4	oGM96	CACCGCTTCCATGTAGAGCACCAG
	oGM97	AAACCTGGTGTCTACATGGAAGC
<i>NLRP5</i> SpCas9 gRNA 1	oGM13	CACCGTTTGGCACAGCTGATTGCTA
	oGM14	AAACTAGCAATCAGCTGTGCCAAAC
<i>NLRP5</i> SpCas9 gRNA 2	oGM15	CACCGATCTTTGTCACTAGATCCC
	oGM16	AAACGGGATCTAGTGACAAAGATC
<i>NLRP5</i> SpCas9 gRNA 3	oGM104	CACCGAAACAAGAGTCTTTCTGGTT
	oGM105	AAACAACCAGAAAGACTCTTGTTTC
<i>NLRP5</i> SpCas9 gRNA 4	oGM17	CACCGATAGTGTTCATATCATCATGT
	oGM18	AAACACATGATGATATGACACTATC
<i>ROSA26</i> SpCas9 gRNA 1	oGM115	CACCGAAGGGATTCTCCCAGGCC
	oGM116	AAACGGGCCTGGGAGAATCCCTTC
<i>ROSA26</i> SpCas9 gRNA 2	oGM113	CACCGGCAGGCTTAAAGGCTAACC
	oGM114	AAACGGTTAGCCTTTAAGCCTGCC
<i>OOEP</i> AsCas12a gRNA 1	oGM80	AGATTGCCACCCAGGCTTCCATGTA
	oGM81	AAAATACATGGAAGCCTGGGTGGCA
<i>OOEP</i> AsCas12a gRNA 2	oGM86	AGATTGAGATTTCCGCTTGGTCTGT
	oGM87	AAAAACAGACCAAGCGGAAATCTCA
<i>OOEP</i> AsCas12a gRNA 3	oGM84	AGATAGCTAGGTTCCCAGAGTTGAC
	oGM85	AAAAGTCAACTCTGGGAACCTAGCT
<i>OOEP</i> AsCas12a gRNA 4	oGM82	AGATCTTTGTGCCAGGCCGCCATGT
	oGM83	AAAAACATGGCGGCCTGGCACAAAG
<i>NLRP5</i> AsCas12a gRNA 1	oGM51	AGATTGGTTGGGACATGATCTTTGT
	oGM52	AAAAACAAAGATCATGTCCCAACCA
<i>NLRP5</i> AsCas12a gRNA 2	oGM53	AGATTCATAGATGGCTTGGATGATA
	oGM54	AAAATATCATCCAAGCCATCTATGA
<i>NLRP5</i> AsCas12a gRNA 3	oGM55	AGATTGGTGGTAATGATGAGAAAGG
	oGM56	AAAACCTTTCTCATCATTACCACCA
<i>NLRP5</i> AsCas12a gRNA 4	oGM59	AGATTCATCATTACCACCAGAAACA
	oGM60	AAAATGTTTCTGGTGGTAATGATGA
<i>ROSA26</i> AsCas12a gRNA 1	oGM210	AGATTAGAAGATGGGCGGGAGTCTTCT
	oGM211	AAAAAGAAGACTCCCGCCCATCTTCTA
<i>ROSA26</i> AsCas12a gRNA 2	oGM212	AGATAGCCTGCCCAGAAGACTCCCGCC
	oGM213	AAAAGGCGGGAGTCTTCTGGGCAGGCT

A.8 crRNAs for SpCas9 ribonucleoprotein

Table A.8 crRNAs and tracrRNA used for SpCas9 ribonucleoprotein.

crRNA	Sequence	crRNA type
crRNA OOEP 1	GACUUCGCACUCGGCCCUGGGUUUUAGAGCUAUGCU	Alt-R CRISPR-SpCas9 crRNA, 2 nmol
crRNA OOEP 4	GCUUCCAUGUAGAGCACCAGGUUUUAGAGCUAUGCU	Alt-R CRISPR-SpCas9 crRNA, 2 nmol
crRNA NLRP5 4	AUAGUGUCAUAUCAUCAUGUGUUUUAGAGCUAUGCU	Alt-R CRISPR-SpCas9 crRNA, 2 nmol
tracrRNA	IDT 67 nt tracr RNA (commercially confidential)	Alt-R CRISPR-SpCas9 tracrRNA, 5 nmol

A.9 Project License for animal work

The following pages contain the cover page from the Project License (PPL) that was granted by the Home Office during my PhD. This PPL will be used for the subsequent generation, breeding and analysis of split gene drive mice. The non-technical summary for the license is also presented. All sensitive information has been blacked out.



Home Office

PROJECT LICENCE

Breeding split gene-drive mice spreading female infertility

Project licence holder

[REDACTED]

This **PROJECT LICENCE** permits the licence holder to carry out a programme of scientific procedures on living animals under the **ANIMALS (SCIENTIFIC PROCEDURES) ACT 1986**.

The project licence holder may carry out the specified programme of work, subject to the restrictions and provisions contained within the Act and any limitations and conditions specified within this licence or by the Secretary of State.

This licence does not authorise the holder or any other person to carry out procedures on any animals unless they hold an appropriate personal licence issued under the Act.

Granted authority

This licence has been granted based on the information provided during the application process.

This licence authorises, only:

- work to meet the specified project aims
- use of specified animals and procedures
- work at the specified places

Retrospective assessment

The Secretary of State has determined that a retrospective assessment of this licence is not required.

Non-technical summary

Aim and duration

What's the aim of this project?

Keep this to a short one or two sentence summary.

This proof-of-concept project aims to test a safeguarded genetic system known as a split gene-drive to spread female infertility through a laboratory-contained mouse population. If successful, this approach may find future application as a pest management tool as after several generations the population would humanely decline due to the lack of fertile females available as breeding partners.

Why is it important to undertake this work?

In agriculture and wildlife, invasive pests are well known: rabbits in Australia, grey squirrel in the UK, and the omnipresent infestation of rodents around the globe. Invasive vertebrate pests impact the environment, economy and society. Current control methods include shooting, poisoning and trapping, which are costly and largely inadequate, and they often lead to unwanted suffering in target and non-target species. A gene drive spreading female infertility may offer a humane, species-specific and cost-effective alternative to current control methods.

What will be the duration of this project?

2 Years 0 Months

Benefits

What outputs do you think you will see at the end of this project?

Outputs can include new information, publications, or products.

The outputs of this project will help assess the feasibility of using genetic approaches to pest management. The findings will assist with technical development of gene drive technology as a potential humane alternative to current pest management tools. Dissemination of the project through mainstream media will encourage public and political discussion on genetic pest control approaches.

Who or what will benefit from these outputs, and how?

The impact of these outputs may be seen in the short-term, or they may not be fully realised until you've completed the project. Consider all timescales in your answer.

Invasive mammalian pests are a global concern that impact the environment, economy and society. Current control methods include shooting, poisoning and trapping, which are costly and largely inadequate, and they often lead to unwanted suffering in target and non-target species. Therefore, the outputs from this project have the potential to benefit environmental conservation efforts, support native predated species, the economy and those within our society that are impacted by mammalian pest, such as farmers and foresters. The outputs also have the potential to reduce suffering in pest species and non-pest species that are affected by currently deployed control methods. This is a proof-of-concept study and these beneficial outputs would likely not be seen for 10 years or more as this technology would require subsequent scientific investigation and a political and public consensus reached before application.

Will this work be offered as a service to others?

No

How will you look to maximise the outputs of this work?

For example, collaboration, dissemination of new knowledge, or publication of unsuccessful approaches.

This project contributes to collaborative research efforts to assess the feasibility of genetic pest control tools in vertebrate species. The outputs of this work will play a key role in this growing pool of knowledge and many within this collaborative research community are eagerly awaiting the results from this project. We will disseminate outputs through the research community via personal communication, invited talks, workshops, conference presentations and scientific publication. Outputs will be disseminated through mainstream media channels to encourage public and political discussion on the potential use of genetic pest control approaches.

Anticipated harms

Explain why you are using these types of animals and your choice of life stages.

The mouse is the preferred mammalian model for genetic research. Mice are cost effective because they are relatively cheap and easy to look care for. The mouse also benefits from a fast generation interval and is small in size, so convenient to house. It is relatively easy to manipulate the mouse genome compared to other mammalian species and this is why it is the preferred animal to study gene drive technology in this project.

As gene drives transmit through sexual reproduction, the mice used in this project will be sexually mature adults

Typically, what will be done to an animal used in your project?

For example, injections and surgical procedures. Include any relevant information about the duration of experiments and the number of procedures.

The gene drive mice used in this project will be mated to generate 16 litters of pups. Within those litters, the percentage of offspring that inherit the gene drive will be quantified. Once the female offspring reach sexual maturity, we will check the fertility status in a select number of female offspring that inherited the split gene-drive system.

What are the expected impacts and/or adverse effects for the animals during your project?

Examples can include pain, weight loss, tumours, or abnormal behaviour. State the estimated duration of these effects on an animal.

This study is expected to have no impact or adverse effects on animal health other than infertility of the female offspring that inherit the split gene-drive system.

What are the expected severities and the proportion of animals in each category (per animal type)?

Infertility of some female offspring is expected to be of mild severity. Loss of pregnancy will occur shortly after conception, at the 2-cell embryo stage. We anticipate no other impact from this study on animal health or wellbeing.

Fate of animals

Will any animals not be killed at the end of this project?

Ensure that any methods of killing to be used are described in the final step of each protocol.

No

Replacement

Why do you need to use animals to achieve the aim of your project?

Gene drives transmit through sexual reproduction, therefore, the only way to investigate the potential of the safeguarded split gene drives in this project is in sexually mature animals. This project will use the mouse to study the potential of gene drive in mammalian animals.

Which non-animal alternatives did you consider for use in this project?

No non-animal systems were considered, as gene drives transmit through sexual reproduction the assessment of these genetic systems is only possible in animal systems.

Why were they not suitable?

Gene drives can't be tested in non-animal alternatives as they transmit through sexual reproduction. At present, there are no non-animal systems for studying the complex biological processes involved in sexual reproduction for the transmission of gene drives.

Reduction

Enter the estimated number of animals of each type used in this project.

Mice: 128

How have you estimated the numbers of animals you will use?

For example, you may have made some power calculations or carried out some statistical modelling.

The only previously reported study on gene drive in mice used a significant different genetic strategy and, therefore, we can not predict the efficiency of the split gene-drive approach in this study based on their findings. The previous study reported up to 5 successful born litters for the gene drive systems tested.

This proof-of-concept study aims to generate 8 litters of pups for each of the two gene drive systems tested.

If we estimate an average litter size of 6 pups, we will generate a total of 96 pups whilst screening offspring. To generate these litters, the project will require approximately 8 split gene-drive breeding males (4 of each genetic system) and the 16 wild type female breeding partners. An additional 8 males will be required to confirm the fertile status of the split gene-drive female offspring once they reach sexual maturity. In total, we estimate this study will use 128 mice.

What steps did you take during the experimental design phase to reduce the number of animals being used in this project?

Include advice taken from local statisticians, online tools (such as the NC3Rs' Experimental Design Assistant), or regulatory requirements.

As this is a proof-of-concept study we have knowingly underpowered the project to reduce animal numbers. For a gene drive to be successful as a pest management tool it will be required to propagate at a rate much greater than 50% (at least 75% based on our computer simulations). We expect to see this effect in the selected sample size. If successful in this initial underpowered study, we will look to undertake additional study to statistically validate the most efficient of the best performing split gene-drive system being tested in this initial project.

