



University  
of Glasgow

Ross, Paul (2016) *Developing novel therapeutic strategies for Rett syndrome*. PhD thesis.

<http://theses.gla.ac.uk/7306/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study

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

# **Developing Novel Therapeutic Strategies for Rett Syndrome**

Paul Ross

BSc (Hons), MRes

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

Institute of Neuroscience and Psychology

College of Medical, Veterinary and Life Science

University of Glasgow

Glasgow, G12 8QQ

UK

March 2016

© Paul Ross

# Abstract

Rett syndrome (RTT) is a rare paediatric disorder of females that leads to lifelong cognitive, motor, and respiratory impairment. In the vast majority of cases the disorder is caused by *de novo* mutations in the X-linked gene *MECP2*. There are currently no treatments, but genetic studies in mice have shown that the disease is reversible, even after the onset of symptoms. Since there remains a fundamental lack of knowledge about the downstream pathways involved in gene function, current therapeutic efforts are focused on targeting the disease at the gene level, mainly using viral based delivery of *Mecp2* gene products. Recent work in mouse models has shown that exogenous delivery of a wild-type (WT) copy of the *Mecp2* gene can lead to significant improvements in RTT-like symptoms, but significant challenges remain, both in the delivery of gene constructs to target cells, and in maintaining gene transcription within physiological tolerance.

The work in this thesis explores an alternative therapeutic approach, using newly developed genome editing technology. There were two major aims of this thesis. The first aim was to examine the role of the peripheral tissues in the development of the RTT phenotype. In order to develop correctly targeted therapies it is crucial to know what tissues are most relevant to disease development. It has been widely assumed that the major RTT symptoms can be explained solely by an absence of MeCP2 from cells in the nervous system. However, this was based on mouse studies in which only a few gross aspects of the disorder were examined. In this thesis a newly created peripheral knock-out (KO) mouse model, in which *Mecp2* transcription is silenced in peripheral tissues but selectively reactivated in the cells of the nervous system, was comprehensively phenotyped in order to determine the role of peripheral MeCP2 in RTT. The second major aim of this thesis was to develop a novel strategy for *Mecp2* mutation repair, using recently developed genome editing tools. Based on the results from the peripheral KO phenotyping, this strategy was designed to overcome the particular challenges associated with genome editing in the nervous system, and involved the insertion of a therapeutic construct directly into a non-coding region of the *Mecp2* gene using TALEN and CRISPR. This construct was designed to splice to upstream *Mecp2* exons in order to replace downstream mutated exons in the final mRNA transcript.

To generate the peripheral KO model (stop-cre), mice in which *Mecp2* transcription was globally silenced by a cre-excisable stop cassette (stop-y) were crossed with a nestin-cre mouse line to selectively reactivate gene transcription in the nervous system. Southern blot

analysis of tissues showed reactivation in a large number of cells (91.9%) in whole brain samples. Reactivation was particularly high in the cerebellum which showed 96.4% efficiency. Robust silencing was shown in peripheral tissues with only very small levels of reactivation in liver (0.9%), spleen (0.5%), skeletal muscle (1.2%) and heart (7.4%). Higher levels were seen in lung (14.3%) and kidney (24.4%) tissue. Peripheral KO mice did not show the early death phenotype seen in global KO mice and showed only very subtle RTT symptoms when examined using a well-established RTT scoring system. These mice also did not display any of the gait, balance, or respiratory dysfunction typical of RTT mouse models. However, peripheral KO mice did show a reduction in activity levels and exercise capacity across a number of tests. In the open-field, spontaneous activity levels were significantly reduced compared with WT (total distance moved =  $3523 \text{ cm} \pm 215 \text{ SEM}$  vs  $4242 \text{ cm} \pm 167$  in WT), on the accelerating rotarod, latency to fall was significantly reduced ( $168 \text{ s} \pm 14.9$  vs  $243.5 \text{ s} \pm 11.5$  in WT) and on an inclined accelerating treadmill, the time lasted before exhaustion was markedly reduced ( $8.7 \text{ min} \pm 1.6$  vs  $16.5 \text{ min} \pm 1.3$  in WT). In addition, peripheral KO mice also displayed the biomechanical abnormalities of bone seen in global KO mice, including reduced cortical stiffness and hardness.

The genome editing mutation repair strategy developed in this thesis required a non-coding target region free of repetitive sequence to be identified upstream of exons 3 and 4, where most of the disease causing mutations occur. A suitable 900 bp region of unique sequence was identified 1.6 – 0.7 kb upstream of the beginning of exon 3. To design TALEN pairs targeting this region, the Cornell University TALEN design tool was used to identify 100s of possible TALEN pairs, which were then filtered based on best practice-TALEN design and for the presence of unique restriction sites at the break site. Four pairs remained after filtering and these were assembled using a two stage cloning process based on the Golden Gate method. The efficiency of each pair was assessed using a restriction digest based assay and an online tool (TIDE) which relied on the decomposition of Sanger sequencing traces. The results showed a range of cutting efficiencies from 2.1% of cells (TALEN # 63) to 42.9% of cells (TALEN # 333). A CRISPR design tool was used to generate CRISPR guide target sequences. The four guides with the lowest predicted off-target effects were selected, synthesised as complementary oligonucleotides, and cloned upstream of a guide RNA scaffold in a CRISPR guide expression plasmid. Cutting efficiency was assessed using TIDE, and the results again showed a range of efficiencies from 60.4% (B52) to 22% (A65).



To assess if the best performing TALEN and CRISPR-Cas9 constructs could successfully target exogenous DNA into intron 2 of the *Mecp2* gene, a repair construct was designed. This contained WT sequence for exons 3 and 4 of *Mecp2* in a minigene format as well as appropriate splice elements and an mCherry fluorescent tag for easy detection. The repair construct was synthesised and cloned into a mammalian expression plasmid, with flanking regions, containing either TALEN or CRISPR target sites, inserted at either end of the construct. For unknown reasons the repair construct was toxic to bacterial cells when flanked by the CRISPR target sites and this version could therefore not be used for transfection experiments. The TALEN repair construct was transfected into cells along with TALEN pair # 333 and successful insertion was assessed using a PCR-based assay. Results showed that the repair construct was successfully inserted into non-coding genomic DNA in the correct location, and that a mutant version of the TALEN construct, designed to increase specificity, led to an increase in the levels of insertion in the correct orientation. To assess if this led to the production of corrected protein, cells were examined for mCherry protein expression using flow cytometry. Results showed that there was no significant increase in mCherry expression in transfected cells, suggesting the repair construct did not successfully splice to upstream *Mecp2* exons.

In summary, the results from this thesis show that RTT is primarily a disorder of the nervous system, and that this should therefore be the main target of new therapies. However, they also show that an absence of MeCP2 from the peripheral tissues leads to a markedly reduced exercise capacity, and is also likely to be the primary cause for the bone dysfunction seen in RTT patients and mouse models. In this thesis, a novel therapeutic strategy for RTT was developed using genome editing tools. A number of TALEN and CRISPR constructs were designed that could successfully target specific non-coding regions of the *Mecp2* gene, and were shown to enable the insertion of an exogenous DNA repair construct into the genome. However, further flow cytometry analysis showed that this did not lead to the expected protein repair suggesting further work is required on the design of the repair construct to enable splicing to endogenous *Mecp2* exons. Overall, the results show that genome editing has a potential role in the treatment of genetic disorders like RTT, but that further work is required to enable successful repair of disease causing mutations.

# Table of Contents

Abstract .....	2
List of Tables .....	10
List of Figures .....	11
Acknowledgements .....	13
Author's Declaration .....	14
Abbreviations .....	15
Chapter 1 .....	17
General introduction .....	17
1.1 Overview .....	17
1.2 Clinical manifestations of RTT .....	17
1.3 <i>MECP2</i> and RTT .....	18
1.4 <i>MECP2</i> structure and DNA binding .....	19
1.5 MeCP2 function .....	21
1.5.1 Repressor of methylated DNA .....	21
1.5.2 Repressor or activator? .....	22
1.5.3 Control of long-gene expression .....	23
1.5.4 Global regulator of chromatin .....	24
1.6 Animal models .....	24
1.6.1 RTT KO mouse models .....	24
1.6.2 RTT knock-in models .....	25
1.7 Reversibility .....	26
1.7.1 Global MeCP2 restoration .....	26
1.7.2 Brain region specific MeCP2 restoration .....	28
1.7.3 Glia and microglia reversal .....	29
1.8 Gene-based therapies .....	30
1.8.1 Unsilencing the inactivated X chromosome .....	30
1.8.2 Gene therapy .....	31
1.8.3 Problems associated with traditional gene therapy approaches .....	33
1.9 Genome editing .....	34
1.9.1 Zinc-finger nucleases (ZFN) .....	34
1.9.2 Transcription activator-like effector nucleases (TALEN) .....	36
1.9.3 CRISPR-Cas9 .....	37

1.9.4 DNA repair pathways .....	38
1.9.5 <i>Ex vivo</i> genome editing .....	40
1.9.6 <i>In vivo</i> genome editing .....	41
1.9.7 Off-target effects .....	42
1.9.8 <i>In vivo</i> delivery .....	44
1.10 Aims .....	45
Chapter 2 .....	47
Material and methods .....	47
2.1 Animal models .....	47
2.1.1 Generation of a peripheral KO mouse model .....	47
2.1.3. Maintenance of mouse colonies .....	47
2.2 Mouse phenotyping and behavioural analysis .....	47
2.2.1 Weight measurement and phenotypic severity scoring .....	47
2.2.2 Nest building .....	47
2.2.3 Whole-body plethysmography .....	50
2.2.4 Open field .....	50
2.2.5 Gait analysis .....	50
2.2.6 Balance beam .....	51
2.2.7 Bone biomechanical tests .....	51
2.2.8 Rotarod .....	51
2.2.9 Exercise tolerance .....	51
2.2.10 Blood biochemistry .....	52
2.3 General molecular biology .....	52
2.3.1 Reagents .....	52
2.3.2 Solutions .....	53
2.3.3 Protocols .....	53
2.3.3.1 PCR primer design .....	53
2.3.3.2 Polymerase chain reaction (PCR) .....	53
2.3.3.3 Measurement of DNA concentration .....	54
2.3.3.4 Restriction digest .....	55
2.3.3.5 Agarose gel electrophoresis and gel extraction .....	55
2.3.3.6 Ligation .....	56
2.3.3.7 Bacterial transformation .....	56

2.3.3.8 Colony PCR .....	57
2.3.3.9 Plasmid isolation.....	57
2.3.3.10 Sequencing.....	57
2.3.3.11 Genomic DNA extraction .....	57
2.4 Cell culture.....	58
2.4.1 Reagents.....	58
2.4.2 Cell culture media .....	58
2.4.3 Cell maintenance.....	59
2.4.3.1 Liquid nitrogen recovery.....	59
2.4.3.2 Cell counting.....	59
2.4.3.3 Cell passage .....	59
2.4.3.4 Freezing cells for long-term storage .....	60
2.4.4 Neon transfection .....	60
2.5 TALEN construction.....	61
2.5.1 Reagents.....	61
2.5.2 Primers .....	61
2.5.3 Vectors .....	61
2.5.4 Protocols .....	61
2.6 Cloning of CRISPR-Cas9 constructs .....	67
2.6.1 Reagents.....	67
2.6.2 Primers .....	67
2.6.3 Vectors .....	67
2.6.4 Protocol.....	68
2.7 <i>Mecp2</i> repair and GFP reporter constructs .....	70
2.7.1 Primers .....	70
2.7.2 Vectors .....	70
2.8 TIDE analysis of cutting efficiency .....	71
2.9 ImageJ band intensity measurement .....	71
2.10 Flow cytometry .....	73
Chapter 3.....	74
Nervous system specific reactivation of <i>Mecp2</i> .....	74
3.1 Introduction.....	74
3.2 Aims.....	75

3.3 Validation of the peripheral KO Mouse Model .....	76
3.4 Absence of early lethality in peripheral KO mice .....	80
3.5 Absence of RTT-like symptoms in peripheral KO mice .....	81
3.6 Peripheral KO mice show reduced bodyweight.....	82
3.7 Peripheral KO mice show no changes in blood biochemistry .....	83
3.8 Peripheral KO mice show mild kidney pathology .....	84
3.9 Peripheral KO mice show mild hypoactivity.....	85
3.10 Peripheral KO mice construct normal nests.....	86
3.11 Peripheral KO mice do not show RTT-like gait defects.....	87
3.12 Peripheral KO mice show normal performance on balance beam.....	89
3.13 Reduced rotarod performance in peripheral KO mice .....	90
3.14 Peripheral KO mice show reduced exercise capacity .....	91
3.15 Absence of RTT-like breathing phenotype in peripheral KO mice .....	92
3.16 Evidence of muscle pathology in peripheral KO mice .....	94
3.17 Peripheral KO mice show RTT-like bone phenotypes .....	95
3.18 Discussion .....	97
Chapter 4 .....	102
Design and synthesis of CRISPR and TALEN constructs.....	102
4.1 Introduction.....	102
4.2 Aims.....	104
4.2 Identification of suitable target region for repair construct insertion .....	105
4.3 Design of TALEN pairs .....	107
4.4 TALEN synthesis.....	108
4.5 Optimisation of electroporation parameters.....	114
4.6 Testing TALEN cutting efficiency .....	117
4.7 Design of CRISPR-Cas9 constructs.....	122
4.8 Synthesis of CRISPR guides.....	124
4.9 Testing of CRISPR cutting efficiency .....	126
4.10 Discussion .....	128
Chapter 5 .....	132
Targeting a repair construct into intron 2 of the <i>Mecp2</i> gene.....	132
5.1 Introduction.....	132
5.2 Aims.....	136

5.3 Design of the repair construct .....	137
5.4 Synthesis of the repair construct .....	138
5.5 Cloning of TALEN repair construct .....	139
5.6 Cloning of CRISPR repair construct.....	141
5.7 Design of PCR system for detecting successful repair construct insertion .....	144
5.8 Repair construct successfully inserted into target site using TALEN # 333 .....	144
5.9 Repair construct insertion with TALEN # 333 does not lead to protein expression .....	149
5.10 Construction of a GFP reporter construct .....	153
5.11 Discussion .....	160
Chapter 6.....	165
General discussion .....	160
6.1 Introduction.....	165
6.2 Major findings and conclusions .....	166
6.2.1 Peripheral contribution to RTT .....	166
6.2.1.1 Findings.....	166
6.2.1.2 Significance.....	167
6.2.1.3 Caveats and technical considerations.....	168
6.2.1.4 Future experiments.....	168
6.2.2 Genome editing as a treatment for RTT .....	169
6.2.2.1 Findings.....	169
6.2.2.2 Significance.....	170
6.2.2.3 Caveats and technical considerations.....	171
6.2.2.4 Future experiments.....	172
6.3 Summary .....	173
References.....	174

## List of Tables

Table 2.1 - General molecular biology reagents .....	52
Table 2.2 - General molecular biology solutions .....	53
Table 2.3 - Components of typical Maxima hotstart green PCR .....	54
Table 2.4 - Components of a typical Phusion high fidelity PCR .....	54
Table 2.5 - Thermocycling conditions for standard PCR .....	54
Table 2.6 - Components of a typical restriction digest .....	55
Table 2.7 - Components of a typical ligation reaction .....	56
Table 2.8 - Cell culture reagents .....	58
Table 2.9 - Composition of cell culture media .....	58
Table 2.10 - Reagents specific for TALEN construction .....	61
Table 2.11 - Primers used for TALEN pair construction and testing .....	61
Table 2.12 - Vectors used for TALEN pair construction .....	61
Table 2.13 - Components of TALEN golden gate # 1 reaction .....	62
Table 2.14 - Components of TALEN golden gate # 2 reaction .....	64
Table 2.15 - Reagents specific for cloning of CRISPR-Cas9 constructs .....	67
Table 2.16 - Primers for cloning target sequence into guide expression plasmid .....	67
Table 2.17 - Plasmid vectors used for CRISPR construction .....	67
Table 2.18 - Overhang design for guide sequence oligonucleotides .....	68
Table 2.19 - Components of oligo annealing reaction .....	69
Table 2.20 - Components of guide expression plasmid digest .....	69
Table 2.21 - Components of ligation reaction for annealed oligos and expression plasmid ...	70
Table 2.22 - Primers for cloning and detecting genomic insertion .....	70
Table 2.23 - Plasmid vectors for repair construct cloning and flow cytometry analysis .....	70
Table 3.1 - Blood biochemistry results .....	83
Table 3.2 - Histopathological screening results .....	84
Table 4.1 - RVD sequence and FokI type for individual TALEN arms .....	113
Table 4.2 - Electroporation parameters used for optimisation .....	116

# List of Figures

Figure 1.1 MeCP2 functional domains .....	20
Figure 1.2 ZFN structure and function .....	34
Figure 1.3 TALEN structure and function .....	36
Figure 1.4 CRISPR-Cas9 structure and function .....	37
Figure 1.5 Repair strategies using alternative DNA repair pathways.....	39
Figure 2.1 Nesting scoring system.....	48
Figure 2.2 Example nests.....	49
Figure 2.3 pFUS plasmid map .....	63
Figure 2.4 pTAL_FokI plasmid map .....	65
Figure 2.5 ImageJ band intensity measurement.....	72
Figure 3.1 Creation of peripheral KO mouse model.....	76
Figure 3.2 Southern blot analysis reveals tissue specific <i>Mecp2</i> reactivation .....	78
Figure 3.3 Immunoblots confirm robust peripheral silencing .....	79
Figure 3.4 Normal survival in peripheral KO mice .....	80
Figure 3.5 Mild RTT-like symptoms displayed by stop-cre mice .....	81
Figure 3.6 Decreased bodyweight in peripheral KO mice.....	82
Figure 3.7 Peripheral KO mice show reduced activity levels.....	85
Figure 3.8 Peripheral KO mice were capable of constructing high quality nests.....	86
Figure 3.9 Majority of stop/y mice incapable of performing on treadmill test.....	88
Figure 3.10 Peripheral KO animals do not show impairment on balance beams .....	89
Figure 3.11 Peripheral KO mice show reduced rotarod performance .....	90
Figure 3.12 Reduced exercise capacity in peripheral KO mice .....	91
Figure 3.13 Absence of respiratory dysfunction in peripheral KO mice.....	93
Figure 3.14 No significant muscle abnormalities in peripheral KO mice .....	94
Figure 3.15 Peripheral KO mice show RTT-like bone phenotype .....	96
Figure 4.1 Identifying a suitable target region for repair construct insertion .....	106
Figure 4.2 DNA target sequence of selected TALEN pairings .....	108
Figure 4.3 Cloning strategy for TALEN arm synthesis .....	109
Figure 4.4 Restriction digest confirmation of correct array assembly .....	111
Figure 4.5 Sequencing confirmation of correct array assembly .....	112
Figure 4.6 Restriction digest confirmation of correct final TALEN arm assembly .....	114



Figure 4.7 Representative plasmid map of final TALEN construct .....	115
Figure 4.8 Optimisation of Neon electroporation parameters .....	117
Figure 4.9 Outline of procedure for testing cutting efficiency of TALEN pairs .....	118
Figure 4.10 Agarose gel of RFLV digestion products .....	120
Figure 4.11 TIDE analysis of TALEN pair # 333 cutting efficiency .....	122
Figure 4.12 DNA target sequence of selected TALEN pairings .....	123
Figure 4.13 Outline of cloning strategy of guide sequence into expression .....	124
Figure 4.14 Sequence confirmation of guide sequence cloning into expression plasmid .....	125
Figure 4.15 Map of Cas9 expressing plasmid .....	126
Figure 4.16 Cutting efficiency of CRISPR guides .....	127
Figure 5.1 Outline of intron 2 repair strategy .....	133
Figure 5.2 Strategy for ensuring correctly orientated insertion of the repair construct .....	135
Figure 5.3 Components of <i>Mecp2</i> repair construct .....	137
Figure 5.4 Map of GeneArt <i>Mecp2</i> repair construct .....	139
Figure 5.5 Plasmid map of final TALEN repair construct .....	140
Figure 5.6 Restriction digest confirmation of cloning of TALEN repair .....	141
Figure 5.7 Map of CRISPR-target region construct .....	142
Figure 5.8 Map of final CRISPR repair construct .....	143
Figure 5.9 Restriction digest confirmation of cloning of CRISPR repair .....	144
Figure 5.10 PCR detection of repair construct insertion .....	145
Figure 5.11 PCR analysis shows TALEN # 333 mediated repair construct insertion .....	146
Figure 5.12 Effect of plasmid concentration on repair construct insertion efficiency .....	147
Figure 5.13 Effect of insert:TALEN arm ratio on genomic insertion efficiency .....	148
Figure 5.14 TALEN cutting efficiency increases with increasing plasmid concentration .....	149
Figure 5.15 Gating of live cells allowed debris and aggregates to be excluded .....	150
Figure 5.16 Flow cytometry analysis of TALEN and repair construct transfected cells .....	152
Figure 5.17 No mCherry expression after TALEN and repair construct transfection .....	153
Figure 5.18 Splice acceptor GFP reporter construct .....	154
Figure 5.19 Plasmid map of final <i>HBB</i> Ex3 SA reporter plasmid .....	155
Figure 5.20 Sequence confirmation of <i>HBB</i> Ex3 SA into GFP reporter .....	156
Figure 5.21 Flow cytometry analysis of GFP reporter construct transfected cells .....	158
Figure 5.22 High background GFP expression from reporter construct .....	159

# Acknowledgements

Firstly, I would like to thank my supervisors Dr. Stuart Cobb, Dr. Mark Bailey and Professor Brian Morris for their guidance and support throughout my PhD, and especially for giving me the freedom to follow my own ideas and to make my own mistakes. I would also like to thank the BBSRC for funding the work carried out in this thesis.

In addition, I am most grateful to Dr. Katherine West and Dr. Adam West for their advice and assistance during all the genome editing work, and to Professor Adrian Bird for his collaboration during the peripheral KO work.

I would like to thank all the people whose contributions made this thesis possible. In particular, Dr. Jacky Guy who made the mouse model used in chapter three, Dr. Noha Bahey for her histological work, and Dr. Bushra ‘Big Barbara’ Kamal for her help with the bone analysis. I would also like to thank Dr. Ralph Hector for his invaluable help with the molecular work and Dr. Kamal Gadalla whose naturally suspicious mind helped me to think like a scientist.

Special thanks goes to John Craig for all his invaluable technical help during my PhD and his assistance during the behavioural testing. It was much appreciated.

I would also like to thank all the people I worked with during my PhD who made it so enjoyable, especially Paul Turko, Faye McLeod, Jake Griffin, Graham Lee, Manuel Denton, Elaine Hunter, Daniela Minchella, Big Fon, and Rebecca Openshaw.

Finally, I would like to thank my family who supported me through more years of education than can be reasonably expected, and in particular my parents who saved me from eating out of bins during my last few months of writing up.

## **Author's Declaration**

I declare that the work presented in this thesis is entirely my own with all exceptions being clearly indicated or/and properly cited in the context.

Signature: .....

Paul Ross

# Abbreviations

(mESC)	Mouse embryonic stem cells
<i>5hmC</i>	5-hydroxymethylcytosine
AAV	Adeno-associated virus
AAV2	Adeno-associated virus serotype 2
AAV8	Adeno-associated virus serotype 8
AAV9	Adeno-associated virus serotype 9
AmpR	Ampicillin resistance gene
ANOVA	Analysis of variance
bGH	Bovine growth hormone
CAG	CMV early enhancer/chicken $\beta$ actin
Cas	CRISPR-associated
CBA	Chicken beta actin
CMV	Cytomegalovirus
CNS	Central nervous system
CREB	cAMP response element binding protein
CRISPR	Clustered regularly interspaced short palindromic repeat
crRNA	CRISPR RNA
CT	Computed tomography
CV	Coefficient of variability
DPBS	Dulbecco's phosphate-buffered saline
DSB	Double strand break
FSC	Forward scatter
GFP	Green fluorescent protein
GI	Gastro-intestinal
H&E	Hematoxylin and eosin
H1	Histone 1
<i>HDAC</i>	Histone deacytelase gene
hGFAP	Human astrocytic fibrillary acidic protein
hmCG	Hydroxymethylated CG
HR	Homologous recombination
HSC	Haematopoietic stem cells
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
KI	Knock-in
KO	Knock-out
LINE	Long interspersed nuclear elements
LTP	Long-term potentiation
MBD	Methyl binding domain
mCA	Methylated CA
<i>MECP2</i>	Human methyl-CpG-binding protein 2 gene
MeCP2	Human or mouse methyl-CpG-binding protein 2 protein

<i>Mecp2</i>	mouse methyl-CpG binding protein 2 gene
<i>MeCP2_e1</i>	Methyl-CpG-binding protein 2 protein isoform 1
<i>MeCP2_e2</i>	Methyl-CpG-binding protein 2 protein isoform 2
NEB	New England Biolabs
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
Ori	Origin of replication
PCR	Polymerase chain reaction
RE	Restriction enzyme
RFLV	Restriction fragment length variations
RTT	Rett syndrome
RVD	Repeat-variable diresidue
SA	Splice acceptor
SCID	Severe combined immunodeficiency disorder
SEM	Standard error of mean
sgRNA	Single guide RNA
SINE	Short interspersed nuclear elements
SmR	Sepctinomycin antibiotic resistance gene
SNPs	Single nucleotide polymorphisms
SSC	Side scatter
STC1	Stanniocalcin 1
T2A	<i>Thoseaasigna</i> virus 2A
TALE	Transcription activator-like effectors
TALEN	Tale nuclease
TIDE	Tracking of indel decomposition
tracrRNA	<i>Trans</i> -activating crRNA
TRD	Transcriptional repressor domain
UBE3A	Ubiquitin protein ligase E3A
UTR	Untranslated region
WT	Wild-type
XCI	X chromosome inactivation
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside
ZFN	Zinc-finger nucleases
ZFP	Zinc-finger protein

# Chapter 1

---

## General introduction

### 1.1 Overview

Rett syndrome (RTT) is a rare paediatric disorder of females that leads to lifelong cognitive, motor, and respiratory impairment (Hagberg, 2002; Neul et al., 2010). In most cases the disorder is caused by *de novo* mutations in the X-linked gene *MECP2* (Amir et al., 1999). There are currently no treatments, but genetic studies in mice have shown that the disease is reversible, even after the onset of symptoms (Guy et al., 2007; Robinson et al., 2012). More recent work has shown that exogenous delivery of a WT copy of the *MECP2* gene can lead to significant improvements in RTT-like symptoms (Gadalla et al., 2012; Garg et al., 2013), but significant challenges remain, both in the delivery of gene constructs to target cells, and in maintaining gene transcription within physiological tolerance. The work in this thesis explores an alternative therapeutic approach, using newly developed genome editing technology such as CRISPR and TALEN. First, a newly created mouse model is used to investigate the organs crucial to the development of the RTT phenotype. This information is then used to develop a novel strategy for mutation repair, involving the insertion of a therapeutic construct directly into a non-coding region of the *MECP2* gene.

### 1.2 Clinical manifestations of RTT

RTT is a relatively rare disorder, occurring in approximately 1 in 10,000 female live births (Neul et al., 2010). The syndrome first manifests as a developmental stagnation after 7 to 18 months, with a failure to meet social and motor milestones, as well as marked microcephaly and growth delay (Hagberg et al., 1983; Hagberg, 2002). This is followed by a rapid regression, with a loss of any previously acquired motor and communication skills, and the replacement of purposeful hand movements with stereotypical behaviour such as clasping (Neul et al., 2010). In addition, there is the appearance of apraxia and an ataxic gait (Engerström, 1992). This regression is followed by a partial recovery and stationary phase which can last from years to decades (Hagberg, 2002). During this phase some locomotor ability is usually retained and communication skills are partially restored. Seizures are common during this period, but decrease in severity after 20 years of age (Steffenburg et al.,

2001). Finally, there is a late motor deterioration phase, characterised by a cessation of ambulation, necessitating complete reliance on the use of a wheelchair (Hagberg, 2002; Roze et al., 2007).

A recent study has shown that 70% of patients survive to at least 45 years of age, suggesting that survival into adulthood is common (Tarquinio et al., 2015). Of the patients who had died, the leading cause of death was cardiorespiratory failure, often due to lung infections. Other risk factors included low bodyweight and seizures. Autonomic dysfunction is common in RTT, including respiratory and gastro-intestinal dysfunction. Daytime breathing is characterised by episodic hyperventilation, breath-holding and apnoeas (Julu et al., 2001) which does not occur during sleep (Glaze et al., 1987; Marcus et al., 1994), while several reports also describe various digestive issues including swallowing dysfunction and GI dysmotility (Hagberg, 2002; Isaacs et al., 2003; Baikia et al., 2014). A majority of RTT patients also display a number of skeletal abnormalities including spinal deformities such as scoliosis, which may require surgical correction (Colvin et al., 2003; Kerr et al., 2003). Other abnormalities include hip deformations, early-onset osteoporosis, and a propensity to low-impact fractures (Guidera et al., 1991; Zysman et al., 2006; Downs et al., 2008; Roende et al., 2011). Over 80% of patients have sleep problems, including teeth grinding, night-time screaming and laughter, and daytime napping (Young et al., 2007; Wong et al., 2015). Studies also show a decrease in the proportion of sleep spent in the rapid eye movement stage (Glaze et al., 1987), as well as an increase in sleep-onset latency and the number of awakenings per night (McArthur and Budden, 1998). The communication impairments associated with RTT make the assessment of recurrent and chronic pain in RTT difficult. A recent study showed that a majority of patients experienced pain in the previous week that was both intense and long-lasting (Barney et al., 2015). The most common source was gastrointestinal pain, followed by musculoskeletal and seizure related pain. In addition to typical RTT, there are a number of recognised atypical forms in which some of the features associated with the disease are either attenuated or more pronounced (Neul et al., 2010).

### **1.3 *MECP2* and RTT**

Since RTT appeared to occur exclusively in females, it was hypothesised that the disorder must be X-linked dominant in origin (Hagberg et al., 1983). This was confirmed in 1999, when candidate region sequencing linked RTT to mutations in the X chromosome gene *MECP2* (Amir et al., 1999). Since then it has been shown that over 95% of typical RTT

patients have *MECP2* mutations (Neul et al., 2008). In the vast majority of cases these mutations are *de novo* (Girard et al., 2001) with a greater proportion occurring in the germline of the father (Wan et al., 1999; Trappe et al., 2001). As only female offspring inherit the male copy of the X chromosome this partially explains the strong female bias for the disorder. Rare cases of familial inheritance can occur due to skewing of X chromosome inactivation in the mother (Villard et al., 2000; Hoffbuhr et al., 2002). In these instances the mutant allele is preferentially silenced during X chromosome inactivation, leading to an asymptomatic carrier mother capable of passing the mutation on to offspring. This is responsible for the rare familial cases in which males inherit the mutant allele (Villard et al., 2000; Zeev et al., 2002). Since males only have one copy of the X chromosome, this leads to extremely severe dysfunction due to a complete lack of functional protein in these patients. Most cases present with infantile encephalopathy and patients die within two years of birth (Villard et al., 2000; Trappe et al., 2001; Zeev et al., 2002; Moretti and Zoghbi, 2006).

The presence of an *MECP2* gene duplication in males presenting with intellectual disability (Van Esch et al., 2005) suggested that over-expression of the gene also leads to dysfunction. Now known as *MECP2* duplication syndrome, this leads to a spectrum of clinical features including intellectual disability, autistic like communication impairment, recurrent infections, gait problems, and seizures (Van Esch et al., 2005; Lugtenberg et al., 2009; Ramocki et al., 2009). Again the disorder is more severe in males, who inherit the duplication from a carrier mother (Van Esch et al., 2005). While females with highly skewed X chromosome inactivation of the mutant allele can be asymptomatic, those with lower skewing can present with a range of psychiatric conditions including anxiety, depression, and compulsions (Ramocki et al., 2009).

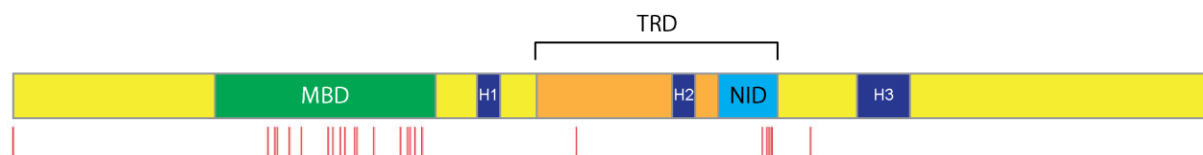
## 1.4 *MECP2* structure and DNA binding

MeCP2 is an abundant nuclear protein, first discovered through its binding affinity for DNA at methylated cytosines in CpG dinucleotides (Lewis et al., 1992). It is concentrated in pericentromeric heterochromatin which contains a particularly high density of methylated cytosines. It is a member of a family of nuclear proteins, all of which contain a methyl-CpG binding domain (MBD) (Hendrich and Bird, 1998). *MECP2* is expressed widely throughout the body, with particularly high levels seen in the nervous system (Shahbazian et al., 2002b; Zhou et al., 2006; Song et al., 2014). The gene contains four exons and has two common splice variants, MeCP2\_e1 and MeCP2\_e2, which encode proteins that differ only at their N-



termini (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004). MeCP2\_e1 has a translational start site in exon 1 and excludes exon 2 from the final protein, while MeCP2\_e2 uses a start site within exon 2. MeCP2\_e1 is thought to be the most abundant form in the brain, where it accounts for over 90% of total MeCP2 (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004; Dragich et al., 2007). The importance of the MeCP2\_e1 isoform is highlighted by the fact that while mutations in exon 1 have been discovered in RTT patients no such mutations have been seen for exon 2 (Mnatzakanian et al., 2004).

MeCP2 contains a number of domains crucial for the function of the protein (**Figure 1.1**). An 85 amino acid methyl-CpG binding domain (MBD) mediates binding of MeCP2 to methylated CpGs (Nan et al., 1993) and is responsible for the association with highly methylated heterochromatin (Nan et al., 1996). The importance of the MBD for DNA binding is made clear by the large scale redistribution of MeCP2 that occurs when the MBD is mutated (Baubec et al., 2013). As well as the known association with methylated CpG dinucleotides, recent evidence suggests that MeCP2 also binds with high affinity with methylated CpA (Chen et al., 2015; Gabel et al., 2015). Interestingly, CpA methylation is common in neurons (Lister et al., 2013; Guo et al., 2014) and levels begin to rise at a similar time-point as MeCP2 levels (Shahbazian et al., 2002b; Skene et al., 2010). Some evidence indicates that the MBD also interacts with 5-hydroxymethylcytosine (5hmC) containing DNA *in vitro* (Mellén et al., 2012; Spruijt et al., 2013), although other studies show that this is only true for CpA methylated DNA but not CpG (Gabel et al., 2015).



**Figure 1.1 – *MeCP2* functional domains**

Diagram outlining the main functional regions of the *MECP2* protein, including the methyl-CpG domain (MBD), three AT-hooks (H1-H3), transcriptional repression domain (TRD), and the NCoR-SMRT interaction domain (NID). Highlighted in red below the domains are RTT causing missense mutations which mostly cluster in the MBD and NID (adapted from (Lyst and Bird, 2015)).

The protein also contains a 104 amino acid transcriptional repressor domain (TRD) that interacts with a number of histone deacetylase (HDAC) containing co-repressor complexes such as mSin3a (Nan et al., 1998) and NCoR/SMRT (Lyst et al., 2013), as well as transcription factor YB1 (Forlani et al., 2010). The association with NCoR/SMRT appears to

be particularly crucial for MeCP2 mediated repression, and mutations in the TRD region that interacts with this co-repressor complex lead to a severe phenotype in mice, although it is milder in human patients (Lyst et al., 2013).

Several recent studies have suggested that not all DNA binding is mediated by the MBD, with a number of regions outside the MBD capable of independently binding DNA (Ghosh et al., 2010b). Several DNA-binding AT-hook motifs, have been discovered in the gene, including a conserved AT-Hook 2 domain in the TRD that can bind DNA and influence the structure of surrounding chromatin (Baker et al., 2013). Mutations in this region strongly influence the onset and symptom severity in male patients, and in mice leads to a reduction in the binding of MeCP2 to DNA and a rapid onset phenotype and early death, similar to that seen in *Mecp2*-null mice. More recently, missense mutations in AT-hook domain 1 have been associated with intellectual disability in males (Bianciardi et al., 2015).

## **1.5 MeCP2 function**

### ***1.5.1 Repressor of methylated DNA***

MeCP2 was first characterised in 1992 by Bird and colleagues, and was described as a DNA binding protein with an affinity for DNA containing methyl-CpG dinucleotides (Lewis et al., 1992). This affinity suggested that MeCP2 may have a role as a repressor of gene transcription, due to the known link between methylation and gene silencing (Bird, 2002). Later experiments in HeLa cell extracts showed that MeCP2, both in its native form and a recombinant version, specifically caused transcriptional repression of genes with methylated promoters but not of those with non-methylated promoters (Nan et al., 1997). In addition, the actual level of repression was found to be related to the density of methylation. These findings were confirmed *in vivo* in *Drosophila*, with MeCP2 causing transcriptional repression of genes with densely methylated promoters (Kudo, 1998). A series of deletion mutants showed that this repression was dependent on the MBD of the protein. A possible mechanism for repression was soon identified with the discovery of a transcriptional-repression domain (TRD) in the protein which interacted with a corepressor complex composed of mSin3A, a transcriptional repressor, and histone deacetylases (HDAC) (Jones et al., 1998; Nan et al., 1998). HDACs are commonly found in corepressor complexes and cause compaction of chromatin structure leading to gene silencing (Delcuve et al., 2012). The importance of HDACs for MeCP2 mediated silencing was highlighted by the absence of

repression in the presence of HDAC inhibitors such as trichostatin A (Jones et al., 1998; Nan et al., 1998). Another study showed that when drugs were used to remove methylation from DNA it led to release of MeCP2, acetylation of histones in the promoter regions and a reactivation of repressed genes (El-Osta et al., 2002). The results from all of these studies suggested a model whereby MeCP2 could link the methylation state of a gene with control of gene transcription, through its interactions with corepressor complexes. Since then a number of other binding partners have been identified, including a number of corepressors like cSki (Kokura et al., 2001) and NCoR/SMRT (Lyst et al., 2013). The NCoR/SMRT co-repressor complex appears to have a particularly crucial role, and a recent study suggest that it is responsible for MeCP2 mediated repression (Lyst et al., 2013). Its importance is supported by the fact that a number of RTT-causing missense mutations in human patients are located in the NCoR/SMRT binding region. When one of the most common of these mutations was modelled in mice, the mice developed a severe RTT-like phenotype similar to KO models (Guy et al., 2001; Lyst et al., 2013). MeCP2 also has a role in the control of mobile genetic elements such as the retroviral-like L1 retrotransposons. In mice, transcription and retrotransposition of these L1 elements are increased in *Mecp2* KO mice (Muotri et al., 2010; Skene et al., 2010). MeCP2 binds to the methylated 5'UTR of L1 elements and repression is relieved when methylation is removed. This is also seen in neuronal progenitor cells derived from RTT patient tissue (Muotri et al., 2010) suggesting a crucial role for MeCP2 in the control of mobile genetic elements and the maintenance of genome integrity.

### ***1.5.2 Repressor or activator?***

Since a number of studies pointed to MeCP2 as a mediator of gene silencing, it was naturally assumed that absence of the protein would lead to large increases in the mRNA levels of its target genes. However, initial gene expression studies failed to show any significant changes, even in the most symptomatic of mice (Tudor et al., 2002). Over the next few years, some studies did show expression changes in a small number of genes, such as those involved in the glucocorticoid stress response (Nuber et al., 2005) or in genes coding for subunits of a mitochondrial respiratory complex (Kriaucionis et al., 2006), but these relatively minor changes couldn't explain the severe and wide-ranging phenotype seen in RTT patients. The picture became more complex when a number of large-scale gene expression studies suggested that, contrary to previous evidence, MeCP2 had a role as a transcriptional activator as well as of a repressor (Chahrour et al., 2008; Ben-Shachar et al., 2009). Using microarrays gene expression levels were examined in *Mecp2* KO and overexpression models, focusing on

either the hypothalamus (Chahrour et al., 2008) or cerebellum (Ben-Shachar et al., 2009). Expression changes were seen in a large number of genes, although the amount of change was relatively subtle. Surprisingly however, the results showed that over 80% of affected genes appeared to be activated rather than repressed by MeCP2. Further support for this came from the fact that ChIP analysis showed MeCP2 was bound to several of the activated genes and interacted with CREB, a well-known activator of transcription (Chahrour et al., 2008). The role of MeCP2 as an activator of transcription was also supported by a study examining gene expression changes in an isogenic human embryonic stem cell model of RTT (Li et al., 2013b). In contrast to previous expression studies, results were normalised to cell number rather than total RNA to account for global shifts in transcription levels that could be masking the effects of MeCP2 loss. Using this approach, a large scale reduction in transcription and protein synthesis was observed in cells obtained from RTT patients, suggesting that the primary role of MeCP2 is as an activator of transcription. However, it is not clear from this study whether this is a primary consequence of MeCP2 loss or secondary to the loss of cell health that results from MeCP2 dysfunction.

### ***1.5.3 Control of long-gene expression***

The brain is composed of a large number of neuronal subtypes as well as large amounts of glia, which have been shown to express low levels of MeCP2 (Ballas et al., 2009). It is perhaps therefore not surprising that large scale changes in gene expression were not detected in initial whole-brain studies. A more recent study overcame this issue by analysing expression changes in selected neuronal sub-types (Sugino et al., 2014). In this case, large changes in gene expression were observed in KO mice. Interestingly, genes which were upregulated in the KO mice were strongly biased towards being long genes whereas this was not the case for those that were downregulated. This was supported by another study which also found an upregulation of long genes in the neurons of KO mice and a reciprocal downregulation in overexpression models (Gabel et al., 2015). Using tissue-specific expression analysis they also show that long genes are particularly associated with neuronal functions and are disproportionately expressed in brains compared to other tissues. An exciting finding in this study was that this repression is achieved by binding to methylated CA (mCA) sites, a type of methylation that accumulates during synaptogenesis and is the most dominant form of methylation in neurons (Lister et al., 2013), perhaps helping to explain the time-course of RTT development. The density of these sites increases as the gene length increases, thus explaining the specific repression of these genes (Gabel et al., 2015). In contrast to the strong

affinity binding for methylated CA sites, MeCP2 shows weak affinity for hydroxymethylated CG (hmCG) sites, the frequency of which increases significantly in neurons at the same time as that of mCA. hmCG is particularly associated with active genes (Mellén et al., 2012; Lister et al., 2013) suggesting a model whereby gene transcription levels are determined by the relative density of these forms of methylation.

#### ***1.5.4 Global regulator of chromatin***

Other studies suggest a role for MeCP2 as a global regulator of chromatin structure rather than a gene specific controller of transcription. MeCP2 is extremely abundant in neurons with levels approaching that of histones and while it roughly tracks the density of methylation it also binds to non-methylated regions (Skene et al., 2010). Even in the absence of methylated DNA MeCP2 can form complexes with nucleosomes that lead to the compaction of chromatin in a manner similar to H1 (Georgel et al., 2003). Both *in vitro* and *in vivo* studies show that MeCP2 competes with histone H1 for binding sites (Nan et al., 1997; Ghosh et al., 2010a) and that H1 levels are almost doubled in neurons from KO mice (Skene et al., 2010) suggesting that the competition between the two for binding sites may determine gene transcription levels.

### **1.6 Animal Models**

#### ***1.6.1 RTT KO mouse models***

After the link between mutations in the *MECP2* gene and RTT was discovered (Amir et al., 1999), a number of KO mouse model were soon generated to enable further study of the pathogenesis of the disorder (Chen et al., 2001; Guy et al., 2001). One of the first models involved the deletion of all of exon 3 of the gene (Chen et al., 2001). Up until four weeks of age male KO mice did not differ significantly from WT, but then began to display slightly laboured breathing and subtle whole-body tremors. The phenotype continued to progress with marked hypoactivity and weight loss, and most mice died by about 10 weeks of age. This early death mirrors the lethality seen in human males born with severe *MECP2* mutations, although in terms of developmental time it is significantly delayed. Female heterozygous KO mice, which were a mosaic of WT and KO cells due to random X chromosome inactivation, appeared normal up until about four months of age, after which they displayed hypoactivity, ataxia, and gait problems. Again, the delayed onset mirrors the situation in female RTT

patients, but as in the case with males it is much later in terms of developmental time point, as the mice are fully adult by this point.

Another KO model involved the deletion of both exons 3 and 4 of the gene (Guy et al., 2001). Similar to the previous model, male KO mice appeared normal for the first 3-4 weeks before developing a severe phenotype including hypoactivity, breathing irregularity, gait abnormalities and hindlimb clasping, and died at around 8-10 weeks. Surprisingly, while mice showed severe weight loss on a C57BL/6 background the reverse was true when crossed to a 129 background, suggesting a modifier gene effect on the weight phenotype. Female mice developed normally until 3-4 months after which they began to display hindlimb clasping and hypoactivity. While about half of the mice showing strong RTT-like symptoms by nine months, including breathing dysfunction, many mice did not show any symptoms even by one year, suggesting a variable phenotype in females.

### ***1.6.2 RTT knock-in models***

In addition to full KO models, a number of knock-in mice, modelling some of the most common human RTT causing mutations, have been created. R168X is the most common RTT causing nonsense mutation and the second most common disease causing mutation overall. It is a truncating mutation that creates a premature stop codon downstream of the MBD but upstream of the TRD. KI male mice with this mutation show severe hypoactivity, gait impairment, weight loss and breathing irregularity by about 6 weeks of age and die between 12-14 weeks of age (Wegener et al., 2014), showing a similar pattern to *Mecp2*-null mice (Chen et al., 2001; Guy et al., 2001). In contrast female mice show very little phenotype, with only rotarod defects at 9 months and some tremor and hindlimb clasping detected. However, when this mutation was modelled on a different background-strain, female KI mice showed reduced bodyweight and breathing abnormalities, highlighting the effect of background strain on phenotype, at least in females.

About 10% of RTT cases are caused by mutations affecting amino acid T158, most commonly altering it to a methionine (T158M) or in some instances to an alanine (T158A). T158 is located within the MBD and is thought to be crucial in stabilising this region (Ho et al., 2008). When this mutation is modelled in KI mice, males develop a range of symptoms including hypoactivity, breathing irregularity, motor dysfunction, and impaired cognition from about four weeks of age (Goffin et al., 2012). The phenotype trajectory is milder than that of *Mecp2*-null mice (Guy et al., 2001) but by about 24 weeks of age 75% of mice are

dead. Females begin to develop a range of symptoms from about four months of age as determined by the widely used combined RTT scoring system (Guy et al., 2007; Gadalla et al., 2012; Robinson et al., 2012), although they weren't extensively phenotyped. Analysis showed that the mutation lead to greatly decreased binding to methylated DNA, and by 90 days there was a 60% decrease in protein levels suggesting a strong decrease in protein stability.

About 5% of RTT cases are due to the mutation R306C, which causes disruption in a conserved region at the end of the TRD responsible for interacting with the NCoR/SMRT corepressor complex (Lyst et al., 2013). When this mutation was modelled in mice, male KI mice showed a severe phenotype similar to that seen in *Mecp2*-null mice, with roughly half of mice being dead by 19 weeks of age (Lyst et al., 2013). This may explain the surprisingly mild phenotype observed in mice engineered to be truncated at amino acid 308 (Shahbazian et al., 2002b), as the protein generated in these mice will still include this crucial NCoR/SMRT interaction domain.

## 1.7 Reversibility

### 1.7.1 Global *MeCP2* restoration

RTT was originally considered a prototypical neurodevelopmental disorder, whose known genetic cause could help to identify a common pathophysiology in these disorders (Neul and Zoghbi, 2004; Percy and Lane, 2005). This suggested that RTT patients may have irreversible dysfunction due to the absence of MeCP2 at a critical time-window. However, a number of animal studies have since revealed that rather than being essential for development MeCP2 is required for mature neuronal function (McGraw et al., 2011), and that restoration of gene expression, even after disease onset, can almost completely reverse symptoms (Guy et al., 2007; Robinson et al., 2012).

A landmark study by Guy and colleagues in 2007 investigated the reversibility of RTT using a conditional KO model. This functional KO model was generated by inserting a lox-stop cassette (Dragatsis and Zeitlin, 2001) into intron 2 of the *Mecp2* gene (Guy et al., 2007). To enable conditional reactivation of gene transcription, a transgene containing Cre recombinase fused to a modified oestrogen receptor was combined with the *Mecp2*-stop allele. This fusion protein remained in the cytoplasm until activated by the oestrogen analogue Tamoxifen, after

which it translocated to the nucleus leading to cre-mediated deletion of the stop cassette. This extremely elegant system allowed reactivation of *Mecp2* transcription at any time-point by repeated injections of tamoxifen. Crucially the gene would be under the control of its own promoter and at its endogenous location, leading to restoration of normal function.

Phenotypically, this functional KO model was very similar to the previously described *Mecp2*-null mice (Chen et al., 2001; Guy et al., 2001), with male KO mice developing symptoms such as hypoactivity, irregular breathing, gait abnormalities, and clasping by 6 weeks of age, and dying by about 11 weeks. Female KO mice developed symptoms such as abnormal gait, hypoactivity, irregular breathing and clasping between 4 to 12 months of age but had a similar life-span to WT. In contrast, male KO mice treated with Tamoxifen before the onset of symptom did not develop any observable RTT-like symptoms and had a similar lifespan to WT. Surprisingly, even after the development of RTT symptoms, treatment with Tamoxifen led to rescue of the RTT phenotype and a lifespan approaching that of WT. Similarly, adult female KO mice treated with Tamoxifen after symptom development showed almost complete phenotype reversal, including restoration of attenuated long-term potentiation levels to WT. High levels of reactivation was achieved in both males and females after repeated injections of Tamoxifen, leading to expression of the gene in about 80% of cells. The very mild symptoms still observed in some males could perhaps be explained by the residual population of cells in which reactivation did not occur, or be due to some aspect of the disorder that leads to permanent dysfunction. This important study showed that the vast majority of the RTT phenotype can be rescued by delayed activation of the gene and suggested that the disorder did not lead to any irreparable damage.

The results of the previous study convincingly showed that the major aspects of the RTT phenotype could be rescued by reactivation of *Mecp2*, however phenotyping of the mice was mostly based on observational scoring and it is possible that more subtle and pervasive symptoms were not rescued. To investigate this, a later study carried out a comprehensive phenotype analysis including brain morphology, breathing, and locomotor function (Robinson et al., 2012). Morphological analysis showed that untreated KO mice showed reduced cortical and white matter tract thickness, reduced soma size, reduced dendritic complexity and dendritic length, and a significant reduction in spine density. However, in mice treated with Tamoxifen after symptom development, these deficits were rescued to almost WT levels. A similar pattern was seen for a wide range of physiological and behavioural phenotypes, with treatment leading to the rescue of respiratory function, balance



and coordination, and muscle strength. However, some phenotypes did not reverse including impaired performance on the forced swim test and gait abnormalities. In this study reactivation of gene expression was achieved in about 70% of cells with Tamoxifen treatment. The failure of these motor defects to reverse after treatment could suggest a particular sensitivity of motor function to absence of MeCP2 in the 30% of cells in which gene expression was not reactivated, or could indicate irreversible damage.

### ***1.7.2 Brain region specific MeCP2 restoration***

MeCP2 is widely expressed throughout the body, with significant levels of protein seen in heart, skeletal muscle, spleen, kidney, liver and lung tissue, but particularly high levels in the post-mitotic neurons of the nervous system (Shahbazian et al., 2002a; Zhou et al., 2006; Song et al., 2014). Several studies have shown that inactivation of *Mecp2* in the nervous system only leads to a phenotype that is indistinguishable from mice in which the gene has been deleted globally (Chen et al., 2001; Guy et al., 2001), suggesting that restoration of gene expression in the nervous system only should be sufficient to completely reverse the disease. This is supported by a study in which *Mecp2* was selectively expressed in post-mitotic neurons under the control of the tau promoter (Luikenhuis et al., 2004). Rescued mice showed a normalisation of bodyweight and lifespan, and activity levels indistinguishable from WT. However, only these gross markers of disease were examined, and it is possible that more subtle aspects of the disorder were not reversed. The effect of restoring *MECP2* expression in selected brain regions has also been examined using cell-specific mouse models. In males, transcription of an inducible *Mecp2* transgene under the control of the CamKII promoter led to robust protein expression in both the cortex and hippocampus (Alvarez-Saavedra et al., 2007), areas thought to be involved in the RTT phenotype (Kaufmann et al., 1997; Dani et al., 2005; Asaka et al., 2006). Surprisingly however, this was not sufficient to prevent the onset of any of the RTT phenotype. Similarly, when expression was restricted to the cerebellum and striatum using the enolase promoter, no aspect of the RTT phenotype was impacted upon (Alvarez-Saavedra et al., 2007). These results suggest that either these are not critical brain regions for the expression of the RTT phenotype, or perhaps that the levels of gene transcription driven by these promoters does not adequately recapitulate the expression pattern seen under the endogenous *Mecp2* promoter. Overall, these results suggests that the loss of *Mecp2* from the brain plays a crucial role in the RTT phenotype, however future studies should comprehensively examine the full extent of disease reversal in brain specific rescues models.

### 1.7.3 Glia and microglia reversal

Previous studies have shown that astrocytes with mutated *Mecp2* could have a non-cell autonomous negative impact on the function and morphology of surrounding neurons (Ballas et al., 2009; Maezawa et al., 2009). To examine the role of astrocytes in RTT, a mouse model was generated which enabled delayed activation of *Mecp2* in astrocytes only, using a tamoxifen inducible cre recombinase under the control of the human astrocytic fibrillary acidic protein (hGFAP) promoter (Liroy et al., 2011). In KO mice, tamoxifen treatment led to a partial restoration of activity levels, a normalisation of breathing, an increase in dendritic complexity, and a dramatically increased life span, with over 80% of mice still alive at 7.5 months. To further investigate this an astrocyte specific KO mice model was generated. While mice showed normal activity levels, dendritic complexity, and lifespans, they did display weight loss, hindlimb clasping and breathing irregularity, suggesting that some aspects of the disorder are impacted by an absence of MeCP2 from glia, and that these cells have to be considered when targeting any future therapies.

To further examine the role of glia in RTT pathology, a recent study investigated the effect of restoring WT microglia to the brain of RTT KO mice (Derecki et al., 2012). To achieve this, the brains of KO mice were irradiated at the time of RTT symptom onset to destroy the resident brain microglia population. Mice were injected with bone marrow extracted from WT animals. Robust detection of a GFP reporter confirmed that this led to repopulation of the brain with WT microglia. Males treated with WT bone marrow showed a significantly increased lifespan (with some still alive at 48 weeks), as well as a normalisation of bodyweight, reduction in apnoeas and breathing irregularity, and improved activity levels compared to mice treated with KO bone marrow. Improvements were also seen in female mice, although improvements in breathing and motor function were very modest. This result was surprising and exciting, particularly as bone marrow transplantation is already a well-established technique, so a multisite effort was made to replicate it (Wang et al., 2015). However, in contrast to the previous study, the results could not be replicated in any of the three RTT models tested. Replacement of mutant microglia with WT did not lead to any extension of lifespan or improvement in weight, apnoeas, activity levels, or gait. This study used much greater numbers and random assignment to treatment groups, and used mice from the same colony as the previous report, as well as two other RTT models, in three independent labs. In addition, when they used a cre line to specifically express *Mecp2* in microglia developmentally there was no extension of lifespan and the phenotype was not

improved suggesting that the presence of WT microglia does not in fact ameliorate the RTT phenotype.

## 1.8 Gene-based therapies

The proven reversibility and monogenic nature of RTT (Guy et al., 2007; Robinson et al., 2012) makes it an attractive target for gene based therapies, particularly since the previously described uncertainty over the functions and targets of MeCP2 makes it difficult to identify any downstream pathways that could be modified pharmacologically.

### 1.8.1 *Unsilencing the inactivated X chromosome*

In females, one copy of the X chromosome is randomly selected for silencing early in development during a process called X chromosome inactivation (XCI). This is thought to provide dosage compensation by preventing the overexpression of X-linked genes in females compared to males (Lyon, 1989). For this reason, RTT patients will have a mixture of cells, with some expressing the WT *MECP2* allele and some the mutant copy. One possible approach to treating RTT would be to unsilence the inactivated chromosome in mutant cells, thus restoring WT MeCP2 protein levels in those cells.

Loss of function mutations in the ubiquitin protein ligase E3A (*UBE3A*) gene causes the severe neurological disorder Angelman syndrome (Kishino et al., 1997). Disease causing mutations occur in the maternal allele because the gene is imprinted, with the paternal allele being permanently silenced by epigenetic mechanisms (Rougeulle et al., 1997). A recent study showed that topotecan, a topoisomerase II inhibitor approved for use in the treatment of certain cancers, could unsilence the paternal allele in a mouse model of Angelman, leading to elevated levels of *UBE3A* mRNA (Huang et al., 2012). While further work is needed to examine the off-target effects on gene expression and long-term efficacy, the demonstration that a well-established and chronically tolerated drug can lead to unsilencing of a target gene suggests that this kind of approach could be relevant for RTT.

XCI relies on *Xist*, a non-coding RNA which coats the inactive chromosome and is essential for both initiation and maintenance of silencing (Kathrin Plath et al., 2002). *Xist* KO mice do not carry out XCI and show embryonic lethality (Marahrens et al., 1997), suggesting that XCI is essential for viability, possibly due to a failure of dosage compensation. If correct, strategies aimed at unsilencing the X chromosome to restore MeCP2 levels would be

deleterious. However, a recent study employing an RNA screening strategy identified 13 genes which prevented XCI when their transcription was blocked (Bhatnagar et al., 2014). Remarkably, a KO mouse model of one of these genes, *STC1*, displayed no abnormal phenotype, despite XCI not occurring in these mice. When gene expression was examined, no upregulation of X-linked gene expression was seen, suggesting the presence of other compensatory measures. However, when the gene was silenced in mouse fibroblasts using shRNA, there was a two-fold increase in mRNA levels, suggesting either that switching off the gene in mature animals leads to altered gene expression or that the compensatory effects take time to develop. Overall, the study provides evidence that inactivating genes involved in XCI, either pharmacologically or by gene therapy, could be a possible treatment strategy for RTT, although confirmation in human derived cells is essential to determine its clinical relevance.

### ***1.8.2 Gene therapy***

Since restoration of WT *Mecp2* leads to a rescue of the RTT phenotype (Guy et al., 2007), one possible therapeutic strategy is to deliver exogenous WT copies of the gene to affected cells using viral vectors. Recent studies have shown that the severity of the RTT phenotype is determined by the proportion of cells expressing WT copies of *Mecp2* (Robinson et al., 2012), and that brain-wide protein expression is required for robust phenotype reversal (Alvarez-Saavedra et al., 2007; Guy et al., 2007). In addition, overexpression of *MECP2*, as seen in duplication patients as well as neuronal overexpression models, leads to a spectrum of motor and cognitive impairments (Luikenhuis et al., 2004; Van Esch et al., 2005; Lugtenberg et al., 2009; Ramocki et al., 2009). These studies suggest that effective gene therapy approaches will need to deliver WT copies of *MECP2* to a large proportion of cells throughout the brain, and that transcription levels must be controlled to prevent toxicity. This is particularly challenging in RTT because random XCI means that half the cells already express WT levels of protein, and are thus particularly vulnerable to overexpression toxicity when transduced with an exogenous transgene.

The earliest gene therapy study for RTT used adenovirus to target a WT *Mecp2* construct to the striatum of KO male mice (Kosai et al., 2005). Symptomatic animals injected with the virus showed a marked improvement in activity levels and motor coordination, showing as a proof of principle that exogenous delivery of functional copy of the gene could impact on the phenotype. In a later study, lentivirus was used to target an *Mecp2* transgene to neurons in

culture (Rastegar et al., 2009), leading to a normalisation of dendritic arborisation. However, Gadalla and colleagues carried out the first comprehensive trial of gene therapy for RTT in a KO mouse model (Gadalla et al., 2012). As a delivery vector they used adeno-associated virus serotype 9 (AAV9). AAV vectors have a number of advantages including wide tropism, low immune response, stable transgene expression, and the ability to target both dividing and non-dividing cells (McCarty, 2008; Samulski and Muzyczka, 2014), but are limited by a cloning capacity of < 5 kb (Wu et al., 2010; Choi et al., 2014), making them unsuitable for the packaging of large constructs. The AAV9 serotype used in this study has a number of particular advantages including high levels of transgene expression, a rapid onset, high neuronal tropism, and the ability to cross the blood brain barrier, enabling IV viral delivery (Foust et al., 2008; Zincarelli et al., 2008; Duque et al., 2009). Direct brain injection of an *Mecp2* construct, under the control of the chicken beta actin (CBA promoter), in neonatal male KO mice led to an almost doubling of lifespan, from a median of 9.3 to 16.6 weeks, with some mice surviving past 38 weeks. In addition, a number of phenotypes showed significant improvement as compared to GFP injected control, including activity levels and motor function. Transduction efficiencies ranged from about 5% in the striatum to almost 40% in the hypothalamus. The markedly increased lifespan in some of the mice suggest that even relatively low levels of transduction can have a significant impact on phenotype. Symptomatic mice injected with the *Mecp2* construct had a much more modest survival benefit, but this was due to an extremely low transduction efficiency in the brain (2-4% of neurons).

Following on from this promising study, a later study tested a similar approach in a female heterozygous KO mouse model (Garg et al., 2013). Again, AAV9 was used to deliver an *Mecp2* construct, this time under the control of the endogenous *Mecp2* promoter. In comparison to the previous study, robust brain transduction was achieved after systemic delivery, with efficiencies of 20-25% in the cortex, hippocampus, and brain stem. Transduction efficiency in the cerebellum was much lower however, with only 7% of cells transduced. KO mice before injection both showed a mild RTT phenotype as assessed by the commonly used RTT scoring system. Mice were then assessed weekly over a period of 25 weeks. Mice that were injected with the *Mecp2* construct showed a reversal of the phenotype over this period, and were almost indistinguishable from WT at the end of the study. In contrast, mice that were injected with control virus showed a steadily worsening phenotype and by the end of the study they showed a severe RTT phenotype. This was confirmed in a

series of behavioural tests which showed robust recovery in motor and cognitive function, although again breathing appeared to be more resistant to reversal.

### ***1.8.3 Problems associated with traditional gene therapy approaches***

While gene therapy has shown promise in the treatment of RTT, there are several disadvantages that could prevent it being a viable long-term option in human patients.

Transgenes delivered using AAV vectors are usually maintained as an extra-chromosomal episome, rather than being integrated into the genome (Penaud-Budloo et al., 2008). This is advantageous as it prevents the insertional mutagenesis seen with randomly integrating viral vectors such as lentivirus, but it is unclear how long expression of the transgene will be maintained in this episomal state. In a primate study, transgene mRNA was still detected six years after a single dose of AAV2 (Bankiewicz et al., 2006), and in a recent clinical trial using AAV8 transgene mRNA was still detected after two years (Nathwani et al., 2011).

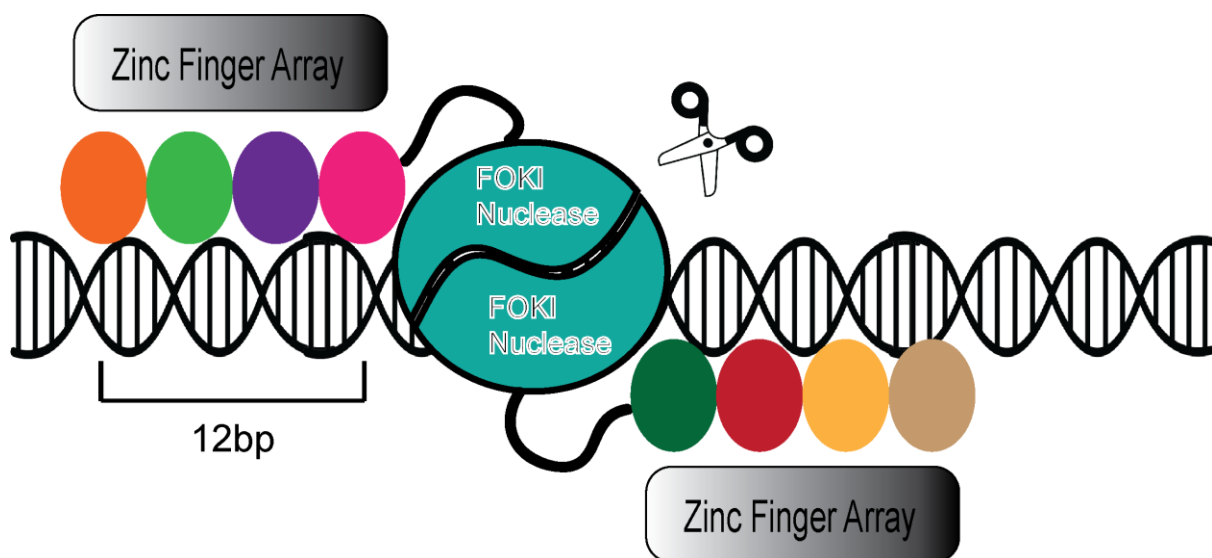
While this suggests that relatively long-term expression is achievable, it is not currently known if AAV delivery will be capable of achieving life-long transgene expression in humans, which will be necessary for the treatment of RTT. In addition, repeated doses may not be possible without immune system silencing, as patients have been shown to develop an immunological response after initial exposure to AAV particles (Mingozzi and High, 2013; Corti et al., 2014). The lack of integration also means that the transgene will not be replicated during cell division. In dividing cells such as glia, which have been shown to contribute to the RTT phenotype (Lioy et al., 2011), this means that the transgene will be quickly diluted in the cell population.

Another crucial issue for RTT gene therapy approaches is the danger of overexpression toxicity. As has been previously described, duplication of the *MECP2* gene leads to a severe neurological phenotype (Ramocki et al., 2010). Since roughly 50% of cells in RTT patients express the WT *MECP2* allele, any WT cells receiving a copy of the transgene will overexpress *MECP2*. In addition, cells can receive multiple copies of the transgene thus exacerbating the toxicity. This is particularly problematic when the virus is delivered systemically, as the liver shows very high levels of transduction and transgene expression after *intravenous* injection of AAV9 (Gadalla et al., 2012). Since liver cells normally express *MECP2* at a very low level (Shahbazian et al., 2002a; Zhou et al., 2006) this large increase in gene expression could be detrimental. At present there is no way to discriminate between WT and mutant cells with viral delivery vectors.

## 1.9 Genome Editing

In the last few years a number of genome editing tools have emerged that have greatly enhanced our ability to make targeted sequence changes to the genome. These tools, such as zing-fingered nucleases (Urnov et al., 2010), transcription activator-like effector nucleases (TALENs) (Bogdanove and Voytas, 2011; Scharenberg et al., 2013), and clustered regularly interspaced short palindromic repeat (CRISPR)-associated nuclease Cas9 (Hsu et al., 2014), are all based on programmable nucleases that allow targeted double-stranded DNA breaks to be made at specific genomic locations. As the cell repairs these breaks various genetic changes can be introduced into the genome. The major advantage of these tools over current gene therapy strategies is that any changes made to the genome are permanent and thus passed on to daughter cells. Genome editing has revolutionised our ability to quickly and easily generate new animal models (Wang et al., 2013) and to probe gene function (Swiech et al., 2014) and has the potential to be a powerful therapeutic tool (Perez et al., 2008; Schwank et al., 2013; Wang et al., 2014a). A number of different programmable nucleases have been developed to enable genome editing, each with their own advantages and disadvantages.

### 1.9.1 Zinc-finger nucleases (ZFN)



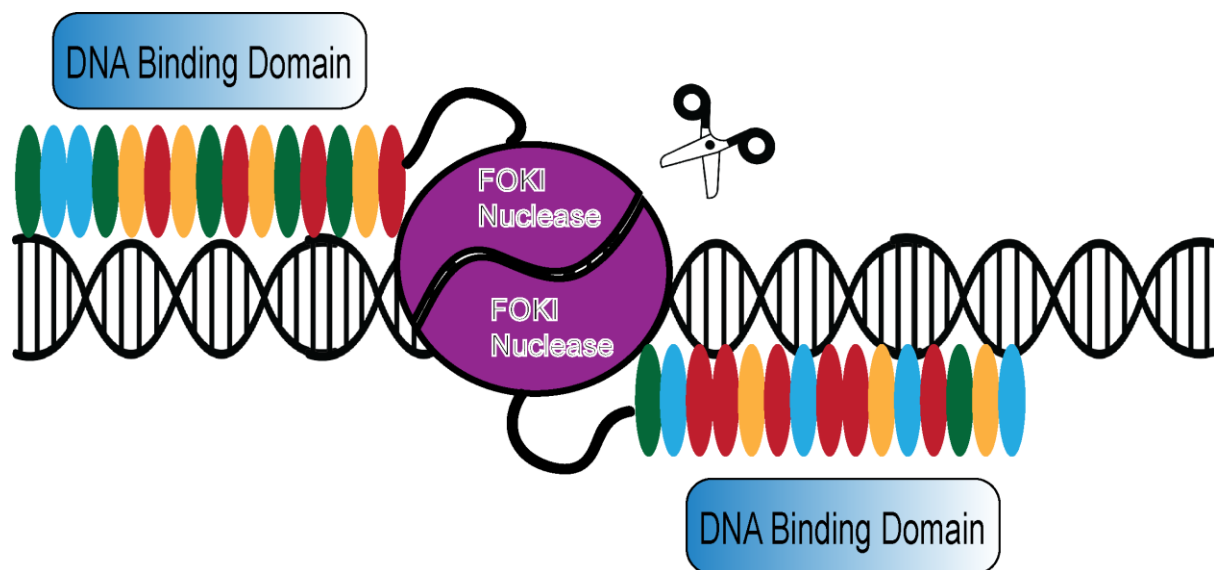
**Figure 1.2 – ZFN structure and function**

ZFNs are composed of a series of zinc-fingers (multi-coloured circles) arranged in a modular fashion to form an array which recognises a sequence of 9-18 bp. DNA cleavage is dependent on the dimerization of two FokI nuclease domains (large blue circle) and is achieved by the binding of two different ZFN monomers to targets on opposing DNA strands.

ZFNs were the first widely used genome editing tool (Kim et al., 1996; Bibikova et al., 2001) and have been used to modify genes in a wide array of organisms including bacteria, plants, and mammals such as mice and pigs (Urnov et al., 2010; Perez-Pinera et al., 2012). A ZFN is composed of two domains (**Figure 1.2**), a zinc-finger protein domain which is responsible for DNA targeting, and a FokI nuclease which cleaves DNA (Kim et al., 1996). Each individual zinc-finger recognises a specific 3 bp DNA sequence, and ZFNs are composed of a series of zinc-fingers arranged in a modular fashion to recognise a sequence of 9-18 bp (Wolfe et al., 2000 p.200). DNA cleavage is dependent on the dimerization of two FokI nuclease domains (Bitinaite et al., 1998). This is achieved by the binding of two different ZFN monomers to targets on opposing DNA strands, with a 5-7 bp spacer region between each monomer. The requirement for dimerization enhances the specificity of ZFNs as it increases the length of the target site thus greatly reducing off-target activity. However, studies have shown that as well as forming heterodimers, ZFNs utilising the WT FokI protein can also form homodimers, leading to off-target toxicity (Kelly Beumer et al., 2006). To overcome this, two mutant variants of the FokI nuclease have been generated that only function when they bind as a heterodimer pair, greatly reducing the levels of off-target cleavage (Miller et al., 2007; Szczeppek et al., 2007). While zinc finger domains for every possible 3bp sequence have been engineered (Mandell and Barbas, 2006; Kim et al., 2009), the binding and cleavage efficiency of ZFNs can be highly unpredictable, with some individual zinc fingers modules only binding reliably in certain sequence contexts (Carroll, 2012), resulting in an average of only one functional pair being obtained for roughly every 100 bp of genomic sequence (Kim et al., 2009). In addition, the synthesis of reliable and functional ZFNs can be extremely challenging. Relatively simple modular approaches have been developed (Kim et al., 2010) but suffer from a low success rate (Ramirez et al., 2008; Kim et al., 2009). More robust methods have subsequently been developed but these can be labour intensive and require significant expertise (Maeder et al., 2008). However, several studies have recently reported methods to improve the ease and success rate of synthesis (Sander et al., 2011; Bhakta et al., 2013).



### 1.9.2 Transcription activator-like effector nucleases (TALEN)



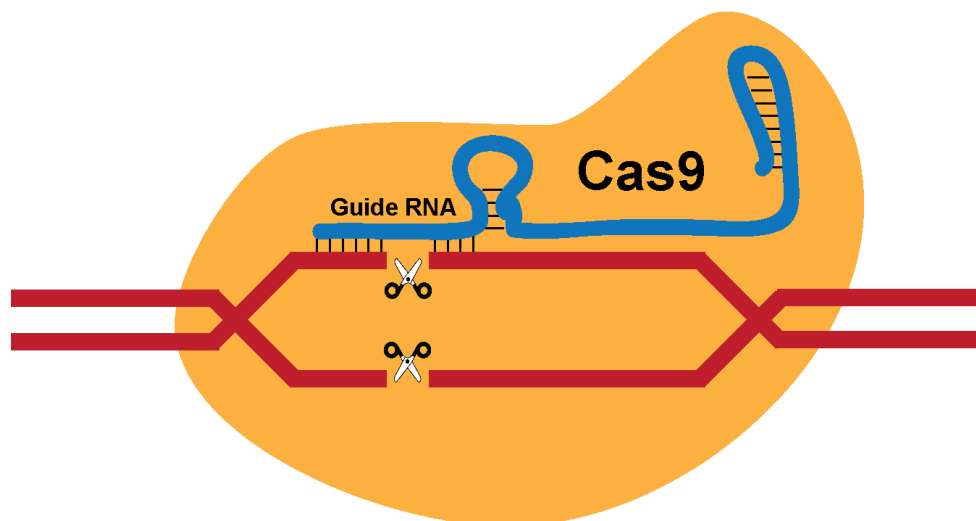
**Figure 1.3 – TALEN structure and function**

TALEN arms recognition sites are composed of a sequential array of 33-35 amino acid repeats (multi-coloured ellipses). DNA cleavage is dependent on the dimerization of two FokI nuclease domains (large purple circle) and is achieved by the binding of two different TALEN arms to targets on opposing DNA strands.

Transcription activator-like effectors (TALEs) are naturally occurring bacterial proteins from the plant pathogen *Xanthomonas* (Bogdanove et al., 2010). They contain a highly specific DNA binding domain which allows them to bind target sequences in the plant cell nucleus and alter gene transcription levels. This domain is composed of a sequential array of 33-35 amino acid repeats (**Figure 1.3**), each of which binds a single DNA base (Boch et al., 2009; Moscou and Bogdanove, 2009). The sequence of each repeat is almost identical except for amino acids 12 and 13, known as the repeat-variable diresidue (RVD), which are variable and determine the nucleotide specificity. Almost 20 different RVDs have been discovered but only four of these are widely used, each targeting a different DNA base (Moscou and Bogdanove, 2009). The discovery of this simple recognition code led to the rapid adoption of the TALE DNA binding domain as an alternative to the zinc-finger motifs used in ZFNs (Christian et al., 2010; Li et al., 2011b), with the targeting domain being fused to the FokI enzyme to create a highly programmable TALE nuclease (TALEN). TALENs have since been used to modify genes in a wide array of organisms including plants, viruses and

mammals (Joung and Sander, 2013; Sun and Zhao, 2013). Similar to ZFNs, DNA cleavage requires dimerization of two FokI nuclease domains (Joung and Sander, 2013), and this is achieved by the binding of two different TALEN arms to target sequences on opposing DNA strands, with cleavage occurring in a 15 bp spacer region between the two binding sites. The simple DNA recognition code makes it easy to design TALENs targeted to almost any DNA sequence giving them a significant advantage over ZFNs, particularly if highly specific targeting is required. However, homology between the amino acid repeats can make the synthesis of TALENs challenging, due to recombination within cells (Holters et al., 2013). To overcome this, methods based on the ‘Golden Gate’ cloning system have been developed which allow the rapid assembly of TALEN arm constructs in as little as a week (Cermak et al., 2011).

### 1.9.3 CRISPR-Cas9



**Figure 1.4 – CRISPR-Cas9 structure and function**

In the CRISPR-Cas9 system a guide RNA (blue) binds to a specific sequence in the genome (red) and guides the Cas9 nuclease protein (yellow) to the target site. The Cas9 enzyme then generates a double strand break within the target region.

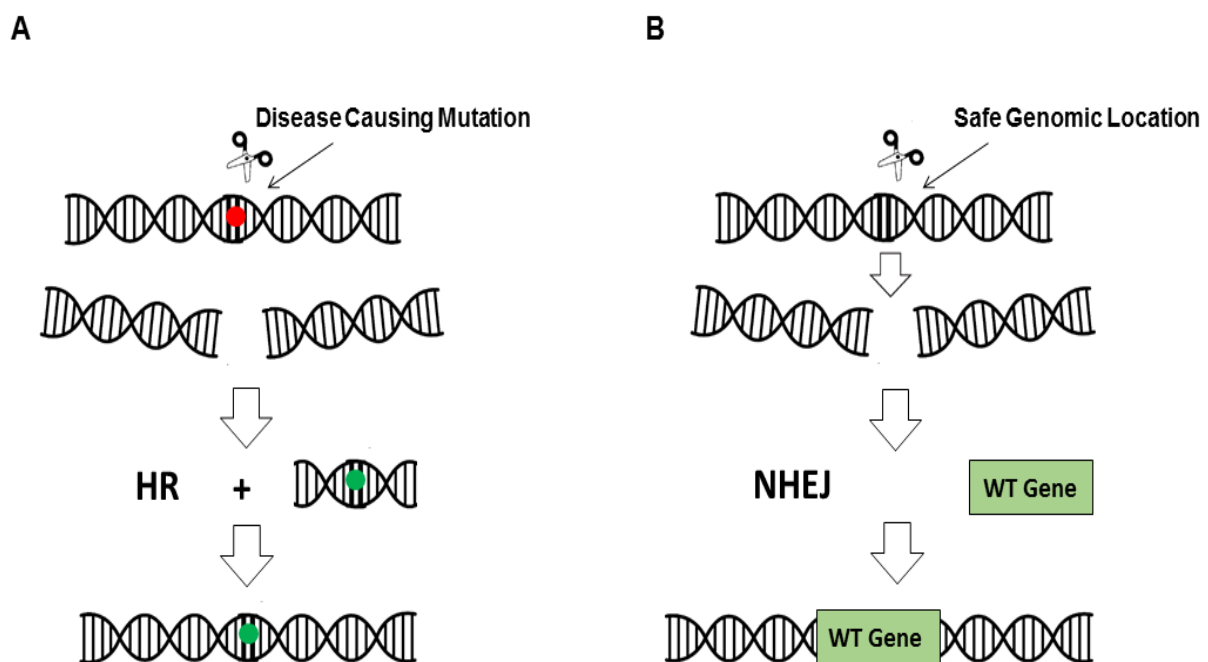
Perhaps the most promising of all the genome editing tools to emerge is the RNA-guided clustered, regularly interspaced, short palindromic repeat (CRISPR) type II system, which is likely to become the dominant tool of choice due to its efficiency, ease of synthesis, and ability to target multiple targets simultaneously (Prashant Mali et al., 2013; Sakuma et al.,

2014). In bacteria, CRISPR along with its associated CRISPR-associated (Cas) proteins provides a form of adaptive immunity against invading viral DNA (Sorek et al., 2008; Philippe Horvath and Rodolphe Barrangou, 2010). A 20 bp fragment of the invading DNA, called a protospacer, is inserted into the bacterial genome to form a CRISPR. These are then transcribed and processed to produce a target specific CRISPR RNA (crRNA). Processing is dependent on a target-independent *trans*-activating crRNA (tracrRNA) which hybridises with the crRNA (Deltcheva et al., 2011). These hybridised RNAs then form a complex with the Cas9 nuclease (Martin Jinek et al., 2012), which is guided to a target DNA sequence by the 20 bp protospacer portion of the crRNA (**Figure 1.4**). This leads to cleavage of the target DNA if the 20 bp target sequence is directly adjacent to a short protospacer adjacent motif (PAM), 5'-NGG in its canonical form, which is recognised by the Cas9 protein itself. After the discovery of this simple DNA targeting system it was quickly realised that it could be adapted for use as a powerful genome editing tool (Martin Jinek et al., 2012). To simplify the synthesis of custom target sequences a single guide RNA (sgRNA) was engineered that combined the crRNA and tracrRNA into a single RNA chimera. The sgRNA is composed of a 20 bp sequence at the 5'-end that determines the target sequence and a hairpin structure at the 3'-end that binds with Cas9. This gives the CRISPR-Cas9 system a huge advantage over both TALENs and ZFNs as changing DNA targets is as simple as replacing the 20 bp sequence at the 5'-end, and does not require any complicated cloning or protein engineering. The target sequence can be easily synthesised as complementary oligonucleotides and cloned into a vector containing the rest of the sgRNA and the Cas9 protein in a single step. It is this simplicity that has led to the rapid adoption of CRISPR-Cas9 technology, with over 2000 CRISPR papers published since 2013 alone. An additional advantage of this system is that multiple guide sequences can be cloned into a single vector allowing parallel editing of several target sites (Cong et al., 2013). This has been used to introduce mutations in up to five different genes in rat, mouse, and zebrafish simultaneously (Jao et al., 2013; Li et al., 2013a; Wang et al., 2013).

### ***1.9.4 DNA repair pathways***

Cells with double stranded breaks use two major pathways to repair damage. The choice of pathway is determined by cell cycle state and the presence of a homologous DNA repair template, such as a sister chromatid. Homologous recombination (HR) is a precise repair pathway that uses a donor template to produce high fidelity repair. The HR pathway can be utilised to introduce targeted sequence changes into the genome using an exogenous DNA

template with homology for the DNA break site (Bibikova et al., 2001; Krejci et al., 2012). If this exogenous template is used during HR repair, any base changes in the template will be incorporated into the genome. This allows for precise genetic changes to be made, including the repair of loss of function mutations (**Figure 1.5-A**). In contrast, the non-homologous end joining (NHEJ) pathway does not require a template but can repair the break by joining the two broken ends together. This usually leads to small deletions at the break site which can lead to loss of gene function (Lieber, 2010). This ability to precisely disrupt gene function has dramatically improved the ease and efficiency of generating animal models of disease (Yang, 2013; Wang et al., 2013).



**Figure 1.5 – Repair strategies using alternative DNA repair pathways**

Double stranded DNA breaks can be repaired using two major repair pathways, homologous recombination (HR) and non-homologous end joining (NHEJ), both of which can be utilised for genome editing. **(A)** In cells in which the HR pathways is active, mutated bases (red) can be replaced with WT bases (green) by recombination with an exogenously supplied repair construct containing homology to the target region. **(B)** In non-dividing cells, a WT copy of the mutated gene (green box) can be permanently ligated into a safe location in the genome using the NHEJ repair pathway, leading to a restoration of gene function.

The NHEJ pathway is thought to be active throughout the cell cycle and is therefore suitable for targeting both dividing and non-dividing cells (Rothkamm et al., 2003; Sharma, 2007). In

contrast, the HR pathway is mostly limited to dividing cells, due to the requirement for a homologous repair template, and is not active in post-mitotic cells such as neurons (San Filippo et al., 2008; Jeppesen et al., 2011). This low level of HR makes precise genetic repair in these cell types particularly challenging and is a major impediment to the use of this technology for loss-of-function nervous system disorders like RTT. One possible solution is to ligate a therapeutic transgene into a defined genomic locus, such as previously identified “safe harbour” sites (**Figure 1.5-B**; Moehle et al., 2007; Barzel et al., 2015). These are sites in the genome that can be disrupted without causing any detectable negative phenotype. Several recent studies have shown that transgenes can be ligated into the genome using the highly active NHEJ pathway, and has been demonstrated using ZFNs, TALENs, and CRISPR-Cas9 (Maresca et al., 2013; Nakade et al., 2014). This has the advantage over traditional gene therapy in that it allows permanent integration of the transgene without the risk of insertional mutagenesis, but does not deal with the issue of overexpression toxicity in cells already expressing the WT allele that is particularly pertinent to RTT. Novel strategies are required to ensure cells only express one functional copy of the *MECP2* gene.

### ***1.9.5 Ex vivo genome editing***

Currently, the most promising genome editing strategies involve *ex-vivo* manipulation of cells. In this strategy target cells are removed from the body, edited, and then transplanted back into the host. The advantage of *ex vivo* therapy is that much higher levels of editing can be achieved due to the availability of a number of efficient transfection systems including electroporation, lipid reagents, and viral vectors. The potential of this approach has been demonstrated for several diseases including HIV and SCID (Perez et al., 2008; Genovese et al., 2014). The strategy for HIV is based on the fact that people with mutations in the viral co-receptor *CCR5* show very high resistance to HIV infection (Liu et al., 1996). Individuals with this mutation appear otherwise completely healthy suggesting that introducing this mutation into the T-cells of affected HIV patients could be an effective treatment strategy. This is supported by the fact that a patient receiving a bone marrow transplant from individuals with this *CCR5* mutation showed undetectable levels of virus and a cessation of the need for antiretroviral therapy (Hütter et al., 2009). Indeed, even 20 months after the transplant viral levels remained undetectable. Since finding immunologically matched *CCR5* mutant donors is not practical for the majority of HIV patients, genome editing provides an alternative approach. Several pre-clinical studies have shown that ZFNs can be used to induce loss-of-function mutations in the *Ccr5* gene of HIV mouse models (Perez et al., 2008; Holt et

al., 2010; Maier et al., 2013). When these modified CD4<sup>+</sup> T cells were transplanted back into mice they led to a reduction in viral load and improved T cell counts. In addition, successfully modified cells were positively selected for due to their HIV resistance allowing them to expand relative to unmodified cells. This strategy has now undergone several clinical trials including a completed Phase I/II study (Clinicaltrials.gov NCT 01252641) which assessed the safety and tolerability of the treatment. Promisingly, treatment led to a significant increase in T cell counts and edited cells showed a survival advantage over non-modified cells (Tebas et al., 2014). A similar strategy has also been used for the correction of mutations in the *IL2RG* gene that cause the immunodeficiency disorder SCID-X1 (Genovese et al., 2014). Target haematopoietic stem cells from human donors were modified using ZFNs and transplanted into mice. Modified cells were successfully engrafted and showed multilineage differentiation. This suggests that this strategy could be appropriate for a wide range of immunodeficiency disorders. However, while *ex vivo* promising has great therapeutic potential for this type of disorder, it is mostly limited to diseases involving adult stem cell populations, as it requires cells to survive outside the body and be amenable to external manipulation and culturing, making it unsuitable for the majority of genetic disorders.

### ***1.9.6 In vivo genome editing***

For diseases involving cells unsuitable for *ex vivo* editing, or those involving multiple organs, it is necessary to edit cells directly *in vivo*. Similar to traditional gene therapy, this usually involves the delivery of genome editing constructs using viral vectors. This makes *in vivo* editing significantly more challenging than *ex vivo* approaches, as these viral vectors tend to be inefficient compared to the more robust *ex vivo* methods such as electroporation. For this reason, this approach has most successfully been demonstrated in diseases in which significant therapeutic effects can be achieved with low levels of gene restoration, or where edited cells have a significant survival advantage over non-edited cells (Li et al., 2011a; Yin et al., 2014). Patients with severe haemophilia B have levels of clotting factor IX that are less than 1% of normal. However, when levels are restored to only 5% of normal the disease becomes a much milder form, suggesting that a small increase in factor IX can have a large effect (High, 2002). In a recent study, the liver of a haemophilia B mouse model was targeted with a factor IX gene specific ZFN along with a repair donor template, using AAV2 (Li et al., 2011a). This led to significant levels of repair construct insertion via the HR repair pathway, with circulating levels of factor IX increased to 3-7% of normal, and this was sufficient for a

normalisation of clotting times. In a more recent study, CRISPR-Cas9 was used to correct mutations in the *Fah* gene of a mouse model of hereditary tyrosinaemia (Yin et al., 2014). This gene codes for the final enzyme in the tyrosine catabolic pathway, and loss-of-function *Fah* mutations lead to the build-up of toxic metabolites and severe liver damage (Paulk et al., 2010). *In vivo* delivery of Cas9, sgRNA, and an ssDNA repair template to the liver, using hydrodynamic injection, led to correction in about 1 in 250 cells (Yin et al., 2014). Despite the relative inefficiency of initial correction, when mouse livers from treated animals were examined 33 days after treatment there was a huge increase in the number of *Fah*<sup>+</sup> cells, suggesting that corrected cells had a strong survival advantage compared to non-edited cells, allowing them to expand quickly as a proportion of the population due to positive selection. While these studies strongly suggest that *in vivo* genome editing is a viable strategy for treating genetic disorders, these particular diseases have features that make them easier to target, including their origin in dividing cells of the liver, which is particularly amenable to both viral and non-viral targeting, and the requirement for only low levels of repair to achieve significant disease reversal. Expanding this approach to treat diseases involving non-dividing cells in organs such as the brain will present a much greater challenge.

### ***1.9.7 Off-target effects***

In order for genome editing to become a viable clinical strategy for severe neurological disorders like RTT several challenges will have to be overcome, the most crucial of which is specificity. Nuclease induced genetic changes are permanent and therefore any off-target editing, particularly those leading to oncogenic mutations, could have devastating effects. As previously described, the specificity of ZFNs was greatly increased by the engineering of an obligate heterodimer version of the FokI nuclease which prevents the formation of unwanted homodimers and subsequent off-target cleavage (Miller et al., 2007; Szczepek et al., 2007). However, in a recent study using a newly developed *in vitro* analysis method, a ZFN pair targeting *CCR5*, which has been developed as an HIV therapy, identified 37 off-target cleavage sites, ten of which were shown to be actually modified when these sites were sequenced from ZFN transfected cultured human cells (Pattanayak et al., 2011). Another method, using the capture of integrase-defective lentiviral vectors at NHEJ induced DSBs, enabled an unbiased genome wide analysis of off-target cleavage and showed a number of non-overlapping off-target cleavages from the same ZFN pair in the same cell line (Gabriel et al., 2011). Crucially, these sites were not identified by the commonly used *in silico* prediction methods used to identify off-target activity, suggesting these do not have predictive utility.

In contrast, TALEN pairs targeting the same loci as a corresponding ZFN have been reported to have greatly reduced off-target activity (Mussolino et al., 2011). However, in a study examining the off-target effects of two TALEN pairs, 76 putative off-target sites were identified using *in vitro* selection, 16 of which showed off-target cleavage in a TALEN transfected human cell line when assessed by high-throughput sequencing (Guilinger et al., 2014). These sites were modified in 0.03-2.3% of cells and were more prominent in TALENs with more repeats, which they suggest was due to excess binding energy. When they reduced this binding energy using amino acid substitution they showed a tenfold decrease in off-target effects. Interestingly, *in vivo* editing in mouse embryos using TALENs appears to show almost no off-target effects, with several different studies reporting no off-target cleavage in predicted off-target sites (Panda et al., 2013; Sung et al., 2013). This is likely due to the fact that in mouse embryos TALENs are delivered as short acting RNA molecules, whereas highly expressing and stable plasmid DNA is used in cell lines. This suggests that off-target cleavage could be greatly reduced through the use of short-acting nucleases and modifications to reduce excess binding energy.

Initial studies characterising the off-target activity of CRISPR-Cas9 showed that a large number of sgRNA showed high levels of off-target activity, with guides containing mismatches of up to 5 bp capable of causing off-target cleavage in a human cell line (Cradick et al., 2013; Fu et al., 2013). In some cases, off-target activity occurred at levels exceeding those of on-target activity. This suggested that CRISPR-Cas9 could be highly non-specific and that the off-target activity varied depending on the particular sequence being targeted. Other studies reported more modest off-target effects, but showed that specificity decreased with highly active guides and with increasing concentrations of sgRNA: Cas9 (Hsu et al., 2013; Pattanayak et al., 2013). To try and reduce this off-target activity several strategies have been developed including the use of truncated guide sequences and modifications to the Cas9 enzyme (Fu et al., 2014; Shen et al., 2014). A recent study has shown that reducing the length of the complementary sgRNA region to 17-18 nt can reduce off-target cleavage by more than 5000-fold, without a substantial reduction in on-target activity (Fu et al., 2014). They hypothesised that by reducing the sgRNA-DNA interface, nucleotides at the 5'-end would not compensate for mismatches at other sites along the interface. To further increase specificity a modified nickase version of Cas9 has been engineered that converts it to a single-strand DNA cleaving enzyme (Shen et al., 2014). In order to induce a double strand break, a pair of nickase sgRNA-Cas9 constructs must generate two single strand nicks closely



adjacent to one another. The requirement for two different cas9 constructs to bind closely to one another greatly reduces off-target cleavage. In support of the effectiveness of these strategies, newly developed, highly sensitive, and unbiased methods, utilising next generation sequencing have shown that the use of truncated guides leads to substantially reduced off-target effects, and that the use of nickase variants can effectively eliminate unwanted off-target cleavage activity (Frock et al., 2015; Wang et al., 2015), suggesting that with careful guide design and the use of suitable Cas9 variants, CRISPR-Cas9 can enable highly precise genome editing.

### ***1.9.8 In vivo delivery***

Similar to gene therapy, the most promising method of delivery for genome editing tools is via viral vectors, in particular AAV. However, one major challenge for this approach is the limited packaging capacity of these vectors (Wu et al., 2010). In addition to delivering a programmable nuclease, precise modification also requires the delivery of a repair construct to act as a template for homologous mediated repair. ZFNs are relatively small and both arms of a ZFN pair can fit comfortably inside a single AAV vector. However, TALENs are significantly larger and each arm will need to be packaged inside a separate vector. This means that multiple plasmids will have to be delivered simultaneously to cells, significantly reducing the likely number of successfully edited cells. In the most commonly used CRISPR-Cas9 system, derived from the *Streptococcus pyogenes*, the Cas9 coding sequence alone is > 4 kb, and when combined with an average sized promoter and polyadenylation signal would exceed 5.3 kb of DNA (Fine et al., 2015), greater than the 4.8 kb packaging limit of AAV (Wu et al., 2010). Functional constructs can be produced containing minimal promoter and polyadenylation sequences (Swiech et al., 2014), but this puts severe limits on the choice of regulatory units and still requires the delivery of a second viral vector containing the sgRNA. One alternative is a *trans*-splice approach, in which the coding sequence of Cas9 is split across two vectors, with the two fragments being expressed separately in the cell and then spliced together to form a functional protein (Fine et al., 2015). However, this is at the expense of a three-fold decrease in activity. More promising is the use of an alternative version of Cas9 from another bacterial species. A very recent study has shown that Cas9 from *Staphylococcus aureus* is more than 1 kb shorter than that from *Staphylococcus pyogenes*, but is still capable of achieving high levels of cleavage (Ran et al., 2015). In addition, the sgRNA and Cas9 sequence can be combined in a single AAV vector, allowing for high efficiency transduction and editing.

As previously described, off-target activity can be reduced by the use of short-acting nucleases, such as delivery of components in the form of RNA, rather than the long-term protein expression provided by viral vectors (Panda et al., 2013; Sung et al., 2013). One approach is to limit transgene transcription using an inducible system such as doxycycline (Dow et al., 2015). However, a more attractive approach would be to deliver short-lived mRNA or protein products (Kormann et al., 2011; Zuris et al., 2015), which would allow much greater control over dosage. As yet, no plausible alternative to viral vectors has been developed for robust brain delivery, however a recent study has shown the potential of exosomes, endogenously derived nano-vesicles that can transport both RNA and proteins (Alvarez-Erviti et al., 2011). When these particles were fused with a neuron-specific peptide they were capable of producing widespread delivery of an siRNA molecule in the brain. This technology has the potential to replace viral vectors as the delivery method of choice and would greatly reduce the toxicity produced by off-target cleavage.