What measures, apart from good experimental design, will you use to optimise the number of animals you plan to use in your project.

This may include efficient breeding, pilot studies, computer modelling, or sharing of tissue.

This is a proof-of-concept (pilot) study. We have chosen to perform an initial small-scale study before considering a larger-scale study that would offer statistical validation of the concept. In this way, we are reducing the number of animals being used.

Refinement

Which animal models and methods will you use during this project?

Explain why these models and methods cause the least pain, suffering, distress, or lasting harm to the animals.

This project will breed and maintain lines of genetically altered mice. Breeding of animals is a natural behaviour and this is anticipated to have no negative impact of animal health or wellbeing. Females may inherit a female infertility trait, which is expected to have minimal impact of animal wellbeing as loss of pregnancy will occur very shortly after conception.

Why can't you use animals that are less sentient?

For example, animals at a more immature life stage, species that are less sentient, or animals that have been terminally anaesthetised?

This project examines the potential of gene drive technology in mammalian pest species. This is the most accessible, genetically sophisticated and cost-effective mammalian model. As gene drives propagate through sexual reproduction it is necessary that sexually mature adult animals are used in this study.

How will you stay informed about advances in the 3Rs, and implement these advances effectively, during the project?

The Roslin Institute is located within the The Easter Bush Campus, which hosts the largest concentration of animal science and animal welfare related expertise anywhere in Europe. The Easter Bush Campus and researchers involved in this project are active participants in 3Rs research and stay informed on the latest advances through scientific publications, social media channels and discussion amongst colleagues and peers. The Roslin Institute is a state-of-the-art facility with the expertise to implement advances in the 3Rs as they evolve.

How will you refine the procedures you're using to minimise the welfare costs (harms) for the animals?

Potential refinements include increased monitoring, post-operative care, pain management, and training of animals.

We anticipate minimal animal suffering during this project. Animals are checked regularly checked (typically daily) and any animal identified to be suffering at a level greater than anticipated will be managed accordingly by experienced staff.

What published best practice guidance will you follow to ensure experiments are conducted in the most refined way?

We will design and carry out our experiments in accordance with the PREPARE and the ARRIVE guidelines. If required, we will consult the Experimental Design Assistant of the NC3Rs (<https://www.nc3rs.org.uk/experimental-design-assistant-eda>).

A.10 CRISPOR data outputs for *in silico* gRNA design

Table A.10.1 CRISPOR predicted gRNA characteristics for SpCas9 gRNAs at *OOEP*, *NLRP5* and *ROSA26* loci.

Name	Guide sequence (20 nt)	PAM	MIT Specificity Score	Predicted efficiency	Out-of-frame	Off-target for 0-1-2 mismatches
<i>OOEP</i> SpCas9 gRNA 1	GACTTCGCACTCGGCCCTGG	TGG	90	60	61	0-0-0
<i>OOEP</i> SpCas9 gRNA 2	GCCCTGGTGGTTCCCAATTC	AGG	67	31	73	0-0-2
<i>OOEP</i> SpCas9 gRNA 3	GGATTACTGACTTCCTGAAT	TGG	77	25	74	0-0-2
<i>OOEP</i> SpCas9 gRNA 4	GCTTCCATGTAGAGCACCAG	AGG	77	66	69	0-0-1
<i>NLRP5</i> SpCas9 gRNA 1	TTTGGCACAGCTGATTGCTA	AGG	76	59	71	0-0-2
<i>NLRP5</i> SpCas9 gRNA 2	GATCTTTGTCAGTAGATCCC	AGG	77	58	66	0-0-3
<i>NLRP5</i> SpCas9 gRNA 3	AAACAAGAGTCTTTCTGGTT	GGG	55	80	74	0-1-2
<i>NLRP5</i> SpCas9 gRNA 4	ATAGTGTCATATCATCATGT	TGG	71	60	53	0-0-2
<i>ROSA26</i> SpCas9 gRNA 1	GAAGGGATTCTCCCAGGCC	AGG	58	38	64	0-1-0
<i>ROSA26</i> SpCas9 gRNA 2	GGCAGGCTTAAAGGCTAACC	TGG	81	47	67	0-0-2

Table A10.2 CRISPOR predicted gRNA characteristics for AsCas12a gRNAs at *OOEP*, *NLRP5* and *ROSA26* loci.

Name	Guide sequence (21 nt)	PAM	Predicted efficiency	Out-of-frame	Off-target for 0-1-2 mismatches
<i>OOEP</i> AsCas12a gRNA 1	TGCCACCCAGGCTTCCATGTA	TTTC	20	55	0-0-0
<i>OOEP</i> AsCas12a gRNA 2	TGAGATTTCCGCTTGGTCTGT	TTTC	5	59	0-0-0
<i>OOEP</i> AsCas12a gRNA 3	AGCTAGGTTCCCAGAGTTGAC	TTTC	18	69	0-0-0
<i>OOEP</i> AsCas12a gRNA 4	CTTTGTGCCAGGCCGCCATGT	TTTT	5	61	0-0-0
<i>NLRP5</i> AsCas12a gRNA 1	TGGTTGGGACATGATCTTTGT	TTTC	7	72	0-1-0
<i>NLRP5</i> AsCas12a gRNA 2	TCATAGATGGCTTGGATGATA	TTTG	10	42	0-0-1
<i>NLRP5</i> AsCas12a gRNA 3	TGGTGGTAATGATGAGAAAGG	TTTC	7	72	0-1-0
<i>NLRP5</i> AsCas12a gRNA 4	TCATCATTACCACCAGAAACA	TTTC	24	75	0-1-0
<i>ROSA26</i> AsCas12a gRNA 1	TAGAAGATGGGCGGGAGTCTTCT	TTTC	8	60	0-0-0
<i>ROSA26</i> AsCas12a gRNA 2	AGCCTGCCCAGAAGACTCCCGCC	TTTA	20	68	0-0-0

A.11 Preliminary PCR screens of engineered ES cells

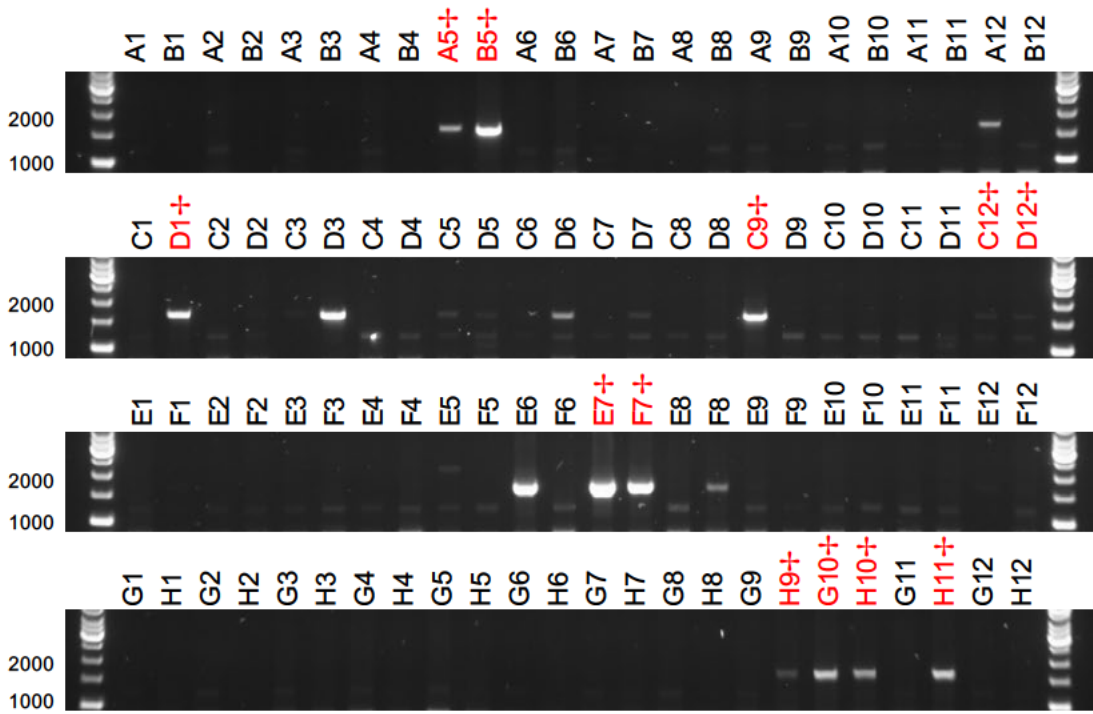


Figure A.11.1 Gel electrophoresis of PCRs from DNA extracted from ROSA26_gRNAs targeted cells in clonal 96 well plate format. PCRs used oGM236 and oGM168 primers flanking the 3' homology arm. Expected product size is 1650 bp. Red labels with + denote those clones that were selected for further screening by ddPCR.

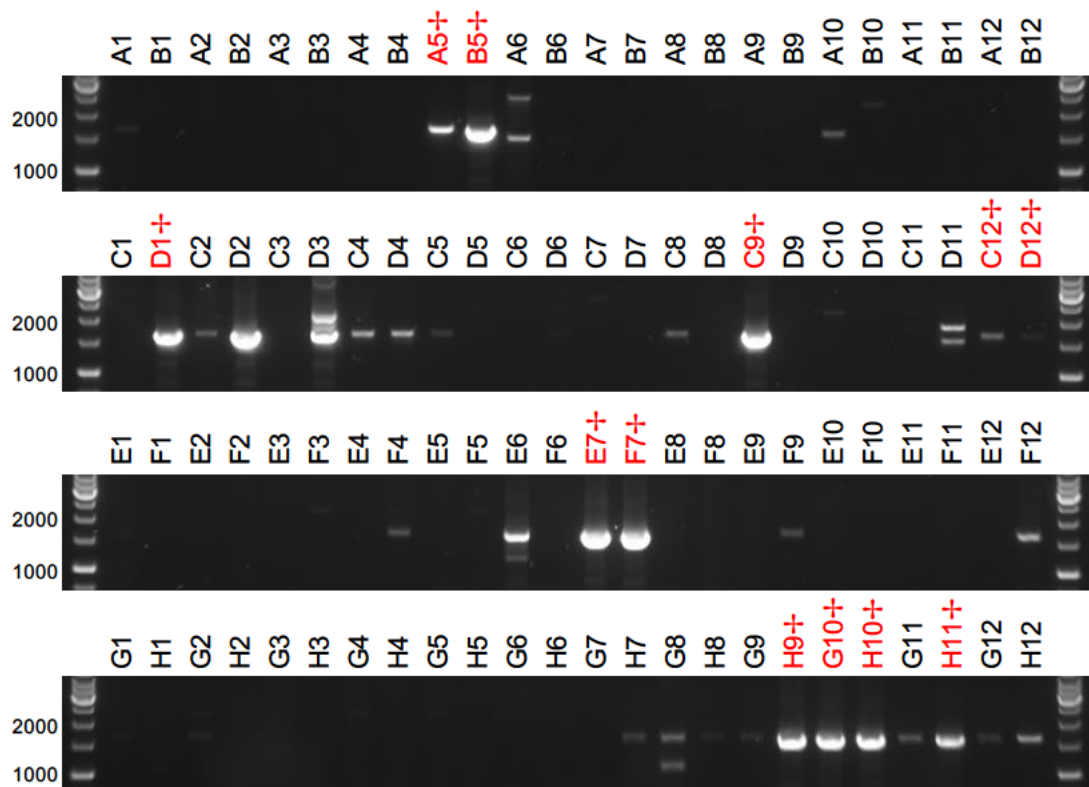


Figure A.11.2 Gel electrophoresis of PCRs from DNA extracted from ROSA26_gRNAs targeted cells in clonal 96 well plate format. PCRs used oGM163 and oGM164 primers flanking the 5' homology arm. Expected product size is 1597 bp. Red labels with + denote those clones that were selected for further screening.