## 1.10 Aims

The overall aim of this thesis was to develop novel therapeutic strategies for the treatment of RTT based on conditional KO mouse models of the disorder which suggest that the disease is reversible even after the onset of symptoms (Guy et al., 2007; Robinson et al., 2012), and recent gene therapy studies demonstrating that significant phenotype rescue is achievable using viral-based delivery of WT copies of the gene to KO mice (Gadalla et al., 2012; Garg et al., 2013). In order to achieve the most effective therapeutic outcomes any therapies must be targeted to the most relevant tissues. With this in mind, the first part of this thesis focuses on the comprehensive phenotyping of a novel mouse model in which *Mecp2* is selectively expressed in neurons and glia. This allows any peripheral contributions to the RTT phenotype to be identified and enable appropriately targeted therapies to be developed. The second part of the thesis focuses on the development of genome editing tools for repairing RTT causing mutations. These novel tools allow targeted changes to be made to the genomic DNA sequence and could help overcome some of the challenges associated with traditional gene therapy approaches such as overexpression toxicity and dilution of the transgene in dividing cells. In this thesis, a novel strategy was employed to ligate an *Mecp2* repair construct into the non-coding regions of the *Mecp2* gene in a mouse cell line, and these cells were then extensively examined for evidence of successful editing. It is hoped that these proof of

principle experiments will aid in the development of alternative therapeutic approaches for RTT.

# Chapter 2

---

## Material and Methods

### 2.1 Mouse model

#### *2.1.1 Generation of a peripheral KO mouse model*

A functionally null (KO) mouse, in which *Mecp2* is silenced by a stop cassette targeted to the *Mecp2* locus (Guy et al., 2007), was crossed with a nestin-cre line (Tronche et al., 1999) to generate a ‘peripheral knockout’ (PKO) mice in which *Mecp2* is selectively reactivated within neurons and glia of the nervous system. Male mice were used in all the experiments carried out in this thesis, and were kindly donated by Professor Adrian Brid (The University of Edinburgh).

#### *2.1.2 Maintenance of mouse colonies*

All mice were housed with their littermates and maintained on a 12 hour light/dark cycle, with access to food and water *ad libitum*. Experiments were carried out in accordance with the European Communities Council Directive (86/609/EEC) and a project licence with local ethical approval under the UK Animals (Scientific Procedures) Act (1986).

### 2.2 Mouse phenotyping and behavioural analysis

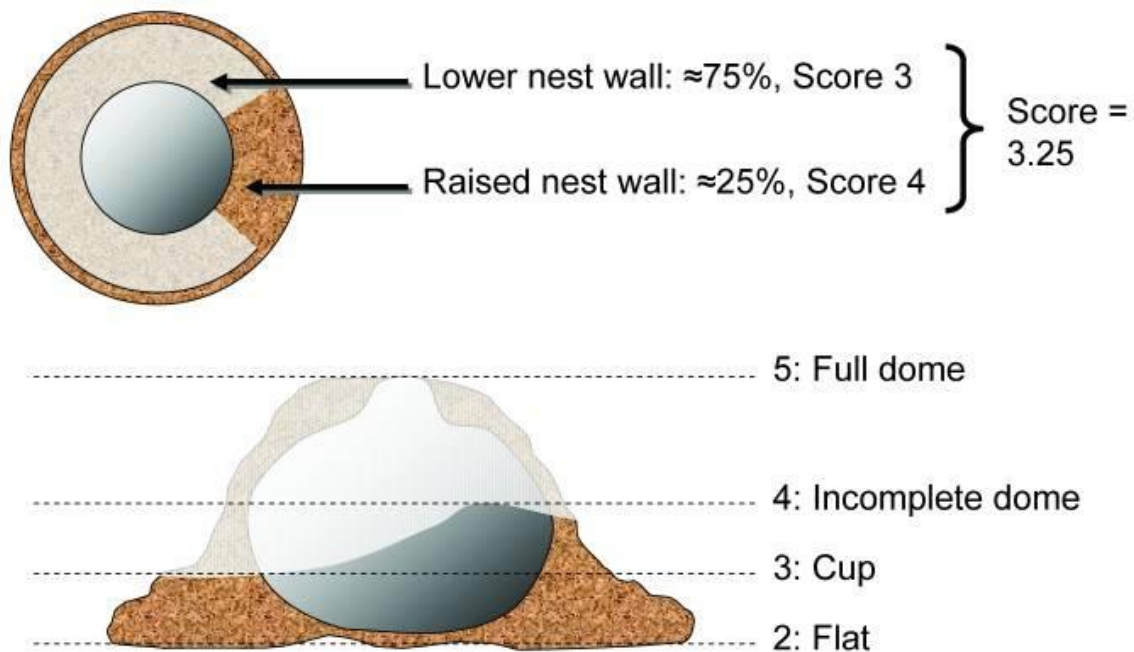
#### *2.2.1 Weight measurement and phenotypic severity scoring*

Each week all mice were weighed and scored for symptoms using an observational scoring system developed specifically for Rett syndrome (Guy et al., 2001). Mice were scored for each of six symptoms (tremor, breathing, hind-limb clasping, gait, mobility, and overall general condition) related to the Rett phenotype and were given a score of either 0 (absent), 1 (present), or 2 (severe). Individual symptom scores were then aggregated to give an overall symptom score. Scoring was carried out blind to genotype.

#### *2.2.2 Nest building*











Home cage nest quality was assessed using a previously described scoring system (Hess et al., 2008). Mice were individually caged overnight and supplied with 8 g of shredded

biodegradable paper strips as a nesting material, distributed evenly over cage floor. Next morning nest quality was assessed using a five point scoring system, which rated nests based on the formation of a central nest hollow with surrounding walls (**Fig. 2.1**). The lowest score of zero was given if the nesting material was undisturbed, and no signs of interaction or manipulation were seen. The maximum score of five was given when the mouse had constructed a fully formed nest, with walls completely enclosing a central hollow (**Fig. 2.2**). In order to assign a score, the lowest point of the nest was identified and scored (**Fig. 2.1**), with an additional 0.25 being added to the score for each quarter of the nest that had a higher wall. Scoring was carried out by two independent scorers blind to genotype.



**Figure 2.1 – Nesting scoring system**

Top and side view demonstrating how scores were assigned for the assessment of nesting behaviour. Scores were determined by the shape and height of the walls surrounding a central hollow, with each quarter of the nest contributing individually to the score (taken from Hess et al., 2008).

Score	View from above	View from side
1		
2		
3		
4		
5		

**Figure 2.2 – Example nests**

Top and side view images of example nests and their corresponding scores (taken from Hess et al., 2008).

### ***2.2.3 Whole-body plethysmography***

Respiratory function was assessed in conscious and unrestrained mice using whole body plethysmography (EMMS, Bordon, U.K.). Mice were placed inside a Plexiglas chamber and left for 20 minutes to become adapted to the environment, after which their breathing was monitored for 30 minutes. A continuous bias airflow supply allowed the animal to be kept in the chamber for extended periods of time. Pressure changes caused by alterations in the temperature and humidity of the air as it enters and leaves the subject's lungs were detected by a pressure transducer. This analogue signal was amplified and converted to a digital display by the custom software to produce a waveform representing the breathing pattern of the animal. The waveform was then exported and analysed using pClamp 10.2 (Molecular Devices inc., California, USA). Respiratory waveforms were analysed for frequency, frequency variability and the presence of apnoeas (expiratory pauses lasting longer than three respiratory cycles).

### ***2.2.4 Open field***

Locomotor function and exploratory behaviour was investigated using an open-field test. Mice were placed in the centre of a 60 cm diameter circular arena and allowed to explore freely for 10 minutes. Experimenters left the room during testing sessions to prevent distraction of the mice. The arena was imaged with an overhead digital camera and the mouse was detected and tracked by background subtraction using Ethovision 3.1 tracking software (Noldus Inc, Leesburg, VA). This produced a digital track that was then analysed by the software and a number of movement related parameters including total distance moved and rearing were calculated. The test was carried out on two consecutive days and the mean result calculated.

### ***2.2.5 Gait analysis***

Gait analysis was carried out using the DigiGait imaging system (Mouse Specifics, Boston, MA). Digital videos were captured of the ventral surface of the mice as they ran at 15cm/s on a transparent motorised treadmill. To ensure mice remained within the viewing field of the camera they were contained within an open-bottomed Plexiglas chamber, with adjustable bumpers at either end to accommodate different sized mice. Videos captured by the camera were analysed with the supplied proprietary DigiGait software which produced values for a large number of gait characteristics.

### ***2.2.6 Balance beam***

Mice were tested for balance and coordination by analysing the time taken to traverse a 50 cm span of a two inclined wooden beams (11 cm, 5 cm). Three trials on each beam were carried out on two consecutive days and the results averaged. Animals that had not crossed the beam within one minute were given the maximum score of 60 seconds.

### ***2.2.7 Bone biomechanical tests***

Biomechanical testing of tibia and micro-hardness testing of polished femur was conducted as described previously (Kamal et al., 2015). Briefly, tibial shafts from each comparison group were scanned by micro-computed tomography (SKYSCAN®1172/A  $\mu$ CT Scanner, Bruker, Belgium) before being subjected to three point bending to test cortical bone. The tests were performed using a Zwick/Roell z2.0 testing machine (Leominster, UK) with a 100 N load cell with tibias placed between supports at 8 mm separation and load applied with a rate of  $0.1 \text{ mm s}^{-1}$  until fracture occurred. Data were analysed to determine values of stiffness, ultimate load and Young's modulus (Kamal et al., 2015). The material properties of bone were assessed by micro-indentation hardness test performed on transverse distal mid-shaft sections of polished femur. Micro-hardness testing was performed using a WilsonWolport Micro-Vickers 401MVA machine (UK), with an applied load of 25 g for 100 seconds. Each sample was tested at seven points and the Vickers pyramid hardness number (HV) calculated.

### ***2.2.8 Rotarod***

Motor learning and coordination was examined using a five-lane accelerating rotarod (Ugo Basile, Italy). During testing mice were placed on a rod which gradually accelerated from 1 to 45 rpm over a period of five minutes. Mice were scored for how long they remained on the rod without falling or passively rotating. Three trials each on two consecutive days were carried out and the mean score of all trials was calculated for each mouse.

### ***2.2.9 Exercise tolerance***

The exercise capacity of mice was investigated using an accelerating elevated treadmill (Kemi et al., 2004). An electrified grid at the base of the treadmill, providing a mild electric shock when stood upon, acted as a deterrent to prevent mice stopping for motivational reasons, in order to ensure true exercise capacity was measured. Mice were placed on the treadmill at an initial speed of 10 cm/s and the speed was increased by 2 cm/s every two minutes until the animals was no longer able to cope with the speed and passively allowed



itself to be shocked on the grid without returning to the treadmill. At this point the trial was terminated and the time the mouse had lasted on the treadmill was noted.

### 2.2.10 Blood biochemistry

Animals were euthanised by CO<sub>2</sub> inhalation and arterial blood samples acquired via terminal cardiac puncture. Samples were quickly transferred to lithium heparin-coated polypropylene tubes, to prevent clotting, and transported to a clinical pathology lab for biochemical analysis.

## 2.3 General molecular biology

### 2.3.1 Reagents

**Table 2.1 – General molecular biology reagents**

Reagent	Supplier	Catalogue #
Maxima hotstart green PCR master mix 2X	Life Technologies	K1061
Phusion high fidelity DNA polymerase	New England Biolabs	M0530S
Gel loading dye, purple (6X)	New England Biolabs	B7024S
Quick ligation kit	New England Biolabs	M2200S
Quick-Load 100 bp Ladder	New England Biolabs	N0647S
Quick-Load 1 KB Ladder	New England Biolabs	N0468S
Deoxynucleotide (dNTP) solution mix	New England Biolabs	N0447S
Subcloning efficiency DH5 $\alpha$ cells	Life Technologies	18265-017
QiaQuick PCR purification kit	Qiagen	28104
QiaPrep spin miniprep kit	Qiagen	27104
QiaQuick gel extraction kit	Qiagen	28704
Plasmid <i>plus</i> maxi kit	Qiagen	12963
PureLink genomic DNA mini kit	Life Technologies	K1820-00
Ethidium bromide solution 10 mg/mL	Sigma-Aldrich	E1510-10ML
Ampicillin sodium salt	Sigma-Aldrich	A0166-5G
Kanamycin sulfate	Life Technologies	11815-024
Spectinomycin dihydrochloride	Sigma-Aldrich	S26447-100MG
Chloramphenicol	Sigma-Aldrich	C0378
Tissue culture flask 75 cm vented	Corning	430641
Tissue culture centrifuge tube (15 ml)	Corning	430790
2.0 ml Feel the Seal cryo-vial tubes	Alpha Labs	LW3534
C-Chip disposable haemocytometer	Labtech	DHC-N01
Trypan blue solution 0.4%	Sigma-Aldrich	T8154
0.2 ml PCR tube, domed cap	Starlab	11402-4300

## 2.3.2 Solutions

**Table 2.2 – General molecular biology solutions**

Solution	Composition
0.5 x TBE Running Buffer pH 8.3	45 mM Tris base 45 mM Boric acid 1 mM EDTA
Luria-Bertani (LB) medium	3% (w/v) Bacto-tryptone 1% (w/v) Bacto-yeast extract 1% (w/v) NaCl

## 2.3.3 Protocols

### 2.3.3.1 PCR primer design

Primers were designed using PrimerQuest (<https://eu.idtdna.com/Primerquest/Home/Index>), an online tool for identifying suitable primers within a given target region. Optimum annealing temperature was set to 62°C with a minimum and maximum of 55 and 65°C respectively. Optimum primer length was set to 22 bp, with a minimum and maximum of 17 and 30 bp respectively. If DNA generated by PCR was required for further downstream cloning, appropriate restriction sites were added to the 5'-ends of forward and reverse primers.

### 2.3.3.2 Polymerase chain reaction (PCR)

For standard reactions Maxima hotstart green mastermix was used. In experiments in which the amplified DNA was required for further downstream cloning, Phusion high fidelity polymerase was used. Reactions were set-up in 0.2 ml thin-walled tubes with components as described in **Table 2.3 & 2.4**. Reactions were run on a thermocycler (Applied Biosystems® Veriti® 96-well fast thermal cycler) and annealing temperature and extension time were adjusted based on primer properties and the length of DNA fragment to be amplified (**Table 2.5**). Amplified DNA was purified from the PCR mixture using QIAquick PCR purification kit (Qiagen, USA) according to the manufacturer's instructions. DNA was eluted into 30 µl of elution buffer.

**Table 2.3 - Components of typical Maxima hotstart green PCR**

Component	Concentration
HotStart green PCR master mix	1X
Forward primer	0.4 $\mu$ M
Reverse primer	0.4 $\mu$ M
Template DNA	10 ng (plasmid) or 25 ng (genomic)
Water	Make up to 50 $\mu$ l

**Table 2.4 - Components of a typical Phusion high fidelity PCR**

Component	Concentration
5X Phusion HF buffer	1X
10mM dNTPs	0.2 mM
Forward primer	0.4 $\mu$ M
Reverse primer	0.4 $\mu$ M
Template DNA	10 ng (Plasmid) or 25 ng (Genomic)
Water	Make up to 50 $\mu$ l

**Table 2.5 – Thermocycling conditions for standard PCR**

	Maxima hotstart		Phusion		
Step	Temp. °C	Time	Temp. °C	Time	Cycles
Initial denaturation	95	4 min	98	30 sec	1
Denaturation	95	30 sec	98	10 sec	35
Annealing	55-65 (primer dependent)	30 sec	55-65 (primer dependent)	20 sec	
Extension	72	1 min/kb	72	30 sec/kb	
Final extension	72	10 min	72	10 min	1
Hold	4	$\infty$	4	$\infty$	

### 2.3.3.3 Measurement of DNA concentration

The concentration of DNA samples was measured by UV absorbance at 260 nm using a Nanodrop 1000 (Thermo Fisher Scientific, USA). A blank measurement was first made using

Qiagen elution buffer. A 1 µl DNA sample was then loaded and measured. DNA purity was calculated using the ratio of 260 nm/280 nm absorbance with a ratio of 1.8 - 2.2 considered as an acceptable purity level.

#### 2.3.3.4 Restriction digest

Restriction digests were used to generate sticky ends for ligations and to check for successful DNA insertions in cloning reactions. Reactions were prepared in 0.2 ml PCR tubes and incubated from 2 hours - overnight in a thermocycler set to 37°C. **Table 2.6** shows components of a typical reaction.

**Table 2.6 - Components of a typical restriction digest**

Component	Concentration
Cut Smart buffer (10X)	1X
DNA	1 µg
Restriction enzyme 1	10 units
Restriction enzyme 2	10 units
Water	Make up to 25 µl

#### 2.3.3.5 Agarose gel electrophoresis and gel extraction

DNA samples were separated by size using agarose gel electrophoresis. Agarose solution was prepared at concentrations varying from 1-2%, depending on the size of fragments, in 0.5x TBE. The solution was heated until all the agarose had dissolved and then allowed to cool. Ethidium bromide DNA stain was then added to a final concentration of 200 µg/ml, and the solution was poured into a gel tray to set for 45 minutes. DNA samples were mixed with 1/5th volume of gel loading dye and added to the wells. For PCR samples 5 µl of sample was loaded, and for digested DNA 25 µl of sample was loaded. In addition, 150 ng of a molecular weight ladder was also loaded onto each gel to allow band size to be determined by comparison against bands of known size. Voltage was set between 70 V – 120 V depending on the size of the gel. Gels were imaged using a UV transilluminator (UV-TM-40, Upland, USA) and pictures were taken with a digital camera (PC1192, Canon, Japan). If DNA was required for further use, the appropriate band was excised using a scalpel and purified from agarose using QIAquick gel extraction kit (Qiagen, USA), according to the manufacturer's instructions. DNA was eluted in 30 µl of elution buffer.

### 2.3.3.6 Ligation

DNA ligase was used to join together DNA inserts and vector backbones during cloning.

Insert and vector were first digested with the same restriction enzymes to produce compatible sticky ends. A 3:1 insert:vector molar ratio was used at all times. For each reaction, 50 ng of digested vector was used along with an amount of insert DNA calculated by the formula:

**desired insert/vector molar ratio x mass of vector (ng) x ratio of insert to vector lengths.**

Components of a typical ligation reaction are shown in **Table 2.7**. Reactions were gently mixed and left at room temperature for half an hour after which they were either used directly for bacterial transformation or stored at -20°C for future use.

**Table 2.7 - Components of a typical ligation reaction**

Component	Amount
Vector DNA	50 ng
Insert DNA	Varies with size of insert
Quick ligase buffer (2X)	10 µl
Quick ligase enzyme	2000 units
Water	Make up to 20 µl

### 2.3.3.7 Bacterial transformation

Sub-cloning efficiency DH5α cells were removed from the -80°C freezer and thawed on ice.

Either 2.5 µl of a ligation reaction or 10ng of plasmid DNA was added to the cells and mixed very gently by stirring with the pipette tip. Cells were incubated on ice for 30 minutes and then heat-shocked at 42°C for 20 seconds. Cells were then cooled on ice for two minutes, and 950 µl of LB was added to the cells, before being transferred to a shaking incubator, set to 37°C and 200 rpm, for one hour. For plasmid amplifications, 200 µl of this cell mixture was spread on an agar plate, containing appropriate antibiotics, using a sterile spreader. For ligation reactions, the cell suspension was concentrated before plating. Cells were centrifuged for two minutes at 10,000 g to pellet the bacteria, 800 µl of the supernatant was removed, and the pellet re-suspended. All of the ~200 µl cell suspension was then spread on the agar plate. Plates were incubated upside down overnight in a 37°C incubator. The next morning plates were wrapped in Parafilm and stored in a refrigerator.

### 2.3.3.8 Colony PCR

Colony PCR was used to identify successful ligation reactions directly from colonies growing on an agar plate. Colonies were picked from agar plates with a 200 µl pipette tip. After picking, the tip was rubbed along the bottom of a PCR tube to provide a template for the PCR. The rest of the colony was pipetted into 3 ml of LB containing appropriate antibiotic. Standard Maxima hotstart green PCR components (**Table 2.3**) were added to the PCR tube along with appropriate primers and the reaction was run on a thermocycler using standard PCR conditions (**Table 2.5**). PCR products were resolved on an agarose gel to identify those which had the correct insert. Positive clones were then sub-cultured overnight in a 37°C shaking incubator for plasmid extraction the next day.

### 2.3.3.9 Plasmid isolation

Plasmid DNA was isolated from successfully transformed bacterial cells by miniprep. Single bacterial colonies were picked, using a sterile 200 µl pipette tip, and cultured overnight in 3 ml of LB (with appropriate antibiotic) in a shaking incubator set to 37°C and 200 rpm. Next morning, the bacterial culture was centrifuged at 13,000 g for three minutes to pellet bacteria. Plasmid DNA was then isolated from pellets using QIAprep spin miniprep kit (Qiagen, USA) following the manufacturer's instructions. DNA was eluted in 30 µl of elution buffer.

### 2.3.3.10 Sequencing

DNA inserts from cloning reactions were checked for correct orientation and absence of mutations by Sanger sequencing. Sequencing primers flanking the insert were designed using PrimerQuest (see 2.3.3.1). For larger inserts, internal primers were also designed to ensure complete sequence coverage. Sequencing reactions were carried out by Source Bioscience UK. For each sequencing reaction 500 ng of DNA and 16pmol of primer was required. Sequence traces were analysed using SnapGene software (GSL Biotech LLC, USA).

### 2.3.3.11 Genomic DNA extraction

Adherent cultured cells were trypsinised using trypsin EDTA 0.25% and centrifuged for 5 minutes at 250 g. The supernatant was removed and the pellet was resuspended in 200 µl of DPBS. Genomic DNA was then extracted using the PureLink genomic DNA mini kit according to the manufacturer's protocol. Extracted DNA was eluted in 50 µl of genomic elution buffer.

## 2.4 Cell culture

### 2.4.1 Reagents

**Table 2.8 - Cell culture reagents**

Reagent	Supplier	Catalogue #
Trypsin-EDTA 0.25% solution	Sigma-Aldrich	T4049-100 ml
Dulbecco's phosphate buffered saline (DBPS)	Life Technologies	14190-144
Trypan blue 0.4% solution	Sigma-Aldrich	T8154
Filter unit 500 ml 0.22 µm polyethersulfone	Merck Millipore	SCGPU05RE
GlutaMAX supplement	Life Technologies	35050-038
Penicillin-streptomycin 10K units / ml	Life Technologies	15140-122
Newborn calf serum, New Zealand origin	Life Technologies	16010-159
Advanced Dulbecco's modified eagle's medium /F12	Life Technologies	12634-010
Dulbecco's modified eagle's medium	Sigma-Aldrich	41965-039
Foetal bovine serum heat inactivated	Life Technologies	1018-157
Dimethyl sulfoxide	Sigma-Aldrich	472301-100 ml

### 2.4.2 Cell culture media

**Table 2.9 - Composition of cell culture media**

Media	Reagent	Concentration
P19 Growth (500 ml)	Advanced Dulbecco's modified eagle's medium /F12	93%
	GlutaMAX supplement	1%
	Penicillin-streptomycin 10K units / ml	1%
	Newborn calf serum, New Zealand origin	5%
Freezing (250 ml)	Dulbecco's modified eagle's medium	39%
	Foetal bovine serum heat inactivated	50%
	Penicillin-streptomycin 10K units / ml	1%
	Dimethyl sulfoxide	10%

### **2.4.3 Cell maintenance**

#### **2.4.3.1 Liquid nitrogen recovery**

Vials of P19 cells (kind gift from Katherine West lab) were stored long-term in liquid nitrogen. When required for experiments, cells were removed from storage and transferred to the lab on dry ice. They were quickly thawed for three minutes in a 37°C water bath and then transferred into a 75 cm<sup>2</sup> cell culture flask containing 10 ml of pre-warmed P19 growth medium (**Table 2.9**). The flask was incubated at 37°C for four hours to allow the cells to attach, after which the growth media was removed and replaced with 10 ml of fresh media.

#### **2.4.3.2 Cell counting**

To determine cell density for plating and transfection reactions, cells were counted using a disposable haemocytometer. A 20 µl sample was taken from the cell suspension and mixed with 20 µl of 0.4% Trypan blue Solution. A 10 µl sample of this mixture was added to each chamber of the haemocytometer, and cells were counted using a light microscope with a 10x objective. Cell counts were made for five squares in each chamber and the mean result calculated. To calculate density per ml of cell suspension the following formula was used:

$$\text{Mean cell count} \times \text{dilution factor} \times 10,000$$

The dilution factor was two as the cell suspension was diluted 1:1 with Trypan blue. Each square counted had a volume of 100 nl and counts were therefore multiplied by 10,000 to produce a value per ml.

#### **2.4.3.3 Cell passage**

In order to keep the cells in the log-phase of growth and to prevent cell death, cells were passaged when 70-80% confluent. For passaging, growth medium was removed and cells were washed gently in 10 ml of DPBS. To detach the cells, 2 ml of trypsin solution was added and the flask was returned to the incubator for three minutes, and 8 ml of serum containing growth medium was added in order to inactivate the trypsin. Cells were pipetted up and down gently using a 10 ml pipette, to prevent clumping, and then 1 ml of this cell suspension was added to 10 ml of pre-warmed growth media in a fresh 75 cm<sup>2</sup> culture flask, and the flask was returned to the incubator.



#### *2.4.3.4 Freezing cells for long-term storage*

For long-term storage, cells were kept in freezing media (**Table 2.9**) in liquid nitrogen. Cells chosen for storage had been passaged as few times as possible. Growth medium was removed and the cells were washed gently with DPBS. Cells were then detached with trypsin, and growth medium was added to inactivate the enzyme. Cell density was first measured using a haemocytometer, and then cells were pelleted by centrifugation at 1500 g for five minutes at room temperature. The supernatant was removed and the pellet re-suspended in freezing medium to a density of 2 million cells per ml. The cell suspension was then added to cryovials, 1 ml per vial, and transferred quickly on dry ice to liquid nitrogen storage.

#### *2.4.4 Neon transfection*

Cells were transfected with DNA plasmids by electroporation using the Neon transfection system (Life Technologies, USA) according to the manufacturer's protocol. Cells were first trypsinised and counted. For a 12-well plate, 200,000 cells per well were required. The appropriate amount of cell suspension was added to a 15 ml falcon tube and the cells were pelleted by centrifugation at 150 g for five minutes at room temperature. The supernatant was removed and the pellet re-suspended in the appropriate amount of buffer R (10  $\mu$ l per 200,000 cells). For each well, 10  $\mu$ l of cell suspension was mixed with 500 ng of each plasmid to be transfected, and then taken up into the Neon pipette tip. The pipette tip was inserted into a Neon tube containing 3 ml of electrolyte buffer E, and the start button was pressed on the neon device to begin the electroporation. Once the process was complete, the pipette was removed from the Neon tube and the cells were pipetted into a well in a 12-well plate containing fresh growth media without any antibiotics. The plate was mixed gently and then returned to the incubator for 48 hours before being used for genomic DNA extraction or flow cytometry analysis.

## 2.5 TALEN construction

### 2.5.1 Reagents

**Table 2.10 - Reagents specific for TALEN construction**

Reagent	Supplier	Catalogue #
T4 DNA ligase	New England Biolabs	M0202S
BsaI	New England Biolabs	R0535S
Esp3I	Life Technologies	FD0454
Plasmid-Safe ATP-dependent DNase	Cambio	E3101K
ATP (10 mM)	New England Biolabs	9804S
X-Gal	Life Technologies	15520-018
Isopropyl-β-D-thiogalactoside (IPTG – 1 M)	Roche Applied Science	10724815001

### 2.5.2 Primers

**Table 2.11 – Primers used for TALEN pair construction and testing**

Primer	Sequence 5' → 3'
pCR8_F1	TTGATGCCTGGCAGTTCCT
pCR8_R1	CGAACCGAACAGGCTTATGT
SeqTALEN_5-1	CATCGCGCAATGCACTGAC
TAL_R2	GGCGACGAGGTGGTCGTTGG
Mecp2_Int2_For	TCTAGCTGCACAACCTTCCA
Mecp2_Int_Rev	GCTGCCTAGGGTTCCAGTAT

### 2.5.3 Vectors

**Table 2.12 - Vectors used for TALEN pair construction**

Vector Name	Source
TALEN golden gate RVD plasmid kit	Addgene Kit #1000000024
pFUS_A	Addgene Kit #1000000024
pFUS_B	Addgene Kit #1000000024
pLR vectors	Addgene Kit #1000000024
pTALv2_FokIWT	Adam West lab
pTALv3-FokI-ELDS	Adam West lab
pTALv3-FokI-KKRS	Adam West lab

### 2.5.4 Protocols

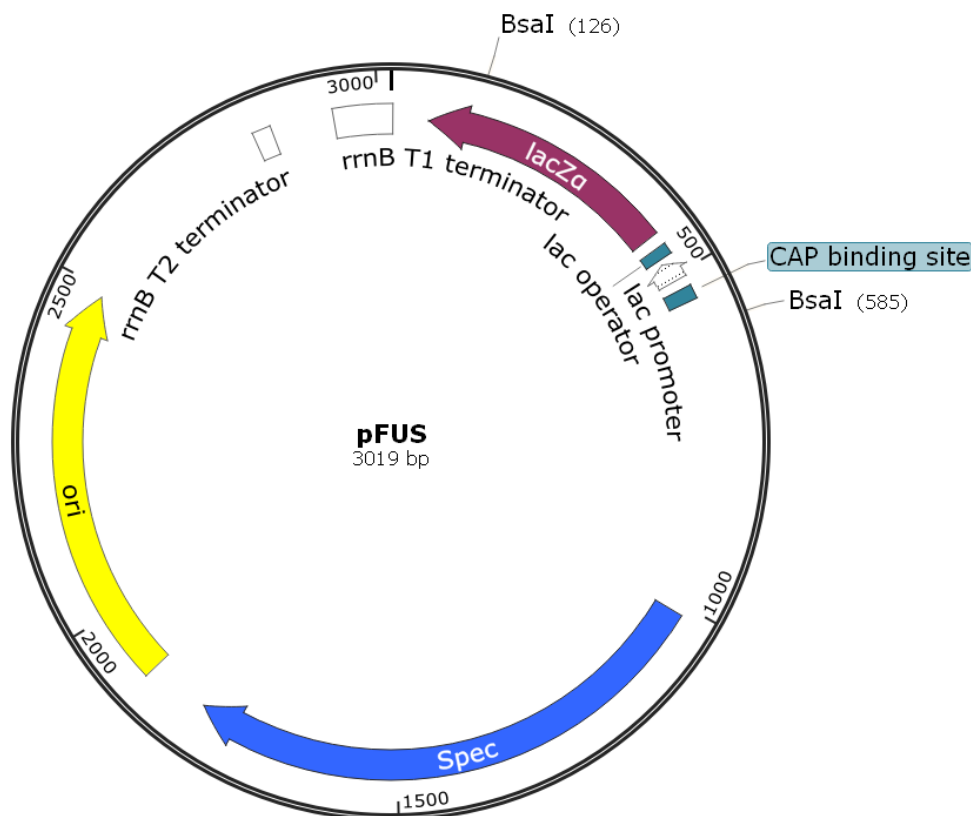
TALEN constructs were assembled using the golden gate cloning method (Engler et al., 2008, 2009). This allowed multiple plasmids to be combined in a defined order. Since TALEN construction involves mixing multiple plasmids together, a picking sheet was used to prevent confusion and errors. TALEN length varied from 17-20 RVDs in length. Assembling this number of fragments together in one step would be inefficient so a two-step process was used (Cermak et al., 2011). Five days were required for full TALEN assembly.

#### Day 1

The first 10 RVDs were assembled in vector pFUS\_A and the remaining RVDs were assembled in pFUS\_B. During assembly RVD inserts replace the lac Z sequence in the plasmid which allows for blue/white screening (**Fig. 2.3**). Appropriate RVDs were picked for each position in the TALEN DNA binding domain. For example, if position one was RVD NI then plasmid pNI1 was added to the tube. If position three was RVD HD then plasmid pHD3 was added. This ensured that the RVDs had the correct overhangs after digestion to allow them to be assembled in the correct order. The RVDs were mixed with the appropriate enzymes and pFUS vector as described (**Table 2.13**) in a 0.2 ml PCR tube. Both digestion and ligation occur within the same reaction.

**Table 2.13 - Components of TALEN golden gate # 1 reaction**

Component	Concentration
Appropriate RVDs depending on TALEN sequence	150 ng of each
pFUS_A or pFUS_B	150 ng
BsaI	10 units
T4 DNA Ligase	1200 units
10X T4 DNA Ligase Buffer	1X
Water	Make up to 20 µl



**Figure 2.4 – pFUS plasmid map**

Plasmid map showing the generic features of the pFUS plasmid used to assemble RVDs on day 1. The lacZ- $\alpha$  sequence codes for  $\beta$ -galactosidase. Digestion with BsaI during RVD assembly removed this sequence, leading to colourless cells, allowing for screening of successful cloning. The plasmid also contains the resistance gene for the spectinomycin antibiotic and an origin of replication (ori).

When all the components were assembled they were transferred to a thermocycler and the following programme was run:

**(5 minutes @ 37°C + 10 minutes @ 16°C) x 10**

**5 minutes @ 50°C**

**5 minutes @ 80°C**

In order to destroy unligated linear DNA fragments, incomplete products, and linearised vector, 1  $\mu$ l of 10 mM ATP and 1  $\mu$ l of Plasmid-Safe nuclease was then added to the tubes and samples were incubated at 37°C for 1 hour. DH5 $\alpha$  cells were then transformed with 5  $\mu$ l of each reaction, and 200  $\mu$ l of the transformation mix was spread on 50 mg/ml spectinomycin agar plates containing 70  $\mu$ l of 20 mg/ml X-gal and 70  $\mu$ l of 0.1 M IPTG. This

allowed blue/white screening of successful reactions, with empty vectors producing blue colonies and vectors with inserts producing white colonies.

## Day 2

Three white colonies were picked from each plate and checked for successful RVD insertion by colony PCR using standard PCR components and thermocycler conditions (see **Table 2.3 & 2.5**). Primers pCR8\_F1 and pCR8\_R1 were used for the reaction, with an annealing temperature of 55°C and an extension time of 1 minute 45 seconds. Clones which had a PCR product of the correct size were cultured overnight in 3 ml of LB containing 50 mg/ml spectinomycin.

## Day 3

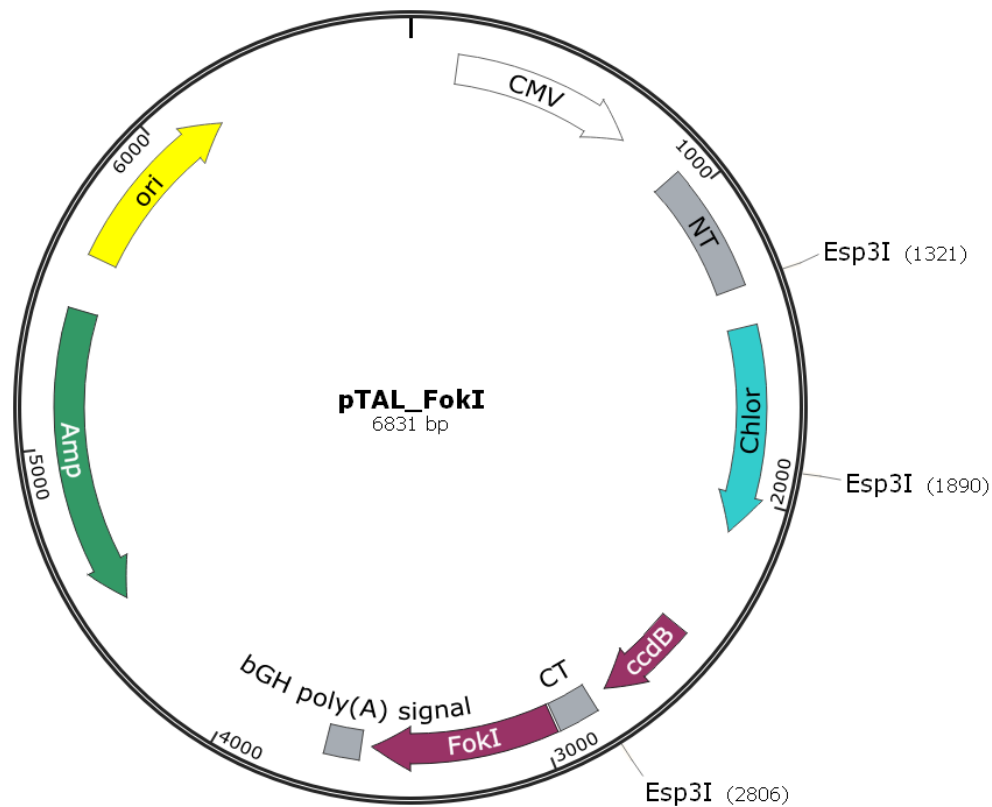
Overnight bacterial cultures were pelleted by centrifugation at 13,000 *g* for three minutes, and plasmid DNA was extracted by miniprep. As an additional check, plasmid DNA was sent for overnight Sanger sequencing to confirm RVDs were inserted in the correct order and without mutations.

To combine the two pFUS plasmids into the final full length TALEN construct, a second golden gate reaction was carried out. pFUS\_A and pFUS\_B were combined with a plasmid containing the final RVD, and a pTAL destination vector, coding for the FokI nuclease, along with appropriate enzymes (**Table 2.14**).

**Table 2.14 - Components of TALEN golden gate # 2 reaction**

Component	Amount
pFUSA	150 ng
pFUSB	150 ng
Appropriate pLR vector (last RVD)	150 ng
Destination vector	75 ng
Esp3I	10 units
T4 DNA ligase	800 Units
10X T4 ligation buffer	1X
Water	Make up to 20 µl

There are five vectors coding for the final RVD, one for each of the five possible RVDs in a TALEN sequence. The destination pTAL plasmid (**Fig. 2.4**) chosen depended on whether the WT or obligate heterodimer mutant version of FokI nuclease was required.



**Figure 2.3 – pTAL\_FokI plasmid map**

Plasmid map showing the features of the final destination pTAL\_FokI plasmid. The final RVD assembly was cloned between the sequences coding for the TALE protein N-terminal (NT) and C-terminal (CT), using *Esp3I*, and replaced the chloramphenicol resistance gene (Chlor) and the *ccdB* gene (*ccdB*). The *ccdB* gene is lethal in standard *E. coli* cells and was used to select against products without an RVD insert. Downstream of the CT is the FokI nuclease sequence (FokI), the bovine growth hormone (bGH) poly(A) signal to terminate transcription, the ampicillin (Amp) resistance gene, and a bacterial origin of replication (ori). TALEN transcription was driven by the cytomegalovirus (CMV) promoter.

Reactions were set-up in a 0.2 ml PCR tube, transferred to a thermocycler, and the following programme was run:

**(5 minutes @ 37°C + 10 minutes @ 16°C) x 5**  
**15 minutes @ 37°C**  
**5 minutes @ 80°C**

DH5 $\alpha$  competent cells were then transformed with 5  $\mu$ l of this reaction product, and then plated on 50 mg/ml ampicillin plates containing X-gal and IPTG (see above).

#### **Day 4**

Two white colonies from each bacterial plate were picked with a 200  $\mu$ l pipette tip and cultured overnight in 3 ml of LB containing 50 mg/ml ampicillin.

#### **Day 5**

Bacterial cultures were pelleted by centrifugation at 13,000  $g$  for three minutes and plasmid DNA was isolated by miniprep. DNA concentration was measured by Nanodrop, and 500 ng of each plasmid was digested with SacI and XhoI and checked for correct sized inserts by agarose gel electrophoresis. Plasmids with correct digestion products were sent for Sanger sequencing using sequencing primers SeqTALEN 5-1 and TAL\_R2, which confirmed RVDs were inserted in the correct order and without mutations.

## 2.6 Cloning of CRISPR-Cas9 constructs

### 2.6.1 Reagents

**Table 2.15 - Reagents specific for cloning of CRISPR-Cas9 constructs**

Reagent	Supplier	Catalogue #
FastDigest Bpil	Life Technologies	FD1014
FastAP thermosensitive alkaline phosphatase	Life Technologies	EF0654
T4 polynucleotide kinase (PNK)	New England Biolabs	M0236S

### 2.6.2 Primers

**Table 2.16 - Primers for cloning target sequence into guide expression plasmid**

Primer	Sequence 5' → 3'
ln2_7SK_g3LA64F	CCTCGCACTGCCCCTCTGAGCTAC
ln2_7SK_g3LA64R	AAACGTAGCTCAGAGGGGCAGTGC
ln2_hH1_g3RB86F	TCCCACTGTGTGCGCACTACAG
ln2_hH1_g3RB86R	AAACCTGTAGTGCGACACAGT
ln2_hH1_g2LA65F	TCCCAAGTAGCAGCTGCCTAT
ln2_hH1_g2LA65R	AAACATAGGCAGCTGCTACTT
ln2_7SK_g2RB52F	CCTCGTCTGGCCTGTAGCTCAGA
ln2_7SK_g2RB52R	AAACTCTGAGCTACAGGCCAGAC
T7F	TAATACGACTCACTATAGG

### 2.6.3 Vectors

**Table 2.17 – Plasmid vectors used for CRISPR construction**

Vector Name	Source
phH1-gRNA-1-2	Adam West lab
h7SK-gRNA-2-2	Adam West lab
pAC84-pCR8-dCas9	Addgene # 48218
pX335-U6-Chimeric_BB-CBh-hSpCas9n(D10A)	Addgene # 42335



### 2.6.4 Protocol

For each CRISPR guide, a 20 bp guide sequence was synthesised as two complementary oligonucleotides (Table 2.16), annealed together, and then cloned into single guide expression plasmids cut with BbsI (Table 2.17).

#### Design of oligonucleotides

Some RNA Polymerase III promoters require a specific nucleotide to start transcription. The human 7SK promoter (*h7SK*) requires a G base to initiate RNA transcription, therefore guides beginning with a G base were cloned into *h7SK\_gRNA*. Guides beginning with an A base were cloned into *hH1\_gRNA* plasmid. Guides beginning with T or C base had an extra G base added at the start of the guide sequence and were cloned into *h7SK\_gRNA*. Extra bases were added to 5' and 3'-ends of oligonucleotides as appropriate to make suitable overhangs for insertion into BbsI cut guide expression plasmids (Table 2.18).

**Table 2.18 - Overhang design for guide sequence oligonucleotides**

Expression Plasmid	Oligo	Overhang (+1 base)	Protospacer (17-20 bases)	Overhang
<i>h7SK_gRNA</i>	Forward	5' -CCTCG	(N) bases	CAAA-5'
	Reverse	3'-C	(N) bases complement	
<i>hH1_gRNA</i>	Forward	5'-TCCCA	(N) bases	CAAA-5'
	Reverse	3'-T	(N) complement	

#### Oligo annealing

Single stranded forward and reverse complementary oligonucleotides were annealed together and phosphorylated for cloning into the expression plasmid. This formed a double stranded DNA molecule with overhangs matching those of the appropriate expression plasmid cut with BbsI. Oligonucleotides were mixed with the appropriate enzyme and buffers in a 0.2 ml PCR tube (Table 2.19), and annealed in a thermocycler using the following conditions:

**30 minutes @ 37°C**

**5 minutes @ 95°C and then ramp down to 25°C at 5°C /min**

**Table 2.19 – Components of oligo annealing reaction**

Component	Concentration
Oligonucleotide 1	100 µM
Oligonucleotide 2	100 µM
10X T4 ligation buffer	1X
T4 PNK	5 units
ddH <sub>2</sub> O	Make up to 20 µl

**Digestion of expression plasmid**

1 µg of appropriate single guide expression plasmid was digested with BbsI RE in 0.2 ml PCR tube (**Table 2.20**) and incubated in thermocycler at 37°C for two hours. Digestion products were then separated by DNA agarose electrophoresis, and the cut vector band was excised and purified using QIAquick gel extraction kit (see 2.3.3.5).

**Table 2.20 - Components of guide expression plasmid digest**

Component	Concentration
Guide expression plasmid	1 µg
FastDigest BbsI	10 units
Fast alkaline phosphatase (AP)	1 unit
10X FastDigest buffer	1X
ddH <sub>2</sub> O	Make up to 20 µl

**Cloning of annealed oligonucleotides and expression plasmid**

To clone the annealed guide sequence into the expression plasmid, a ligation reaction was set-up in a 1.5 ml Eppendorf tube (**Table 2.21**) and incubated at room temperature for 10 minutes. DH5α cells were then transformed with 2.5 µl of this reaction, and the cells were spread on 50 mg/ml ampicillin plates.

**Table 2.21 - Components of ligation reaction for annealed oligos and expression plasmid**

Component	Concentration / Amount
BbsI digested plasmid	50 ng
Annealed oligonucleotides	40 nM
2X Quick ligation buffer	1X
Quick ligase enzyme	2000 units
ddH <sub>2</sub> O	Make up to 11 µl

## 2.7 Mecp2 repair and GFP reporter constructs

### 2.7.1 Primers

**Table 2.22 - Primers for cloning and detecting genomic insertion**

Primer	Sequence 5' → 3'
Genomic_FOR	AGAGGCCTGCATTCTTAAC TAC
Genomic_REV	TTGGGT CACATGGGTCTTTAC
C1	GACTCGAGCTTTACATAGAGCG
C2	TCCCACAACGAGGACTACA
TALEN_KI_FOR	CGGAAGGCCGTC AAGG
TALEN_KI_REV	CCAGTCTGGCCTGTAGCTC
CRISPR_KI_FOR	CCAGAGGTCTCACATGCTGTG
CRISPR_KI_REV	GGTACCCACTTT CACAGAGAG
HBB_Ex3_SA_FOR	CTAGACTGACCTCTTCTCTTCTCC CACAGGG
HBB_Ex3_SA_REV	AATTCCCTGTGGGAGGAAGAGAAGAGGTCAGT
T7F	TAATACGACTCACTATAGGG

### 2.7.2 Vectors

**Table 2.23 – Plasmid vectors for repair construct cloning and flow cytometry analysis**

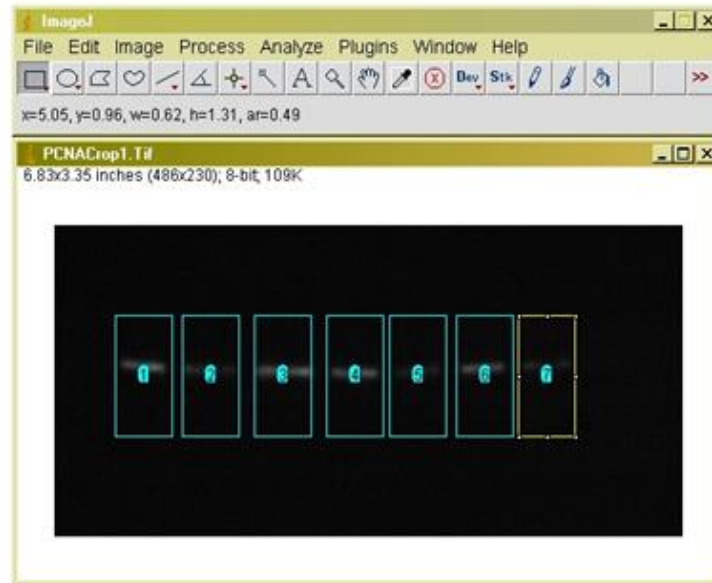
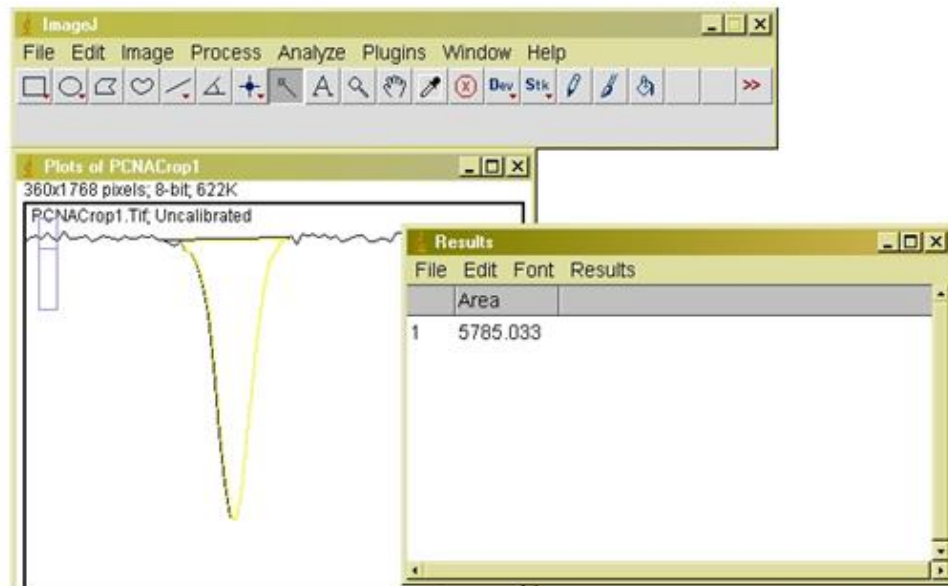
Vector Name	Source
pENTR-1A	Addgene #17398
Collectis_4829	Collectis (France)
p-CMV-mCherry	Joe Mountford lab
pLenti-PGK-GFP	Kind donation Kamal Gadalla

## 2.8 TIDE analysis of cutting efficiency

An online tool called Tracking of indel decomposition (TIDE- <http://tide.nki.nl/>) was used to quantify cutting efficiency of TALEN and CRISPR constructs. Genomic DNA was extracted from cells transfected with constructs, as well as from untransfected controls, and PCR amplified using primers *Mm\_Mecp2\_Int2\_For* and *Mm\_Mecp2\_Int2\_Rev*. PCR products were then sent for Sanger sequence analysis. Sequence traces from control and transfected samples were uploaded onto the TIDE software, the TALEN or CRISPR target sequence was inputted and the software then automatically calculated cutting efficiency and characterised indel composition.

## 2.9 ImageJ band intensity measurement

ImageJ software (NIH, USA) was used to measure the band intensity of PCR products on agarose gels. Images files were opened and the “rectangular select” tool was used to select an area around the first gel band. This was marked as the first band by pressing CTRL + 1. The mouse cursor was used to drag the box to the next band which was then selected by pressing CTRL + 2. The box was dragged to each band in turn and the bands selected each time by pressing CTRL + 2. When the final band was selected CTRL + 3 was pressed, which displayed histograms representing the intensity of each selected band. For each histogram, the “draw line” button was selected and a line drawn across the histogram from where the curve begins to drop until where it levels out again. The “magic wand” button was then selected and an area inside the histogram was clicked using the left mouse button. This selected area then became highlighted in yellow and an intensity value appeared in a new “results” window. This process was repeated for each histogram until intensity values for all the selected bands was acquired.

**A****B**

**Figure 2.4 – ImageJ band intensity measurement**

The intensity of DNA bands on agarose gels were measured using ImageJ. **(A)** Bands to be measured were selected using the “rectangular select” tool (blue and yellow numbered rectangles). **(B)** Histograms, representing the intensity of each band, were generated and the straight line tool was used to enclose the area of the peak. This enclosed area (inside yellow lines) was then selected using the “magic wand” tool and the area of the peak was measured. This measurement allowed the relative density of the peaks to be compared. (Diagram taken from (<http://www.di.uq.edu.au/sparqimagejblots>)).

## 2.10 Flow cytometry

Cells expressing GFP or mCherry were detected, and fluorescence levels quantified, by flow cytometry. To prepare cells for analysis, cells were trypsinised with 0.25% trypsin solution, centrifuged at 1500 *g* for 5 minutes, and then resuspended in PBS. For initial neon optimisation experiments (see Chapter 4), GFP fluorescence was detected using a blue 488 nm laser and 530/30 nm emission filter on an Attune acoustic focusing cytometer (Applied Biosystems, USA). For knock-in detection (**see Chapter 5**), fluorescence was detected using a MACSQuant VYB flow cytometer (Miltenyibiotec, Germany). GFP was detected using a blue 488 nm laser with 525/50 nm filter and mCherry was detected using a yellow 561 nm laser with 615/20 nm filter. Enough cell solution was analysed until 30,000 viable cells were detected. To detect viable cells, a gate was drawn to exclude events with low forward and high side scatter. To eliminate false positive fluorescence detection, background fluorescence levels in cells was determined using untransfected control cells. Data was analysed using FlowJo software (FlowJo, LLC, USA).

# Chapter 3

---

## Nervous system specific reactivation of *Mecp2* reveals peripheral phenotypes of Rett syndrome

### 3.1 Introduction

MeCP2 is expressed in most tissues of the body. An early study (Shahbazian et al., 2002a) quantifying MeCP2 protein levels using western blot, showed particularly high levels in the brain, lungs, and spleen, with moderate levels in the heart and kidney and barely detectable levels in the stomach, small intestine and liver. Another study (Zhou et al., 2006), again using western blot, showed MeCP2 present in all tissues looked at including testis, thymus, pancreas, and with higher levels in the small intestine than seen in the previous study. A more recent study (Song et al., 2014) comprehensively investigated tissue steady state protein levels of MeCP2 protein using immunostaining on tissue cryosections. In total they characterised 60 different cell types from 16 different tissues. MeCP2 protein expression was seen in the vast majority of cell types with high levels seen in brain, spleen, heart, skeletal muscle and, in contrast to earlier studies, liver. Although there are disagreements on the levels in particular tissues, it is clear from these studies that MeCP2 is expressed in a wide range of tissues throughout the body.

Despite being an almost ubiquitously expressed protein, several studies have shown that mice in which *Mecp2* is deleted only from the nervous system are indistinguishable from mice in which *Mecp2* has been deleted globally (Chen et al., 2001; Guy et al., 2001). Mice in which exons 3 and 4 of *Mecp2* were deleted specifically in neurons and glia using nestin-cre mediated excision (Guy et al., 2001) showed hypoactivity, abnormal gait, low weight and the early lethality seen in global KO mice. Similarly, mice in which exon 3 was deleted in neurons and glia (Chen et al., 2001) showed the reduced brain weight and early lethality seen in mice in which the exon was deleted globally.

Results from these studies would suggest that the RTT phenotype can be attributed to the absence of functional MeCP2 from the neurons and glia only. However in both cases, a comprehensive phenotyping was not carried out, with only certain aspects of the phenotype examined such as gait, brain size, body-weight, and early lethality. In addition, the early

lethality of these mice means that any subtler phenotypes, or those that take longer to develop, could be masked in these animals. More recent animal studies have identified a number of novel phenotypes including cardiovascular abnormalities ((McCauley et al., 2011; Panighini et al., 2013), lung phenotypes (De Felice et al., 2010), bone and skeletal muscle abnormalities (O'Connor et al., 2009; Conti et al., 2015; Kamal et al., 2015) and altered cholesterol biosynthesis (Buchovecky et al., 2013; Segatto et al., 2014) which potentially have a peripheral component.

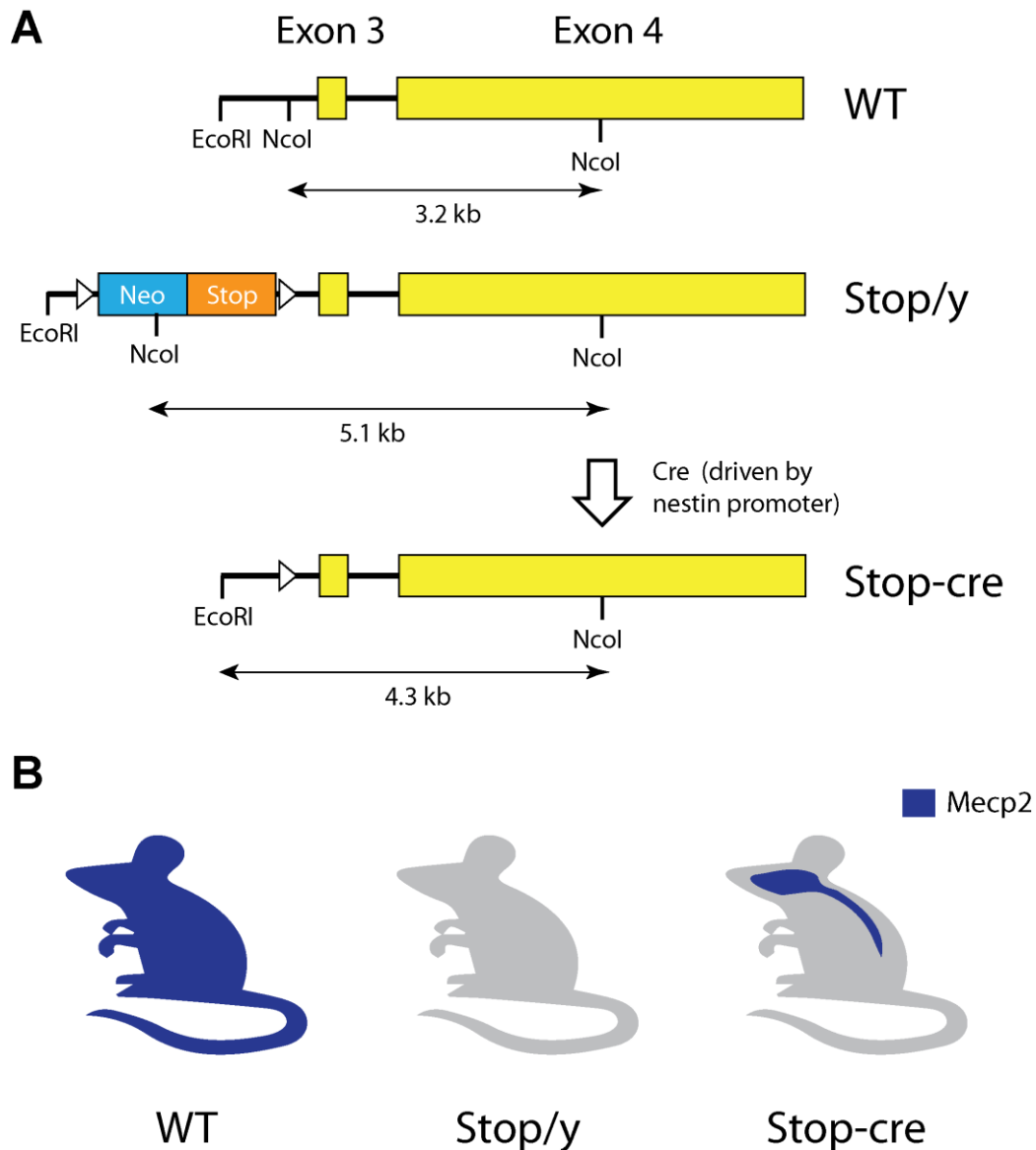
### 3.2 Aims

The overall aim of the work in this chapter was to comprehensively phenotype a novel male RTT mouse model in which *Mecp2* was selectively expressed in neurons and glia of the CNS and PNS, but silenced in all other cell types, in order to investigate the peripheral contribution to the major RTT-like phenotypes, as well as to more recently identified and less well-studied aspects of the disorder. The objectives of this chapter were to:

- (1) To quantify MeCP2 levels in the novel mouse model, both in peripheral and neuronal tissues in order to assess the validity of the model
- (2) To determine the effect of nervous system-specific expression of MeCP2 on survival, weight and gross RTT-like symptoms
- (3) To carry out an extensive battery of behavioural, functional, and structural tests to identify novel peripheral phenotypes



### 3.3 Validation of the peripheral KO mouse model

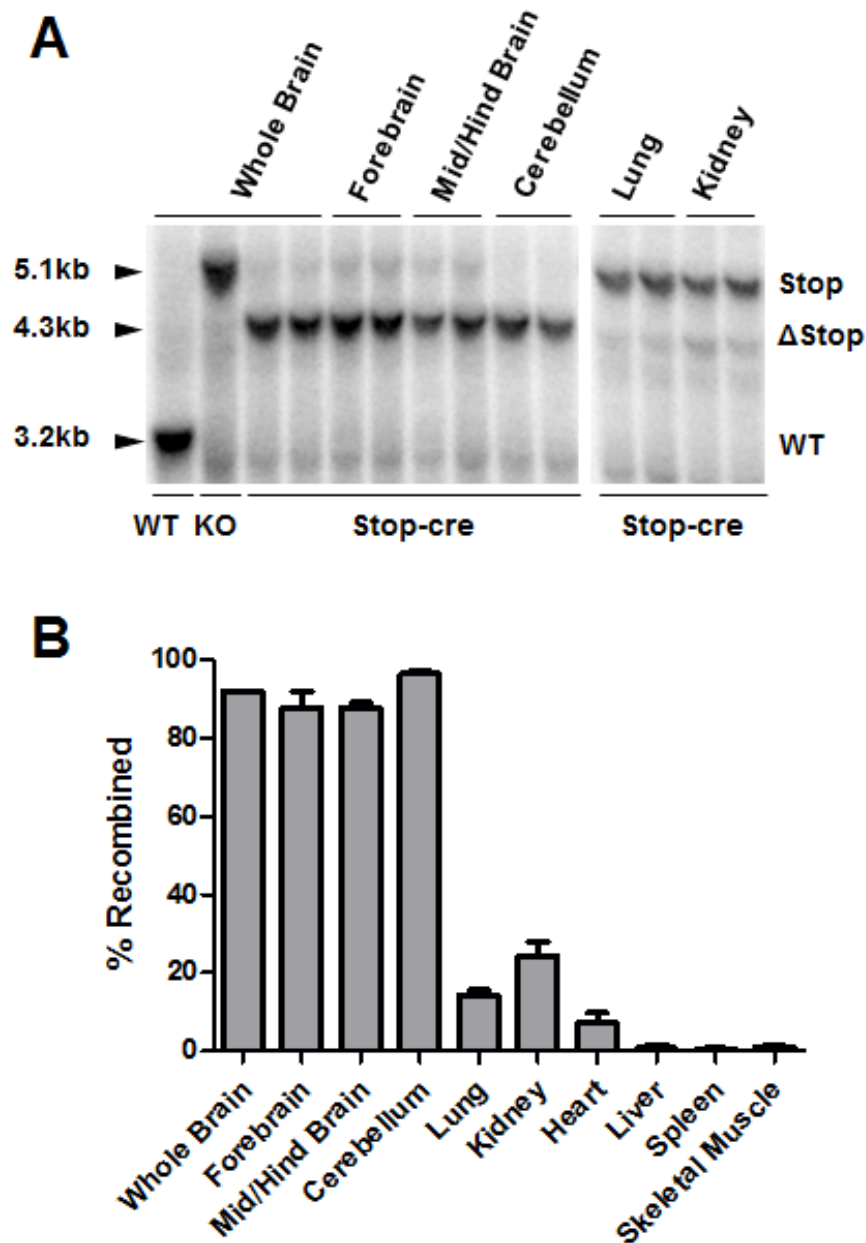


**Figure 3.1 – Creation of peripheral KO mouse model**

Summary of the mouse models used in this study. (A) A global KO mouse *Mecp2*<sup>stop/y</sup> (referred to as stop/y) was created by insertion of a Neo-stop cassette into intron 2 of the *Mecp2* gene. *Mecp2* was specifically reactivated in nervous system cells in peripheral KO mice *Mecp2*<sup>stop-cre/y</sup> (referred to as stop-cre) by excision of stop cassette in nestin expressing cells. Digestion with EcoRI and NcoI produces different sized products in the three models which can be detected by Southern blot. (B) Illustration of the MeCP2 protein expression pattern in mouse models, showing presence (blue) of MeCP2 in nervous system and absence (grey) from rest of body in stop-cre mice.

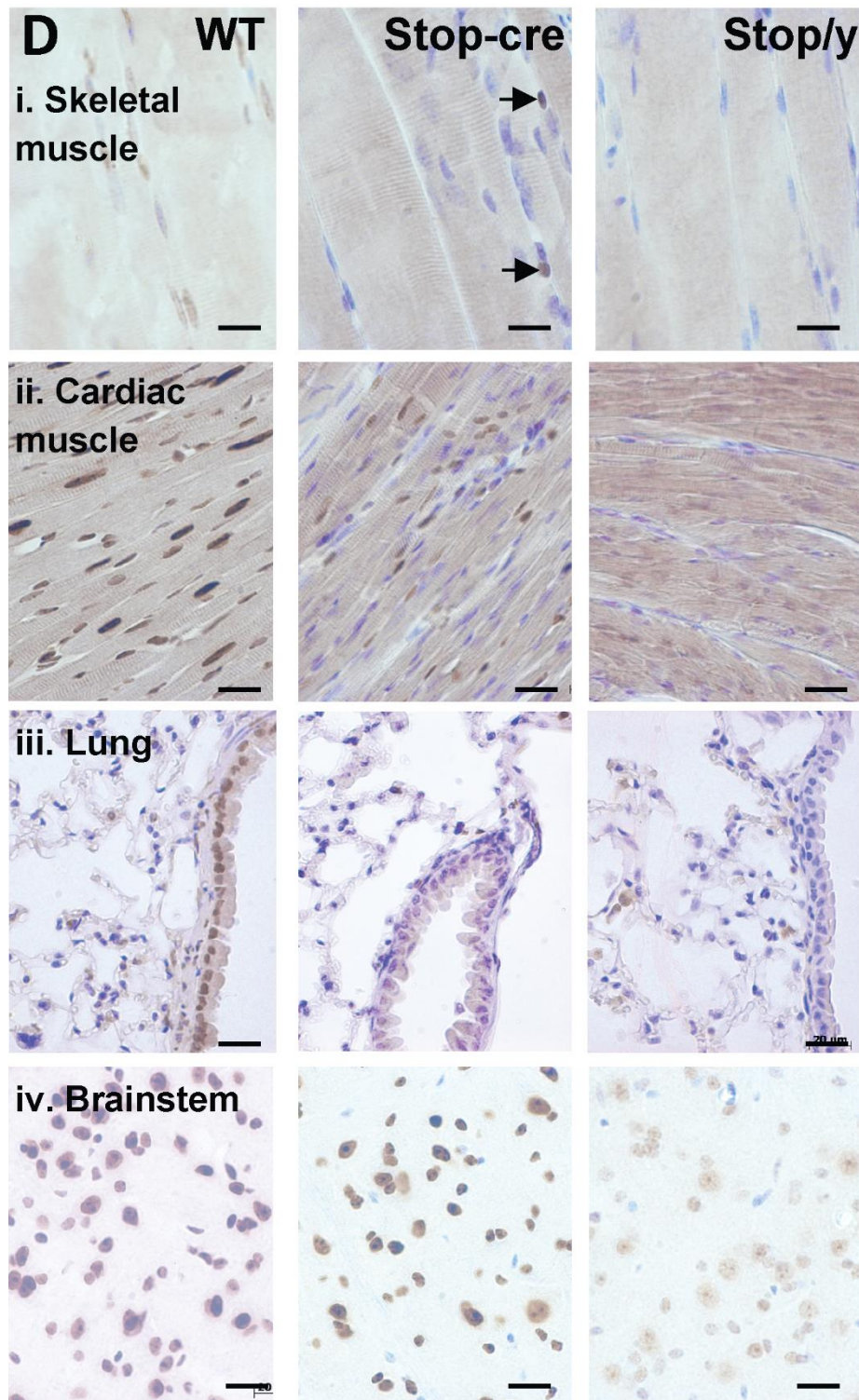
In order to investigate the role of central versus peripheral MeCP2 protein expression in RTT, a tissue-specific KO mouse model (stop-cre) was generated compared to a global KO model (stop/y), in which gene transcription is silenced by the insertion of a stop cassette into intron 2 of the gene (Guy et al., 2007); **Fig. 3.1-A**), as well as WT mice. The stop cassette can be excised using cre-recombinase, allowing gene transcription to be reactivated. To generate a peripheral KO model, stop/y mice were crossed with mice that expressed cre under the control of the nestin promoter. In this line, *Mecp2* is selectively activated in nestin-expressing cells of the CNS and PNS, including neurons and glia, and remains silenced in all other cell types (**Fig. 3.1-B**).

To confirm that MeCP2 expression was indeed limited to cells of the nervous system, Southern blot analysis of a panel of tissues was conducted by Dr. Jacky Guy, Edinburgh University (**Fig. 3.2**). Genomic DNA was extracted from tissues and digested with EcoRI and NcoI. Southern blots were then probed with a fragment covering the coding region of exon 4. Distinct band sizes are produced depending on whether cells have WT MeCP2 expression (3.2 kb), a stop cassette present (5.1 kb), or a stop cassette excised by cre mediated recombination (4.3 kb; **Fig. 3.1-A**). The results showed that in stop-cre mice the stop cassette had been successfully excised in all brain tissue examined, as evidenced by the presence of 4.3 kb bands on the blot (**Fig. 3.2-A**). Very faint bands at 5.1 kb in forebrain and midbrain/hind brain samples indicate a small proportion of cells in which the stop cassette was not removed. In contrast, when peripheral tissue was analysed, examples from lung and kidney tissue are shown (**Fig. 3.2-A**), strong bands at 5.1 kb were seen, indicating the presence of the stop cassette. Faint bands at 4.3 kb indicate a small proportion of cells in which stop cassette has been excised. When band intensities were quantified using ImageJ (**Fig. 3.2-B**), recombination frequency was shown to be extremely high in the brain, with 91.9% efficiency in whole brain samples. Recombination was particularly high in the cerebellum which showed 96.4% efficiency. Robust silencing was shown in peripheral tissues with only very small levels of recombination in liver (0.9%), spleen (0.5%), skeletal muscle (1.2%) and heart (7.4%). Higher recombination levels were seen in lung (14.3%) and kidney (24.4%) tissue. To confirm the results seen in the Southern blot, tissue sections were prepared and immunolabelled with anti- MeCP2 antibody (**Fig. 3.3**). Results showed strong staining in brainstem as expected, and minimal staining in skeletal muscle and lung. Staining in cardiac tissue was slightly higher than expected from Southern blot data, but the reasons for this were not clear.



**Figure 3.2 – Southern blot analysis reveals tissue specific *Mecp2* reactivation**

Silencing in peripheral tissues confirmed by southern blot of restriction digested genomic DNA. (A) Representative blots from genomic DNA extracted from various tissues. DNA extracts probed with fragment covering the coding region of exon 4 of the *Mecp2* gene. Black arrows indicate location of expected 3 bands: WT (3.2 kb), stop-cassette (5.1 kb), excised stop cassette (4.3 kb). (B) Plot of mean recombination frequency  $\pm$  S.D in stop-cre mice (n = 2-4 mice) as determined by band intensity on Southern blots. (Work carried out by Dr. Jacky Guy – Edinburgh University)

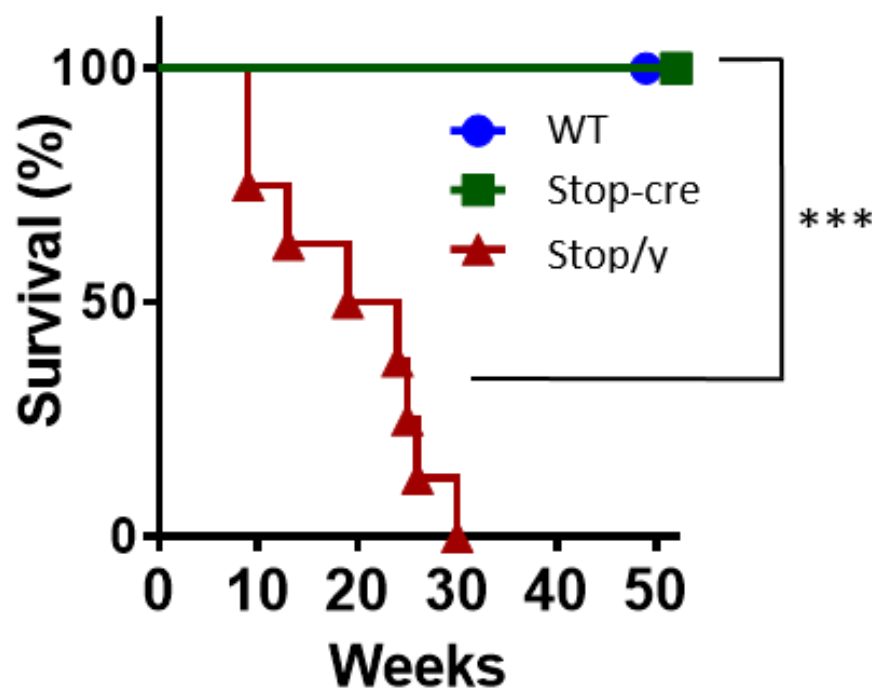


**Figure 3.3 – Immunoblots confirm robust peripheral silencing**

Representative bright field images from stop-cre tissue sections showing anti-MeCP2 peroxidase labelled nuclei (brown). Tissues shown are (i) skeletal muscle (ii) cardiac muscle (iii) lung and (iv) brainstem. Arrows indicate nuclei expressing MeCP2 in peripheral tissue. Scale bar is 20  $\mu$ m. (Images taken by Dr. Noha Bahey – Glasgow University)

### 3.4 Absence of early lethality in peripheral KO mice

Previous studies have shown that global KO of *Mecp2* expression in male mice, using a stop cassette, leads to early lethality (Guy et al., 2007; Robinson et al., 2012). Similarly, in this study, stop/y mice showed decreased survival (median survival = 150 days) compared to both WT and stop-cre animals ( $p < 0.001$ ; **Fig. 3.4**). In contrast, stop-cre animals did not show this early lethality, with all mice surviving until the end of the study.

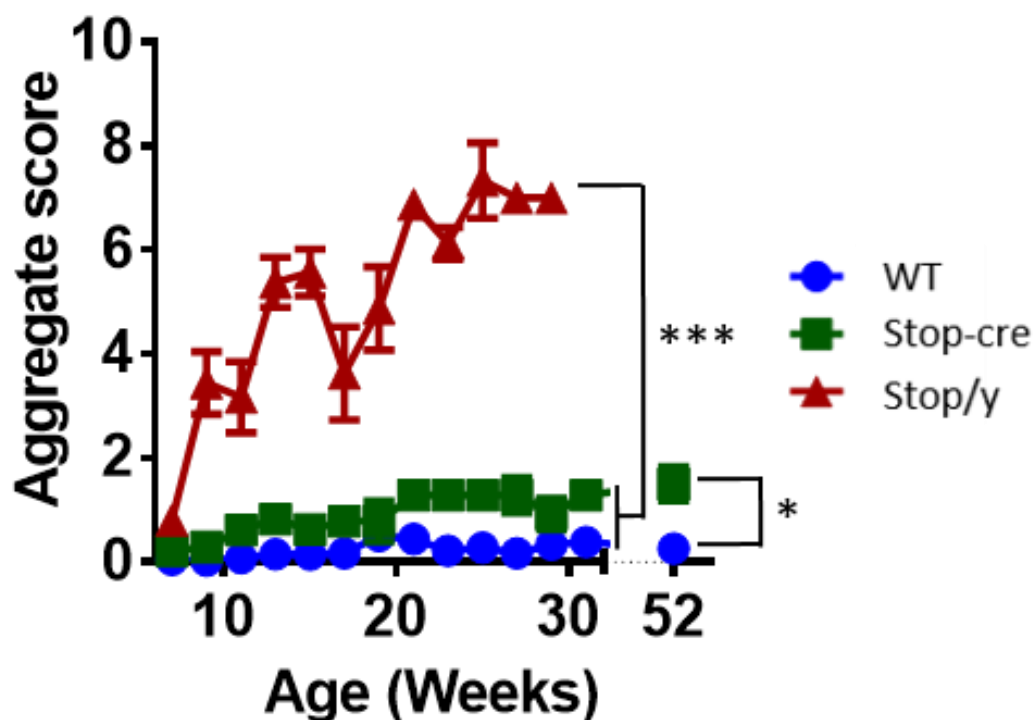


**Figure 3.4 – Normal survival in peripheral KO mice**

Mice were monitored for early lethality. Results show percentage survival of WT (blue circle), stop-cre (green square) and stop/y (red triangle). Mice were monitored over the course of a 52 week study. Stop/y mice showed significantly decreased survival compared to both WT and stop-cre ( $p < 0.001$ ; Kruskal-Wallis test and Dunn's *post hoc* analysis). WT ( $n = 17$ ); stop-cre ( $n = 7$ ); stop/y ( $n = 8$ ). \*\*\* $p < 0.001$ .

### 3.5 Absence of RTT-like symptoms in peripheral KO mice

To determine if peripheral KO mice developed any of the symptoms commonly associated with RTT mouse models, mice were monitored over a period of 52 weeks for gross symptoms of the disorder using a well-established severity scoring system (Guy et al., 2007; Daniel T. Lioy et al., 2011; Gadalla et al., 2012). In stop/y mice, symptoms became overt at 10 weeks of age and increased over time until death (**Fig. 3.5**). At the time of behavioural testing (15 weeks; **Fig. 3.7-3.12**) stop/y mice differed from both WT and stop-cre ( $p < 0.001$ ) but stop-cre mice did not differ from WT. However, when mice were scored at 52 weeks, stop-cre mice had developed some mild RTT-like symptoms, mostly related to hypoactivity, that were not seen in WT mice (aggregate score: WT =  $0.3 \pm 0.015$ ; stop-cre =  $1.5 \pm 0.3$ ;  $p < 0.001$ ).

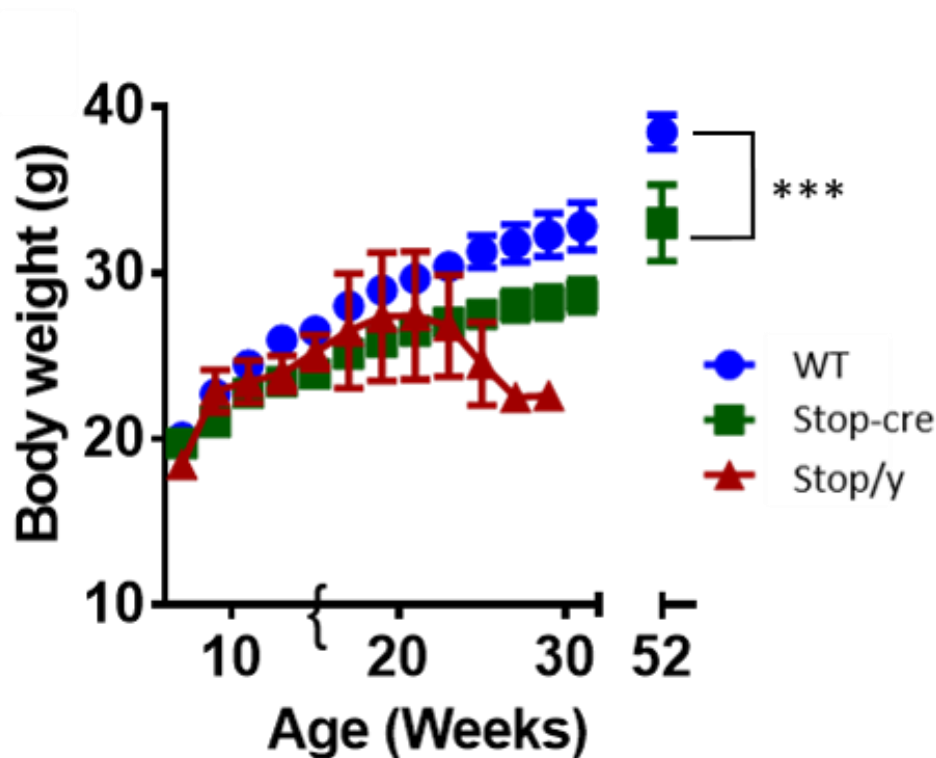


**Figure 3.5 – Mild RTT-like symptoms displayed by stop-cre mice**

Mice were scored weekly for RTT-like symptoms. Results show mean composite score  $\pm$  SEM of WT (blue circle), stop-cre (green square) and stop/y (red triangle). Mice were scored for the first 30 weeks and then again at 52 weeks. Groups were compared using Kruskal-Wallis test and Dunn's *post hoc* analysis at 15 weeks and Mann-Whitney U at 52 weeks. WT ( $n = 17$ ); stop-cre ( $n = 7$ ); stop/y ( $n = 8$ ). \*\*\* $p < 0.001$

### 3.6 Peripheral KO mice show reduced bodyweight

To assess bodyweight changes, animals were weighed weekly over the course of the study. Analysis of the results revealed no difference between the genotypes at time of behavioural testing (15 weeks), but by one year stop-cre mice showed a decreased bodyweight compared to WT mice (**Fig 2C**; mean bodyweight: WT =  $38.5 \pm 1$ ; stop-Cre =  $33 \pm 2.3$  g;  $p < 0.05$ ).



**Figure 3.6 - Decreased bodyweight in peripheral KO mice**

Plot showing bodyweight changes over the course of the study. Results show mean bodyweight  $\pm$  SEM of WT (blue circle), stop-cre (green square) and stop/y (red triangle). Also indicated is the time-point of behavioural testing ({}). Mice were weighed for the first 30 weeks and then again at 52 weeks. Groups were compared using one-way ANOVA with Tukey's *post hoc* analysis at 15 weeks and two-tailed Student's unpaired t-test at 52 weeks. WT (n = 17); stop-cre (n = 7); stop/y (n = 8). \* $p < 0.05$



### 3.7 Peripheral KO mice show no changes in blood biochemistry

In addition to gross symptoms and bodyweight, mice from each genotype were also examined for routine blood biochemistry (**Table 3.1**). Blood serum measures showed modest changes in a small number of markers in the stop/y group compared to WT.

However, there were no significant differences between stop-cre and WT samples across all measures.

**Table 3.1. Blood biochemistry results**

	Mean value $\pm$ S.E.M		
	WT	Stop-cre	Stop/y
Sodium (mmol/l)	151.7 $\pm$ 0.64	153.4 $\pm$ 0.82	153.1 $\pm$ 1.80
Potassium (mmol/l)	7.22 $\pm$ 0.22	7.83 $\pm$ 0.33	8.43 $\pm$ 1.01
Sodium:potassium ratio	21.29 $\pm$ 0.70	20.86 $\pm$ 0.40	18.78 $\pm$ 2.04
Chloride (mmol/l)	110.9 $\pm$ 0.45	109.4 $\pm$ 0.54	108.1 $\pm$ 1.25 *
Calcium (mmol/l)	2.48 $\pm$ 0.02	2.49 $\pm$ 0.04	2.48 $\pm$ 0.07
Phosphate (mmol/l)	2.64 $\pm$ 0.16	2.95 $\pm$ 0.10	3.81 $\pm$ 0.68 *
Urea (mmol/l)	10.72 $\pm$ 0.58	10.8 $\pm$ 0.52	20.6 $\pm$ 9.40
Creatinine (umol/l)	30.5 $\pm$ 0.90	31.55 $\pm$ 0.86	41.22 $\pm$ 13.11
Cholesterol (mmol/l)	3.26 $\pm$ 0.09	3.65 $\pm$ 0.15	3.01 $\pm$ 0.23
Triglyceride (mmol/l)	2.53 $\pm$ 0.27	2.54 $\pm$ 0.33	2.20 $\pm$ 0.25
Total bilirubin (umol/l)	2.57 $\pm$ 0.31	4.55 $\pm$ 1.12	3.14 $\pm$ 0.67
ALK phos (U/l)	245.9 $\pm$ 12.3	247 $\pm$ 28.89	297.4 $\pm$ 30.59
AST (U/l)	127.3 $\pm$ 15.53	164 $\pm$ 55.85	201.7 $\pm$ 43.33
ALT (U/l)	44.79 $\pm$ 7.27	44.67 $\pm$ 4.47	153 $\pm$ 73.11
GGT (U/l)	Not detectable	Not detectable	Not detectable
Total protein (g/l)	53.64 $\pm$ 0.78	56.09 $\pm$ 1.16	58.11 $\pm$ 1.33 *
Albumin (g/l)	29.79 $\pm$ 0.54	31.27 $\pm$ 0.45	33 $\pm$ 1.08 *
Globulin (g/l)	23.86 $\pm$ 0.43	24.82 $\pm$ 0.80	25.11 $\pm$ 0.59
Albumin: globulin ratio	1.25 $\pm$ 0.026	1.27 $\pm$ 0.03	1.32 $\pm$ 0.04

Arterial blood was sampled at 15 weeks for biochemical analysis. Data show mean values  $\pm$  SEM. Groups were compared using one-way ANOVA with Tukey's *post hoc* analysis. Group sizes were WT = 14, stop-cre = 11, stop/y = 9. \*  $p < 0.05$



### 3.8 Peripheral KO mice show mild kidney pathology

Mice were also examined for histopathological changes across a large panel of tissues and organs (**Table 3.2**). Tissue sections were H&E stained and examined, blind to genotype, by a qualified veterinary pathologist. This histopathological evaluation revealed no gross structural or histopathological changes for the majority of tissues examined. However, kidney sections from stop-cre and stop/y mice revealed mild to moderate vacuolation in renal tubule epithelium, potentially indicative of lipid accumulation. Moderate to diffuse vacuolation was also observed in liver but this was seen sporadically and across all genotypes.

**Table 3.2. Histopathological screening results**

	Wild Type			Stop-cre			Stop/y		
	1	2	3	1	2	3	1	2	3
Heart	X	X	X	X	X	X	X	X	X
Lungs	X	X	X	X	X	X	X	X	X
Liver	X	X	★	★	X	X	★	★	X
Salivary Glands	X	X	X	X	X	X	X	X	X
Mandibular Nodes	X	X	X	X	X	X	X	X	X
GI System	X	X	X	X	X	X	X	X	X
Pancreas	X	X	X	X	X	X	X	X	X
Kidneys	X	X	X	⌘	⌘	⌘	⌘	X	⌘⌘
Adrenal Gland	X	X	X	X	X	X	X	X	X
Spleen	X	X	X	X	X	X	X	X	X
Skin Dorsal	X	X	X	X	X	X	X	X	X
Skin Inguinal	X	X	X	X	X	X	X	X	X
Thymus	X	X	X	X	X	X	X	X	X
Mesenteric Nodes	X	X	X	X	X	X	X	X	X
Femur & Knee Joint	X	X	X	X	X	X	X	X	X
Sternum	X	X	X	X	X	X	X	X	X
Cranium	X	X	X	X	X	X	X	X	X
Brain	X	X	X	X	X	X	X	X	X
Tail	X	X	X	X	X	X	X	X	X

H&E stained tissue sections for organs listed above were assessed for histopathological changes by a veterinary pathologist. 3 animals were examined per genotype. Results of analysis are indicated using key below.

X No gross of histopathological changes observed

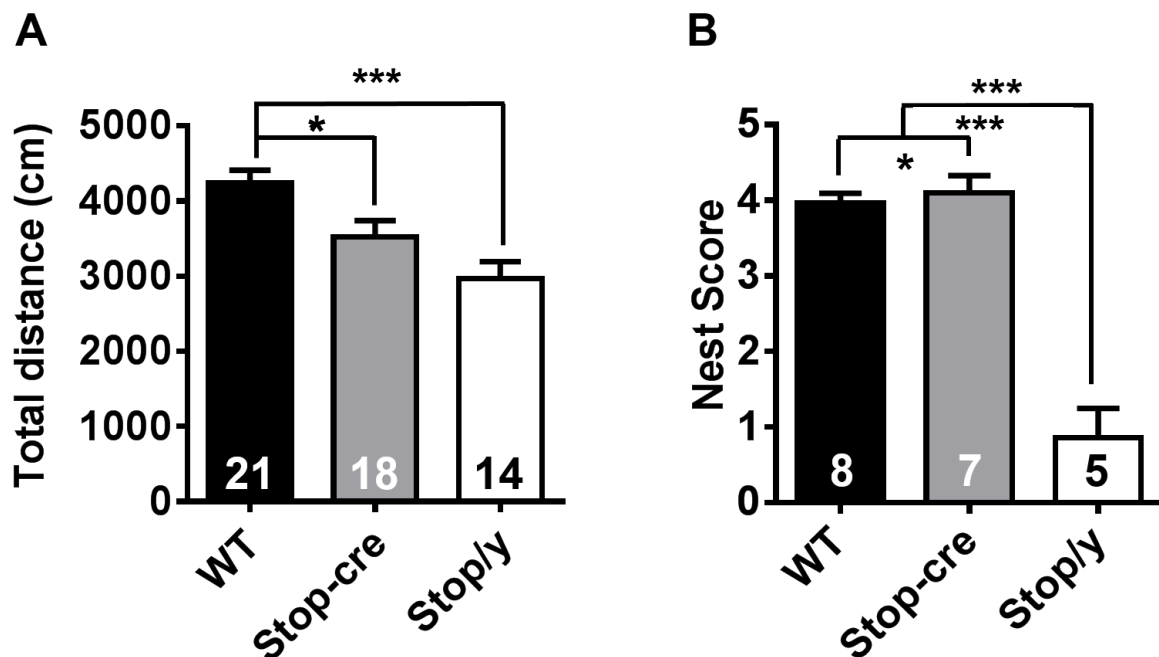
★ Moderate to diffuse coarse cytoplasmic clearing/vacuolation

⌘ Very mild vacuolation in the renal tubule epithelium

⌘⌘ Moderate vacuolation in the renal tubule epithelium

### 3.9 Peripheral KO mice show mild hypoactivity

Decreased spontaneous activity is a commonly observed feature seen in RTT mouse models (Shahbazian et al., 2002b; Goffin et al., 2012; Lyst et al., 2013). In order to assess activity levels, mice were monitored whilst ambulating freely in an open-field arena. Both stop-cre and stop/y mice showed a reduction in the total distance moved during the trial compared to WT mice (distance moved in 10 mins; WT =  $4242 \pm 167$ ; stop-cre =  $3523 \pm 215$ ; stop/y =  $2963 \pm 230$  cm;  $p < 0.05$ ; **Fig 3.7-A**). The reduction was more pronounced in stop/y mice compared to stop-cre mice. Both stop-cre and stop/y mice also showed a reduction in the number of rearing instances (regarded as a measure of exploratory behaviour) compared to WT mice, although again the difference was more moderate in stop-cre mice compared to stop/y (rearing instances per session; WT =  $35.00 \pm 3.5$ ; stop-cre =  $22.36 \pm 2.9$ ; stop/y =  $15.00 \pm 2.2$ ;  $p < 0.05$ ; **Fig 3.7-B**).

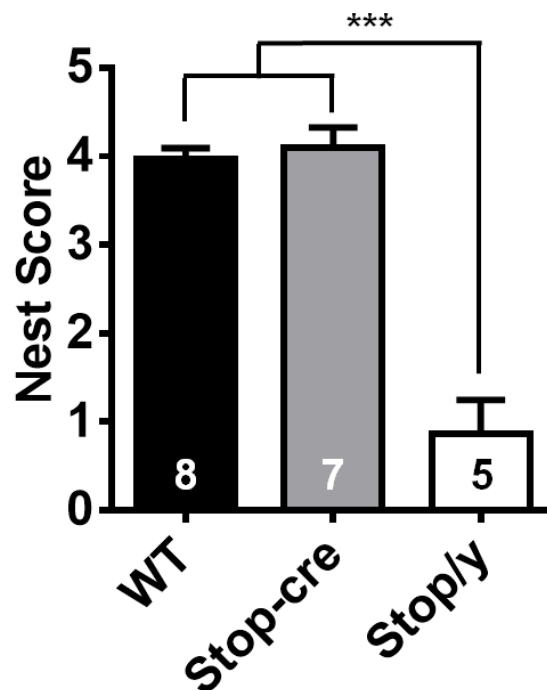


**Figure 3.7 – Peripheral KO mice show reduced activity levels**

Spontaneous activity levels were assessed using the open-field test. Results show (A) total distance moved during the session and (B) number of rearing events per session. Results displayed as mean value  $\pm$  SEM. Number of animals per group displayed in results columns. Mice tested between 14-16 weeks of age. Groups were compared using one-way ANOVA with Tukey's *post hoc* comparisons. \* $p < 0.05$ , \*\*\* $p < 0.001$ . No significant difference between stop-cre and stop/y mice.

### 3.10 Peripheral KO mice construct normal nests

A number of factors may influence performance in the open field including motivational and cognitive factors, motor function and anxiety. In an attempt to discriminate between these factors, mice were scored for their nest-building behaviour, a test that can reveal brain-wide deficits, including motor defects (Deacon, 2006). In agreement with previous studies of *Mecp2*-null mice (Moretti et al., 2005), stop/y mice showed a profound reduction in nest-building behaviour compared to WT (**Fig. 3.8**; nesting score; WT =  $4 \pm 0.4$ ; stop-cre =  $4.1 \pm 0.25$ ; stop/y =  $0.9 \pm 0.15$ ;  $p < 0.001$ ). In contrast, stop-cre mice did not differ from WT in their ability to construct complex nest structures ( $p > 0.05$ ).

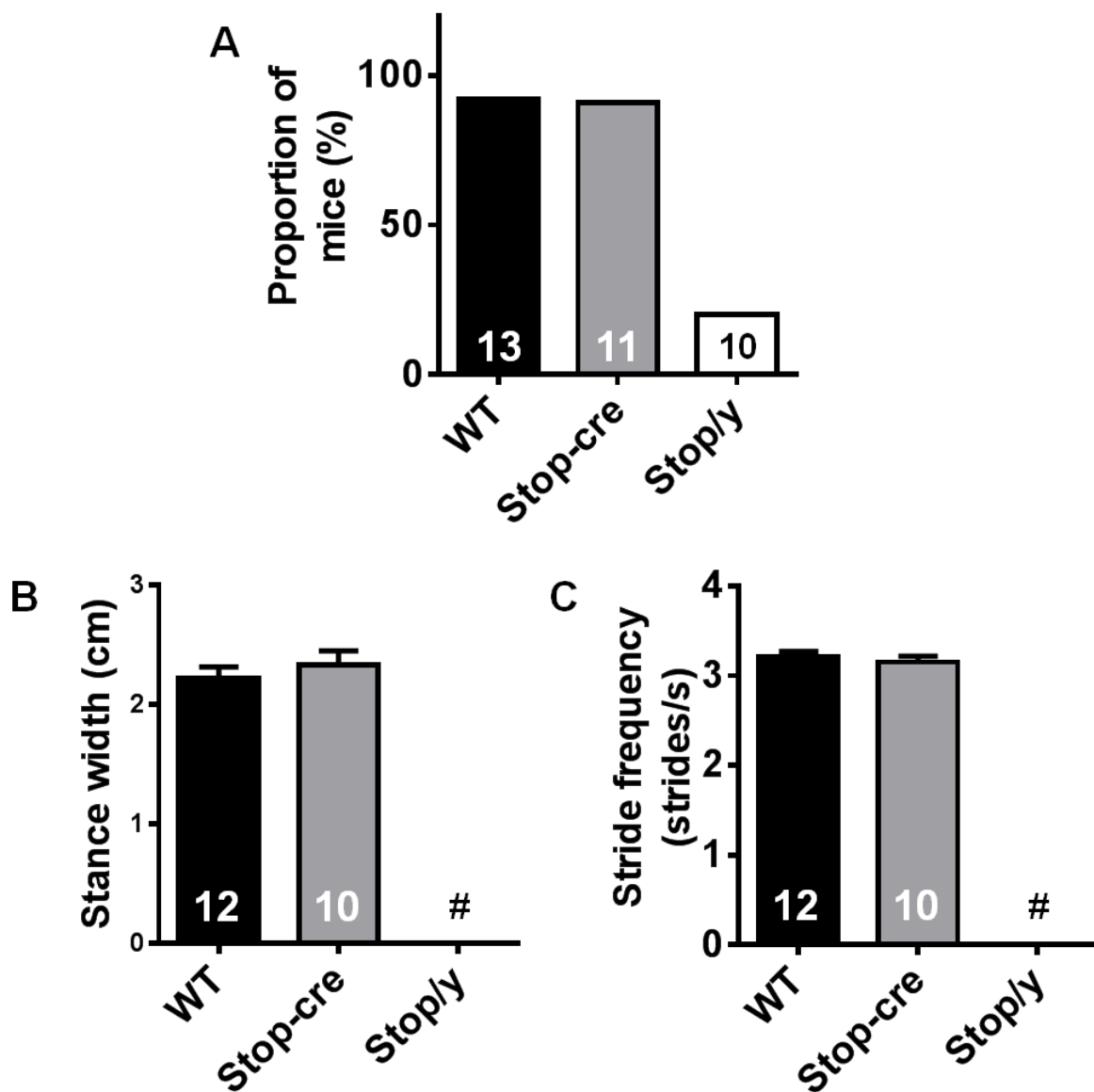


**Figure 3.8 – Peripheral KO mice were capable of constructing high quality nests**

Nesting behaviour was assessed using a five-point scoring system. Results show mean nesting score  $\pm$  SEM. The number of animals per group is displayed in the results columns. Mice were tested between 14-16 weeks of age. Groups were compared using one-way ANOVA with Tukey's *post hoc* comparisons. \*\*\* $p < 0.001$

### 3.11 Peripheral KO mice do not show RTT-like gait defects

Gait abnormalities are a characteristic feature of RTT mouse models and form part of the observational scoring system described previously. In order to comprehensively examine gait, mice were assessed using an elevated treadmill-based system. A previous study utilising this approach has shown stop/y mice to develop a number of gait disturbances that increase in severity as the disease progresses between four and ten weeks of age (Gadalla et al., 2014). To acquire data for this test, mice must be capable of moderate running (10 cm/s) on the treadmill for several seconds. It was found that at the time of testing in this current study (14-16 weeks of age), the majority of symptomatic stop/y mice (8 of 10) were incapable of performing to criterion (**Fig. 3.9A**). As a result, only WT and stop-cre mice could be compared using this test. When tested, stop-cre mice showed no significant differences from WT across a wide range of gait parameters, including stride frequency (**Fig. 3.9-B**) and stance width (**Fig 3.9-C**).