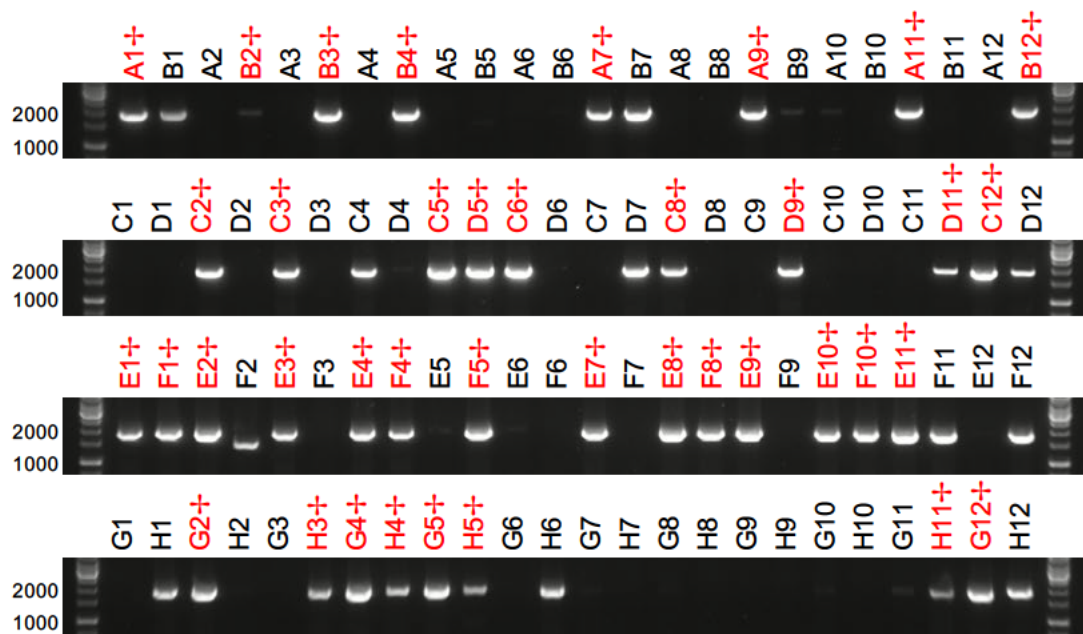


Figure A.11.3 Gel electrophoresis of PCRs from DNA extracted from OEP_VASA targeted cells in clonal 96 well plate format. PCRs used oGM189 and oGM190 primers flanking the 3' homology arm. Expected product size is 1888 bp. Red labels with + denote those clones that were selected for further screening.

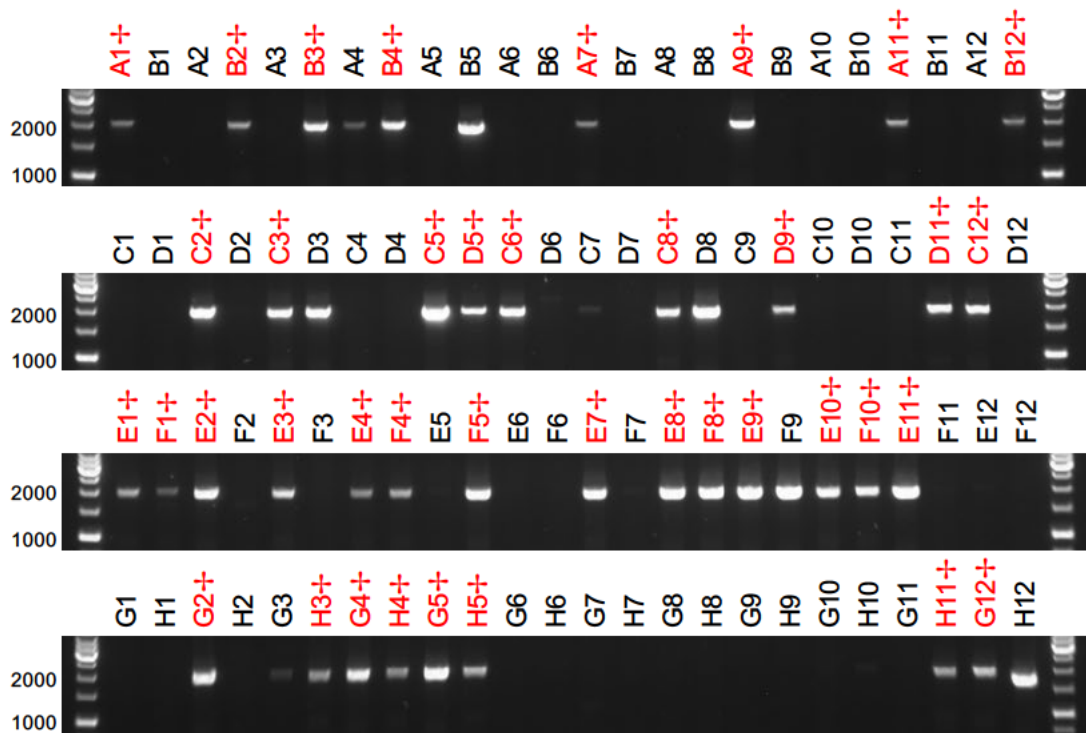


Figure A.11.4 Gel electrophoresis of PCRs from DNA extracted from OOEP_VASA targeted cells in clonal 96 well plate format. PCRs used oGM185 and oGM186 primers flanking the 5' homology arm. Expected product size is 1981 bp. Red labels with + denote those clones that were selected for further screening.

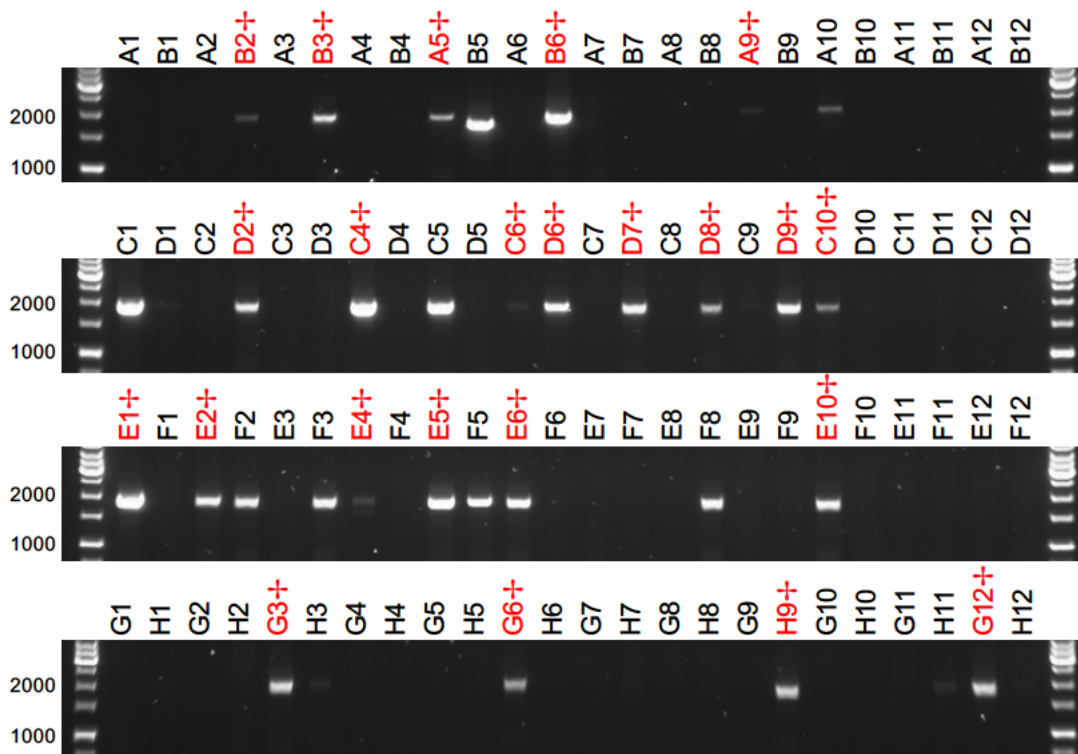


Figure A.11.5 Gel electrophoresis of PCRs from DNA extracted from OOEP_CAG targeted cells in clonal 96 well plate format. PCRs used oGM189 and oGM190 primers flanking the 3' homology arm. Expected product size is 1888 bp. Red labels with + denote those clones that were selected for further screening.

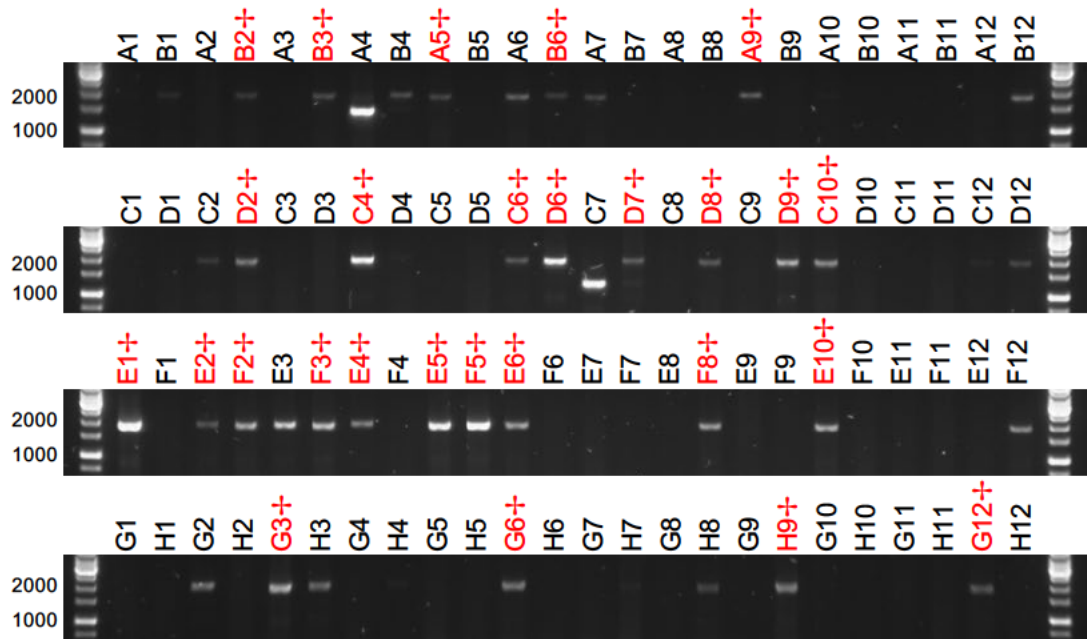


Figure A.11.6 Gel electrophoresis of PCRs from DNA extracted from OOEP_CAG targeted cells in clonal 96 well plate format. PCRs used oGM185 and oGM186 primers flanking the 5' homology arm. Expected product size is 1981 bp. Red labels with + denote those clones that were selected for further screening.