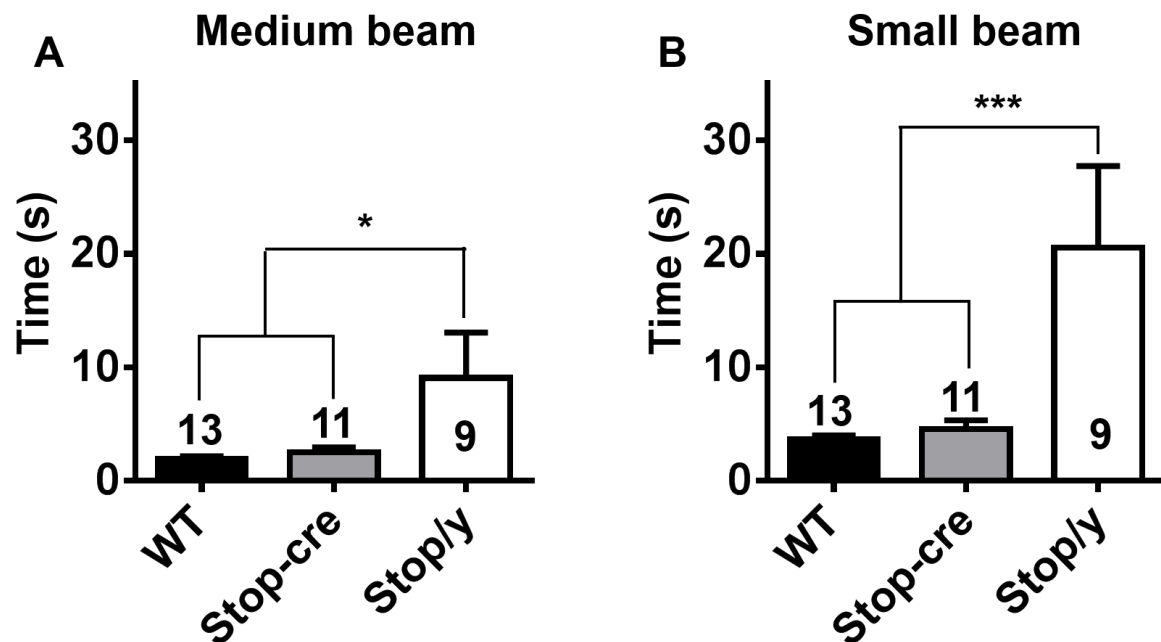


**Figure 3.9 – Majority of stop/y mice incapable of performing on treadmill test**

Mice were tested for their ability to run on a motorised treadmill at a speed of 10 cm/s. Results show (A) Proportion of mice capable of running on the treadmill for at least two seconds (B) stance width and (C) stride frequency. For B and C, results are mean value  $\pm$  SEM and groups were compared using unpaired t-test. No significant difference seen between WT and stop-cre mice. Number of animals per group is displayed in the results columns. Mice were tested between 14-16 weeks of age. # Mice were not tested as they couldn't meet performance criteria.

### 3.12 Peripheral KO mice show normal performance on balance beam

To further assess locomotor ability, balance was assessed in the mice using an inclined beam. Mice were tested using either a medium width (11 mm) or a more challenging narrow width (5 mm) beam. Compared to WT and stop-cre mice, stop/y mice showed a significant increase in the time taken to traverse both the medium (11 mm; WT =  $1.9 \pm 0.25$ ; stop-cre =  $2.5 \pm 0.44$ ; stop/y =  $9.1 \pm 4$  s;  $p < 0.05$ ; **Fig. 3.10-A**) and narrow beam (5 mm; WT =  $3.7 \pm 0.3$  s; stop-cre =  $4.6 \pm 0.8$  s; stop/y =  $20.6 \pm 7.2$  s;  $p < 0.01$ ; **Fig. 3.10-B**). In contrast, stop-cre mice did not differ significantly from WT animals on either beam.

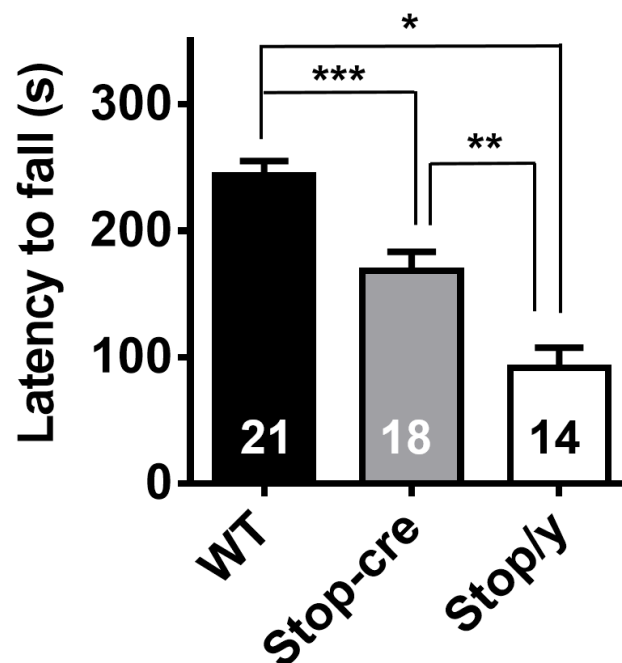


**Figure 3.10 – Peripheral KO animals do not show impairment on balance beams**

Mice were tested for balance using an inclined balance beam test. Results show time taken to traverse a (A) medium or (B) narrow width beam. Data shows mean  $\pm$  SEM. Number of animals per group is displayed in results columns. Mice tested between 14-16 weeks of age. Groups were compared using one-way ANOVA with Tukey's *post hoc* comparisons. \* $p < 0.05$ , \*\*\* $p < 0.001$ .

### 3.13 Reduced rotarod performance in peripheral KO mice

A more challenging test of balance and coordination is the rotarod test. Again, stop/y mice showed a significantly reduced performance compared to both WT and stop-cre mice (latency to fall; WT =  $243.5 \pm 11.5$ ; stop-cre =  $168 \pm 14.9$ ; stop/y =  $91.5 \pm 16.1$  s;  $p < 0.001$ ; **Fig. 3.11**). However, stop-cre mice also showed a significantly reduced performance in comparison to WT mice, although this reduction was more moderate than that for stop/y mice. Although the rotarod is mostly commonly used to test balance and coordination, it also sensitive to endurance fatigue. This is because the rod accelerates over time, and consequently the longer the mouse lasts without falling, the more energetically demanding the task becomes.

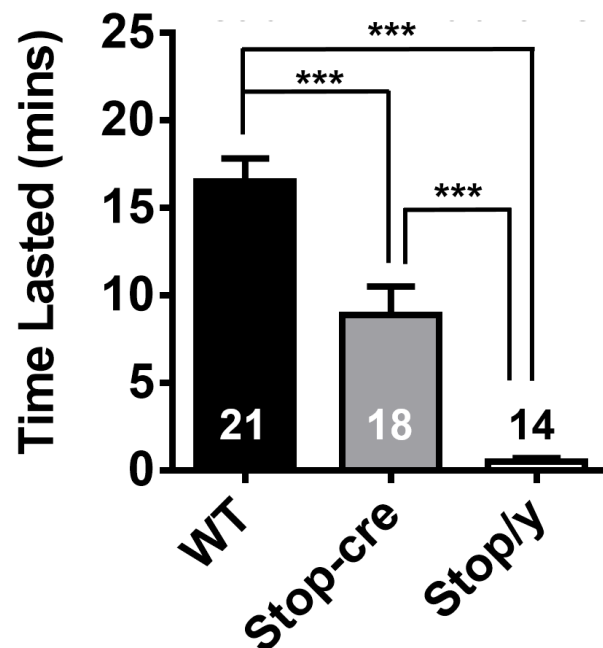


**Figure 3.11 – Peripheral KO mice show reduced rotarod performance**

Mice were tested using an accelerating rotarod. Results show mean latency to fall  $\pm$  SEM. Number of animals per group displayed in results columns. Mice tested between 14-16 weeks of age. Groups were compared using one-way ANOVA with Tukey's *post hoc* comparisons. \*\*\* $p < 0.001$ .

### 3.14 Peripheral KO mice show reduced exercise capacity

To investigate if stop-cre mice were indeed prone to exercise fatigue, they were tested on an accelerating inclined treadmill (Narkar et al., 2008). To ensure that mice performed to their genuine capacity, an electric grid was placed at the bottom of the treadmill. This grid administered a mild electric shock to the paws of mice who attempted to come off the treadmill. Although mild, this aversive stimulus motivated mice to carry out running on the treadmill until fatigued. This helped to remove motivational confounding factors. Stop/y mice showed a profoundly reduced capacity for exercise compared to WT and stop-cre animals (time lasted: WT =  $16.5 \pm 1.3$ ; stop-cre =  $8.7 \pm 1.6$ ; stop/y =  $0.5 \pm 0.2$  min;  $p < 0.001$ ; **Fig. 3.12**). Interestingly, stop-cre mice also showed marked impairment, with an average reduction of 46% in time lasted on the treadmill compared to WT mice ( $p < 0.001$ ). These results, in combination with those from the rotarod test suggest a significant impairment in exercise capability in peripheral KO mice.



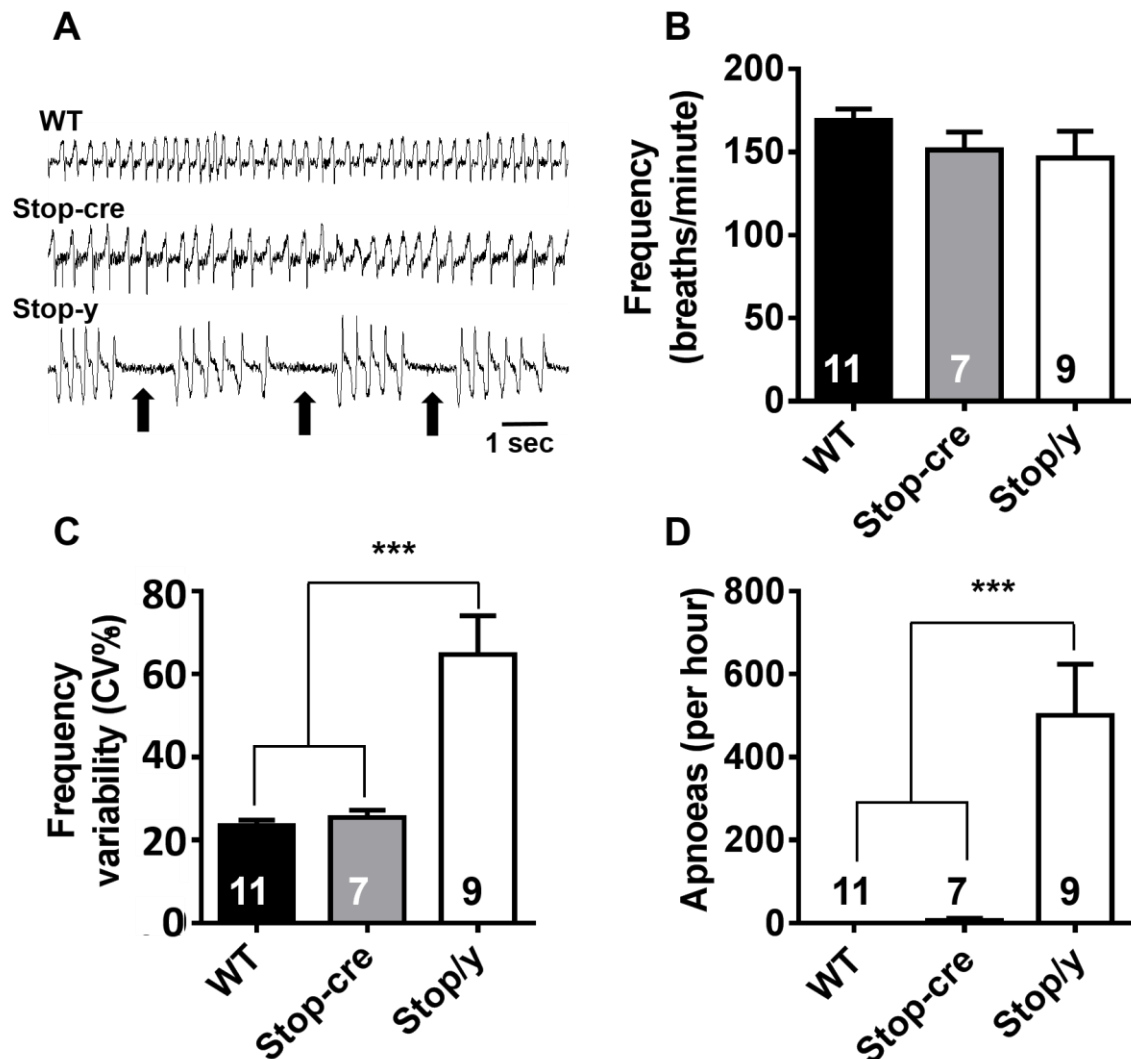
**Figure 3.12 – Reduced exercise capacity in peripheral KO mice**

Exercise capacity was assessed using an elevated, accelerating treadmill. Results show mean time lasted on the treadmill  $\pm$  SEM. Number of animals per group displayed in results columns. Mice tested between 14-16 weeks of age. Groups compared using one-way ANOVA with Tukey's *post hoc* comparisons. \*\*\* $p < 0.001$ .



### 3.15 Absence of RTT-like breathing phenotype in peripheral KO mice

Respiratory dysfunction, including breath holding and apnoeas, is a hallmark feature of RTT in human patients (Elian and Rudolf, 1991; Marcus et al., 1994; Weese-Mayer et al., 2006) and is also seen in *Mecp2* KO mouse models (Viemari et al., 2005; Ogier et al., 2007; Stettner et al., 2007). While these abnormalities are typically attributed to disturbed autonomic function, MeCP2 is highly abundant in lung tissue and pulmonary lesions have been observed in both RTT patients and mouse models (De Felice et al., 2010). This suggests that absence of MeCP2 from lung tissue could contribute to the respiratory phenotype. To investigate this further, we used plethysmography to assess breathing in our mice. Respiratory traces (**Fig. 3.13-A**) were analysed for breathing frequency and regularity, as well as for apnoeas. There was no difference in breathing frequency between the three genotypes (**Fig. 3.13-B**). Stop-cre mice showed a very regular breathing pattern (mean CV% =  $26 \pm 2$ ; **Fig 3.13-C**) that did not differ from WT (mean CV% =  $26 \pm 1$ ). In contrast, stop/y mice showed a highly irregular breathing pattern (mean CV% =  $65 \pm 9$ ) compared to both WT and stop-cre mice ( $p < 0.001$ ). A similar pattern was seen when traces were examined for the presence of apnoeas. Both WT and stop-cre mice showed very little occurrence of apnoeas during testing (apnoea number: WT = 0; stop-cre =  $7 \pm 4.61$  / hr), while stop/y animals showed an extremely high incidence of such events (apnoea number:  $501 \pm 123.7$  / hr) that differed significantly from both WT and stop-cre mice ( $p < 0.001$ ). In addition to this functional testing, we also looked for the presence of pulmonary lesions. Lung tissue biopsies were taken (three per genotype) and sent to a veterinary pathologist for comprehensive histopathological examination. However, no abnormalities or pathological signs were seen in any of the genotypes (**Table 3.2**).

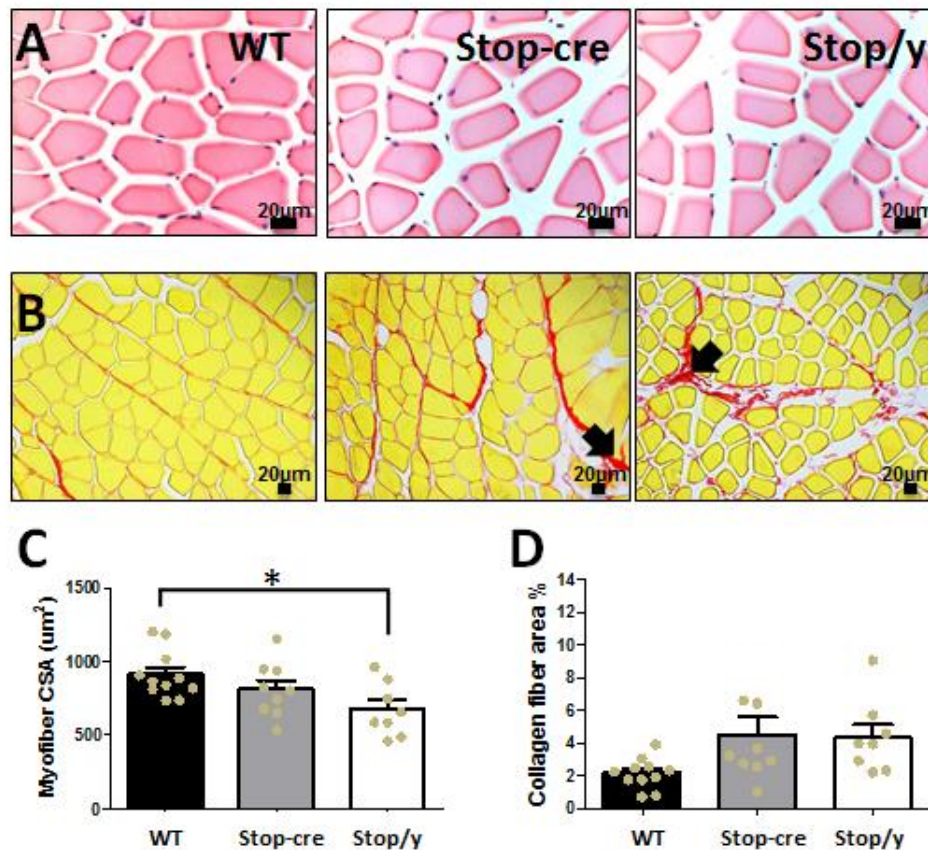


**Figure 3.13 – Absence of respiratory dysfunction in peripheral KO mice**

Respiratory function assessed using whole-body plethysmography. Results show (A) representative respiratory traces from each genotype, (B) baseline breathing frequency, (C) breathing frequency variability (D) apnoea frequency. Arrows indicate occurrence of apnoeas in stop/y mice. Results for B, C & D show mean  $\pm$  SEM. Number of animals per group displayed in results columns. Mice tested between 14-16 weeks of age. Groups compared using one-way ANOVA with Tukey's *post hoc* comparisons. \*\*\* $p < 0.001$ .

### 3.16 Evidence of muscle pathology in peripheral KO mice

Having assessed some of the core phenotypes that define RTT, peripheral tissue systems that have been previously shown to be altered in RTT were then examined. Recent reports suggest altered morphology of skeletal muscle in *Mecp2* KO mice including reduced fibre diameter (Conti et al., 2015). Histological examination of gastrocnemius muscle showed a similar pattern of reduced fibre cross sectional area in stop/y mice compared to WT and stop-cre ( $p < 0.05$ ; **Fig. 3.14-A&C**). Sections were also stained for collagen and revealed evidence of fibrosis in some stop-cre and stop/y mice (**Fig. 3.14. B&D**), although the overall increased proportion of tissue section area showing picosirius red staining did not achieve statistical significance from WT ( $p = 0.08$  WT vs stop-cre;  $p = 0.07$  WT vs stop/y).

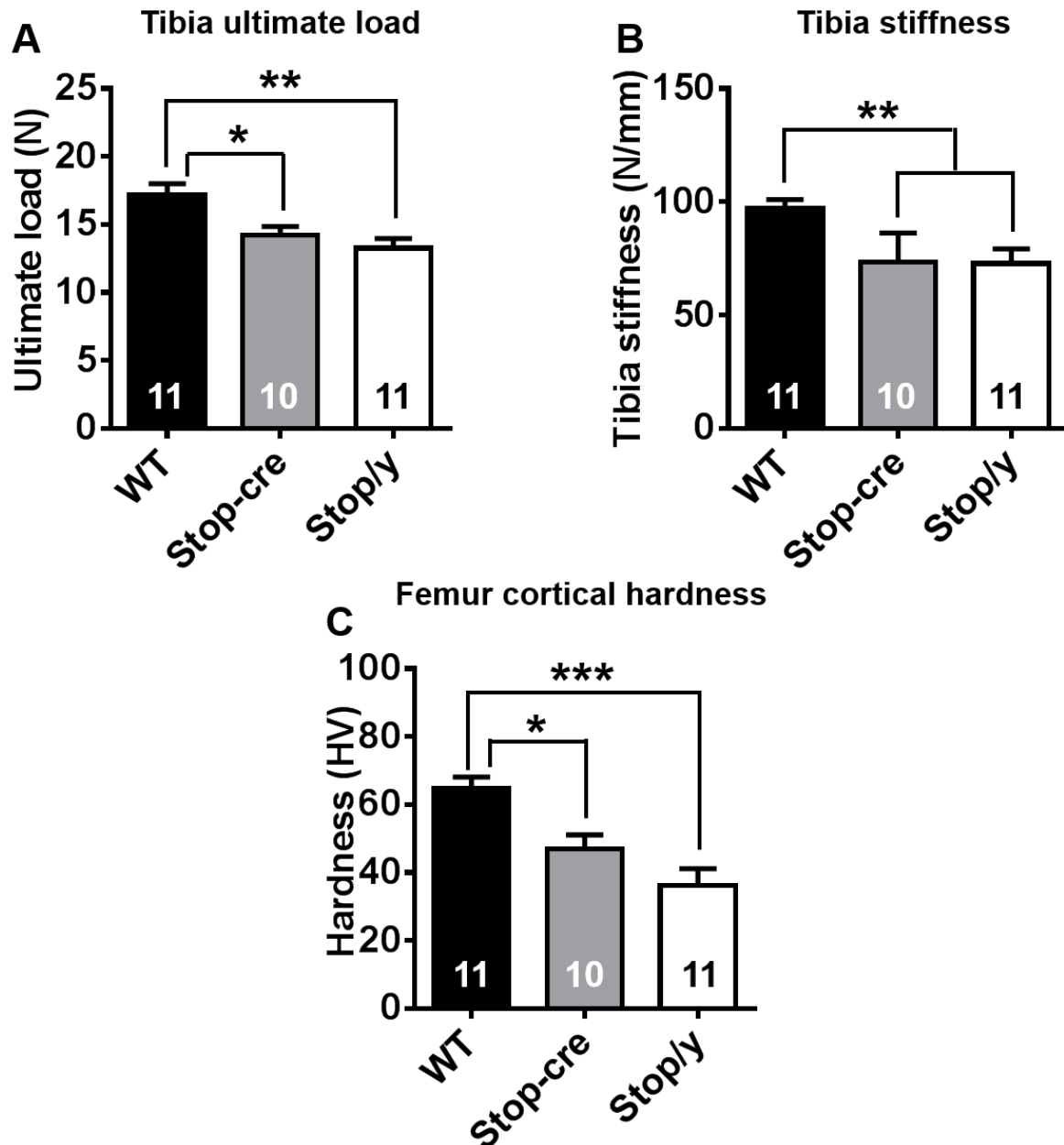


**Figure 3.14 - No significant muscle abnormalities in peripheral KO mice**

Representative images of gastrocnemius muscle cross section stained with (A) haematoxylin and eosin (H&E) or (B) picosirius red for collagen fibers. Black arrows indicate areas of concentrated collagen (fibrosis) (C) myofiber cross sectional area for each genotype calculated from H&E sections. (D) collagen fibre area % from each genotype calculated from PSR stained sections. Results show mean  $\pm$  S.E.M. Each black circle represents mean value for an individual animal. Animals per genotype are WT = 11, stop-cre = 9, and stop/y = 8. Groups were compared using one-way ANOVA and Tukey's *post hoc* comparisons. \* $p < 0.05$

### 3.17 Peripheral KO mice show RTT-like bone phenotypes

RTT patients show a number of skeletal anomalies including spinal deformity, osteopenia, osteoporosis, and an increased vulnerability to low-impact fractures (Keret et al., 1988; Guidera et al., 1991; Leonard et al., 1999; Zysman et al., 2006; Downs et al., 2008; Percy et al., 2010). Studies in RTT KO mouse models have shown structural and functional defects in bone tissue (O'Connor et al., 2009; Kamal et al., 2015), but since these studies used global KO models it was not clear whether the defects were central or peripheral in origin. To answer this question we carried out biomechanical testing on long bone samples. Functional assessment of the maximum force that the tibial bones could resist revealed that both stop-cre and stop-y mice showed a reduced ultimate load compared to WT mice (ultimate load: WT =  $15.83 \pm 0.56$ , stop-cre =  $13.93 \pm 0.85$ , stop/y =  $12.22 \pm 0.67$  N;  $p < 0.05$ ; **Fig. 3.15-A**). A similar pattern was seen when tibia stiffness was assessed using a three-point bending test, with stop/y and stop-cre mice showing significantly reduced stiffness (tibia stiffness: WT =  $97.1 \pm 4.0$ , stop-cre  $73.31 \pm 4.1$ , stop/y =  $72.9 \pm 6.4$  N / mm;  $p < 0.01$ ; **Fig. 3.15-B**). Further biomaterial testing revealed the same pattern in femur, with a reduction in cortical bone hardness in both stop-cre and stop/y mice compared to WT (cortical hardness: WT =  $64.7 \pm 3.4$ ; stop-cre =  $47.0 \pm 4.1$ , stop/y =  $36.3 \pm 4.9$  HV;  $p < 0.05$ ; **Fig. 3.15-C**). Across all these tests stop-cre mice did not differ significantly from stop/y mice.



**Figure 3.15 – Peripheral KO mice show RTT-like bone phenotype**

The biomechanical properties of bone was assessed using functional tests. Results show (A) tibia ultimate load (B) tibia stiffness (C) cortical bone hardness in femur. Results displayed as mean value  $\pm$  SEM. Number of animals per group displayed in results columns. Mice tested between 14-16 weeks of age. Groups compared using one-way ANOVA with Tukey's *post hoc* comparisons. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . (Tests carried out by Dr. Bushra Kamal – Glasgow University).

### 3.18 Discussion

Recent studies have shown that gene based therapies can deliver significant therapeutic benefits in RTT mouse models (Gadalla et al., 2012; Garg et al., 2013). In order to be as effective as possible it is crucial that therapies are targeted to the correct location. It is widely assumed that the RTT phenotype is due solely to an absence of MeCP2 from the brain. This is surprising since MeCP2 is thought to be widely expressed throughout the body, including high levels of protein in post-mitotic cells of the heart and lungs (Shahbazian et al., 2002b; Song et al., 2014). While it is clear that absence of MeCP2 from the brain only, leads to a severe phenotype, including the early death seen in male KO mice, these studies have not carried out a vigorous phenotyping and have relied on gross markers of disease such as reduced brain size and weight loss (Chen et al., 2001; Guy et al., 2001; Giacometti et al., 2007). Thus the presence or absence of more subtle phenotypes has not been comprehensively examined. In addition, the severe phenotype trajectory and very early death seen in these models means that phenotypes which take longer to develop will not be detected. In order to systematically investigate the role of MeCP2 outside the nervous system, and to identify peripheral contributions to the RTT phenotype, a novel mouse model was generated (Dr. Jacky Guy, The University of Edinburgh), in which *Mecp2* is selectively expressed in neurons and glia of the CNS and PNS, but silenced in peripheral tissues, and an extensive phenotyping of male mice was carried out. Results from this phenotyping revealed that the majority of the RTT-like phenotype can indeed be attributed to a loss of functional MeCP2 in the brain, but that an absence of MeCP2 from peripheral tissues leads to a markedly reduced exercise capacity and defective bone properties.

The crucial role of the nervous system in RTT is shown by the lack of overt RTT-like symptoms in peripheral KO mice. There is a complete absence of the early lethality seen in male global KO mice (Chen et al., 2001; Guy et al., 2001). This is in agreement with a previous study that showed that deletion of *Mecp2* from the nervous system only leads to reduced survival similar to that seen in global KO mice. Comparing mice using the well-established observational scoring system revealed that stop-cre mice are almost indistinguishable from WT mice, although a mild hypoactivity can sometimes be detected. This is also observable when handling the stop-cre mice, as they appear much more passive and less active than WT. While no difference in bodyweight was observed in the first 30 weeks of the study, after which all stop/y mice had died, when mice were weighed again at 12 months there was a significant difference.

Further detailed phenotyping revealed that other major aspects of the RTT-like phenotype such as breathing abnormalities (Ogier et al., 2007; Voituren et al., 2009; Ramirez et al., 2013), and balance and gait disturbances (Santos et al., 2010; Robinson et al., 2012; Gadalla et al., 2014) were not detected in peripheral KO mice. In this study we rigorously assessed gait using a treadmill-based approach. A wide range of parameters were assessed and stop-cre animals did not differ from WT in any of these. Previous work has shown that pronounced gait defects can be detected using this treadmill system and these defects become more severe as the overall phenotype progresses over time (Gadalla et al., 2014). Indeed, at the relatively late time-point used in this study our global KO mice were not even capable of being assessed using this test as they were incapable of running on the treadmill. With regards to breathing, RTT mouse models typically show highly disturbed breathing patterns, including irregular breathing frequency, interspersed with periods of apnoea, which are also seen in human RTT patients (Viemari et al., 2005; Neul et al., 2010). These disturbances can be rescued by global reactivation of *Mecp2* expression in mice (Robinson et al., 2012) and are therefore a viable therapeutic target. A number of studies report disordered GABAergic and serotonergic control of brain respiratory networks as the major cause of the disrupted breathing and apnoeas that characterise RTT in both patients and mouse models (Viemari et al., 2005; Abdala et al., 2010; Chao et al., 2010; Voituren and Hilaire, 2011). However, one study (Bissonnette and Knopp, 2006) has reported global KO of *Mecp2* to cause increased incidences of apnoea in response to hypoxia induced hyperventilation, an effect that was not seen in nervous system-specific KO mice. Whilst such findings may suggest a non-neuronal element for MeCP2 in respiratory function, in the current study we did not specifically assess the response of mice to hypoxia. Instead, under baseline conditions, we observed a complete rescue of apnoeas and episodic breathing in stop-cre mice confirming that these characteristic RTT-like features, at least under resting conditions, are due to nervous system dysfunction. Previous reports in human patients have also suggested that global KO of *Mecp2* leads to the presence of pulmonary lesions (De Felice et al., 2010), as detected by CT imaging. This was supported by a study in mice, which showed the presence of diffuse inflammatory infiltrates in about half of the *Mecp2* KO mice examined (De Felice et al., 2014). We investigated this in our mice to see if this was due to absence of MeCP2 from lung tissue, or a secondary consequence of nervous system dysfunction. In contrast to these previous reports, no histopathological changes were seen in any of our mouse models, including the global KO stop/y mice. One reason for this could be that the stop/y mouse model is less severe than the null model used in the previous study (Robinson et al., 2012; De Felice et al., 2014). In

addition, only 50% of mice in the previous study showed the phenotype suggesting that this is a variable phenotype, possibly secondary to other dysfunction.

The results of the histopathological and blood serum biochemical screens showed that there were few differences between genotypes confirming a lack of widespread and overt tissue pathology as a result of MeCP2 deficiency in peripheral tissues. The vast majority of organs showed no signs of gross structural or pathological changes with the exception of the kidney where there was some evidence of tubular epithelium vacuolation in both stop-cre and stop/y mice, a feature often associated with lipid accumulation and disordered lipid regulation (Sastre et al., 2013; Zhou et al., 2015). Previous observations in *Mecp2* KO mice (Buchovecky et al., 2013) and in RTT patients (Segatto et al., 2014) have suggested that the absence of MeCP2 leads to altered cholesterol biosynthesis and an increase in serum cholesterol levels. There was no significant difference observed in serum cholesterol levels between genotypes in the current study. However, altered cholesterol metabolism is not seen in all lines of *Mecp2* KO mice (Buchovecky et al., 2013), with hypomorphic KO mouse lines not showing this phenotype. Previous work using the stop/y mouse line (Robinson et al., 2012) has shown that the mouse line is also hypomorphic, with background MeCP2 levels ~ 2-4% of WT being expressed, which may explain the lack of a cholesterol phenotype in our mice.

A particularly robust finding in this study was the marked reduction in exercise capacity and vulnerability to fatigue displayed by peripheral KO animals. Whilst levels of spontaneous activity in the open field test was only moderately reduced in comparison to WT, when animals were challenged by more intensive tasks (such as the elevated treadmill) the deficit was more pronounced, with an almost 50% decrease in performance compared to WT mice. This was less than the deficit seen in stop/y mice but nevertheless suggests an exercise fatigue phenotype that may be a true consequence of peripheral MeCP2 deficiency. Inertia and reluctance to movement is widely reported in *Mecp2* KO mice (Guy et al., 2007) and could explain the fatigue phenotype. However, the use of aversive stimulation on the exercise treadmill is designed to negate against such confounding motivational differences between mice. The exact cause of the exercise fatigue is therefore difficult to assess. We consider it unlikely to be due to any overt cardiorespiratory dysfunction as stop-cre did not differ from WT when assessed for a range of cardiac (data not shown) and respiratory measures. However, these parameters were only tested under baseline conditions and it is possible that future studies assessing cardiac and respiratory responses under exercise conditions may



uncover peripheral MeCP2-mediated phenotypes which could contribute to the observed fatigue phenotype.

Previous reports have shown altered mitochondrial ultrastructure and impaired mitochondrial function in skeletal muscle biopsies from RTT patients and in *Mecp2* KO mice (Gold et al., 2014). More recently, a study in mice demonstrated structural alterations in muscle fibre cross sectional area in global *Mecp2* KO mice, but this effect was not seen when *Mecp2* was selectively deleted in muscle tissue (Conti et al., 2015). The results of this study agree with these findings and suggest that these structural abnormalities disrupted nervous system function or aberrant skeletal innervation. We also looked for evidence of disorganisation and fibrosis in muscle tissue sections from mice in the current study by staining for collagen. While there was a clear trend towards increased collagen deposition in some mice in both stop-cre and stop/y cohorts but this did not reach statistical significance. While some individual mice showed 3-6 x average WT levels, others showed much lower levels, suggesting this is not a consistent phenotype, and may be secondary to other factors.

It has been shown previously that global absence of MeCP2 in mice leads to impairments in the structural and biomechanical properties of bone, and that these defects can be reversed by global reactivation of *Mecp2* expression (Hess et al., 2008; Goffin et al., 2012). This agrees with studies that show bone abnormalities, such as early osteoporosis and a vulnerability to low-energy fractures, in RTT patients (Guidera et al., 1991; Zysman et al., 2006; Downs et al., 2008). However, it was not possible to determine from these studies whether the primary cause of these defects was due to central or peripheral absence of MeCP2. Results from this study suggest that it is the absence from peripheral tissue that is primary cause of the dysfunction, as both stop-cre and stop/y mice show similar levels of dysfunction. This is important as it suggests that any therapies targeting this aspect of the RTT phenotype will need to be targeted outside the nervous system.

A criticism that could be levelled at the approach taken in this study is that, because 100% cre-mediated reactivation was not achieved in the brain, some of the moderate phenotypes detected could be due to the effect of the small amount of nervous system cells not expressing functional MeCP2. Recombination levels achieved in this study were extremely high, with ~90% of cells from whole-brain samples showing successful reactivation. However, previous studies in which *Mecp2* expression was reactivated globally, using a tamoxifen-cre strategy, achieved only 60-70% recombination in the nervous system (Robinson et al., 2012; Kamal et

al., 2015). Despite this lower efficiency, functional testing of these animals revealed a reversal of the majority of defects seen in KO animals, including full reversal of motor defects and bone abnormalities. In addition, in this study the overtly brain-specific measures such as balance and innate nest building behaviour in stop-cre mice were indistinguishable from WT measures. While nestin expression is mostly confined to the nervous system, expression has been detected in other organs, including kidney and pancreas (Delacour et al., 2004; Sclafani et al., 2006; Bertelli et al., 2007). In this study, a relatively high level of recombination in the kidneys, and above background levels in the lungs were observed. Other tissues such as the heart, liver and skeletal muscle showed very little recombination showing that *Mecp2* expression was very effectively silenced in these tissues. The combination of extremely high brain recombination levels coupled with robust peripheral silencing suggests that this is a valid model for examining the peripheral contribution to the RTT phenotype, and supports the interpretation that defects in phenotypes seen in the stop-cre mice are a consequence of the absence of MeCP2 in peripheral tissues.

Overall, the results from this chapter show that the majority of the RTT phenotype is due to loss of functional MeCP2 from the brain, and that this should be the major focus for targeting therapies. However, the results also suggest that a subset of phenotypes may be peripheral in origin and benefit from systemic treatments.

# Chapter 4

---

## Design and synthesis of CRISPR and TALEN constructs for targeting non-coding regions of *Mecp2*

### 4.1 Introduction

As discussed in Chapter 1, studies in mouse models of RTT have shown that the major symptoms of the disorder can be reversed, even after the onset of symptoms (Guy et al., 2007; Robinson et al., 2012). This suggests that novel therapies designed to replace or repair the dysfunctional *Mecp2* gene could be a potential treatment. Work in the previous chapter and other studies (Chen et al., 2001; Guy et al., 2001) has shown that it is the absence of functional MeCP2 from cells of the nervous system that is the cause of the majority of the RTT phenotype. This, combined with the fact that post-mitotic neurons have particularly high levels of MeCP2 (Shahbazian et al., 2002a), indicates that neurons should be the main target of any therapies. Since there is currently a lack of understanding of the downstream targets of MeCP2, the most attractive therapeutic strategy is to target the disorder at the gene level (Gadalla et al., 2011).

New genome editing technologies developed in the last few years enable targeted DNA changes to be made (Zhang et al., 2011; Cong et al., 2013), and their potential as therapeutic tools has been demonstrated in several studies (Perez et al., 2008; Schwank et al., 2013; Wang et al., 2014a; Yin et al., 2014). Most studies using these tools for therapy seek to repair disease causing mutations, or confer protective mutations, via precise homologous recombination based editing (Perez et al., 2008; Schwank et al., 2013; Yin et al., 2014). TALEN or CRISPR constructs are used to make targeted cuts near mutated DNA, and during the repair process mutated DNA is replaced by WT DNA supplied by an exogenous repair template. However, this repair pathway is not active in post-mitotic cells such as neurons (San Filippo et al., 2008; Jeppesen et al., 2011). The predominant repair pathway in these cells is non-homologous end joining (NHEJ), a repair mechanism which usually leads to mutagenic changes due to small insertions or deletions (Lieber, 2010). Therefore, targeting mutated exonic DNA directly in neurons would only lead to further disturbance of the gene's

coding sequence. This is particularly problematic in the RTT female mosaic brain, as mutagenic changes in cells expressing WT copies of the gene would lead to further exacerbation of the disease.

A recent study by Maresca and colleagues suggests an alternative approach that may be more suitable for treating the nervous system (Maresca et al., 2013). The study showed that large DNA fragments can be inserted into precise genomic locations using TALENs, and that this insertion is dependent on the NHEJ pathway that predominates in neurons. The DNA to be inserted is flanked by the same TALEN-binding sequence as the genomic DNA target region, and delivered as an exogenous plasmid along with the TALEN arm coding constructs. Both the insert and genomic DNA are cut by the TALEN pair, and the insert can then be captured at the genomic break site, leading to incorporation of the sequence into the genome at the intended location. Modified versions of the FokI nuclease have been generated that only cut DNA when the two different TALEN arms of a pair function as a heterodimer, compared to the WT version in which two of the same arm can function as a homodimer (Michal Szczepek et al., 2007; Yannick Doyon et al., 2010). These obligate heterodimers lead to reduced off-target effects and toxicity compared to the WT version of the protein, although at the expense of reduced activity levels. Maresca and colleagues designed a strategy to exploit these obligate heterodimer mutants and ensure that DNA insertion occurs in the correct orientation. By careful design of the targeting sites on the exogenous plasmid, insertion in the correct orientation can prevent re-cutting by the TALEN pair but not when it is inserted in the reverse orientation. As of yet there is no equivalent CRISPR-Cas9 system for biasing insertion in the correct orientation, and it is therefore not clear if CRISPR-Cas9 based targeting is similarly capable of generating a significant amount of correct insertion.

RTT-causing mutations occur in all three coding exons of the major brain isoform of MeCP2, but the most common disease causing mutations are clustered in the portions of exon 3 and exon 4 that code for the methyl-binding (MBD) and transcriptional repressor (TRD) domains of the protein (<https://www.rettsyndrome.org/research/rettbase>). Replacing the protein coding sequences of exon 3 and 4 in mutant cells with exogenously delivered WT copies is an attractive therapeutic strategy, suitable for a wide range of disease causing mutations.

To investigate if this could be achieved using genome editing tools, a novel strategy was developed to ligate an *Mecp2* repair construct, containing a WT copy of the protein coding region of exon 3 and 4, upstream of the endogenous sequence of these exons. With the use of appropriate splice sites, this repair construct will be spliced to exon 1 of the gene during mRNA processing. A transcriptional terminator will stop transcription at the end of the repair construct sequence, thus excluding the endogenous exons 3 and 4 from the final mRNA transcript. If successful this will lead to the restoration of the WT coding sequence and the production of functional protein. The major advantage of this strategy over traditional gene therapy approaches is that gene transcription levels remain under the control of the endogenous promoter, and thus maintains normal spatial, temporal, and quantitative regulation.

In this chapter the design, synthesis, and functional assaying of TALEN and CRISPR-Cas9 constructs, capable of targeting appropriate non-coding regions of the *Mecp2* gene, is described. In the next chapter, the most effective TALEN and CRISPR-Cas9 constructs will then be assessed for their ability to successfully target an *Mecp2* repair construct into the target region.

## 4.2 Aims

The overall aim of the work presented in this chapter was to develop genome editing tools capable of making targeted DNA breaks in a non-coding region of the *Mecp2* gene. These tools will then be used to investigate the therapeutic strategies described in chapter 5. The objectives for this chapter were to:

- (1) Identify an intronic target region suitable for insertion of *Mecp2* repair construct
- (2) Design and synthesise TALEN and CRISPR constructs that will target genomic DNA in the identified region
- (3) Assay cutting efficiency of all constructs to determine most effective constructs for future work

## 4.2 Identification of suitable target region for repair construct insertion

The repair strategy employed here relies on inserting a DNA repair construct into a specific non-coding region of the *Mecp2* gene. Since the majority of RTT causing mutations are in exons 3 and 4 of the gene (**Fig. 4.1A**) the intronic region upstream of these exons was targeted. Intron 2 of *Mecp2* is extremely large (~60 kb) and contains a large fraction of repetitive DNA unsuitable for TALEN and CRISPR targeting, as it would lead to multiple sites being targeted. To locate a suitable target region in this intron, the University of California, Santa Cruz Genome Browser (<https://genome.ucsc.edu>), in combination with the repeat masker tool (<http://www.repeatmasker.org/>), was used to identify areas of the intron free of repetitive DNA sequences. An ~ 900 bp region of unique sequence was identified 1.6 – 0.7 kb upstream of exon 3 (**Fig. 4.1B**).

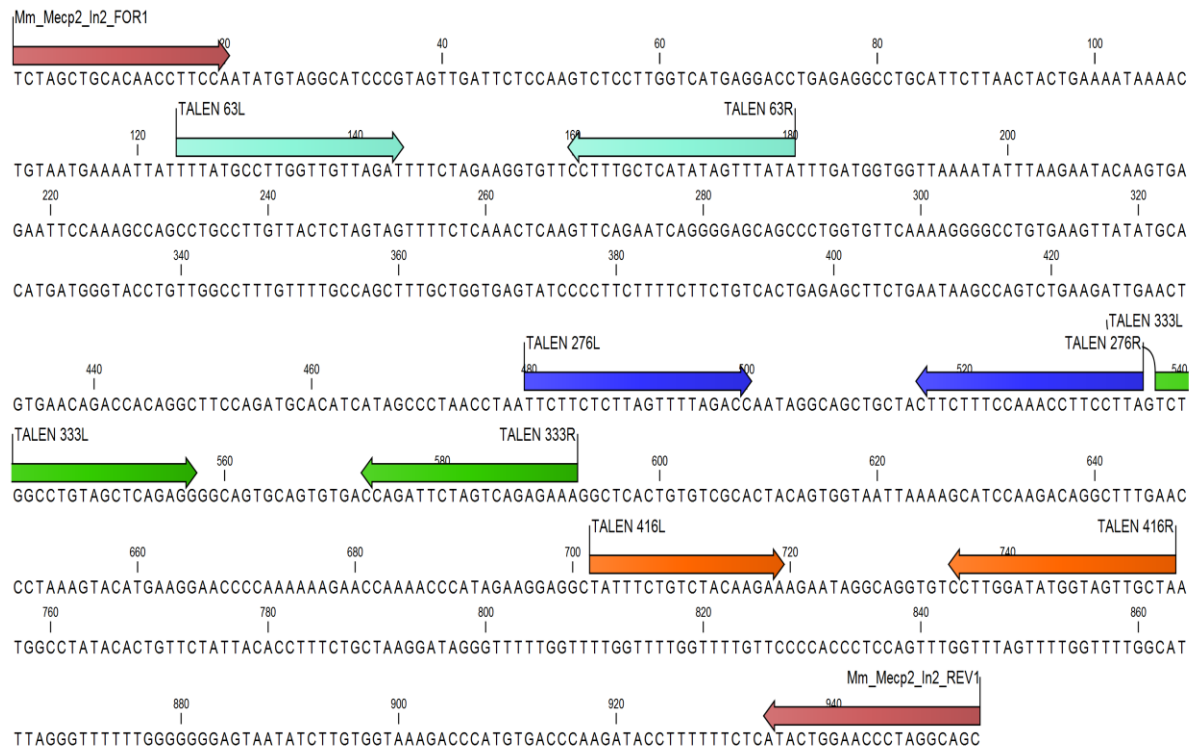


### 4.3 Design of TALEN pairs

The Cornell University TALEN design tool (<https://tale-nt.cac.cornell.edu/node/add/talen>) was used to identify 100s of TALEN pairings that could target sequences in the previously identified target region (see 4.2). The software parameters were set to only identify TALENs whose cutting efficiency would be easy to assay, by limiting the results to those in which a unique restriction enzyme (RE) was found in the spacer region between the binding sites of the two TALEN arms of a pair. In the absence of a homologous template, cells will mostly repair TALEN-induced double-stranded DNA breaks using the mutagenic NHEJ repair pathway. This leads to small insertions and deletions at the break site. Any RE recognition sites overlapping this area will therefore be destroyed during the repair process. This can be detected by PCR amplification of the target region followed by RE digestion of the PCR product, known as a restriction fragment length variations (RFLV) assay.

Results from the software were exported to an Excel spreadsheet and furthered filtered to identify the pairings that could be assayed using REs available in our lab. Lastly, candidates were filtered to only include designs with 15 bp or 16 bp spacer regions between TALEN recognition sites, and TALENs with large ratios of purine bases, as studies have found that these properties confer highest performance levels (Miller et al., 2011; Mussolino et al., 2014). After filtering there were four TALEN pairings remaining for use in our target region (Fig. 4.2). PCR primers flanking the target region were then designed to enable selective amplification of the region for RFLV analysis.





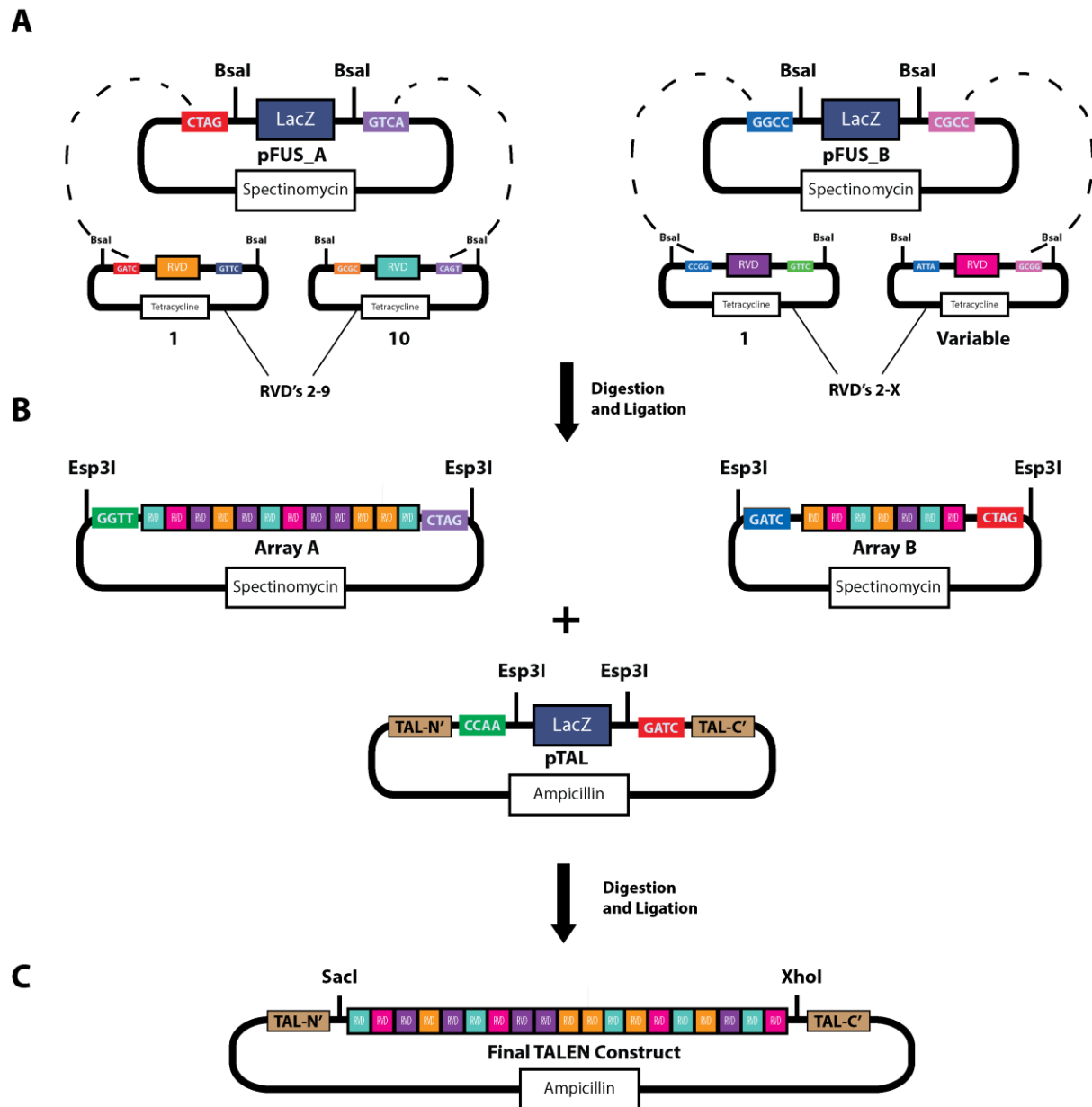
**Figure 4.2 – DNA target sequence of selected TALEN pairings**

DNA sequence is intron 2 target region previously identified. Pink arrows at either end show the sequence of primers designed to PCR amplify the target region. Remaining coloured arrows show the target sequence of the four TALEN pairings that remained after filtering of the initial candidate list. Each pairing consists of a left and right targeting arm.

## 4.4 TALEN synthesis

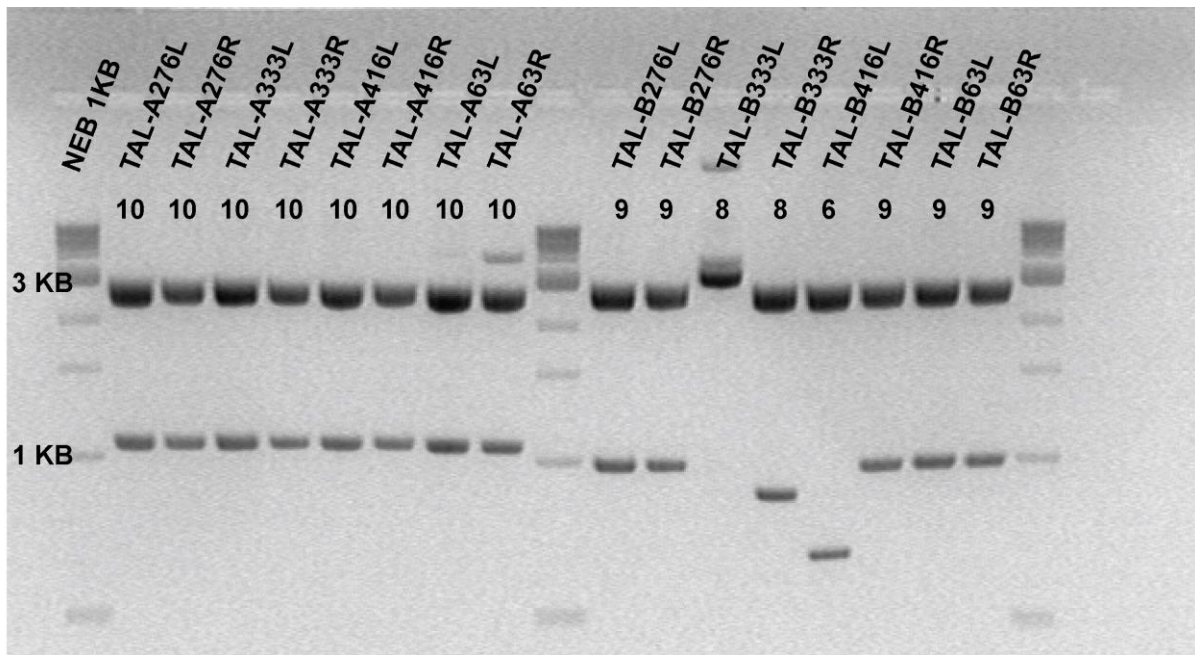
In this study we used the TALEN assembly method developed by Voytas and colleagues (Cermak et al., 2011), which utilises the golden gate method of cloning (Engler et al., 2008, 2009). This cloning method uses type IIS restriction enzymes. These enzymes cleave DNA outside of their recognition site leading to the creation of unique 4 bp overhangs, which are determined by the sequence surrounding the recognition site. Using this method, multiple fragments of DNA can be ligated together in a precise order in a single reaction.

TALEN assembly is a two-stage process. The first stage involves the assembly of DNA targeting repeat modules into ordered arrays (**Fig. 4.3A-B**).



The four TALEN pairs identified by the design software had target sites ranging from 17 to 20 bp. Since the efficiency of Golden Gate Cloning decreases as the number of DNA fragments increases, the TALE repeat modules were first assembled into two intermediate arrays, each containing a maximum of ten repeats. The first intermediate array (pFUS\_A) always contained ten repeats and the number in the second (pFUS\_B) varied depending on the length of recognition sequence for each arm.

In total, eight individual TALEN arms had to be synthesised, a left and right arm for each of the four TALEN pairs. For each arm, the appropriate repeat module plasmids along with a pFUS\_A or pFUS\_B destination plasmid were put together in a single reaction tube, digested with BsaI, and fragments were ligated together to form the intermediate arrays. Reactions were transformed into DH5 $\alpha$  cells and plated on spectinomycin-containing agar plates. Correct assembly involved removal of the LacZ sequence from pFUS\_A and pFUS\_B plasmids, allowing blue/white screening of cells. White colonies were picked, cultured overnight, and plasmid DNA was extracted by mini-prep. To confirm correct assembly, plasmid DNA was digested with AflIII and XbaI, to cut out the array of fused repeat modules, and digest products were separated by agarose gel electrophoresis. Correct array fragments were sized ~60 bp + 100 bp for each repeat module in the array. Digestion products from all reactions, except one, showed the expected band sizes, confirming correct assembly of the intermediate arrays (**Fig. 4.4**). TAL-B333L was a failed digestion. A correct sized fragment was confirmed on repeat digestion. To confirm that all repeats were assembled in the correct order and without mutations, sequencing primers were designed (pCR8\_F1 and pCR8\_R1) flanking the region containing the repeat array. All pFUS\_A and pFUS\_B plasmids were then sent, along with sequencing primers, for Sanger sequence analysis. Sequencing results confirmed that all arrays were correctly assembled and free of mutations.



**Figure 4.4 – Restriction digest confirmation of correct array assembly**

Plasmid DNA was digested with AflIII and XbaI and separated by agarose gel electrophoresis. The large fragment in each lane is vector backbone. The smaller fragment is a repeat array whose size varies depending on the number of repeats. Expected sizes: All TAL-A arrays = 1060 bp (10 repeats); TAL-B276, TAL-B416R, TAL-B63 = 960 bp (9 repeats); TAL-B333 = 860 bp (8 repeats); TAL-B416L = 660 bp (6 repeats). Numbers above bands indicate number of repeats for each arm. Ladder was NEB 1 KB Quick.

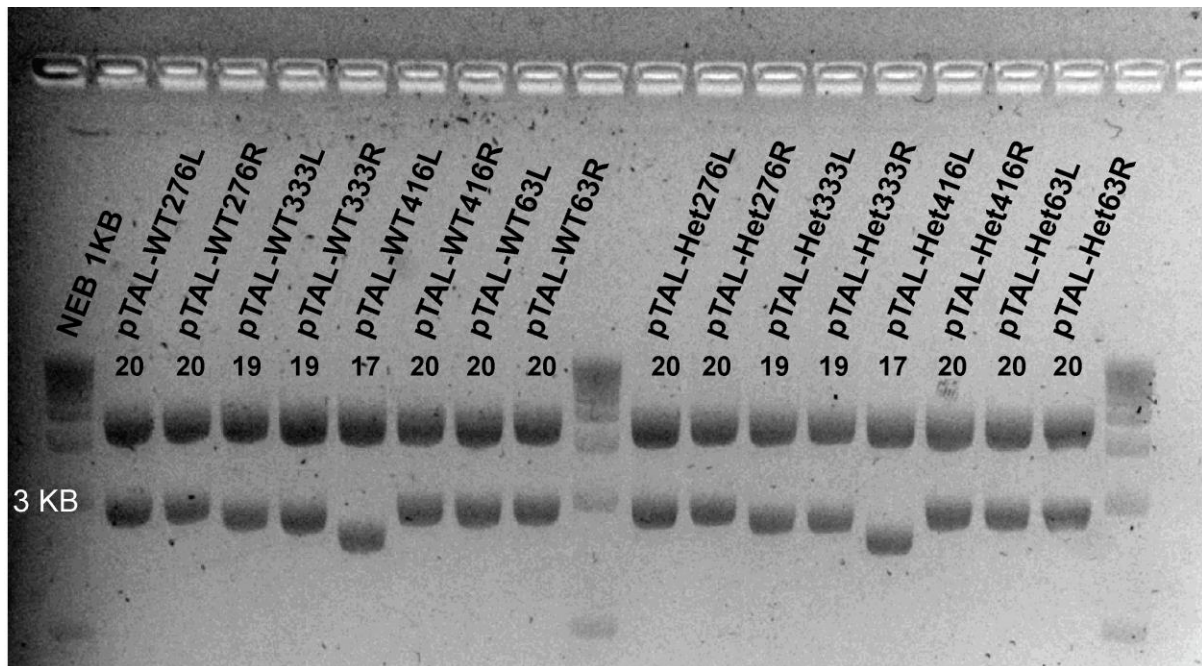
In the second stage of the assembly process (**Fig. 4.3B-C**) pFUS\_A and pFUS\_B plasmids for each TALEN arm were combined into a plasmid backbone containing the C' and N' regions of the TALE protein as well as the FokI nuclease sequence. In this study two versions of the FokI nuclease were used. As well as the WT version, a modified version was used in which the FokI nuclease has been modified to ensure that DNA cutting only occurs when a left arm and right arm come together to form a heterodimer. In order to compare cutting efficiency between these two versions, all 8 TALEN arms were assembled into two different plasmid backbones, containing either the WT (pTALv2\_FokIWT) or the obligate heterodimer mutant FokI (pTALv3-FokI-ELDS for left arm and pTALv3-FokI-KKRS for the right arm).

For each arm, the appropriate pFUS\_A and pFUS\_B array plasmids from stage one were mixed in a single tube along with a backbone plasmid containing either WT or obligate heterodimer FokI. Plasmids were digested with Esp3I and the fragments ligated together to form the final TALEN arms (**Table 4.1**).

**Table 4.1 – RVD sequence and FokI type for individual TALEN arms**

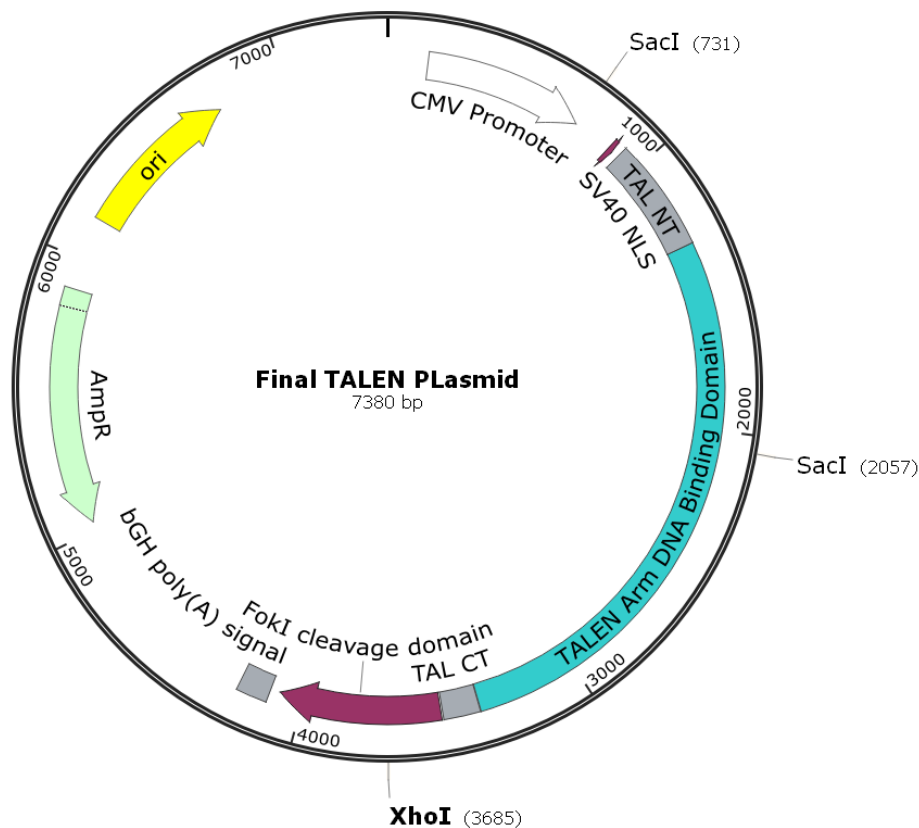
Arm name	Pair	Target Sequence	FokI type
pTAL-WT276L	276	TTCTTCTCTTAGTTTTAGACC	WT
pTAL-Het276L			Obligate het.
pTAL-WT276R		TAAGGAAGGTTTGGAAAGAAG	WT
pTAL-Het276R			Obligate het.
pTAL-WT333L	333	TCTGGCCTGTAGCTCAGAGG	WT
pTAL-Het333L			Obligate het.
pTAL-WT333R		TTTCTCTGACTAGAATCTGG	WT
pTAL-Het333R			Obligate het.
pTAL-WT416L	416	TATTTCTGTCTACAAGAA	WT
pTAL-Het416L			Obligate het.
pTAL-WT416R		TAGCAACTACCATATCCAAGG	WT
pTAL-Het416R			Obligate het.
pTAL-WT63L	63	TTTATGCCTTGGTTGTTAGAT	WT
pTAL-Het63L			Obligate het.
pTAL-WT63R		TATAAACTATATGAGCAAAGG	WT
pTAL-Het63R			Obligate het.

Reactions were transformed into Dh5 $\alpha$  cells and plated on ampicillin containing agar plates. Blue/white screening was again used to identify correct clones and white colonies were picked, cultured overnight, and plasmid DNA extracted by mini-prep. To confirm correct assembly, plasmid DNA was digested with SacI and XhoI to cut out the full array of fused repeat modules, and digest products were separated by agarose gel electrophoresis. Correct array fragment band sizes were ~930 bp + 100 bp for each repeat module in the array. Digestions from all 16 reactions showed the expected product, confirming correct assembly of final TALEN arm constructs (**Fig. 4.6**). A representative final TALEN construct plasmid map is shown in **Fig. 4.7**.



**Figure 4.6 - Restriction digest confirmation of correct final TALEN arm assembly**

Plasmid DNA was digested with XhoI and SacI, to cut out full array of fused repeat modules and restriction fragments were separated by agarose gel electrophoresis. Large fragment in each lane is vector backbone. Smaller fragment is repeat array, size varies depending on the number of repeats. Expected sizes: pTAL-276, pTAL416R and pTAL-63 = 2960 bp (20 repeats); TAL-333 = 2960 bp (19 repeats); TAL-416L = 2760 bp (17 repeats). Ladder was NEB 1 KB Quick.



**Figure 4.7 – Representative plasmid map of final TALEN construct**

Plasmid map showing standard features of a final TALEN arm construct. Transcription is driven by a cytomegalovirus (CMV) promoter. A simian virus 40 nuclear localisation signal (SV40 NLS) ensures TALEN protein is targeted to the nucleus. The DNA binding domain is cloned between the N-terminal (TAL NT) and C-terminal (TAL CT) of the TALE protein, and the sequence for the FokI nuclease is directly downstream of the TAL CT. Transcription is terminated by the bovine growth hormone (bGH) poly (A) signal. The ampicillin resistance gene (AmpR) allows for antibiotic selection of the plasmid and the bacterial origin of replication (ori) enables replication of the plasmid in bacterial hosts.

## 4.5 Optimisation of electroporation parameters

To test the ability of our TALENs to make targeted DNA cuts we tested them on P19 cells, a well-established male embryonic carcinoma cell line (McBurney and Rogers, 1982). In order to cut DNA, plasmids coding for both TALEN arms of a pair needed to be present in the same cell. To achieve as high a transfection efficiency as possible we used the Neon electroporation device, which has been shown to achieve high levels of transfection in multiple mouse cell lines (Abolhassani et al., 2010; Hsu and Meng, 2010).

Optimum Neon parameters such as pulse voltage (v), pulse width (ms), and pulse number have to be empirically determined for each individual cell line. Recommended settings



specific for P19 cells, as well as those for mouse embryonic stem cells (mESCs) were obtained from the Life Technologies Neon website

(<http://www.lifetechnologies.com/uk/en/home/life-science/cell-culture/transfection/transfection---selection-misc/neon-transfection-system/neon-protocols-cell-line-data.html>) (Table 4.1).

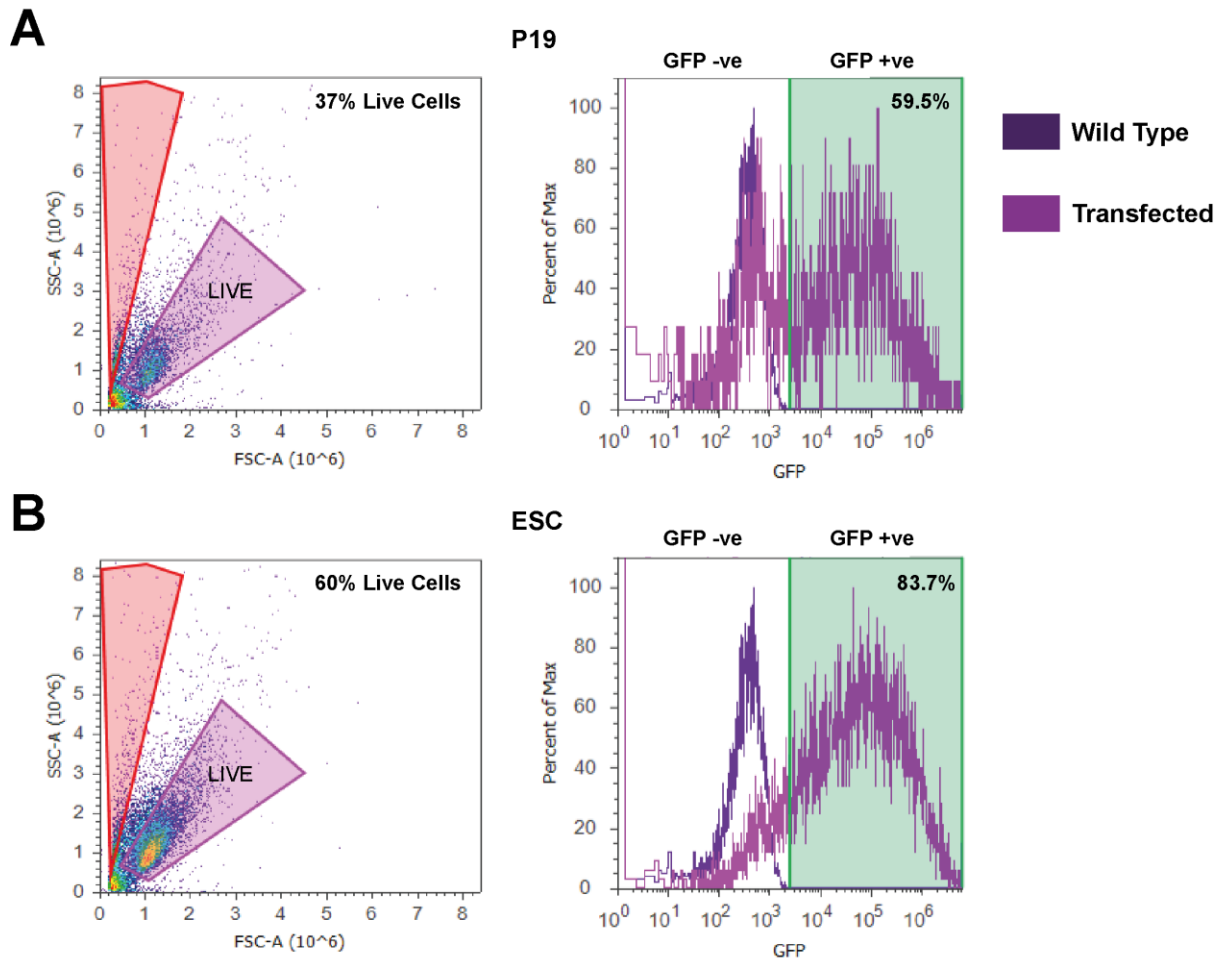
**Table 4.2- Electroporation parameters used for optimisation**

<b>Recommended cell type</b>	<b>Pulse voltage (V)</b>	<b>Pulse width (ms)</b>	<b>Pulse number</b>
<b>P19</b>	1275	30	1
<b>Mouse embryonic stem cells (mESC)</b>	1400	10	3

To determine if either of these settings produced high transfection efficiency in P19 cells a GFP reporter plasmid was used (KpLenti-PGK-GFP). Since cells need to be transfected with two different TALEN arm plasmids for cutting, this reporter was transfected along with a single TALEN arm, in order to ensure the settings were optimal for a double plasmid transfection. For each transfection 400,000 cells were electroporated with a total of 4 µg of DNA, 2 µg of each plasmid, using either P19 or mESC settings. Transfected cells were then plated into a well in a 6-well dishes and incubated for 48 hours.

Transfection efficiency was assessed using an Attune flow cytometer (Life Technologies, USA). Forward and side scatter plots from untransfected P19 cells were used to create gates to identify the population of live cells (R2), and a further gate was set to identify cells within this live population that expressed GFP above the background levels seen in control cells (R4). Cells in this second population were those successfully transfected with the GFP reporter plasmid. Cells transfected using the recommended P19 cell settings had 37% of cells identified as live and 59.5% of these live cells identified as GFP positive. In comparison cells transfected with mESC settings had 60% of cells identified as live and 83.7% of these live cells identified as GFP positive (**Fig. 4.8**). As the mESC settings produced lower cell death and extremely high transfection efficiency we used these settings for all subsequent P19 transfections.



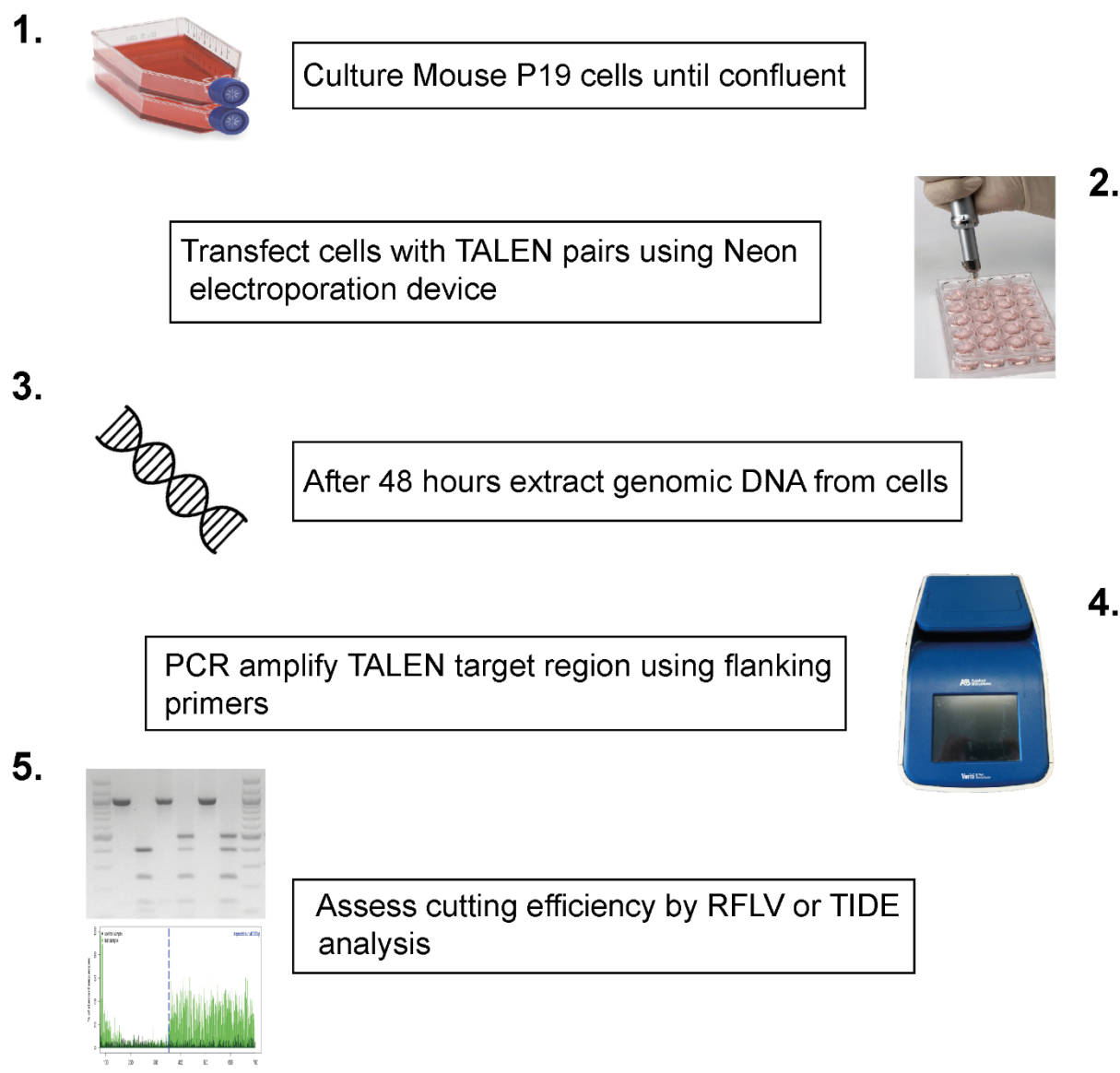


**Figure 4.8 – Optimisation of Neon electroporation parameters**

GFP protein levels and cell death for either (A) P19 (1275(v), 30(ms), one pulse) or (B) mouse embryonic stem cell (ESC; 1400(V), 10(ms), 3pulses) recommended settings were assayed using Attune Flow Cytometry. Gates were drawn to identify both live and GFP positive (+ve) cells. Left panels show histogram of forward scatter (FSC-A) versus side scatter (SSC-A) used to identify live cells (purple rhombus). The percentage of cells gated as live is shown in each panel. Right panels show histogram plotting GFP fluorescence intensity versus percent of max (a normalised representation of the number of events) in the live population. Results from WT (dark purple) and transfected cells (light purple) are shown superimposed over each other. The percentage of GFP positive cells (green shaded area) are shown in each panel.

## 4.6 Testing TALEN cutting efficiency

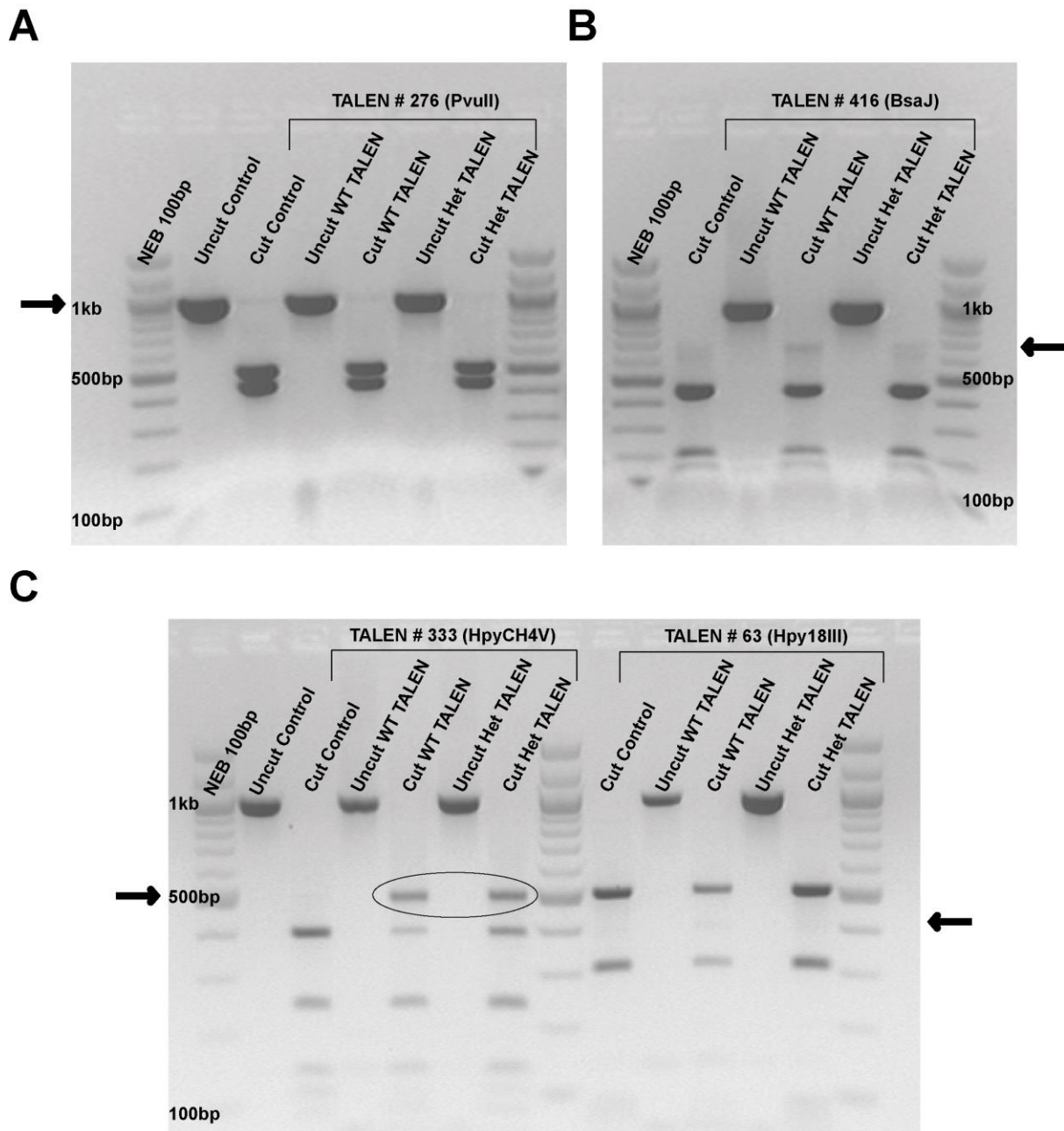
Once optimal transfection conditions were determined, all 4 TALEN pairs were assessed for their ability to make targeted double stranded breaks as outlined in **Fig. 4.9**.



**Figure 4.9** – *Outline of procedure for testing cutting efficiency of TALEN pairs*

P19 cells were cultured and then transfected with TALEN pair plasmids. After 48 hours, genomic DNA was extracted from these cells and the TALEN target region was amplified using intron 2 specific primers (Mm\_*Mecp2*\_In2\_For1 and Mm\_*Mcp2*\_In2\_Rev1; **Table 2.11**). This PCR product was then used to quantify cutting efficiency for each TALEN pair, using the RFLV assay described previously. Small insertions and deletions, introduced during NHEJ repair of DNA breaks, lead to destruction of the unique RE recognition site contained in the spacer region between the two TALEN arm binding sites. For each TALEN pair transfection, the PCR product was digested with the appropriate RE, and the resulting digestion fragments were separated via agarose gel electrophoresis (**Fig. 4.10**). Successful cutting was indicated by the appearance of unique bands in digestion products from DNA extracted from TALEN pair transfected cells, compared to DNA from untransfected controls. The intensity of these unique bands gave an indication of the efficiency of cutting.

Successful cutting with TALEN pair # 276, leads to destruction of a PvuII RE recognition site, and the appearance of a new band at 953 bp on the agarose gel. Results from the RFLV assay (**Fig. 4.10-A**) showed that no band of this size was present, either with the WT or mutant version of FokI, indicating that no cutting had occurred. Cutting with TALEN pair # 416 leads to the appearance of a new band at 650 bp. RFLV results (**Fig. 4.10-B**) showed faint bands of this size on the gel, with a stronger intensity band for WT FokI. Cutting with TALEN pair # 333 leads to the appearance of a new band at 494 bp. In contrast to results from the previous two pairs, results from this RFLV (**Fig. 4.10-C**) showed the presence of very intense bands of this size, indicating efficient cutting by this TALEN pair. Cutting with TALEN pair # 63 leads to the appearance of a new band at 388 bp. RFLV results for this pair (**Fig. 4.10-D**) showed no evidence of cutting.



**Figure 4.10 – Agarose gel of RFLP digestion products**

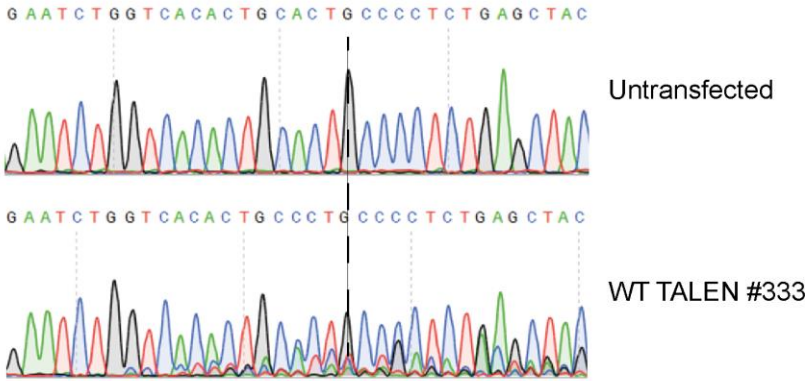
Gel image of digestion products of RFLV assay from cells transfected with (A) TALEN #276 (B) TALEN #416 (C) TALEN #333 or TALEN #63 as indicated. For each TALEN pair either the WT or the obligate heterodimer mutant (Het) version of the FokI nuclease was used. Each digestion product was run alongside uncut samples of the same PCR product. In addition cut and uncut products from cells not transfected with any TALEN pairs are shown. Black arrows indicate the size of unique band expected in cut samples successfully cut with TALENs. Black circle highlights intense bands of expected size in digested samples from TALEN #333 transfected cells, indicating high cutting efficiency in this pair.

Although RFLV analysis allows the relative cutting efficiency of different TALEN pairs to be assessed, it is limited by the fact that it relies on a suitable RE recognition site to be directly overlapping the TALEN cut site, otherwise it may not be sensitive to very small insertions or deletions that do not disturb the recognition site. In addition, it may not always be possible to find a suitable unique recognition site anywhere within the spacer region. Recently, a new method for quantifying cutting events has been developed (Brinkman et al., 2014). This method, called tracking indels by decomposition (TIDE), relies on Sanger sequencing of the PCR products from transfected and non-transfected cells. The resulting sequencing traces are then analysed using an online software tool (<http://tide.nki.nl/>). Since repair of TALEN induced DNA breaks is error prone, PCR products from TALEN transfected cells are a heterogeneous mixture of indels, leading to a composite sequence trace. The TIDE tool decomposes this composite trace via multivariate non-negative linear modelling, and uses the control sequence from untransfected controls to model the type and proportion of indels. This enables both the cutting efficiency of TALENs or CRISPRs, and the nature of the indels produced, to be determined.

To test this new tool, genomic DNA was extracted from each of the TALEN pair cells, as well as from untransfected control cells, and the intron 2 target region was PCR amplified using flanking primers (Mm\_*Mecp2*\_In2\_For1 and Mm\_*Mcp2*\_In2\_Rev1). These PCR products were then sent for Sanger sequencing and the resulting sequence traces analysed using the TIDE tool (**Fig. 4.11**). Results of TIDE analysis were in agreement with those from the RFLV assays. For WT FokI versions, TALEN #333 showed the highest cutting efficiency (42.9%) and #416 the second highest (19.6%). TALEN #276 and #63 showed low cutting (8% and 2.1% respectively). As expected, use of the mutant obligate heterodimer FokI nuclease led to reduced cutting efficiency with TALEN #333 reduced to 20.8% and #416 dramatically reduced to 2.9%. Since TIDE analysis proved to be quick, quantitative, and was in agreement with the results from the RFLV assays, it was used instead of the RFLV assay for the subsequent CRISPR Cas-9 experiments described in this chapter.

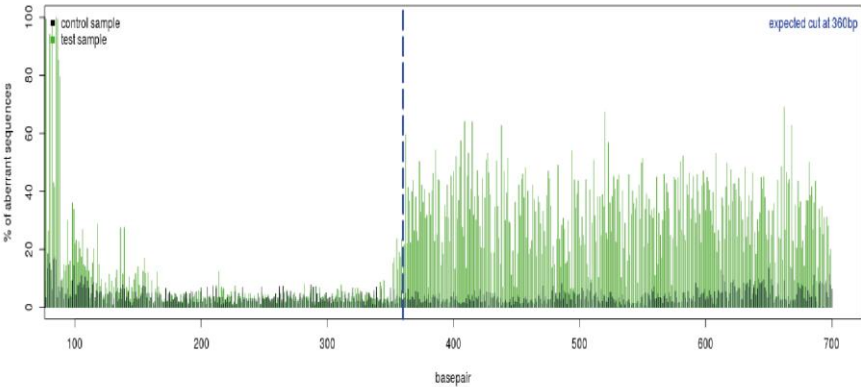
**A**

Sequence trace comparison



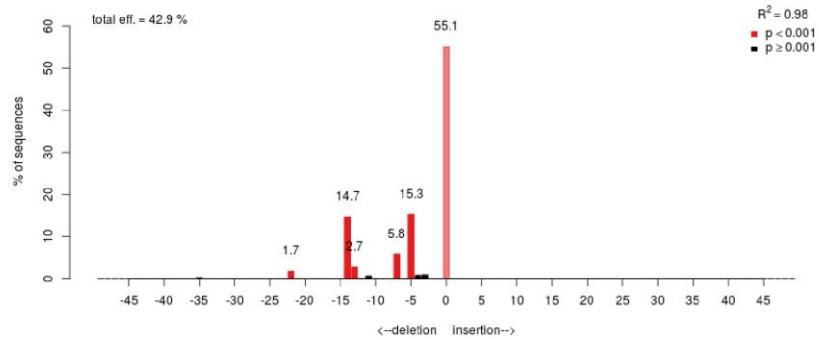
**B**

Aberrant sequence signal



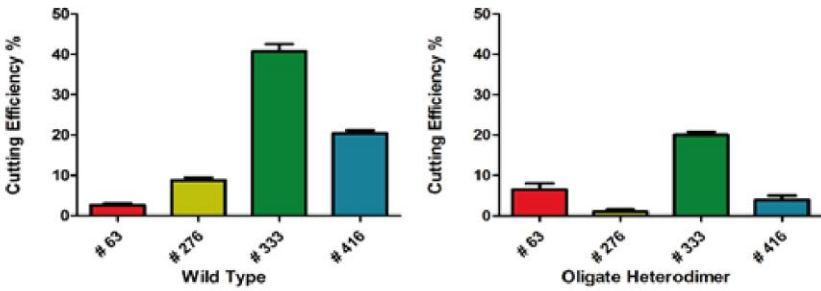
**C**

Indel spectrum



**D**

Cutting Efficiency



### Figure 4.11 – *TIDE* analysis of TALEN pair # 333 cutting efficiency

Cutting efficiency of TALEN pairs was quantified using *TIDE* analysis. (A) Sequence traces from untransfected (top trace) and WT TALEN # 333 transfected (bottom trace) P19 cells. Dashed black line indicates the expected break site. Breakdown of sequence before the expected cut site in WT TALEN # 333 indicates the presence of small deletions in the genomic sequence introduced after the repair of TALEN-mediated DNA breaks (B) Visualization of aberrant sequence signal between control (black) and transfected sample (green). Also shown is expected cut site (blue dashed line). (C) Representative image of WT pair # 333 indel spectrum and total cutting efficiency %. Goodness of fit assessed by  $R^2$  calculation. The *P*-value associated with the estimated abundance of each indel was calculated using a *t*-test of the variance-covariance matrix of the standard errors. (D) Graphs of estimated cutting efficiency for TALEN pairs as determined by the software. Results show mean  $\pm$  SEM of two replicates.

## 4.7 Design of CRISPR-Cas9 constructs

Results from the initial TALEN experiments showed that TALEN pairs could be designed to efficiently induce targeted double-stranded DNA breaks in the *Mecp2* target region. However, the TALEN pairs were very time consuming to synthesise, and cutting was extremely variable between the different pairs. To determine if the CRISPR-Cas9 system could be used to more efficiently target the same region, a set of CRISPR-Cas9 single guide constructs was synthesised and assayed. Since TALEN pair #333 showed much greater cutting efficiency than the other three pairs, we focused our CRISPR-Cas9 designs on the small region surrounding the #333 target site. A CRISPR design tool (<http://crispr.mit.edu>) was used to generate CRISPR guide target sequences. This tool has the advantage of identifying likely off-target binding sites, and to score potential guides by the inverse likelihood of off-target binding. The higher the score, the lower the predicted off-target levels by the software. The top four scoring single guide target sequences were selected (Figure 4.12).



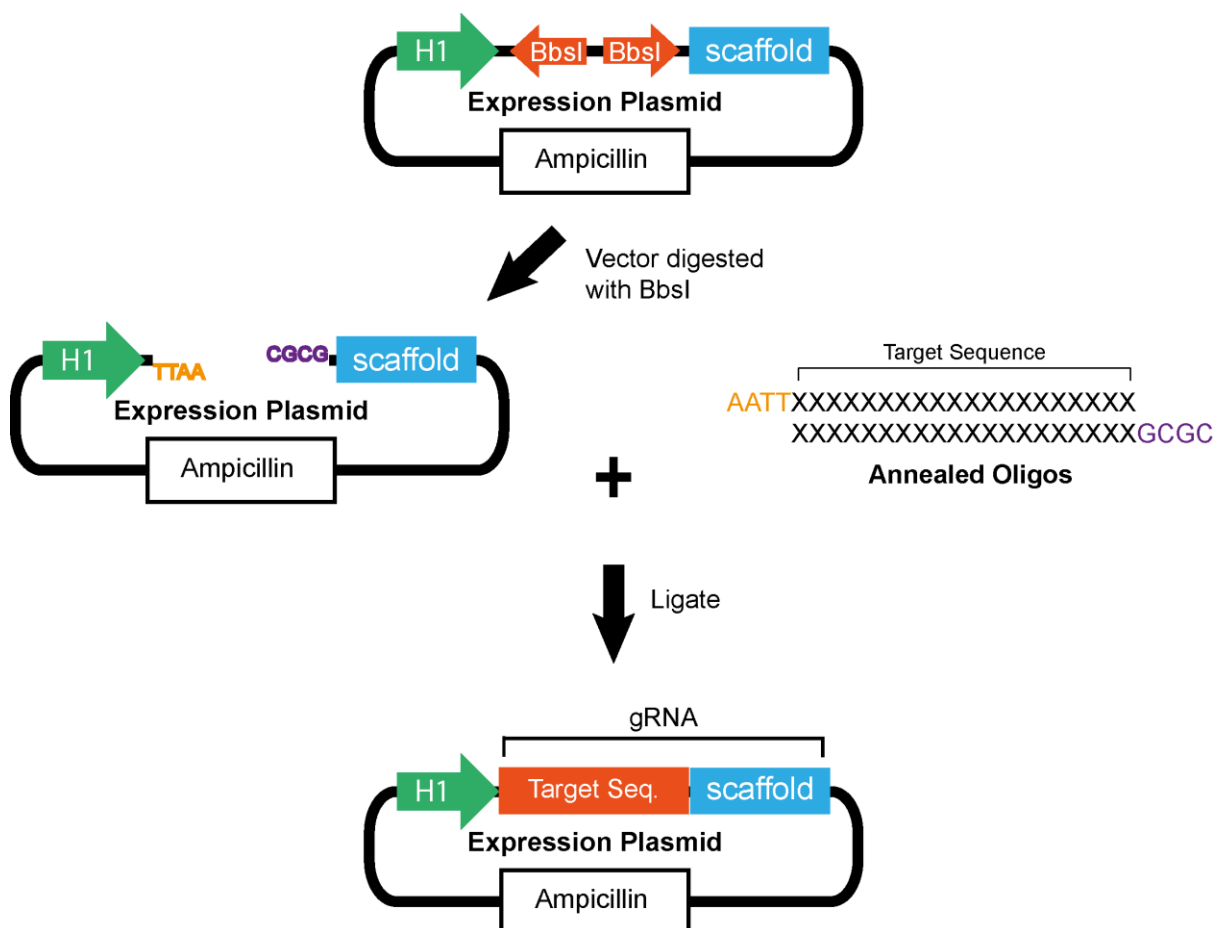
**Figure 4.12 – DNA target sequence of selected TALEN pairings**

DNA sequence of intron 2 target region. Pink arrows at either end show the sequence of PCR primers designed to amplify the target region. Remaining coloured arrows show the target sequence of the four top scoring CRISPR guides. Red underlined sequence shows the target sequence for the arms of TALEN pair # 333.