A.12 ddPCR quality control data

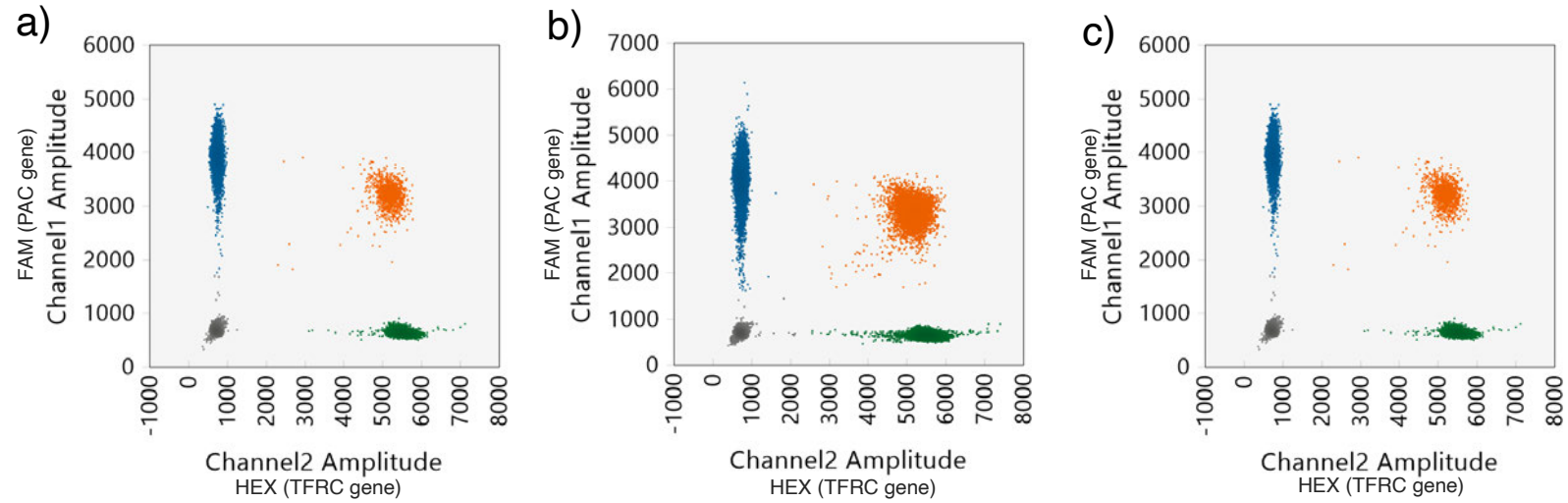


Figure A.12.1 Quality control for ddPCR calls in genome edited cell lines. Each plot represents an overlay of all samples processed by ddPCR for a) OOEP_CAG, b) OOEP_VASA and c) ROSA26_gRNAs cells.

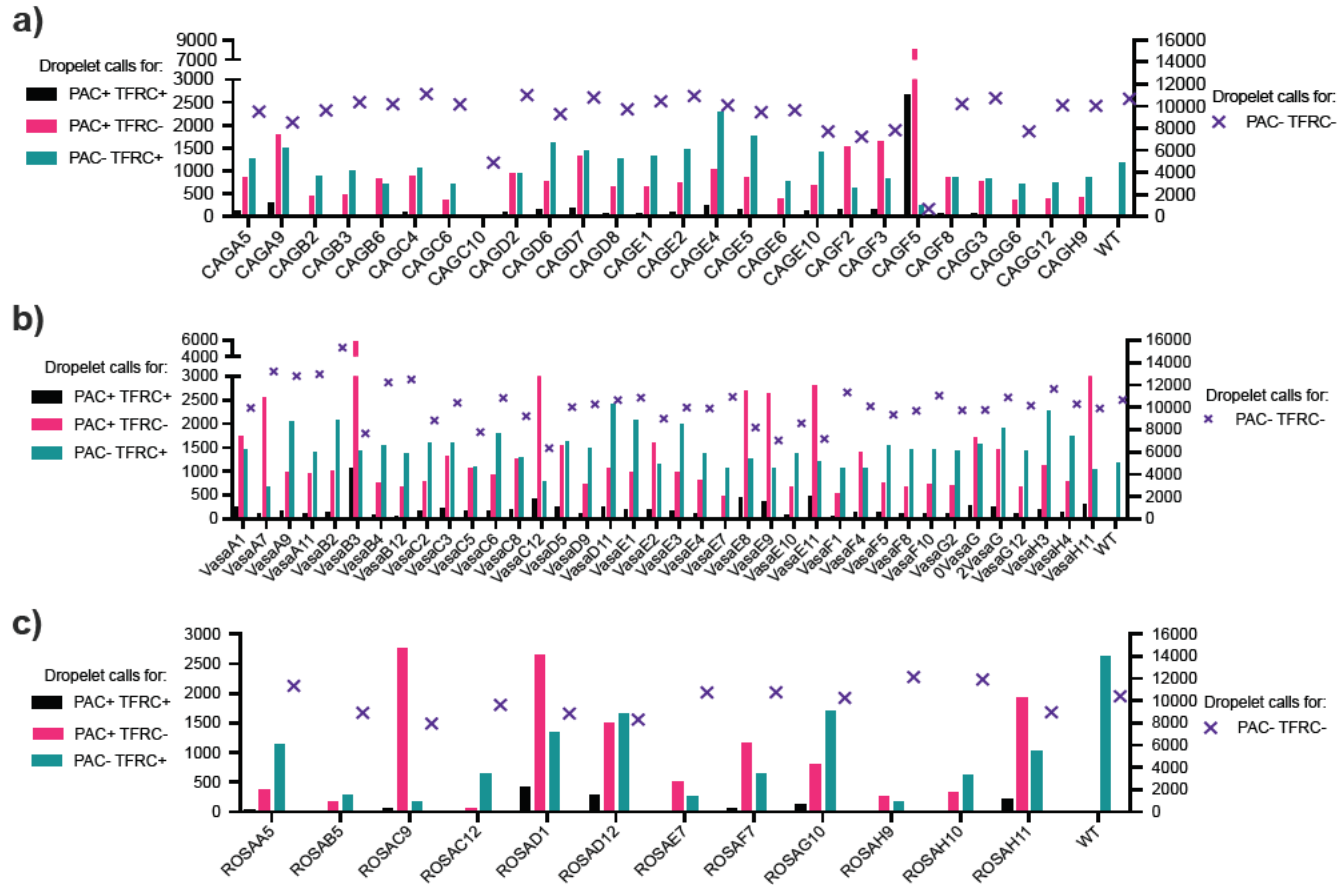


Figure A.12.2 Quality control graphs for droplet digital PCR experiments. X-axis labels clone of interest. Lefthand y-axis is total droplet calls for *PAC* positive (+) and *TFRC* +, *PAC* + and *TFRC* negative (-), and *PAC* - and *TFRC* + droplets. While the righthand y-axis is total droplet calls from droplets that are negative for both *PAC* and *TFRC* fluorphores. a) is OOEP_CAG targeted cells; b) OOEP_VASA targeted cells; and c) is ROSA26_gRNAs targeted cells.

A.13 Publication: On-farm genome editing of livestock

The following pages contain a paper I co-authored during my PhD project and has been referred to within this document (including section 5.4.2).



On-Farm Livestock Genome Editing Using Cutting Edge Reproductive Technologies

Gus R. McFarlane*, Hamish A. Salvesen, Anna Sternberg and Simon G. Lillico

Whitelaw Group, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:

Yiping Qi,
University of Maryland, College Park,
United States

Reviewed by:

Mark Paul Running,
University of Louisville, United States
Agnieszka Barbara Najda,
University of Life Sciences of
Lublin, Poland

*Correspondence:

Gus R. McFarlane
gus.mcfarlane@ed.ac.uk

Specialty section:

This article was submitted to
Crop Biology and Sustainability,
a section of the journal
Frontiers in Sustainable Food Systems

Received: 12 August 2019

Accepted: 30 October 2019

Published: 15 November 2019

Citation:

McFarlane GR, Salvesen HA,
Sternberg A and Lillico SG (2019)
On-Farm Livestock Genome Editing
Using Cutting Edge Reproductive
Technologies.
Front. Sustain. Food Syst. 3:106.
doi: 10.3389/fsufs.2019.00106

The global demand for animal-based food products is anticipated to increase by 70% by 2050. Meeting this demand in a way that has minimal impact on the environment will require the implementation of advanced technologies. Genome editing of livestock is a tool that will allow breeders to improve animal welfare, performance and efficiency, paving the way to a more sustainable future for livestock agriculture. Currently, genome editing of livestock is limited to specialized laboratories due to the complexity of techniques available for the delivery of genome editing reagents into zygotes and reproductive cells. The emergence of three cutting-edge reproductive technologies—(i) zygote electroporation, (ii) zygote transduction of recombinant adeno-associated virus (rAAV), and (iii) surrogate sire technology—will provide livestock breeders with a new toolkit of delivery strategies for genome editing. The simplicity of these technologies will enable widespread on-farm application in major livestock species by seamlessly integrating into current breeding systems. We believe it is timely to highlight these three cutting-edge reproductive technologies for genome editing and have outlined pipelines for their implementation in on-farm settings. With a nuanced regulatory framework these technologies could fast-track livestock genetic gain and help secure a sustainable future for livestock.

Keywords: genome editing, CRISPR, genetic engineering, pigs, cattle, livestock, sustainability

INTRODUCTION

Preparing to feed a balanced and nutritious diet to the projected 9.7 billion people on the globe by 2050 will be one of the greatest challenges humanity has ever faced. The FAO estimates demand for animal-based food products will increase by 70% in this time (Alexandratos and Bruinsma, 2012). Increasing reliance on plant-based diets and artificial meat production will contribute to improving food security and the sustainability of commercial agriculture, however outright omission of animal protein from human diets risks nutritional deficiencies and malnutrition, particularly in developing regions. Meeting the anticipated increase in demand for animal food products in a way that has minimal impact on the environment and ensures high animal welfare standards will likely require the implementation of advanced technologies, including genome editing and cutting-edge reproductive technologies. Considering the huge potential of these technologies, it would be negligent not to examine their inherent possibilities further.

Traditional livestock breeding is restricted by genetic linkage and the available genetic variation within a breed. Genome editing allows animal breeders to overcome these biological impediments

and introduce polymorphisms that are not present in the gene pool of elite brood stock, or even create novel changes predicted to result in improved gain (Laible et al., 2014). This powerful technology allows animal breeders to specifically and efficiently alter an animal's DNA. Precise genetic alterations to remove deleterious mutations, such as recessive lethal genetic variants, or introduce desirable traits, such as hornlessness, heat tolerance and disease resistance; without affecting other genetic characteristics of their herd (Davis et al., 2017; Mueller et al., 2019; Proudfoot et al., 2019).

The genome editor toolbox currently contains variants of ZFNs, TALENs and CRISPR-Cas, and these have been successfully used in livestock species such as pigs, cattle, sheep, goats, and chickens (Tait-Burkard et al., 2019). These genome editors all work on the principle of introducing double strand DNA breaks (DSBs) at a user-defined target site in the genome, stimulating endogenous cellular DNA repair pathways to make modifications. DSBs will most frequently be repaired by non-homologous end joining (NHEJ), which is error prone and can result in the introduction of nucleotide insertions or deletions. However, if a homologous DNA repair template (HDRT) is provided, then the repair can then occur via the homology-directed repair (HDR) pathway. In this manner, DNA sequences can be precisely modified or introduced into the genome (Fernandez et al., 2017).