## 4.8 Synthesis of CRISPR guides

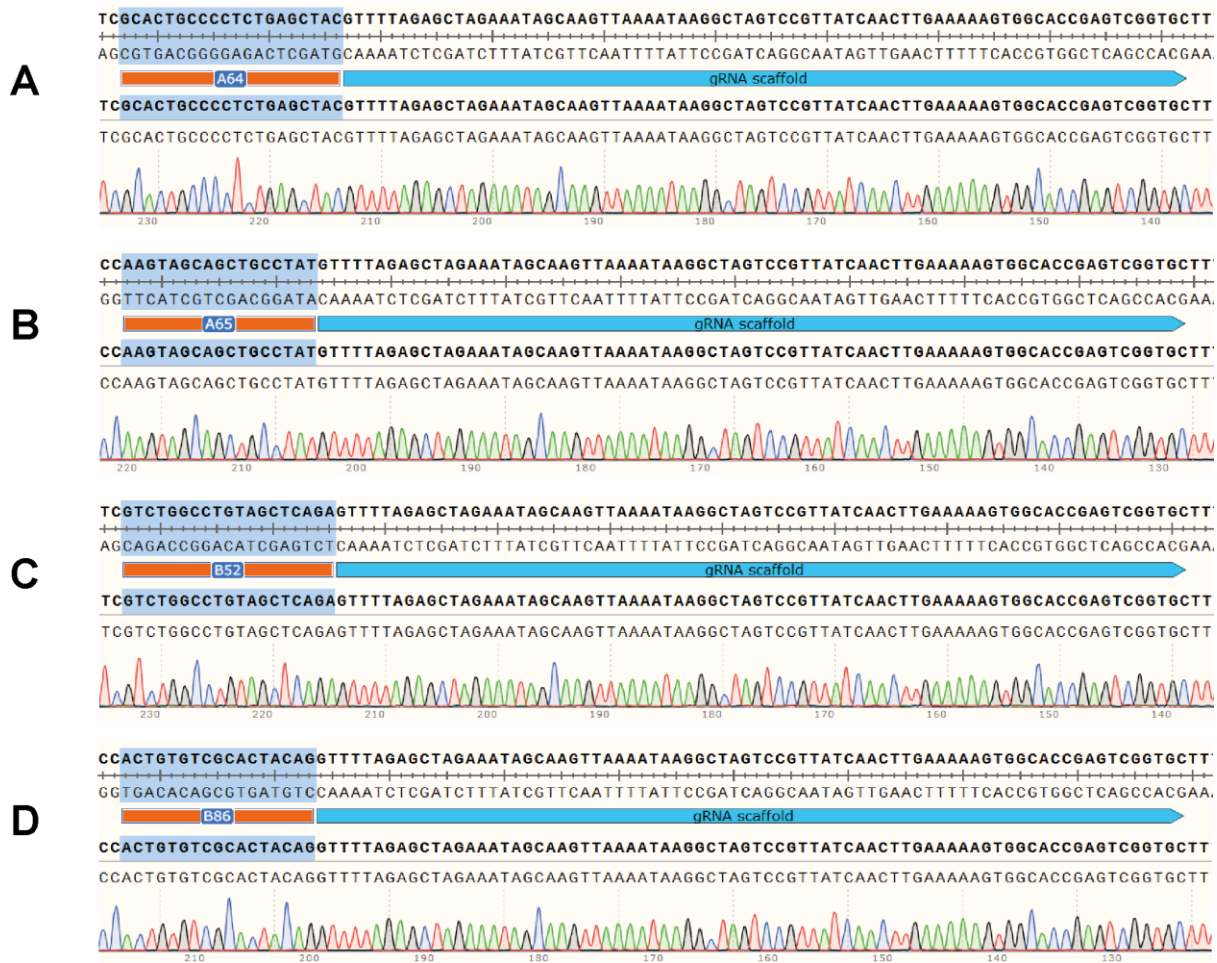
The selected guide target sequences were then synthesised as two complementary oligonucleotides, annealed together, and cloned upstream of the gRNA scaffold in a CRISPR guide expression plasmid (**Fig. 4.13**).



**Figure 4.13 – Outline of cloning strategy of guide sequence into expression plasmid**

Guide target sequences were synthesised as complementary oligonucleotides, annealed together, and ligated into a BbsI cut expression plasmid upstream of tracrRNA scaffold to form the guide RNA (gRNA). Extra bases (yellow and purple) were added to 5'-ends of oligonucleotides to create matching overhangs to BbsI cut plasmid. Transcription of the gRNA was driven by the RNA polymerase III promoter human H1 (green arrow). The plasmid also contained the ampicillin resistance gene (white box).

Transcription of the guide is driven by an RNA polymerase III promoter. Oligonucleotides were synthesised so that when annealed they would have overhangs complementary to those generated by BbsI digestion of the expression plasmid. This allowed easy cloning of the guide sequence into the plasmid. Successful cloning of guide sequences was confirmed by Sanger sequence analysis (**Fig. 4.14**).

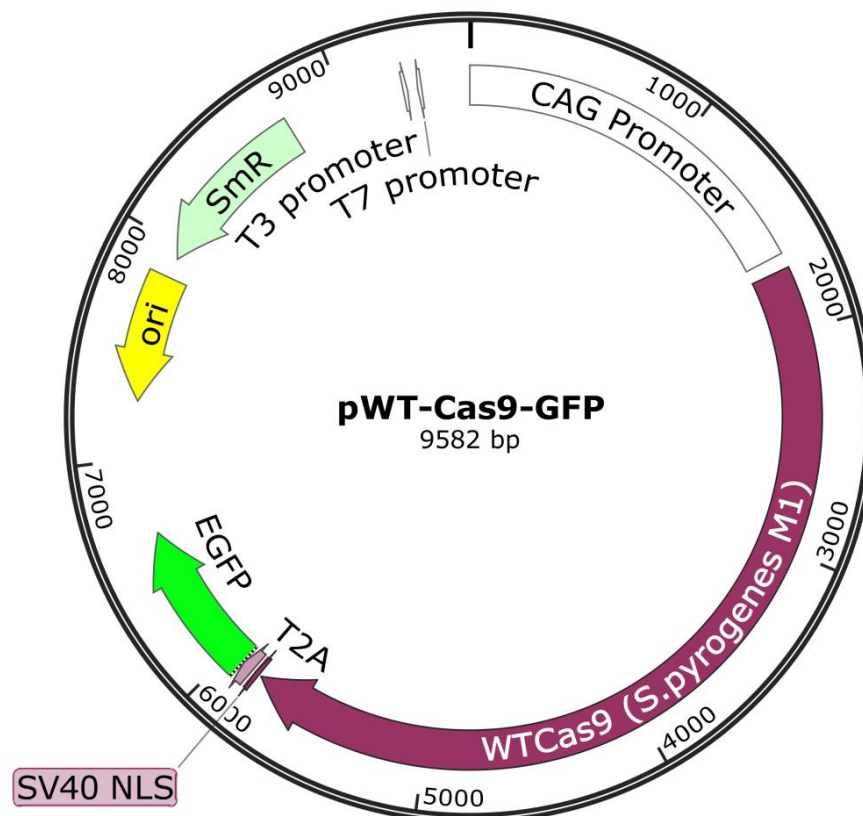


**Figure 4.14 – Sequence confirmation of guide sequence cloning into expression plasmid**

Shows extract of sequence traces from expression plasmids confirming successful cloning of CRISPR guide target sequences (**A**) A64 (**B**) A65 (**C**) B52 (**D**) B86, upstream of the gRNA tracrRNA scaffold. Guide target sequence is highlighted 20 bp section. For each guide the nucleotide sequence extracted from sequencing traces (bottom two strands) is compared against the expected final nucleotide sequence (top two strands). All expression plasmids showed an exact match between the two, confirming successful cloning. Chromatograms are shown below sequence data.

## 4.9 Testing of CRISPR cutting efficiency

Cutting efficiency for each of the four selected guides was tested by a process similar to that described previously for TALENs (**Fig 4.9**). Successful cutting requires a cell to be transfected with the guide plasmid as well as a separate plasmid expressing the Cas9 nuclease (**Fig 4.15**).



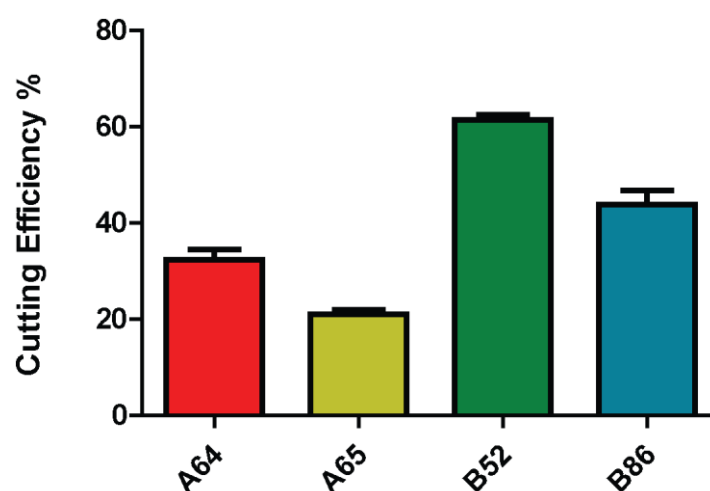
**Figure 4.15 – Map of Cas9 expressing plasmid**

Plasmid map showing features of the Cas9 expression plasmid. Transcription of WT Cas9 and EGFP was driven by the CMV early enhancer/chicken  $\beta$  actin (CAG) promoter. A *Thomomys* 2A peptide (T2A) sequence between the Cas9 and GFP sequence caused the ribosome to skip the synthesis of the peptide bond at the C-terminus of the 2A peptide, leading to cleavage of the downstream GFP peptide. The SV40 nuclear localisation signal (NLS) ensured that the Cas9 nuclease was targeted to its site of action in the nucleus. Other features include a bacterial origin of replication (ori) and the spectinomycin antibiotic resistance gene (SmR).

This Cas9-expressing plasmid also expresses GFP, under the control of the same CAG promoter that drives transcription of the Cas9 gene, enabling transfection efficiency to be easily assessed. The T2A sequence inserted between Cas9 and EGFP allowed the proteins to be expressed as two separate peptides, which prevented the potential detrimental effects of GFP on Cas9 function that could occur if they were made as a fusion protein.

For each CRISPR guide to be tested, P19 cells were transfected with equimolar amount of guide and Cas9 plasmids and incubated for 48 hours. Genomic DNA was extracted and the target region was PCR amplified using flanking primers (Mm\_*Mecp2*\_In2\_For1 and Mm\_*Mcp2*\_In2\_Rev1). The PCR product was then sent for Sanger sequence analysis and the resulting sequencing traces were used to quantify cutting efficiency using TIDE analysis.

TIDE analysis results (**Fig. 4.16**) showed that all CRISPR guides were capable of significant cutting, with some guides showing extremely high efficiency. Guides B52 and B86 showed the highest efficiency (60.4% and 46.8% respectively) and both performed better than the best performing TALEN pair. Guides A64 and A65 showed lower efficiency (34.5% and 22% respectively) but these still performed better than the other three TALEN pairs.



**Figure 4.16 – Cutting efficiency of CRISPR guides**

Graph summarising the cutting efficiency of the top four selected CRISPR guides as determined by TIDE analysis. Results show mean  $\pm$  SEM of two replicates.

## 4.10 Discussion

Ligation of a repair construct into non-coding regions of the *Mecp2* gene represents a novel therapeutic strategy for RTT. In order to test the potential of this approach, TALEN and CRISPR-Cas9 constructs capable of targeting non-coding regions of *Mecp2* are required. Since the majority of RTT causing mutations are located in the protein coding regions of exons 3 and 4, replacing these two exons could provide an attractive single therapeutic strategy for a majority of patients. Attention was therefore focused on the large intron upstream of these exons. In this chapter, a number of CRISPR-Cas9 constructs targeted at this intron were designed, synthesised, and tested for their ability to induce double-strand DNA breaks. The results showed that several TALEN and CRISPR designs were capable of making targeted breaks at a high level of efficiency.

In order to find a suitable target region, the sequence of *Mecp2* intron 2 was examined using the genome browser tool. It was important to avoid regions with a high density of repetitive DNA, as it is essential that the TALEN or CRISPR-Cas9 constructs have a unique target site. On examination, the intron contained large stretches of highly repetitive DNA, but a suitable region of roughly 1 kb was identified slightly upstream of endogenous exon 3. Genome browser data also showed that there was a number of single nucleotide polymorphisms known in this region, which could affect the design of the TALEN and CRISPR target sequences, but sequence analysis on genomic DNA extracted from the P19 cells to be used in these experiments confirmed that the sequence in this region matched the reference sequence. Overall these results suggested that this stretch of DNA would be a suitable target region for insertion of the repair construct.

One disadvantage of using TALEN constructs is that they are much more difficult to synthesise than CRISPR. This is because the DNA binding domain is composed of an array of repeating protein domains (Boch et al., 2009; Moscou and Bogdanove, 2009). The different repeat domains differ by only two amino acids meaning that the coding sequence is highly repetitive, and unsuitable for DNA synthesis (Deng et al., 2012; Mak et al., 2012). In this study, a Golden Gate cloning approach was used based on a previously published method (Cermak et al., 2011). This allowed a large numbers of repetitive DNA fragments to be ligated efficiently in a single reaction. The number of repeats in the four TALEN pairs designed for this study ranged from 16-19. As this was too many to ligate together in one reaction they were synthesised as two intermediate fragments which were then ligated

together into a final expression plasmid. Using this strategy, it is possible to efficiently synthesise TALEN arm plasmids in as short as five days. However, in this study 16 different TALEN arms were required and the cloning took significantly longer, mainly due to several reactions failing during the first attempt. There was also an initial expense with this approach, as it required purchasing a kit containing over 80 different plasmids, as well as various different enzymes. In addition, it was necessary to make high quality plasmid preps of all the plasmids in the kit to make them suitable for use in the protocol, which took a significant amount of time. In contrast to the multi-step process required for TALEN constructs, CRISPR-Cas9 constructs can be cloned very easily and quickly. Target sequences were simply synthesised as two complementary oligonucleotides, annealed together, and cloned into an expression plasmid, and could be generated in as little as three days at little cost. This offered a huge advantage over TALEN in terms of cost, time, and simplicity of design and synthesis.

Once the TALEN and CRISPR-Cas9 constructs had been synthesised they were tested for cutting efficiency in P19 cells. Although neurons will be the ultimate target cell, they are more difficult to culture and to transfect with plasmids, usually requiring the use of viral vectors for high level transfection (Karra and Dahm, 2010). Since this study involved the testing of 8 different TALEN constructs as well as four CRISPR constructs, it was not feasible to do the initial screening assay in neurons. For both TALEN and CRISPR approaches, successful cutting relied on a cell receiving copies of two different plasmids. Using standard transfection methods this would have led to low numbers of cells being transfected, making it difficult to assay the different constructs quickly and effectively. In contrast, the Neon electroporation device enabled very high rates of transfection in P19 cells. Using optimised parameters, transfection rates of over 80% were achieved, providing a robust screening platform for identifying the most effective TALEN and CRISPR constructs. One disadvantage to this approach was the cost of transfections, approximately £5 per well. This was due to the expense of the neon pipette tips which are not reusable.

Initially, cutting efficiency was measured using a restriction digest based approach. Insertions and deletions introduced during NHEJ repair of DNA breaks led to destruction of restriction enzyme recognition sequences in TALEN and CRISPR target sites. PCR amplification across the target region, and digestion of the resulting product with the appropriate restriction enzyme, generated digestion fragments which could be separated on an agarose gel. The

appearance of a unique band in transfected cells compared with untransfected cells was an indication that cutting had occurred. However, this approach had several disadvantages. First, it relied on finding a unique restriction site within the TALEN or CRISPR target site. To ensure that these sites were present, the initial TALEN designs in this study were filtered for the presence of a unique restriction site in the target region. Whilst this ensured the TALENs were suitable for restriction analysis, it meant that potentially more efficient TALENs may have been discarded as they didn't meet this criteria. Second, it was expensive, as a different restriction enzyme needed to be purchased for each TALEN or CRISPR-Cas9 construct. Third, it was time consuming to carry out the restriction digests and run the agarose gels. Recently, an alternative approach, called TIDE analysis, has been developed, which enables the cutting efficiency of TALEN and CRISPR constructs to be quickly and easily assessed (Brinkman et al., 2014). In this approach PCR products spanning the target region are sent for sequence analysis, and specialised software (<http://tide.nki.nl/>) uses the sequence traces produced to calculate cutting efficiency. Crucially, this method is quantitative and does not rely on there being a unique restriction enzyme site located in the target region, so it was suitable for all TALEN and CRISPR constructs. In addition it was also cheaper and required less hands on bench-time. To validate the approach for use in this study, TALEN constructs were first assessed using the standard restriction digest method and then then assessed using TIDE analysis. The TIDE results showed the same pattern as those from the restriction digest method, showing that this new method is suitable for the assessment of cutting efficiency. Since the TIDE approach was cheaper and quicker, it was used for all future constructs.

When assessed by both restriction digest and TIDE analysis, the cutting efficiency of the different TALEN pairs varied dramatically, suggesting that the ability of TALENs to cut is heavily dependent on the particular sequence targeted. Two of the pairs (# 63 and # 276) showed extremely low cutting efficiency, while a third (# 416) showed only moderate activity. However, TALEN # 333 did show robust cutting. One important question to be answered in this study was the effect on cutting efficiency of using the obligate heterodimer mutant version of the FokI nuclease instead of the WT version. This mutated version is designed to increase the specificity of targeting, by ensuring that cutting only occurs when a left and right arm dimerise to form a heterodimer, in comparison with the WT version which is also capable of forming homodimers (Michal Szczeppek et al., 2007; Yannick Doyon et al., 2010). This forced heterodimerisation is also a crucial feature in ensuring that the repair construct is ligated into the genome in the correct orientation (Maresca et al.,

2013). In contrast, WT FokI nuclease is capable of forming homodimers as well as heterodimers. This can lead to off-target cutting and substantial toxicity (Kelly Beumer et al., 2006). However, this increased specificity comes at the expense of a substantial reduction in cutting efficiency in the mutated version, although recent modifications have improved efficiency significantly (Yannick Doyon et al., 2010). In this study, the use of the obligate heterodimer FokI mutant led to a large decrease in cutting efficiency. For example, the best performing TALEN (#333) showed a ~50% decrease in efficiency compared to the WT version, and other pairs showed an even greater fall in efficiency. However, this needs to be weighed against the substantial increase in specificity which will be crucial when thinking about therapies, as well as the potential for controlling the orientation of the insertion.

CRISPR technology, which was only developed in the last few years (Martin Jinek et al., 2012), is becoming the genome editing tool of choice due to a combination of low cost, high efficiency and easy synthesis (Prashant Mali et al., 2013; Sakuma et al., 2014). In this study, all four of the CRISPR guides designed led to robust cutting. The best performing guide, B52, led to over 60% of cells being cut in the target region. This figure is in fact likely to underestimate the actual efficiency, as not all cells will have been successfully transfected with both plasmids, suggesting that CRISPR is capable of achieving extremely high cutting efficiency. It also suggest that when targeting the same region CRISPR-Cas9 constructs will lead to substantially higher levels of cutting than TALENs, perhaps indicating that they are less affected by the particular sequence that they are targeting. However, despite this advantage in efficiency, there is as yet no equivalent of the obligate heterodimer mutant for the Cas9 enzyme. This means that there is no way to bias insertion towards the correct orientation, and that insert will be vulnerable to re-cutting and excision. This will be explored in the next chapter.

The work in this chapter has shown that specific non-coding regions of the *Mecp2* gene can be successfully targeted by both TALEN and CRISPR constructs and that robust cutting can be achieved. The best performing TALEN (#333) and CRISPR (B52 and B86) constructs achieved particularly high efficiency and were therefore used for the work described in the next chapter.



# Chapter 5

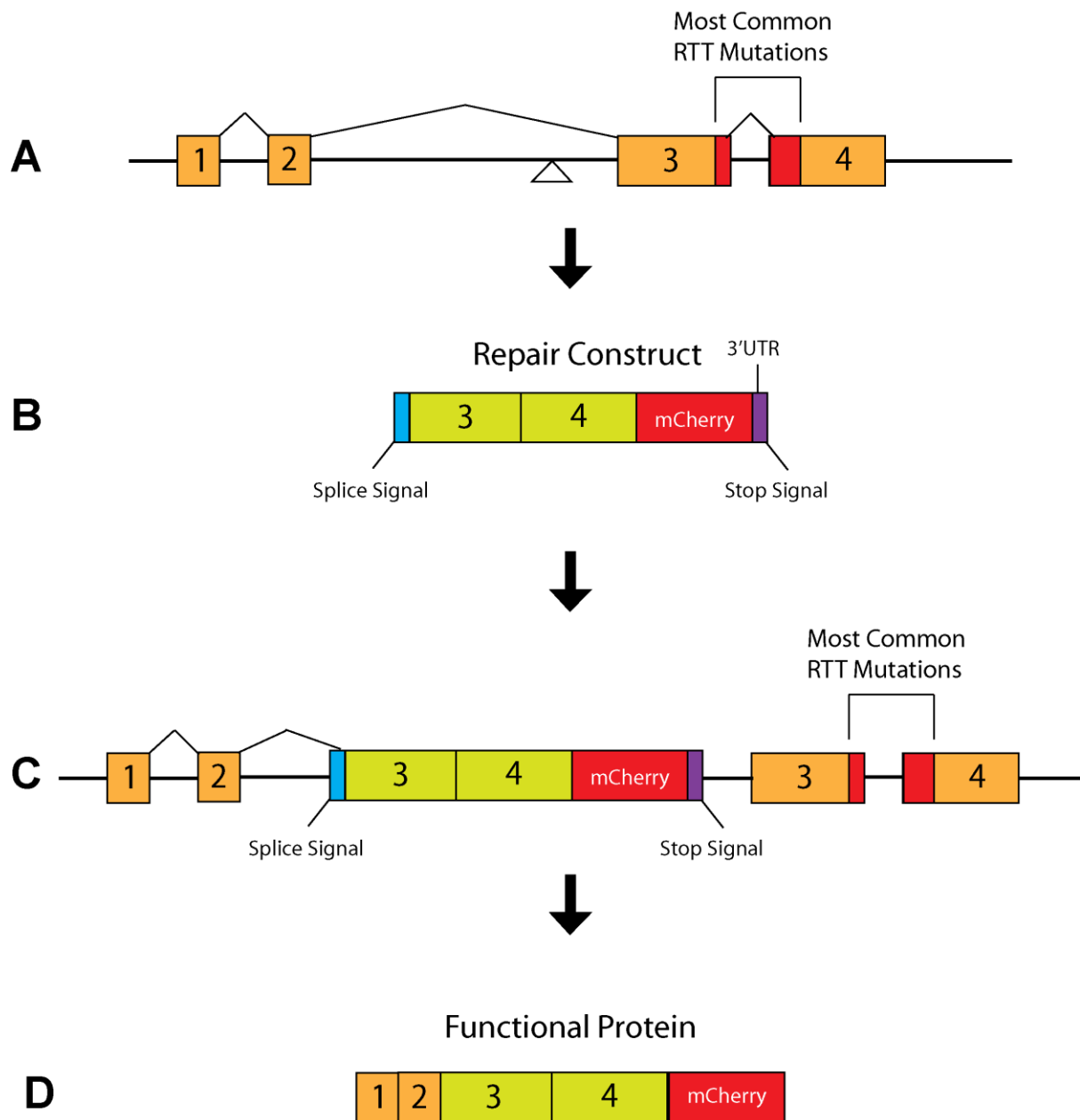
---

## Targeting a repair constructs into intron 2 of the *Mecp2* gene

### 5.1 Introduction

While genome editing tools show great promise in tackling human disease (Schwank et al., 2013; Mussolino et al., 2014; Wang et al., 2014a, 2014b; Smith et al., 2015), they have so far not been utilised in the treatment of nervous system disorders. As described previously (see section 4.1), RTT presents many challenges for this type of approach, the most significant of which is the low levels of activity in neurons of the precise homologous recombination (HR) repair pathway (San Filippo et al., 2008; Jeppesen et al., 2011). The reliance on the mutagenic non-homologous end joining (NHEJ) pathway in these cells means that any attempts to directly modify mutations in coding regions of the gene is likely to lead to further damage rather than repair.

To overcome this issue, a novel strategy to repair RTT causing mutations was developed based on the approach of Maresca and colleagues (Maresca et al., 2013), who showed that exogenous DNA could be efficiently inserted into the genome at TALEN induced DNA break sites. Building on this insight we designed an *Mecp2* repair construct that could be ligated into non-coding regions of the *Mecp2* gene to replace downstream mutated DNA (**Fig. 5.1**). In this approach, WT copies of *Mecp2* exons 3 and 4, the site of the vast majority of RTT causing mutations, are ligated into intron 2 of the gene, upstream of the equivalent endogenous exons. By including suitable splice acceptor sites and transcription terminators these inserted exons should splice to upstream exons and replace the downstream endogenous exons in the final transcript. If successful, this approach would provide a single strategy for the repair of most RTT causing mutations. To be successful the repair construct must contain the 3'-splice acceptor elements necessary for splicing to upstream exons 1 and 2. These include an almost invariant AG dinucleotide sequence at the end of the intron, known as the 3'-splice acceptor site, and an upstream branch point that includes an adenine base (Black, 2003; Faustino and Cooper, 2003). In vertebrates, the branch point sequence is partially conserved and is located 10-50 bp upstream of the 3'-splice site (Green, 1991; Gao et al., 2008).

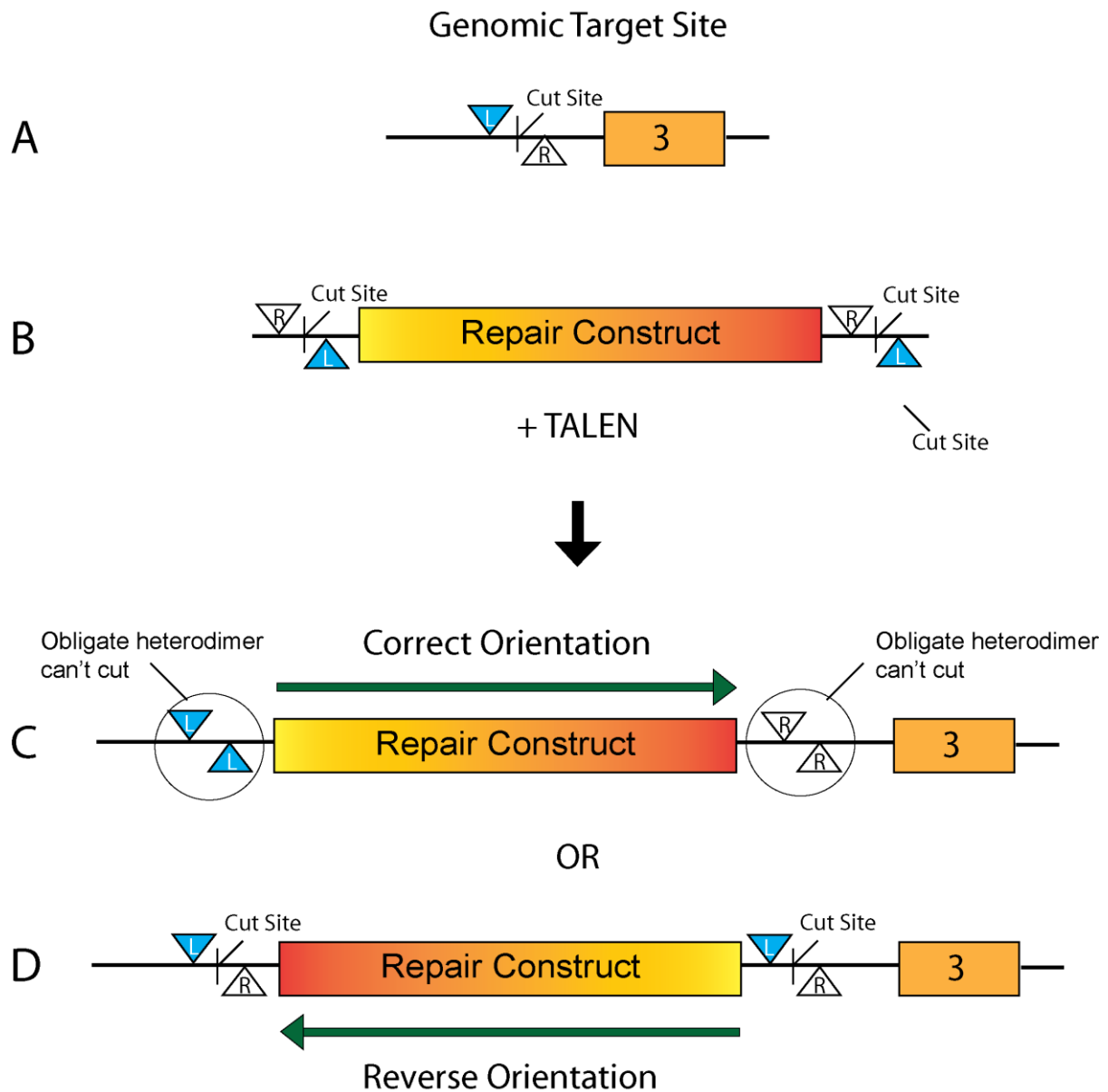


**Figure 5.1 – Outline of intron 2 repair strategy**

Summary diagram outlining the genome editing repair strategy for RTT. To aid clarity components are not drawn to scale. **(A)** *Mecp2* gene diagram showing exons (yellow boxes) and introns (black horizontal lines) and site of most common RTT causing mutations (red areas). White arrowhead indicates TALEN and CRISPR target site for repair construct insertion. **(B)** Diagram of the basic components of the repair construct. **(C)** *Mecp2* gene diagram after successful insertion of repair construct into the target site. Splice signal (blue box) will cause exon 2 to splice to the beginning of the repair construct (black diagonal lines) and transcription stop signal (purple box) will prevent endogenous downstream exons being included in final transcript. **(D)** Final protein encoded by endogenous exons 1 and 2 and coding region of repair construct, including WT copies of exons 3 and 4.

The importance of the branch point is illustrated by a study that showed that mutations in this region lead to a dramatic reduction in splicing efficiency (Reed and Maniatis, 1988). In addition, vertebrate genes often contain a 10-20 bp polypyrimidine tract that is located between the 3'-splice site and the branch point (Ruskin 1988). This suggests that the last 50 bp of intron 2 should contain all the elements necessary for splicing of the repair construct to upstream exons.

The ligation repair approach has two major obstacles that prevent efficient insertion of a repair construct. First, if the orientation of the repair construct cannot be controlled then 50% of the insertions will be in the reverse orientation. Second, since ligation of the insert into the genome can reconstitute the TALEN or CRISPR binding sites, the insert could be repeatedly cut back out and re-ligated, leading to DNA loss. To overcome these obstacles, Maresca and colleagues developed a TALEN based strategy that would promote insertion in the correct orientation and prevent the knocked-in repair construct from being retargeted for cutting. Crucial to this strategy was the use of the obligate heterodimer version of the FokI nuclease, which only functions when a left arm and right arm come together to form a dimer (Michal Szczypek et al., 2007; Yannick Doyon et al., 2010). It works by inverting the TALEN arm recognition sites in the repair construct, such that they are in the opposite orientation from the genomic target sites (**Fig. 5.2-A&B**). This means that when the repair construct is ligated into the genome in the correct orientation, the original TALEN target sites are not reconstituted and the insert is not vulnerable to re-cutting (**Fig 5.2-C**). In comparison, when insertion is in the reverse orientation, the TALEN target sites are reconstituted and can be re-cut from the genome (**Fig. 5.2-D**). This combination strongly biases insertion in the correct orientation. As of yet, no comparable strategy has been developed for CRISPR-Cas9 mediated insertion.



**Figure 5.2 –Strategy for ensuring correctly orientated insertion of the repair construct**

Summary diagram outlining the modified TALEN approach for biasing insertion of the repair construct in the correct orientation. **(A)** Partial *Mecp2* gene diagram showing TALEN arm target sites (blue arrow = left arm, white arrow = right arm) upstream of exon 3 (yellow box). **(B)** Inverted TALEN target sites flank either end of the repair construct (yellow box). **(C)** Insertion of the repair construct in the correct orientation leads to two left arm target sites at the 5'-end and two right arm sites at the 3'-end. These combinations will not be cut by obligate heterodimer FokI. **(D)** Insertion in reverse orientation leads to left and right arm target sites at either end, leaving the insert vulnerable to being re-cut by FokI nuclease.

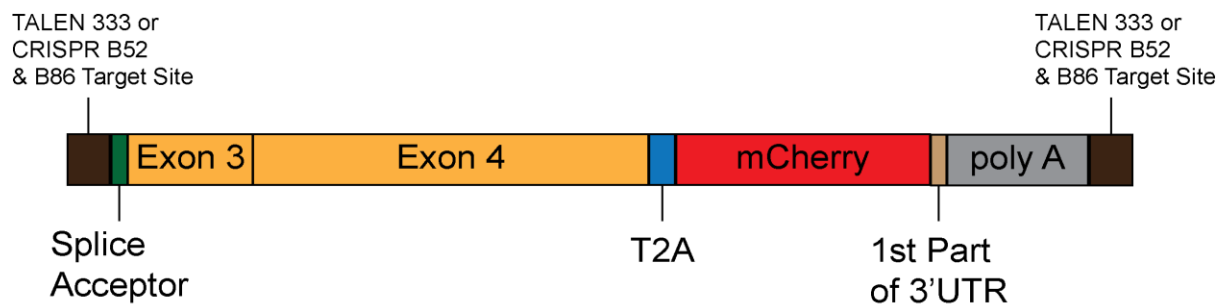
## 5.2 Aims

The results from the previous chapter (**Chapter 4**) showed that several TALEN and CRISPR-Cas9 constructs were capable of selectively targeting and cutting a specific region of *Mecp2* intron 2. In this chapter, these constructs were then tested for their ability to mediate the insertion of an *Mecp2* repair construct upstream of exons 3 and 4 of the gene. The objectives of this chapter were to:

- (1) To design and synthesise an *Mecp2* repair construct capable of repairing mutations in exons 3 and 4 of the gene
- (2) To test the ability of TALEN and CRISPR-Cas9 constructs developed in **Chapter 4** to mediate insertion of the repair construct into the genome
- (3) To assess if the inserted construct is capable of producing repaired MeCP2 protein

### 5.3 Design of the repair construct

To assess if the best performing TALEN and CRISPR-Cas9 constructs (see **Chapter 4**) could successfully target exogenous DNA into intron 2 of the *Mecp2* gene, a repair construct was designed (**Fig. 5.3**).



**Figure 5.3 – Components of the *Mecp2* repair construct**

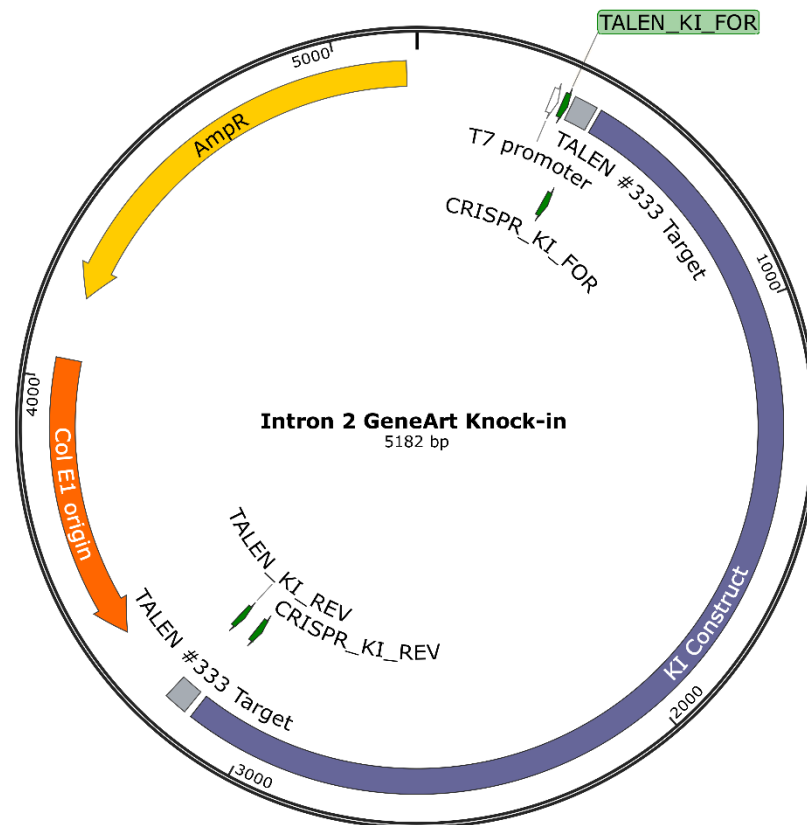
Detailed diagram showing the components of the *Mecp2* repair construct (components are drawn to scale). The construct contained WT copies of exons 3 and 4 of *Mecp2* (yellow boxes) and was flanked at either end by TALEN or CRISPR target sites (brown boxes). Upstream of the exons was the last 50 bp of intron 2 of *Mecp2* containing the 3'-splice acceptor elements of intron 2 (green box). A *Thoseaasigna* virus 2A sequence (T2A - blue box) was inserted between exon 4 and mCherry (red box). Downstream of the mCherry sequence was the first 33 bp of the *Mecp2* 3'UTR (pink box) the endogenous distal polyadenylation signal of the *Mecp2* gene, with surrounding sequence (grey box).

This construct contained WT sequence for exons 3 and 4 of *Mecp2* in a minigene format. If successfully integrated into *Mecp2* intron 2, these WT exons would replace the endogenous versions, and thus any RTT-causing mutations in these regions would be bypassed. The coding sequence of mCherry fluorescent protein was then cloned downstream of these exons to enable easy detection of protein expression from the repair construct after insertion. Since it was not clear if directly fusing a reporter to the MeCP2 protein would perturb normal function, a *Thoseaasigna* virus 2A (T2A) sequence was inserted between *Mecp2* exon 4 and the mCherry sequence. This sequence leads to ribosomal skipping during protein synthesis (Daniels et al., 2014) and enables mCherry to be synthesised as a separate protein, but still under the control of the same *Mecp2* promoter and thus expressed at the same levels as the *Mecp2* transcript. The last 50 bp of *Mecp2* intron 2 was cloned directly upstream of exon 3. As described previously, this region was predicted to contain all of the elements necessary for splicing to upstream exons (Green, 1991; Gao et al., 2008). At the 3'-end of the construct was

the sequence for the distal part of the *Mecp2* 3'UTR, including the polyadenylation signal necessary for terminating transcription and preventing downstream exons from being included in the final *Mecp2* transcript. This was crucial, as the mCherry sequence contained a stop codon, which would be recognised as aberrant by the nonsense mediated decay system if the transcript contained further downstream exons (Chang et al., 2007). This would lead to mRNA degradation and prevent protein expression. Finally, to enable the repair construct to be cut simultaneously with the genomic target DNA, the construct was flanked at either end with target sites for either TALEN # 333 or CRISPR B52 and B86, the best performing constructs as described in the previous chapter. Crucially, the construct did not contain a mammalian promoter or an in-frame start codon. This ensured that the protein coding regions of the repair construct could only be expressed if the construct was ligated into the genome, and correctly spliced to endogenous exons.

## 5.4 Synthesis of the repair construct

Since the repair construct contained multiple different components, it was decided to get the construct commercially synthesised (GeneArt, Life Technologies). The construct was delivered as a plasmid in the pMA-RQ GeneArt backbone and contained the repair construct sequence as well as the elements necessary for replication and antibiotic selection, but did not include a mammalian promoter sequence (**Fig. 5.4**). Initially, it was hoped that this plasmid could be used directly to test for TALEN mediated insertion. However, large scale plasmid DNA preps produced extremely low yields, suggesting that the insert was toxic when expressed in bacteria. Examination of the plasmid sequence showed that it contained the promoter from the T7 bacteriophage, which is capable of driving high transcription rates in *E. coli* cells. In order to prepare a suitable yield of high quality plasmid DNA, primers were designed to PCR amplify the repair construct for subsequent cloning into an alternative plasmid, which did not have a bacterial promoter.



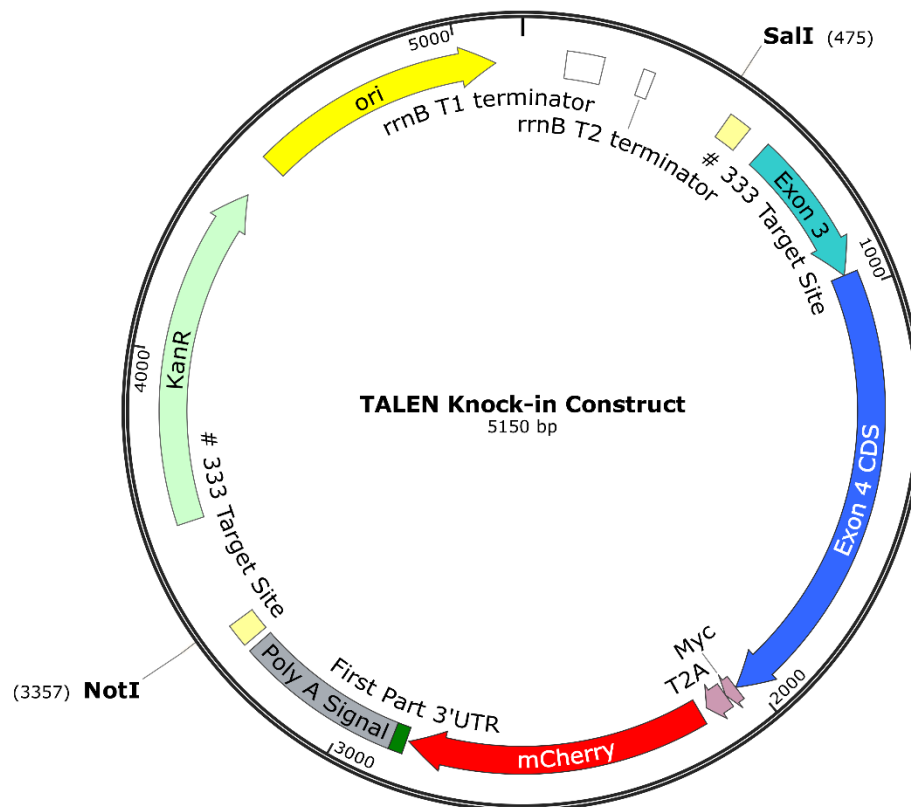
**Figure 5.4 – Map of GeneArt *Mecp2* repair construct**

*Mecp2* repair plasmid synthesised by GeneArt (Life Technologies). Repair construct sequence (purple box) was flanked at either end by target sites for TALEN #333 (grey boxes) in pMA-RQ backbone. Green arrows show primers for PCR amplification of construct for downstream cloning. Col\_E1 origin (orange arrow) is a bacterial origin of replication and antibiotic selection was via the ampicillin resistance gene (AmpR – yellow arrow). The T7 promoter (white arrow) was upstream of the repair construct.

## 5.5 Cloning of TALEN repair construct

The commercially synthesised repair construct was designed to be flanked by target sites for TALEN # 333. PCR primers TALEN\_KI\_FOR and TALEN\_KI\_REV, containing SalI and NotI primer tails respectively, were used to amplify the entire repair construct sequence, for subsequent cloning into plasmid pENTR-1A (**Fig. 5.5**). This plasmid was chosen as it contained no bacterial promoter, and in addition strong transcriptional terminators were present upstream of the repair construct insertion site, ensuring that there was no residual transcription due to the upstream kanamycin promoter, thus preventing any bacterial toxicity.

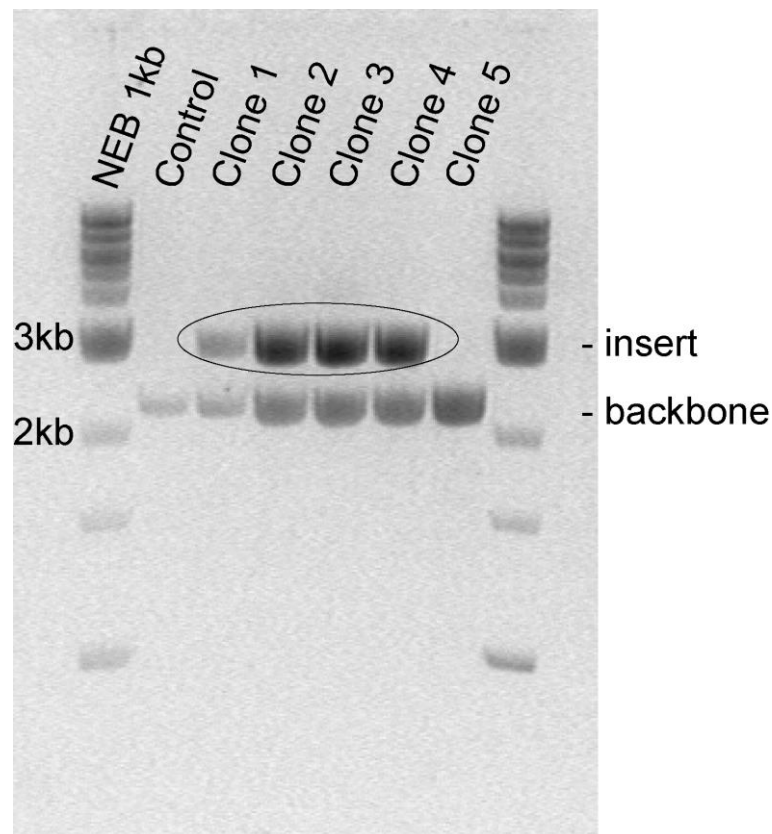




**Figure 5.5 – Map of the final TALEN repair plasmid**

Final TALEN repair plasmid. Shows main components of the TALEN construct (previously described) cloned between SalI and NotI restriction sites. Plasmid replication was enabled by bacterial origin of replication (ori – yellow arrow). Bacterial selection was via the kanamycin resistance gene (KanR – green arrow). *E. coli* rrnB transcriptional terminators T1 and T2 (white boxes) ensured that there was no background transcription of the repair construct.

To identify successfully cloned plasmids, several bacterial colonies were picked and cultured overnight. Plasmid DNA was then extracted by mini-prep, digested with SalI and NotI, and the products separated by agarose gel electrophoresis, to identify any positive clones (**Fig. 5.6**). Four of the five clones tested showed the correct insert size and this was confirmed by sequence analysis. A large scale plasmid DNA prep of one of the successful clones produced a high yield of high-quality DNA, consistent with the previous poor yield being due to bacterial toxicity.

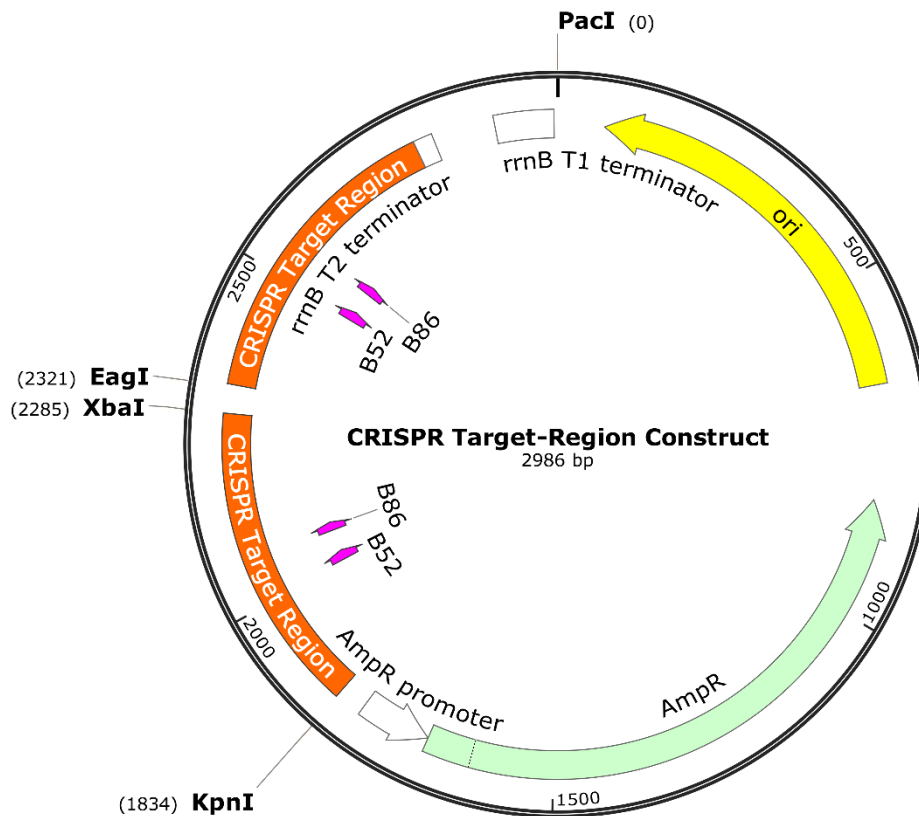


**Figure 5.6 – Restriction digest analysis of cloning of TALEN repair construct**

Plasmid DNA was digested with SalI and NotI and separated by agarose gel electrophoresis. Successful cloning was indicated for clones 1-4 by the presence of a 2882 bp band (black circle). Clone 5 was unsuccessful as indicated by the presence of the vector band (2268 bp) only. The control lane was cut pENTR-1A plasmid only. Ladder was NEB 1 KB Quick.