Of the genome editing tools currently available, CRISPR-Cas has quickly become the “go to” technology due to its ease-of-use, high efficiency and low cost. Since it's repurposing into a genome editor in 2012, there has been a flurry of research activity around the technology and an ever expanding toolbox of CRISPR-Cas reagents, including base-editors and nickase systems, which can drive editing without the need for introducing DSBs (Pickar-Oliver and Gersbach, 2019). Due to the simplicity and range of CRISPR-Cas tools now available, it is foreseeable that a significant number of genome edited livestock will be produced over the next decade. Within this article, we describe livestock genome editing strategies based on CRISPR-Cas, although ZFNs and TALENs could also achieve the same results.

A precondition for applying CRISPR-Cas in livestock are reproductive technologies that enable efficient delivery of CRISPR-Cas reagents into zygotes or reproductive cells. Although genome editors have rapidly developed, the reproductive technologies for delivering CRISPR-Cas remain complex and inefficient at a large scale. Somatic cell nuclear transfer (SCNT) and zygote microinjection are the conventional techniques. Both are technically challenging, costly, labor-intensive, and require expert skills with bulky micromanipulation equipment; restricting their use to a small number of specialized laboratories (Sheets et al., 2016).

New reproductive technologies that simplify the delivery of CRISPR-Cas reagents into livestock reproductive cells are needed to disseminate the benefits of genome editing beyond research institutes and corporate biotechnology enterprises. The emergence of three cutting-edge reproductive technologies—(i) zygote electroporation, (ii) zygote transduction with recombinant adeno-associated virus (rAAV), and (iii) surrogate sire technology (SST)—will

provide animal breeders with a new toolkit for delivering CRISPR-Cas to reproductive cells. The simplicity of these approaches will allow livestock genome editing to occur on-farm.

We believe it is timely to highlight these three cutting-edge reproductive technologies and have outlined pipelines for their implementation in on-farm settings. It is hoped these cutting-edge reproductive technologies will disperse the capacity to genome edit farm animals, fast-tracking genetic gains and helping to secure a sustainable future for livestock agriculture.

ZYGOTE ELECTROPORATION

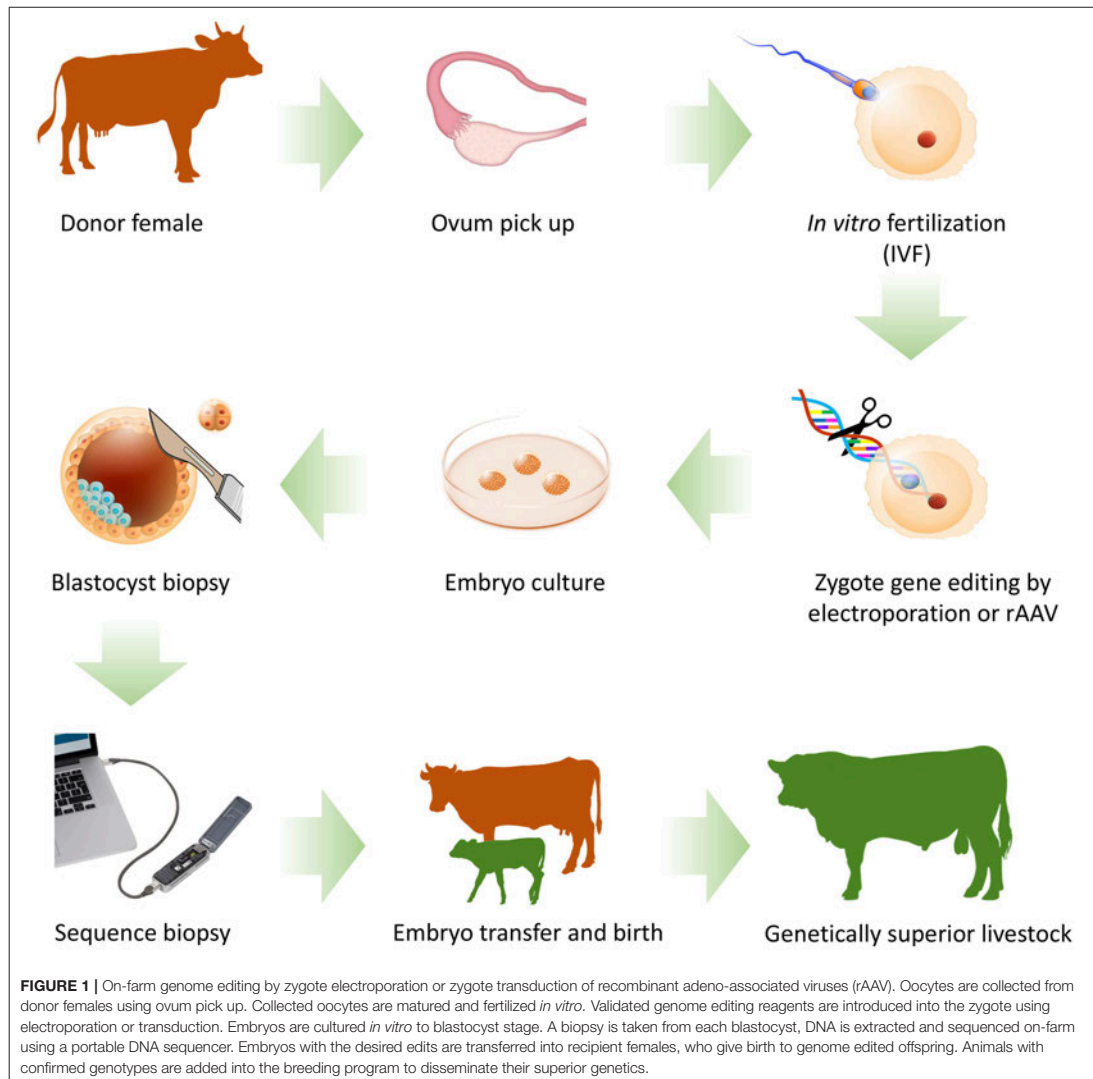
Zygote electroporation is a recently developed method that overcomes many of the shortcomings of conventional SCNT and microinjection delivery approaches. This technique allows direct introduction of genome editing tools into zygotes by application of voltage to zygotes suspended in a medium containing CRISPR-Cas reagents. The pulses of electricity cause pores to form in the zygote membrane, allowing the genome editing tools in the suspension to pass through the pores in the zygote membrane and into the nucleus where CRISPR-Cas can begin genome editing activity (Miao et al., 2019).

Electroporation is a well-established method for introducing reagents into mammalian cells but has only recently been refined for application to zygotes. Initial efforts required the enzymatic removal of the zona pellucida, a protective membrane surrounding the zygote. However, the removal of the zona pellucida makes zygotes sticky and difficult to work with, which restricted adoption of zygote electroporation. Optimization of the technique now makes it possible to introduce CRISPR-Cas into zona-intact zygotes, making it significantly more attractive to users.

Several livestock genome editing facilities have now employed zona-intact zygote electroporation to successfully edit the genomes of pigs and cattle (Laible, 2018; Miao et al., 2019; Namula et al., 2019). The simplicity of this approach allows for zygote electroporation to be incorporated into on-farm embryo transfer (ET) programs, which are increasingly common in commercial livestock farming.

In on-farm settings, we foresee a pipeline where donor females are super-ovulated and oocytes collected for *in vitro* fertilization - as per a conventional ET program. After fertilization, the zygotes undergo electroporation to introduce CRISPR-Cas reagents. The genome edited embryos are then matured *in vitro* and the editing of each zygote is confirmed by portable biopsy sequencing. Validated embryos are then transferred into recipient females to give birth to genetically superior animals (Figure 1).

Equipment, expertise, throughput and efficiency inhibits SCNT or zygote microinjection from being routinely applied on-farm. One of the key benefits of electroporation is that CRISPR-Cas reagents and HDRTs can be designed on a basic computer, ordered online and supplied ready-to-use in a short time frame. Compared to conventional microinjection-based approaches, electroporation is compatible



with high throughput, potentially editing hundreds of zygotes simultaneously with reduced physical damage and improved embryo survival rates. The portable equipment and minimum training requirements would allow zygote electroporation to be integrated into established ET programs with little disruption.

Although electroporation has many benefits over SCNT and zygote microinjection, there are still several limitations (Laible, 2018). Primarily, only short HDRTs have been successfully used in zygote electroporation. This limits the genetic alterations

to <1 kb in length. Secondly, like zygote microinjection, electroporation can be associated with mosaicism in the genome edited offspring. This occurs when the genome editing does not occur prior to the first zygotic cleavage but when the embryo has progressed to the 2 cell stage or beyond. Optimization of the timing of genome editor delivery following fertilization will likely result in a reduction in the frequency of mosaic offspring. Although mosaicism is undesirable, it is not a major concern as it can be bred out in a single generation.

ZYGOTE TRANSDUCTION WITH RECOMBINANT ADENO-ASSOCIATED VIRUS (rAAV)

Due to its non-pathogenic nature, rAAVs can be used as vehicles to deliver CRISPR-Cas and HDRTs. This strategy has proven effective in editing the genome of mice zygotes but like electroporation it has been used on other mammalian cell types for many years (Kaulich and Dowdy, 2015). To generate genome edited livestock using rAAV, oocytes are collected from donor mothers and fertilized. The fertilized zygotes are then bathed in a solution containing rAAVs that enter the zygote and drive genome editing by expressing CRISPR-Cas and providing HDRT. Once validated by biopsy sequencing, edited embryos are then transferred to recipient females to develop into genetically superior livestock (Figure 1).

Many of the benefits of rAAV transduction are similar to zygote electroporation when compared to SCNT and microinjection. Transduction with rAAV does not require physical damage to the zygotes which significantly improves embryo survival rates. The technique increases throughput and does not require any additional equipment or skills within on-farm ET programs. Furthermore, as rAAV is non-pathogenic, it can be handled safely at biosafety level 1, making it suitable for most on-farm settings.

Although this approach is yet to be applied to livestock, in mice, rAAV has been applied to zona-intact zygotes to successfully generate genome edited pups. It has very high embryo survival rates with editing in up to 100% of offspring (Yoon et al., 2018). One rAAV vector can comfortably accommodate a 3.25 kb HDRT, and even larger DNA sequences could be integrated if multiple rAAV vectors are designed to sequentially integrate HDRTs (Bak and Porteus, 2017). No other technique exhibits such simplicity for precise integration of large DNA sequences, and this is likely where rAAV transduction will be of most value in livestock genome editing.

The single stranded DNA genome of rAAV integrates into non-homologous host genomic sites at very low frequencies (~0.1%). rAAV is also quickly diluted as the cells undergoes multiple rounds of division, making it ideally suited for manipulating the genome of zygotes (Yoon et al., 2018). However, like most genome editing delivery approaches, mosaicism has been observed in rAAV genome edited mice. We expect optimization of embryo collection and the timing of rAAV transduction will reduce the frequency of mosaicism in rAAV edited offspring.

Customized rAAV vectors can be ordered from commercial suppliers as ready-to-use reagents (Sandoval et al., 2019). The design does require knowledge of viral genetics and is time-consuming, however online tools are available to assist and standardized rAAV kits could be developed for specific breeds and traits. Genome editing zygotes by rAAV transduction would not require the purchase of any additional equipment over a standard ET program and adding rAAV to zygotes in culture is a relatively straightforward procedure. The low cost and skill-level required by operators could see this technology widely adopted in on-farm settings in the near future.

SURROGATE SIRE TECHNOLOGY

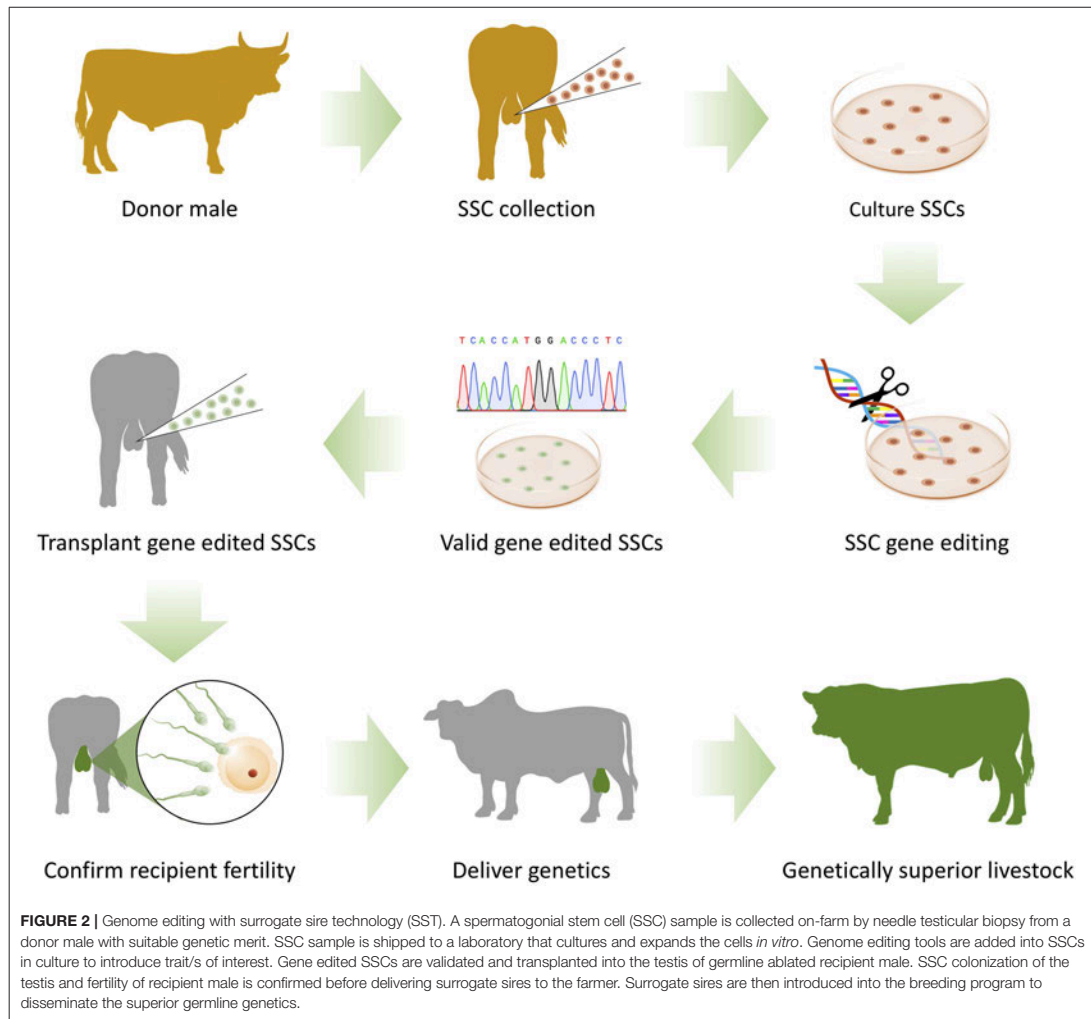
Surrogate sire technology (SST) describes the creation of male animals lacking endemic germline stem cells, and therefore their ability to produce mature sperm. Spermatogonial stem cells (SSCs) from a donor male can then be transplanted into recipient testes of a surrogate sire, providing a source of self-propagating stem cells which can produce mature sperm containing the genetic information of the donor (Giasetti et al., 2019). The recipient male can then disseminate the genetics of the donor by natural breeding, operating as an ambulatory artificial insemination system.

Commercial application of SST relies on a supply of males lacking their own germline stem cells. To achieve this, researchers have generated male mice and pigs lacking a functional copy of NANOS2, a highly-conserved mammalian gene that plays an essential role in the maintenance of SSCs (Park et al., 2017). The resulting animals are physiologically healthy, displaying no abnormal phenotypes other than a complete lack of native SSCs. The transplantation of donor SSCs into the otherwise physiologically normal testicular environment of these animals leads to the establishment of the donor cell SSC population and eventually to the production of donor-derived sperm. Researchers are now working to establish NANOS2 knockout cattle, sheep and goats.

To integrate genome editing into established SST procedures, CRISPR-Cas and HDRTs can be delivered whilst donor SSCs are expanded during *ex vivo* cell culturing. CRISPR-Cas has been used to successfully edit the genome of mouse SSCs during *ex vivo* cell culture. The edited mouse SSCs were then transplanted into the testes of recipient males and lead to the generation of offspring harboring the edited donor genetics (Wang et al., 2017). These findings suggest the same will be achievable for mammalian livestock species, enabling the dissemination of genome-edited livestock genetics through surrogate sires.

Genome editing with SST would be particularly valuable for disseminating beneficial traits through expansive ranch-style operations or in developing regions where other genome editing approaches are limited by equipment and expertise availability or the scale of the operation. For on-farm settings, a vet could collect a needle testicular biopsy from a donor animal and ship the sample to a laboratory. SSCs would be expanded from the biopsy, with the desired genome edits introduced during the cell culture process. The cells would then be transplanted into juvenile sire recipients, and delivered to the farmer for integration into their breeding operation (Figure 2). With this genome editing pipeline, SST keeps on-farm equipment and expertise requirements to a minimum and empowers the farmer to maintain their own herd blood lines whilst benefiting from the technology.

In regions such as Africa and South America, robust bulls of indigenous breeds could be used as surrogate sires to carry elite genome edited semen. Within the pig industry, SST boars could rapidly disseminate advantageous genome edited traits. This technology allows livestock breeders to achieve breeding objectives in less time. Using genomic prediction tools, elite SSCs from sexually immature males could be obtained, genome edited



and transplanted into maturing germline ablated recipients within months. Instead of waiting until breeding maturity for each animal, generation intervals could be skipped, accelerating the rate of genetic gain.

RISKS ASSOCIATED WITH UNINTENDED OR OFF-TARGET EDITING

Until recently, genome editing approaches in livestock were commonly performed using an editor to introduce a DSB at the target site, with a double-stranded DNA (dsDNA) molecule, typically a plasmid, supplied as a repair template. Although this

approach has seen success, it is not uncommon for the plasmid repair template to insert elsewhere in the genome. This issue was recently highlighted when the US Food and Drug Administration (FDA) found additional plasmid sequence proximal to the target site in genome edited hornless bulls (Norris et al., 2019). There are straightforward methods available to screen for unintended repair template integration and once identified the unwanted integration can be bred out using standard breeding strategies. This was the approach taken with the genome edited hornless bulls (Young et al., 2019). Most livestock genome editing efforts have now transitioned to using single stranded DNA (ssDNA) repair templates, which have a significantly reduced frequency of unintended genomic integration. Despite the reduced risks with

ssDNA repair templates, comprehensive screening of founder livestock is important to maintain public and political trust in the technology.

Off-target editing occurs when a genome editor cuts at an unplanned site in the genome. It was a significant concern with early genome editing experiments as the impact of off-target effects remained contentious. However, the latest genome editing technologies have improved and quantified specificity, reducing off-target effects, and concerns. The goal of improving CRISPR-Cas reagents and increasing genome editing specificity has been primarily driven through biomedical research looking to apply genome editing to treat or cure human disease. The work done in the biomedical arena has significantly expanded the toolkit of CRISPR-Cas reagents available, which now includes base editors that can edit single nucleotides in the genome without the need to induce DSBs (Eid et al., 2018). Although current CRISPR-Cas reagents are adequate for on-farm application, further improvements in the specificity and efficiency of CRISPR-Cas will deliver reagents that carry a risk of off-target effect substantially lower than the frequency of spontaneous mutations naturally occurring in animal genomes.

REGULATIONS

The science is clear: genome editing could improve animal welfare and performance while reducing the environmental footprint of livestock production. Furthermore, dietary DNA is generally regarded as safe to consume, as naturally occurring DNA variations are a routine ingredient in the food products we consume. What remains unclear is the regulatory pathway to bring genome edited animal-derived foods to market (Zhou et al., 2019). Currently, regulations vary substantially between geo-political regions. Harmonizing the regulations associated with genome editing in food species is imperative to allow livestock farmers access to genome editing tools that could increase global food security in a sustainable manner.

Genome editing provides an opportunity to align the interests of producers and consumers. Despite the foreseeable benefits, the EU applies an inhibitory regulatory framework on genome edited foods and the US currently mandates premarket new animal drug regulatory evaluation for all genome edited food animals (Van Eenennaam et al., 2019). Although the EU and the US have an oversupply of food, the prohibitive regulatory frameworks they currently implement may have detrimental knock-on effects in developing countries who stand to benefit the most from genome editing technology.

As the tide of public acknowledgment to the benefits and safety of genome editing appears to be turning, the technology will continue to gain media attention and public debate. In the US and EU there are movements from within government and the scientific community to modernize regulatory frameworks. In the US, President Trump recently signed an “Executive Order on Modernizing the Regulatory Framework for Agricultural

Biotechnology Products,” while in the EU a report by the European Academies’ Science Advisory Council on “Genome Editing: Scientific opportunities, public interests, and policy options in the EU” was delivered to the European Parliament and European Commission to prompt a rethink of EU’s stance on genome edited foods.

Outside of the EU and US, countries such as Japan, Argentina, Brazil, Canada, and Australia have to varying extents deregulated genome editing of livestock. These proactive countries are likely to gain a competitive advantage and leave other non-subscribing nations scrambling to prevent genome edited foods entering their supply chain. Unlike previous transgenic technologies, the genetic alteration in genome edited foods are often “scarless,” in that they contain no foreign DNA. Without any plausible methodology for discriminating “naturally occurring” from intentionally edited DNA variations, regulators will have difficulty enforcing importation restrictions on genome edited foods.

CONCLUDING REMARKS

The amalgamation of genome editing and cutting-edge reproductive technologies offers a powerful tool for improving the livestock breeding landscape. Success in creating precise and heritable germline edits in diverse livestock species for a plethora of traits has demonstrated the potential benefits of this technology. However, to spread the beneficial impacts across economies, geographic regions and societies, strategies of translating established genome editing protocols into livestock breeding systems are necessary. Zygote electroporation and rAAV transduction of genome-editing reagents evades the associated, costs, labor, and facilities required by traditional methods. SST converged with genome editing could offer a commercially valuable tool to farmers with natural breeding programs. This array of cutting-edge reproductive technologies makes it technically plausible to apply genome editing in on-farm settings to rapidly improve productivity, fertility, sustainability, and animal welfare with minimal infrastructure and moderate fiscal inputs. The key to unlocking these benefits now lays in the hands of regulators.

AUTHOR CONTRIBUTIONS

GM and HS envisaged the article and contributed to writing. AS contributed to writing the manuscript. SL assisted with editing and conceptual development. GM generated the figures.

FUNDING

GM was supported by the UK Commonwealth Scholarship Commission. HS was supported by Edinburgh Global Scholarship from Edinburgh University. AS was supported by Erasmus+. This work was part funded by the BBSRC through ISP BB/P013759/1.

REFERENCES

- Alexandratos, N., and Bruinsma, J. (2012). *World Agriculture Towards 2030/2050: The 2012 Revision*. ESA Working paper No. 12-03. Rome: FAO.
- Bak, R. O., and Porteus, M. H. (2017). CRISPR-mediated integration of large gene cassettes using AAV donor vectors. *Cell Rep.* 20, 750–756. doi: 10.1016/j.celrep.2017.06.064
- Davis, S. R., Spelman, R. J., and Littlejohn, M. D. (2017). BREEDING AND GENETICS SYMPOSIUM: Breeding heat tolerant dairy cattle: the case for introgression of the “slick” prolactin receptor variant into *Bos taurus* dairy breeds. *J. Anim. Sci.* 4, 1788–1800. doi: 10.2527/jas.2016.0956
- Eid, A., Alsharaf, S., and Mahfouz, M. M. (2018). CRISPR base editors: genome editing without double-stranded breaks. *Biochem. J.* 475, 1955–1964. doi: 10.1042/BCJ20170793
- Fernandez, A., Josa, S., and Montoliu, L. (2017). A history of genome editing in mammals. *Mamm. Genome.* 28:6299. doi: 10.1007/s00335-017-9699-2
- Giasetti, M. I., Ciccarelli, M., and Oatley, J. M. (2019). Spermatogonial stem cell transplantation: insights and outlook for domestic animals. *Annu. Rev. Anim. Biosci.* 7, 385–401. doi: 10.1146/annurev-animal-020518-115239
- Kaulich, M., and Dowdy, S. F. (2015). Combining CRISPR/Cas9 and rAAV templates for efficient gene editing. *Nucleic. Acid. Ther.* 25, 287–296. doi: 10.1089/nat.2015.0545
- Laible, G. (2018). “Production of transgenic livestock: overview of transgenic technologies,” in *Animal Biotechnology 2*, eds H. Niemann and C. Wrenzycki (Berlin: Springer), 95–121. doi: 10.1007/978-3-319-92348-2_6
- Laible, G., Wei, J., and Wagner, S. (2014). Improving livestock for agriculture – technological progress from random transgenesis to precision genome editing heralds a new era. *Biotech. J.* 10, 109–120. doi: 10.1002/biot.201400193
- Miao, D., Giasetti, M. I., Ciccarelli, M., Lopez-Biladeau, B., and Oatley, J. M. (2019). Simplified pipelines for genetic engineering of mammalian embryos by CRISPR-Cas9 electroporation. *Biol. Reprod.* 101, 177–187. doi: 10.1093/biolre/iox075
- Mueller, M. L., Cole, J. B., Sonstegard, T. S., and Van Eenennaam, A. L. (2019). Comparison of gene editing versus conventional breeding to introgress the POLLED allele into the US dairy cattle population. *J. Dairy Sci.* 5, 4215–4226. doi: 10.3168/jds.2018-15892
- Namula, Z., Wittayarat, M., Hirata, M., Hirano, T., Nguyen, N. T., Le, Q. A., et al. (2019). Genome mutation after the introduction of the gene editing by electroporation of Cas9 protein (GEEP) system into bovine putative zygotes. *In Vitro Cell. Dev. Biol. Anim.* 55, 237–242. doi: 10.1007/s11626-019-00385-w
- Norris, A. L., Lee, S. S., Greenlees, K. J., Tadesse, D. A., Miller, M. F., and Lombardi, H. (2019). Template plasmid integration in germline genome-edited cattle. *bioRxiv [Preprint]*. doi: 10.1101/715482
- Park, K.-E., Kaucher, A. V., Powell, A., Waqas, M. S., Sandmaier, S. E. S., Oatley, M. J., et al. (2017). Generation of germline ablated male pigs by CRISPR/Cas9 editing of the NANOS2 gene. *Sci. Rep.* 7:40176 doi: 10.1038/srep40176
- Pickar-Oliver, A., and Gersbach, C. A. (2019). The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* 20, 490–507. doi: 10.1038/s41580-019-0131-5
- Proudfoot, C., Lillico, S., and Tait-Burkard, C. (2019). Genome editing for disease resistance in pigs and chickens. *Anim. Front.* 9, 6–12. doi: 10.1093/af/vfz013
- Sandoval, I. M., Collier, T. J., and Manfredsson, F. P. (2019). “Design and assembly of CRISPR/Cas9 lentiviral and rAAV vectors for targeted genome editing,” in *Viral Vectors for Gene Therapy*, eds F. Manfredsson and M. Benseky (New York, NY: Humana Press), 29–45. doi: 10.1007/978-1-4939-9065-8_2
- Sheets, T. P., Park, C.-H., Park, K.-E., Powell, A., Donovan, D. M., and Telugu, B. P. (2016). Somatic cell nuclear transfer followed by CRISPR/Cas9 Microinjection results in highly efficient genome editing in cloned pigs. *Int. J. Mol. Sci.* 17:2031. doi: 10.3390/ijms17122031
- Tait-Burkard, C., Doeschl-Wilson, A., McGrew, M. J., Archibald, A. L., Sang, H. M., Houston, R. D., et al. (2019). Livestock 2.0 – genome editing for fitter, healthier, and more productive farmed animals. *Genome Biol.* 19:204. doi: 10.1186/s13059-018-1583-1
- Van Eenennaam, A. L., Wells, K. D., and Murray, J. D. (2019). Proposed U.S. regulation of gene-edited food animals is not fit for purpose. *Npj Sci. Food* 3:3. doi: 10.1038/s41538-019-0035-y
- Wang, Y., Ding, Y., and Li, J. (2017). “CRISPR-Cas9-mediated gene editing in mouse spermatogonial stem cells,” in *RNAi and Small Regulatory RNAs in Stem Cells*, ed B. Zhang (New York, NY: Humana Press), 293–305. doi: 10.1007/978-1-4939-7108-4_20
- Yoon, Y., Wang, D., Tai, P. W. L., Riley, J., Gao, G., and Rivera-Perez, J. A. (2018). Streamlined *ex vivo* and *in vivo* genome editing in mouse embryos using recombinant adeno-associated viruses. *Nat. Comm.* 9:412. doi: 10.1038/s41467-017-02706-7
- Young, A. E., Mansour, T. A., McNabb, B. R., Owen, J. R., Trott, J. F., Brown, C. T., et al. (2019). Genomic and phenotypic analyses of six offspring of a genome-edited hornless bull. *Nat. Biotechnol.* doi: 10.1038/s41587-019-0266-0. [Epub ahead of print].
- Zhou, J., Lai, L., Ji, W., and Zhou, Q. (2019). Genome editing in large animals: current status and future prospects. *Nat. Sci. Rev.* 6, 402–420. doi: 10.1093/nsr/nwz013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 McFarlane, Salvesen, Sternberg and Lillico. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A.14 Publication: Accelerating Evolution

The following pages contain a paper I co-authored during my PhD project and has been referred to within this document (section 6.10).

Accelerating evolution

Artificial gene drives are a way of spreading genetic changes through a population quickly – for example, to make large numbers of wild mosquitos infertile or unable to carry malaria. **Bruce Whitelaw** and **Gus McFarlane** explain how this technology works, and how it could be used in conservation and agriculture as well as public health



Gene drives may offer a more cost-effective, humane and species-specific alternative for controlling invasive pests

It was the work of Gregor Mendel and Charles Darwin that revealed the fundamental mechanisms of genetic inheritance. Mendel observed consistent patterns of inheritance from parents to offspring and Darwin's theory of natural selection illustrated how traits are selected over time based on fitness.

However, we now know inheritance does not always follow the rules. In some cases the odds of a particular gene being transmitted to the next generation can be heavily skewed. Genetic elements that manage to beat the odds by being passed on more than half the time are known as 'gene drives', and this numerical advantage enables them to spread through populations even if they reduce an individual animal or plant's fitness.

Gene drives evolve in sexually reproducing species due to the genetic conflict that occurs between genes in an organism, enabling them to gain an advantage over their genetic competition and increase their likelihood of being passed to future generations. While most genes are transmitted to 50% of the offspring, gene drives manage to be transmitted to more than 50% of the offspring – that could be just 50.01% or a full 100%.

Broadly speaking there are two main strategies genes can use to achieve 'drive'. The first is interference, whereby the gene gains an advantage by disrupting the transmission of an alternative gene. For example, a gene in a diploid organism could reduce the fitness of the 50% of sperm or eggs that receive the alternative gene during the process of gametogenesis, thereby increasing

its own chance of success and reproduction in the next generation.

The second method is by overreplication. These gene drives bias their transmission to the next generation by being replicated more than other genes in an animal. Homing endonuclease genes are a good example: these cut DNA and get copied into the cut site as part of the repair process, increasing their abundance in the animal's genome¹.

LEARNING TO DRIVE

It was the study of naturally occurring gene drives that lead Austin Burt of Imperial College London to first propose that synthetic gene drives could be used to alter or suppress wild animal populations².

That was over a decade ago, and the genome editing tools available at the time were not useful in many organisms or suffered from evolutionarily instability. Fast forward to 2019 and the blockbuster genome editing tool known as CRISPR has enabled scientists to engineer synthetic gene drives in animals that could profoundly benefit public health, agriculture and the environment.

CRISPR gene editing uses a Cas enzyme, which works like a pair of a molecular scissors to cut DNA. A guide RNA guides the Cas enzyme to a targeted DNA sequence in the genome³. In 2014 Kevin Esvelt and his team at MIT outlined a technically feasible way to develop CRISPR-based gene drives⁴ and this approach has led to a flurry of research activity and media attention.

The most investigated form of CRISPR-based gene drive is known as a homing gene drive. This works by cutting a target DNA sequence in an animal's genome and then copying or 'homing' itself into the cut site. A CRISPR-based homing gene drive is essentially an engineered overreplication gene drive.

To build this system an animal is engineered with a gene drive cassette that encodes the two components of CRISPR: the Cas enzyme and guide RNA (see diagram, p21). These components, when expressed from one allele, then cut the sister allele on the other (homologous) chromosome.

After the sister allele on the homologous chromosome is cut, a cellular process known as homologous recombination results in the gene drive cassette (and any additional genetic payload included in it) being copied onto the homologous chromosome.

This means that both sets of chromosomes have copies of the gene drive cassette – and any animal with two copies of a gene is guaranteed to pass it on to all their offspring. As well as the CRISPR machinery, the gene drive cassette can be engineered to deliver additional DNA sequences with a desired trait.

In the offspring generation, the homing process occurs again and will do so in all subsequent generations, resulting in the gene drive spreading through the targeted population⁴.

BEYOND MOSQUITOES

To date CRISPR-based homing gene drives have been developed in yeast, fruit flies, two-species of malaria-carrying mosquitoes and most recently in mice^{5,6}. Most of the research on

CRISPR gene drives thus far has focused on controlling populations of mosquito as vectors of human disease and significant progress has been made towards achieving this goal. Demonstrations of the technology in other species have primarily been proof-of-concept experiments, which have widened the scope of the technology's potential applications.

Another potential application of gene drives is to control invasive pests. Invasive vertebrate pests, such as cane toads and rabbits in Australia, grey squirrels in the UK, possums in New Zealand and rodents around the globe cause significant damage to the environment, and incur economic and social costs. Control measures include shooting, poisoning, trapping and the release of biological agents, which are largely inadequate and costly, and often lead to unwanted suffering in both target and non-target species.

Gene drives may offer a more cost-effective, humane and species-specific alternative. In most pest species, female procreative capacity is responsible for maintaining the overall population size and so an efficient means of population suppression is to bias the sex ratio in favour of males. A grossly male breeding population will result in a population decline, while an all-male breeding population would lead to eradication.

FEMALE FERTILITY

A homing-based gene drive can achieve this with a gene drive cassette targeted to disrupt the coding sequence of an essential female fertility gene⁵. Males would be unaffected by loss of the female fertility gene, while all females that inherit the drive would be infertile.

Gene drives could also be used to rescue threatened or endangered animals. For instance chronic wasting disease (CWD) is a form of prion disease in cervids that affects primarily deer, elk and moose. The infectious prions responsible for CWD ▶

Scientists are working to develop gene drives that are inherently self-limiting – that is, they only stay in a population transiently and could be localised to targeted populations

Gus McFarlane AMRSB (left) and Bruce Whitelaw (FRSB) from the Roslin Institute



are caused by misfolding and aggregation of a normal, cellular prion protein (PrP^C) and are spread through saliva, urine and faeces. Once an animal is infected, misfolded prion protein builds up in the brain, causing neuronal loss and eventually leading to death. In some states in northern US, prevalence of CWD in deer is up to 35% in free-ranging populations and 90% in captive deer.

Studies in rodents and large animals have shown that animals do not require PrP^T and that gene drives that remove or 'knock out' the PrP gene could be used to rescue infected cervid populations from this highly infectious disease. However, the long generation interval in cervids would mean the drive spreads far more slowly through the population than in insect applications, for example.

RESCUING AMPHIBIANS

Gene drives could also be applied to rescue threatened amphibians from chytridiomycosis. Chytridiomycosis is the disease caused by the fungus *Batrachochytrium dendrobatidis*, which has been implicated in mass die-offs and the extinction of hundreds of frog species since the 1990s, and threatens up to 50% of amphibians.

As the fungus reproduces mostly asexually, it cannot be controlled with a gene drive itself. However, resistance to *B. dendrobatidis* infection varies both within and among amphibian species. Major histocompatibility complex (MHC) peptides play an important role in the innate immune system of vertebrates and a specific MHC allele in the lowland leopard frog has been shown to increase survival of infected individuals⁹.

Gene drive could be used to spread the lowland leopard frog's resistant MHC allele through threatened amphibian populations, rescuing species from this killer fungus and potentially saving them from extinction. Careful consideration in the gene drive design would be required to ensure the gene

POTENTIAL GENE DRIVES IN AQUA- AND AGRICULTURE

In Norway the total cost of sea lice to the aquaculture industry is around £400m per year, and treatments involve chemicals that damage marine ecosystems, affect fish appetite and growth, and mean treated fish can't be sold for several weeks after treatment. There is also resistance among lice to the three major classes of chemicals used. Self-limiting gene drives could be deployed to locally suppress lice populations in fish farms without affecting the broader marine ecosystem.

Another pest species that has significant economic and welfare implications for animal agriculture is the New World screwworm (NWS) fly. NWS causes myiasis, a parasitic infestation of warm-blooded animals as a result of the flies' larvae growing inside the host and feeding on its flesh. The screwworm is found in all South American countries except Chile and has an overwhelming impact on livestock farming



The salmon louse (*Lepeophtheirus salmonis*), a species of sea louse

throughout the region. Wildlife is also threatened, especially endangered species. The US is considered NWS-free after a 50-year eradication programme estimated to have cost £1bn. As a potentially quicker and cheaper solution the nations of South America could deploy self-limiting gene drives to locally suppress NWS or a self-perpetuating drive to eradicate the pest from the entire continent, potentially saving wildlife from NWS infestation and increasing agricultural productivity.

Gene drives could be applied to protect cervids such as moose from chronic wasting disease



drive does not impair the function of key immune system genes.

The main concern with homing gene drive strategies is that these self-perpetuating genetic elements can, in theory, spread indefinitely. There are, however, scenarios and strategies to prevent the duration and extent of this.

First, the spread would be limited by naturally arising resistant alleles that prevent CRISPR from recognising and cutting its target site. These resistant alleles could exist in the population prior to release or evolve when an error in the DNA repair process occurs and a CRISPR-mediated cut is repaired with a new functional DNA sequence (which the CRISPR guide RNA in future generations then does not recognise).

If a gene drive was not stopped by naturally arising resistance, there are 'stop' buttons available, theoretically. Gene drives can be inactivated by the release of animals with engineered resistant alleles or even with a 'reversal gene drive' that immunises the animal against the original drive.

However, these reactive stop buttons are not ideal. A more proactive approach is needed and scientists are working to develop gene drives that are inherently self-limiting – that is, they only stay in a population transiently and could be localised to targeted populations.

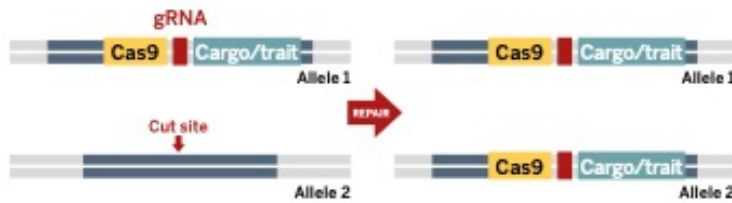
Several self-limiting gene drive concepts have been proposed and it is hoped these systems will soon be realised⁹. One type is a so-called 'daisy-chain

THE HOMOLOGOUS GENE DRIVE

Ensuring genetic modifications are on both chromosomes of a pair means it will spread rapidly through a population

A 'gene drive cassette' is inserted into an organism's genome on one chromosome of a homologous pair.

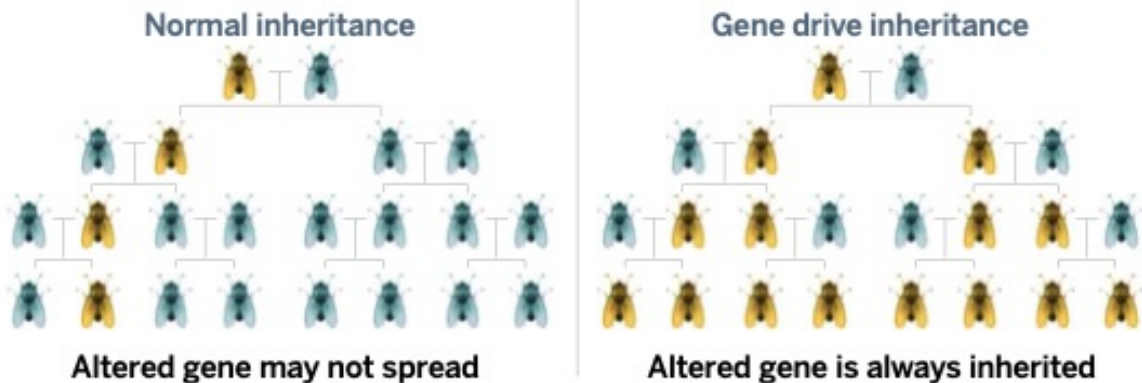
When the Cas9 and guide RNA are expressed, together they make a cut at the same location on the other chromosome of the pair.



The cell repairs the cut by copying the sequence at this location from the other chromosome, including the cassette.

With the cassette now on both chromosomes, it is almost certain to be passed on to all offspring.

The process repeats in the cells of the next generation, with the cassette copying itself so it is on both the paternal and maternal chromosome.



drive': with this approach the CRISPR components of the gene drive are split up and scattered throughout the genome so that none of them can be copied on its own. Element A can only copy itself if element B is present, and element B can only copy itself if element C is present. And element C, crucially, cannot copy itself at all – it can only be passed on to offspring at the normal rate of 50%.

If daisy-chain animals carrying all three elements were released, all the offspring will inherit element A and B, but only half will inherit element C. In the following generation element B and A will spread, but C will gradually die out. Once it does, B will start to disappear and finally A will too.

FUTURE POTENTIAL

If robust self-limiting gene drives can be developed, it puts the power in the hands of communities to deploy a gene drive locally, avoiding the need for highly complex geopolitical agreements that will be required for self-perpetuating gene drive systems.

With the realisation of self-limiting gene drives, this technology would be more amenable to applications in animal agriculture and aquaculture, such as the pressing issue of sea lice in fish farming (see 'Potential gene drives in aqua- and agriculture', above left) or the eradication of the New World screwworm.

Although gene drive technology still faces significant scientific, political and social hurdles, we are optimistic of its future potential and as such have

chosen to highlight prospective applications of the technology in this article. However, it must be noted that before any of these proposed applications are deployed there is a requirement for in-depth analysis of the ecological implications, as well as the need for broad community engagement with those that may be affected by the release of a gene drive.

Each potential application should be considered on a case-by-case basis, but with the rapid progress in this field, it is hoped the risks associated with current homing gene drive strategies are reduced with the realisation of self-limiting gene drive approaches.

For the first time we have the makings of a technology that could allow us to overcome the limitations imposed by conventional genetic inheritance and utilise molecular biology to spread beneficial traits through wild animal populations. Both Mendel and Darwin would certainly be astounded by how far our understanding has come and the benefits gene drives could offer our society and the environment.

Bruce Whitelaw is deputy director and director of partnerships at the Roslin Institute and professor of animal biotechnology at the Royal (Dick) School of Veterinary Studies. He is currently establishing robust methodology for genome editing in animal systems.

Gus McFarlane is a doctoral student at the Roslin Institute focused on developing CRISPR-based gene drive strategies for controlling invasive vertebrate pests. He is interested in the application of genome editing tools to improve the welfare and performance of livestock.

REFERENCES

- 1) Burt, A. & Trivers, R. *Genes in conflict: the biology of selfish genetic elements*. Harvard University Press (2009).
- 2) Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society B: Biological Sciences* 270(1518), S21–S28 (2003).
- 3) Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6095), 816–821 (2012).
- 4) Esvelt, K. et al. Concerning RNA-guided gene drives for the alteration of wild populations. *Elife* 3, e03401 (2014).
- 5) McFarlane, G. et al. CRISPR-based gene drives for pest control. *Trends in Biotechnology* 36(2), 130–133 (2018).
- 6) Grunwald, H. et al. Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline. *Nature* 566(742), 105–109 (2019).
- 7) Richt, J. et al. Production of cattle lacking prion protein. *Nature Biotechnology* 25(1), 132–138 (2007).
- 8) Saavedra, A. & Zamudio, K. MHC genotypes associate with resistance to a frog-killing fungus. *Proceedings of the National Academy of Sciences* 108(40), 16705–16710 (2011).
- 9) Min, J. et al. Harnessing gene drive. *Journal of Responsible Innovation* 8(3), 540–565 (2018).