## 5.6 Cloning of CRISPR repair construct

To generate a repair construct flanked by target sites for CRISPR B52 and B86, a two-step cloning process was carried out. Due to restriction site incompatibility with the previously used pENTR-1A plasmid, an alternative plasmid (Collectis-4829) with features similar to pENTR-1A was chosen for cloning. For the CRISPR target site, a 400 bp region containing CRISPR 52 and B86 recognition sites, as well as surrounding sequence, was identified. A DNA insert, consisting of two copies of the target region as well as strong transcriptional terminators, was synthesised commercially (Integrated DNA technologies), with the insert flanked by KpnI and PacI restriction sites. XbaI and EagI sites were synthesised between the two target region copies to allow future cloning of the repair construct sequence (**Fig. 5.7**).

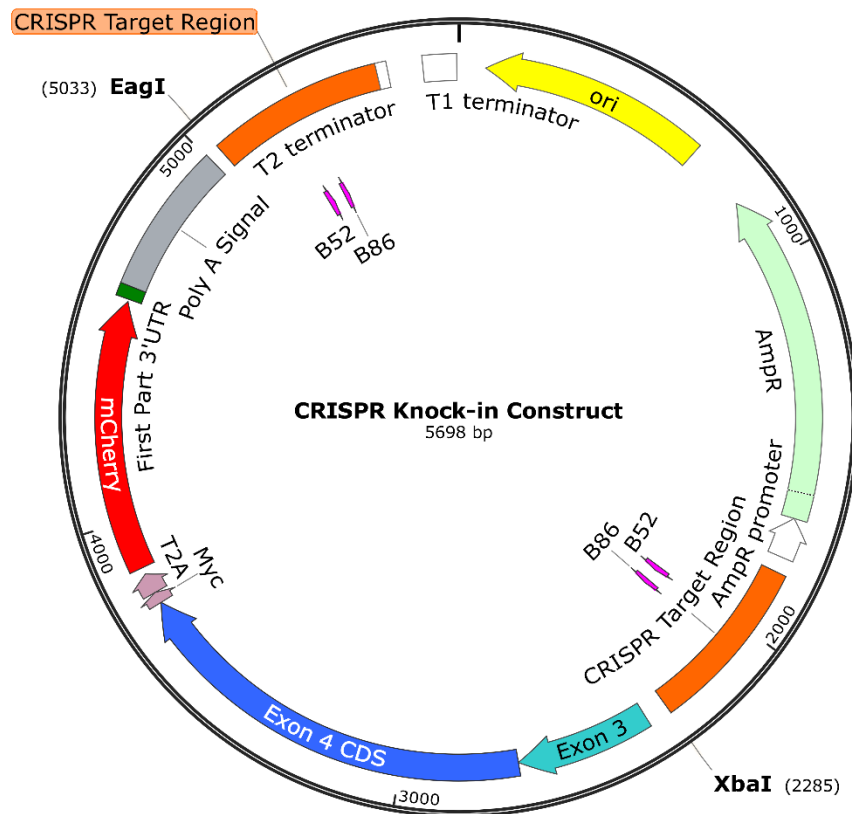


**Figure 5.7 – Map of CRISPR-target region construct**

Intermediate construct containing two copies of the 400 bp intron 2 CRISPR target region (orange boxes) cloned between KpnI and PacI restriction sites. EagI and XbaI sites between the two copies of the target region allowed for future cloning of the repair construct between the two regions. Pink arrows show the location of the CRISPR guide B52 and B86 target sites. Plasmid also contained *E. coli* transcriptional terminators T1 and T2 (white boxes), a bacterial origin of replication (ori – yellow arrow) and the ampicillin resistance gene (AmpR – green arrow).

The DNA construct and the Collectis\_4829 plasmid were both digested with KpnI and PacI, and the resulting products were ligated, transformed into DH5 $\alpha$  cells, and plated on ampicillin-containing agar plates. Several colonies were picked, cultured overnight, and plasmid DNA was extracted by mini-prep. Plasmid DNA from each colony was then digested with KpnI and PacI, and the products separated by agarose gel electrophoresis to identify any clone containing the correct insert. All clones tested contained the correct sized insert.

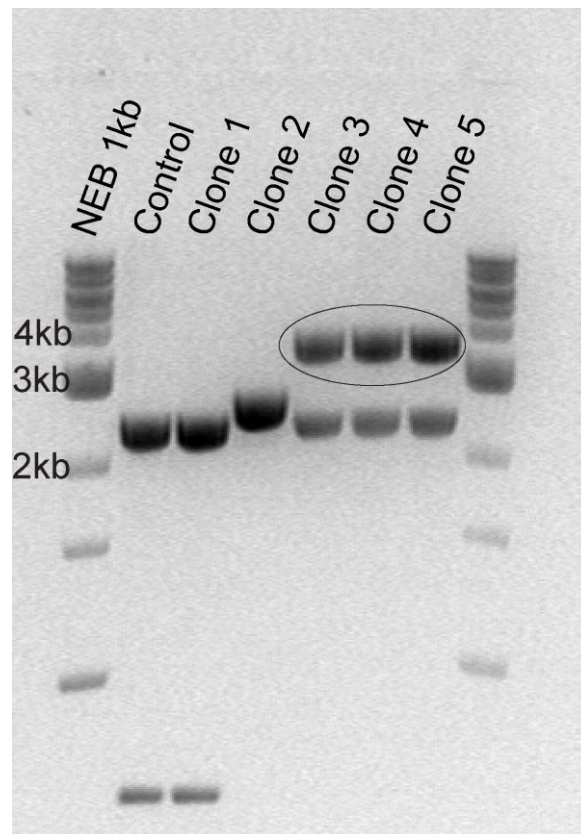
To create the final CRISPR repair construct (**Fig. 5.8**), the repair construct fragment was first PCR amplified from the GeneArt plasmid (**Fig. 5.4**) using primers CRISPR\_KI\_FOR and CRISPR\_KI\_REV, which contained XbaI and EagI tails respectively.



**Figure 5.8 – Map of final CRISPR repair construct**

Final CRISPR repair plasmid map. The main components of the CRISPR construct (previously described) are cloned between XbaI and EagI restriction sites and flanked by the CRISPR target region (orange box). Plasmid replication is enabled by the bacterial origin of replication (ori – yellow arrow) and bacterial selection is via the ampicillin resistance gene (AmpR – green arrow). *E. coli* rrnB transcriptional terminators T1 and T2 (white boxes) ensured that there was no transcription of the repair construct.

The resulting PCR product and CRISPR target-region plasmid (**Fig. 5.7**) were then digested with XbaI and EagI, ligated, transformed into DH5α cells, and plated on ampicillin-containing agar plates. Several colonies were picked and cultured overnight, and plasmid DNA was extracted by mini-prep. Plasmid DNA was then digested with XbaI and PacI and the products separated by agarose gel electrophoresis, to identify any clone containing the correct repair construct insert (**Fig 5.9**). Three of the five clones showed the expected digestion product. However, despite being contained in a plasmid without a bacterial promoter large scale plasmid DNA preps still produced extremely low yields that were unsuitable for transfection. Time restraints meant that it was not possible to further investigate this failure, so the TALEN # 333 approach was utilised in later experiments.

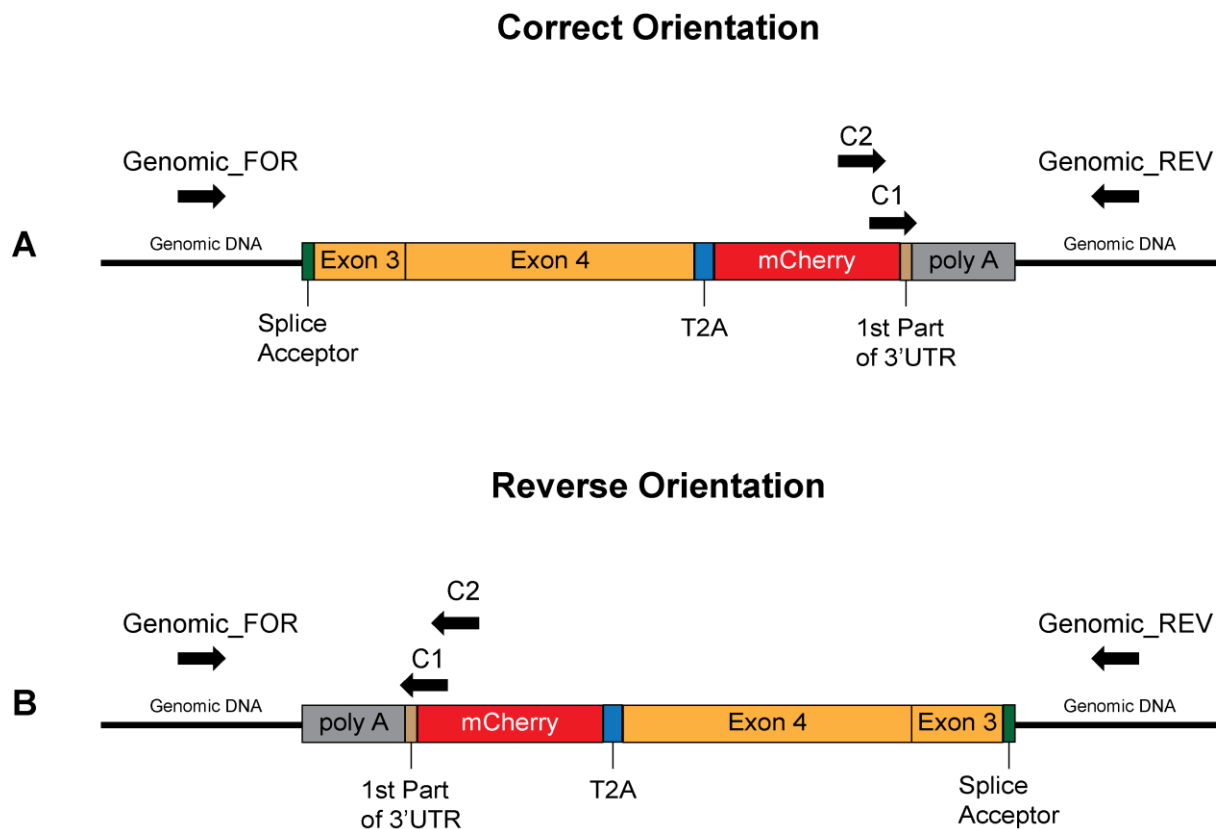


**Figure 5.9 – Restriction digest analysis of cloning of CRISPR repair construct**

Plasmid DNA was digested with XbaI and PacI and separated by agarose gel electrophoresis. Successful cloning was indicated for clones 3-5 by the presence of a 3413 bp band (black circle). Clones 1 and 2 did not contain the correct insert as indicated by the presence of the vector band (2285 bp) only, although the band in clone 2 ran slightly slower than other vector bands for unknown reasons. The control lane was cut CRISPR target region plasmid (**Fig. 5.7**) only. Ladder was NEB 1 KB Quick.

## 5.7 Design of PCR system for detecting successful repair construct insertion

In order to detect insertion of the repair construct into the *Mecp2* intron 2 target site, a PCR assay was designed (**Fig. 5.10**). The primers consisted of two genomic primers flanking the target site (Genomic\_FOR and Genomic\_REV) as well as two partially nested insert specific primers (C1 and C2). This combination of primers allowed both successful insertion and the orientation of the insert to be detected.



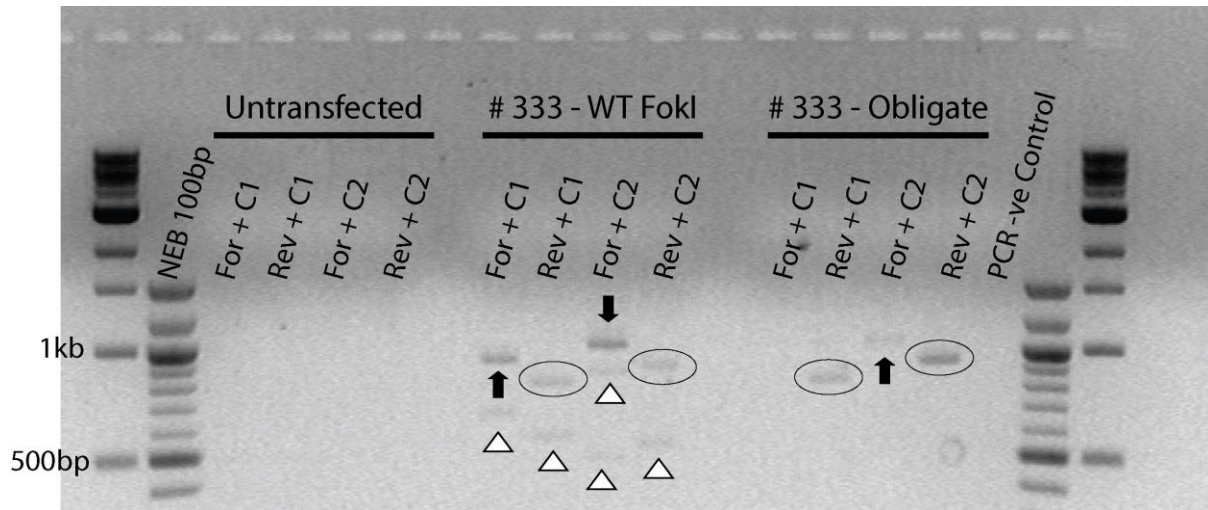
**Figure 5.10 – PCR detection of repair construct insertion**

Diagram outlining PCR-based system for detecting intron 2 insertion events. The design of the primers allowed the orientation of the repair construct insertion to be determined. **(A)** Correct insertion of the construct should produce a PCR product only with a combination of the reverse flanking genomic primer (Genomic\_REV) and either of the two insert specific primers (C1 & C2). **(B)** Insertion of the construct in the reverse orientation should produce a PCR product only with a combination of the forward flanking genomic primer (Genomic\_FOR) and either of the two insert specific primers (C1 & C2).

## 5.8 Repair construct successfully inserted into target site using TALEN # 333

To test if TALEN # 333 could facilitate the insertion of the repair construct into the intron 2 target region, an easy to transfect embryonic carcinoma cell line, P19 cells, were co-transfected with 500 ng each of both TALEN arms and the TALEN repair construct (TALEN arm:repair construct molar ratio of 1:1.5). Cells were incubated for 48 hours and genomic DNA was extracted. The genomic DNA from each transfection was then assessed for

insertion using the PCR assay described above (**Fig. 5.10**). PCR products from this assay were separated on an agarose gel. (**Fig. 5.11**).



**Figure 5.11 – PCR analysis shows TALEN # 333 mediated repair construct insertion**

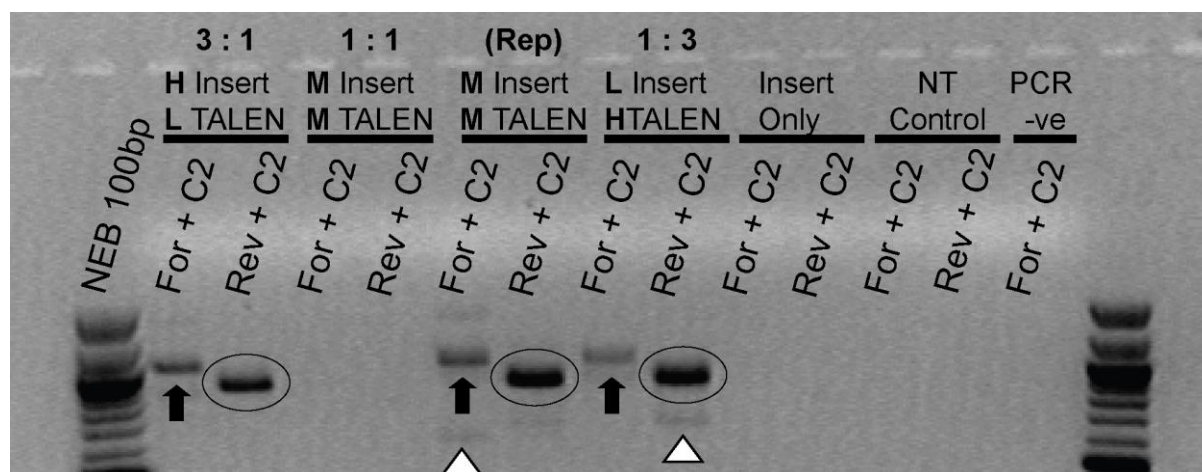
Genomic DNA was extracted from both transfected and untransfected cells, and assessed for successful repair construct insertion using PCR. Agarose gel shows resulting PCR products from untransfected cells (left), WT #333 + repair construct transfected cells (middle), and obligate heterodimer #333 + repair construct transfected cells (right). Presence of 830 bp and 920 bp bands in REV + C1 and REV + C2 lanes respectively (black circles) indicates insertion of the repair construct in the correct orientation. 960 bp and 1050 bp bands in FOR + C1 and FOR + C2 lanes respectively (black arrows) indicates insertion of repair construct in the wrong orientation. Bands of unexpected size (white arrowheads). Ladder is NEB Quick 100 bp and 1 kb.

The presence of PCR products of the expected size suggested that both TALEN # 333 constructs, containing either the WT or obligate heterodimer version of FOKI, were capable of mediating insertion. However, while the WT version led to insertion in both the correct and reverse orientation, the obligate heterodimer version was strongly biased towards insertion in the correct orientation only. In addition, the WT version also led to the insertion of various truncated PCR products in both orientations, while the obligate heterodimer did not.

P19 cells were transfected with varying concentrations of repair construct and TALEN arm plasmids to assess the effect on efficiency of insertion. Since results from the previous experiment showed that the obligate heterodimer version of TALEN # 333 led to the repair construct being inserted preferentially in the correct orientation, it was decided to focus on



this version for the following experiments. Cells were transfected with 3.5  $\mu\text{g}$  DNA, with an insert: TALEN arm molar ratio of 3:1, 1:1, or 1:3, and incubated for 48 hours. Genomic DNA was then extracted from the cells and assessed for insertion using the PCR assay.



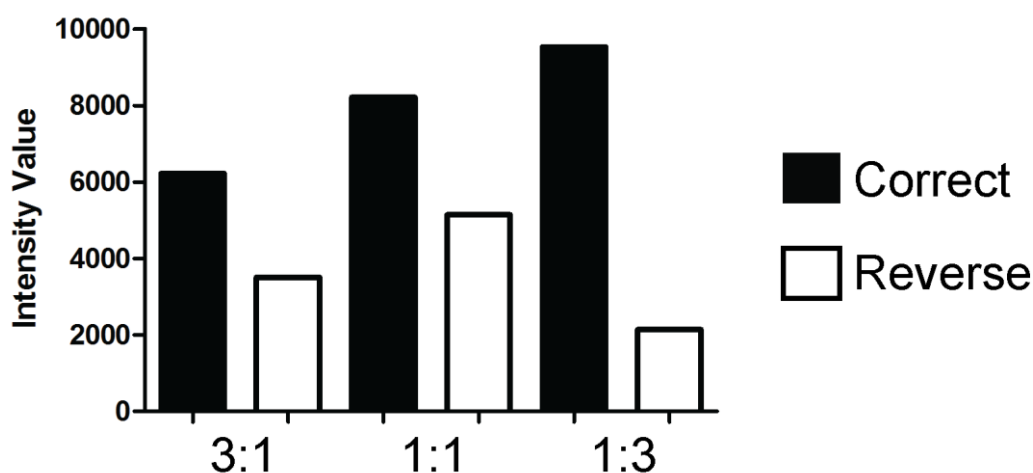
**Figure 5.12 – Effect of plasmid concentration on repair construct insertion efficiency**

Genomic DNA was extracted from both transfected and untransfected cells, and assessed for successful repair construct insertion using PCR. Agarose gel shows PCR products from transfections of various concentrations of insert and obligate heterodimer TALEN # 333 plasmid including; high insert (2  $\mu\text{g}$ ) / low TALEN arm (750 ng each); medium insert (1.5  $\mu\text{g}$ ) / medium TALEN arm (1  $\mu\text{g}$ ); low insert (1  $\mu\text{g}$ ) / high TALEN arm (1.25  $\mu\text{g}$ ); as well as insert only (no TALEN plasmid) and WT non-transfected controls. Also included is a no template negative PCR control. Molar insert:TALEN arm ratios are shown above the lane labels. Highlighted are PCR products indicating repair construct insertion in correct (black circles) or reverse (black arrows) orientation. Bands of unexpected size (white arrowheads) indicate insertion of truncated versions of the repair construct. Ladder is NEB Quick 100 bp. M insert / M TALEN transfection was repeated, as first one was suspected to have failed due to an air bubble in the Neon pipette during electroporation.

The assay results showed that all ratios of insert: TALEN arm produced strong bands of the expected size in the REV + C2 reactions, indicating insertion in the correct orientation (**Fig. 5.12**). However, in contrast to the previous experiment (**Fig. 5.11**) bands were also produced from the FOR + C2 reactions, indicating some insertion in the reverse orientation. There was not a linear relationship between TALEN concentration and insertion efficiency, with the medium TALEN concentration producing the highest amount of total insertion compared to the low and high concentrations as determined by the measurement of band intensity using ImageJ (intensity value; low TALEN = 9680; med TALEN = 13,405; high TALEN = 11,433; **Fig. 5.13**). However, the combination of a high TALEN concentration and low repair



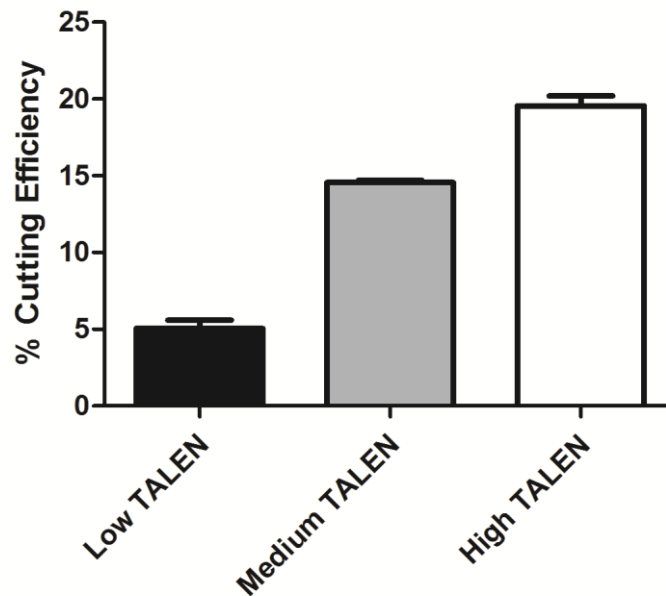
construct concentration did lead to the highest level of overall insertion in the correct orientation (intensity value; low TALEN = 6315; med TALEN = 8361; high TALEN = 9332) as well as a greater proportion of the total repair construct insertions being correctly orientated (proportion; low = 65%; med = 62%; high = 82%) .



**Figure 5.13 – Effect of insert:TALEN arm ratio on genomic insertion efficiency**

Repair construct insertion was assessed using a PCR assay (see **Fig. 5.13**). The band intensity of PCR products was measured using ImageJ in order to quantify relative insertion efficiency. Columns show the intensity value of the PCR product band for each insert:TALEN arm ratio. Black columns show the intensity value for products in correct orientation (Rev + C2) while white columns show value for those in reverse orientation (For + C2).

To examine the relationship between insertion rates and cutting efficiency, genomic DNA from each of the different insert:TALEN ratios was assessed by TIDE analysis (**Fig. 5.14**). As expected, cutting efficiency increased as the concentration of TALEN increased, ranging from 5% efficiency for the lowest concentration to 19.5% for the highest.



**Figure 5.14 – TALEN cutting efficiency increases with increasing plasmid concentration**

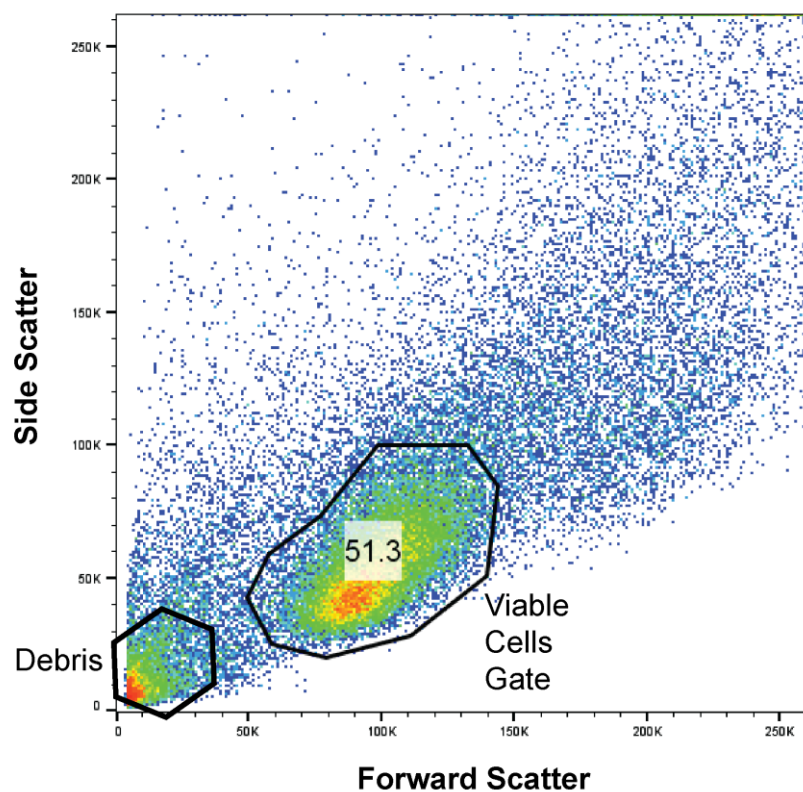
TALEN cutting efficiency for three different TALEN arm plasmid concentrations (see **Figure 5.13**) was estimated by TIDE analysis. Columns show mean % cutting efficiency  $\pm$  SEM for two replicates of either low (500 ng – black box), medium (1  $\mu$ g – grey box), or high (1.5  $\mu$ g) plasmid amount.

## 5.9 Repair construct insertion with TALEN # 333 does not lead to protein expression

The previous experiments showed that obligate heterodimer TALEN # 333 was capable of mediating insertion of the repair construct into intron 2 of the *Mecp2* gene. To determine if this insertion led to subsequent protein expression, flow cytometry was used to quantify the number of cells expressing the mCherry tag. Since the repair construct plasmid contains no promoter of its own, the mCherry fluorescent reporter could only be expressed if the repair construct was successfully spliced to exons 1 and 2 of *Mecp2*.

P19 cells were first transfected with a total of 3.5  $\mu$ g DNA, with insert: TALEN arm ratios of 3:1, 1:1, or 1:3, as previously described (see **Fig. 5.12**). As a positive control, cells were transfected with 500 ng of an mCherry expressing plasmid (p-CMV-mCherry). This allowed the transfection efficiency to be measured, and also enabled the flow cytometry laser settings to be optimised for mCherry detection. To control for any background transcription from the repair construct, cells were also transfected with 500 ng of repair construct only. Forward and

side scatter plots from untransfected P19 cells were used to exclude debris and cell aggregates, and to identify a population of viable cells to be used for fluorescence analysis. A polygon gate was drawn around this population (**Fig. 5.15**) and this gate was used for all TALEN # 333 transfection experiments.



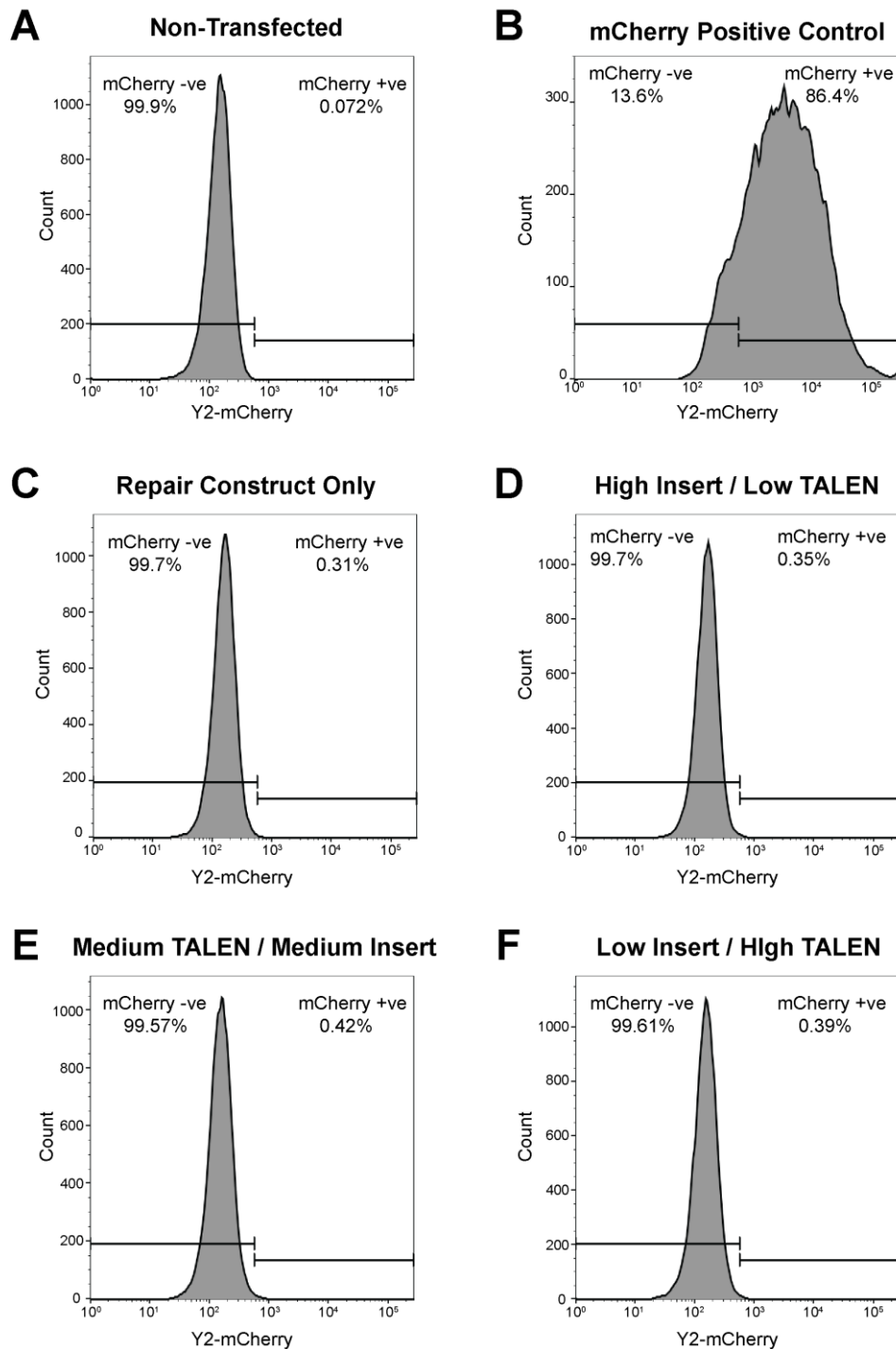
**Figure 5.15 – Gating of live cells allowed debris and aggregates to be excluded**

Flow cytometry forward versus side scatter intensity plots from untransfected P19 cells. The plot was used to exclude cell debris and aggregates, and to identify a population of viable cells (large black polygon) to be used for fluorescence analysis. This ‘viable cells’ gate was used for all future FACS experiments involving P19 cells. Value within viable gate indicates % of total cells contained within this region.

Untransfected P19 cells were also used to determine the cellular background auto-fluorescence levels. Fluorescence intensity was measured using a yellow 561 nm laser with a 615/20 nm filter, which is optimised for mCherry detection. Intensity levels plotted on a histogram, allowing two gates to be created which split cells into mCherry positive and mCherry negative populations (**Fig. 5.16-A**), with any cells expressing mCherry fluorescence

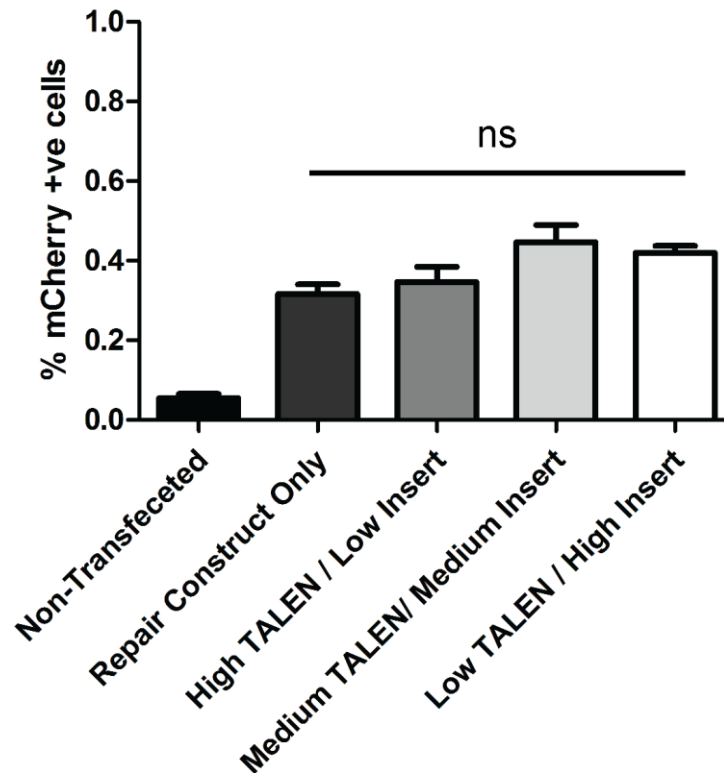
levels above the background levels seen in untransfected cells being classed as mCherry positive. Cells from all transfections were then assessed and gated using the same settings.

Results from the control transfections showed that transfection with the mCherry positive control plasmid (**Fig. 5.16-B**) led to a large proportion of cells expressing mCherry protein (86.4% of cells positive). This confirmed that the transfection efficiency was high and that the laser settings used were suitable for detecting mCherry protein expression. In contrast, transfection with the TALEN repair construct (**Fig. 5.16-C**) led to only a tiny proportion of cells expressing mCherry (0.31% of cells positive), showing that there was no endogenous expression of mCherry from the repair construct plasmid. Surprisingly, transfection of the repair construct along with either of the three insert:TALEN arm ratios (**Fig. 5.16-D-F**) also led to only a tiny proportion of cells expressing mCherry (% of cells mCherry positive; 3:1 = 0.35%; 1:1 = 0.42%; 1:3 = 0.39%), and these reactions did not differ significantly from the repair plasmid only transfection, or from each other ( $p < 0.05$ ; **Fig. 5.17**). This suggests that insertion of the repair construct into intron 2 of the *Mecp2* gene did not lead to subsequent expression of repaired MeCP2 protein.



**Figure 5.16 – Flow cytometry analysis of TALEN and repair construct transfected cells**

Representative histograms of mCherry protein expression as determined by flow cytometry. Shows results from viable population of (A) WT non-transfected cells (B) mCherry positive control plasmid (C) repair construct only (D) high insert (2  $\mu$ g) / low TALEN arm (500 ng each) (E) medium insert (1.5  $\mu$ g) / medium TALEN arm (1  $\mu$ g) (F) low insert (750 ng) / high TALEN arm (1.5  $\mu$ g). x-axis shows mCherry fluorescence intensity and the y-axis shows the number of cells. Non-transfected cells (A) were used to generate gates for determining cell populations positive and negative for mCherry expression (horizontal black lines). The % value of each population, as a proportion of the total viable cell population, is shown in top left and right of each histogram.



**Figure 5.17 – No mCherry protein expression after TALEN and repair construct transfection**

Summary of flow cytometry data from TALEN and repair construct transfected P19 cells. Results show the mean % of mCherry positive cells  $\pm$  SEM from the total viable cell population of three replicates for a variety of insert: TALEN ratios, as well as controls (see **Fig. 5.16**). Groups were compared using one-way ANOVA. ns = non-significant.

## 5.10 Construction of a GFP reporter construct

The previous experiments showed that while TALEN # 333 enabled insertion of the repair construct into the genome, this was not sufficient to produce detectable protein expression. Previous work using anti-MeCP2 antibodies has shown that undifferentiated P19 cells express detectable levels of MeCP2 protein when assessed by western blot (Hwang et al., 2007). It was therefore hypothesised that the lack of protein expression in the current experiment was due to a failure of correct splicing, rather than an inability of the *Mecp2* promoter to drive transcription in P19 cells. The repair construct used in the previous experiments contained the last 50 bp of intron 2 of the *Mecp2* gene, the region in which the 3'-elements vital for splicing are normally found (Green, 1991; Gao et al., 2008). However, the precise location of splice elements has not been determined for intron 2 of *Mecp2*. To

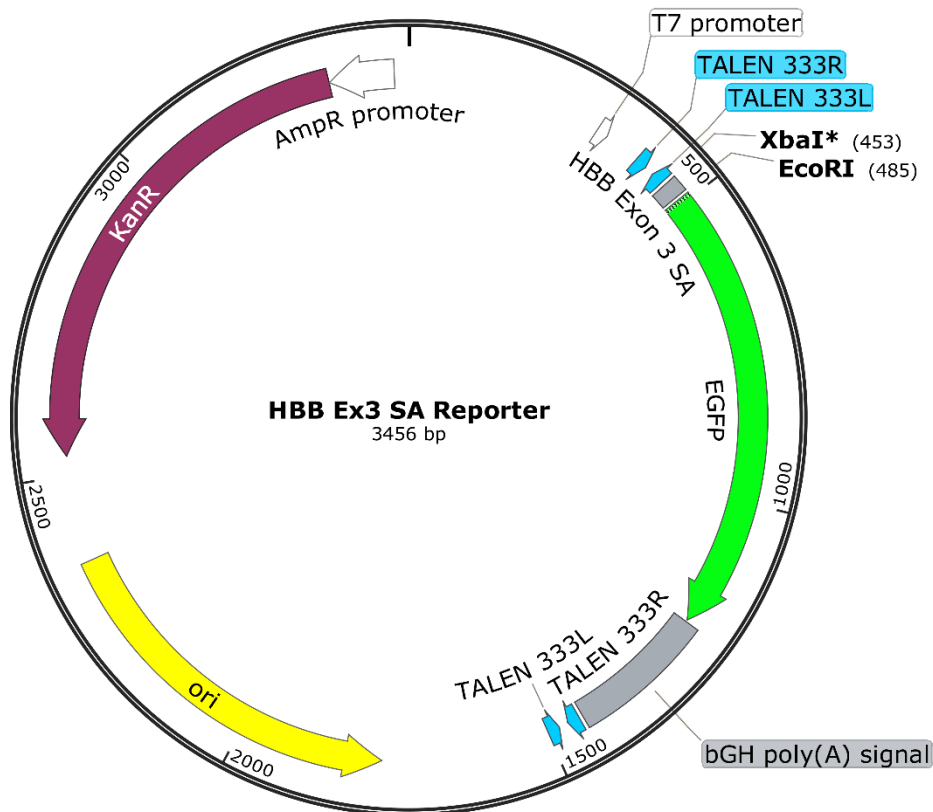
determine if the failure of protein expression was due to a lack of a functional splice acceptor site in the repair construct, a simple GFP reporter construct was designed (**Fig. 5.18**).



**Figure 5.18 – Splice acceptor GFP reporter construct**

Diagram showing the components of the GFP reporter construct (components are drawn to scale). The construct contains eGFP (green box) to allow insertion to be detected by flow cytometry. Upstream of the eGFP is a splice acceptor (yellow box), composed of either the last 70 bp of *Mecp2* intron 2 or the exon 3  $\beta$ -globin acceptor site (*HBB* SA), while downstream is the bovine growth hormone polyadenylation sequence (bGHpA – grey box) to terminate transcription. The construct is flanked at either end by the target site for TALEN # 333 (blue boxes).

The reporter contained the coding sequence of eGFP, cloned upstream of the bovine growth hormone polyadenylation sequence. Upstream of eGFP, the last 50 bp of *Mecp2* intron 2, as previously used in the repair construct, or the previously established  $\beta$ -globin exon 3 splice acceptor (*HBB* SA), was cloned. The *HBB* SA sequence has previously been used to successfully induce splicing of a construct inserted into the AAV 1 *sage* locus (Hockemeyer et al., 2009). The construct was flanked at either end by target sites for TALEN # 333 to enable insertion into the genome. To speed up the cloning process, the reporter was synthesised commercially (GeneArt, Life Technologies). This synthesised version contained the last 50 bp of *Mecp2* intron 2 and was cloned into the GeneArt pMH-RQ backbone, which did not contain a mammalian promoter. To generate a plasmid containing the alternative *HBB* SA sequence (**Fig. 5.19**), the *HBB* SA sequence was first synthesised as two complementary oligonucleotides. These oligonucleotides were synthesised so that when annealed they formed overhangs complementary to those generated by *Xba*I and *Eco*RI digestion of the reporter plasmid.



**Figure 5.19 – Plasmid map of final *HBB Ex3 SA* reporter plasmid**

Map showing main components of the final *HBB* ex3 SA GFP reporter plasmid. The  $\beta$ -globin Exon 3 splice acceptor (*HBB* ex3 SA) was cloned between *Xba*I and *Eco*RI restriction sites. Downstream is the EGFP coding sequence (green arrow) and the bovine growth hormone polyadenylation site (grey box) to terminate transcription. The reporter construct is flanked by target sites for TALEN # 333 arms (blue arrows). Plasmid replication is enabled by bacterial origin of replication (*ori* – yellow arrow) and bacterial selection is via the kanamycin resistance gene (*KanR* – purple arrow).

The annealed oligos and digested plasmid were then ligated together, transformed into DH5 $\alpha$  *E. coli* cells, and plated on kanamycin-containing agar plates. Three colonies were picked and cultured overnight, and plasmid DNA was extracted by mini-prep. Successful insertion of the *HBB* SA sequence was confirmed by Sanger sequence analysis of plasmid DNA. All three clones showed the correct insert (**Fig. 5.20**).





**Figure 5.20 – Sequence confirmation of *HBB* ex3 SA into GFP reporter**

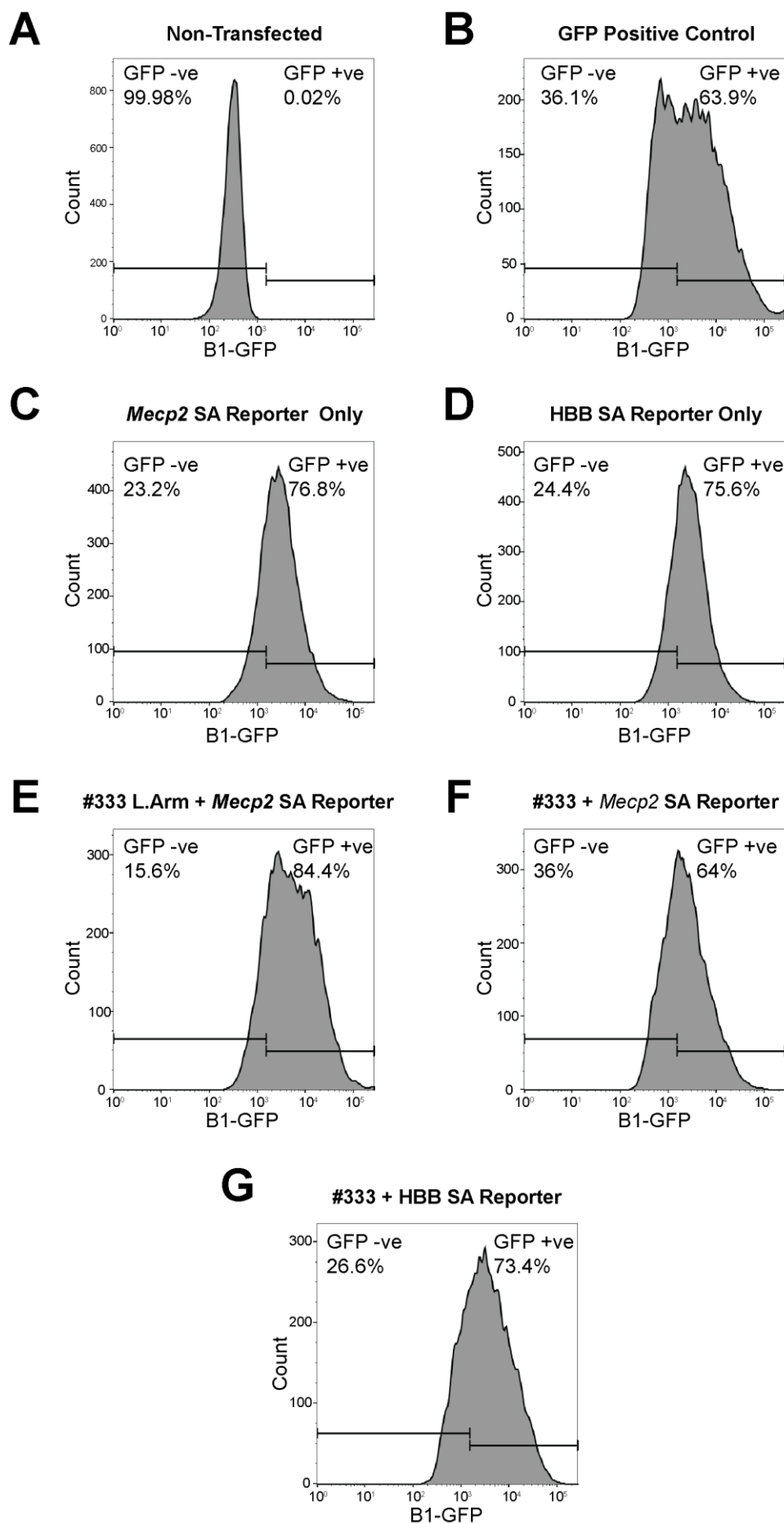
Partial sequence traces confirming successful cloning of the  $\beta$ -globin exon 3 splice acceptor (*HBB* ex3 SA) into the GFP reporter construct. Plasmid DNA from 3 clones was assessed and all showed the correct insert (blue highlighted sequence). (A) Double stranded sequence of intended final product. Shows *HBB* ex3 SA (grey box) cloned between XbaI and EcoRI restriction sites. Also shown are upstream TALEN target site (dark green box) and downstream eGFP (light green box). (B) Sequencing data for three clones. For each clone the bottom row shows the sequencing chromatogram and the top row shows the DNA sequence generated from the chromatogram. The T7F primer was used in all sequencing reactions.

P19 cells were transfected with a total of 3.5  $\mu$ g DNA, 1  $\mu$ g of the reporter plasmid and 1.25  $\mu$ g of each TALEN # 333 arm. As a positive control, cells were transfected with 500 ng of an eGFP expressing plasmid (pLenti-PGK-GFP). This allowed the transfection efficiency to be measured, and also enabled the flow cytometry laser settings to be optimised for eGFP detection. To control for any background transcription from the reporter constructs, cells were also transfected with 500 ng of each of the reporter construct only, without any TALEN arm constructs. Background GFP fluorescence intensity was measured using a blue 488 nm laser with 525/50 nm filter, which is optimised for eGFP detection. Intensity levels were then plotted as a histogram, allowing two gates to be created which split cells into GFP-positive and GFP-negative populations (**Fig. 5.21-A**), with any cells expressing GFP fluorescence levels above background being classed as GFP positive. Cells from all transfection reactions were then assessed and gated using the same settings.

Results from the control transfections showed that transfection with the GFP positive control plasmid (**Fig. 5.21-B**) led to a large proportion of GFP expressing cells (63.9% of cells positive). This confirmed that the transfection efficiency for the reactions was high and that the laser settings used were suitable for detecting GFP protein expression. Surprisingly

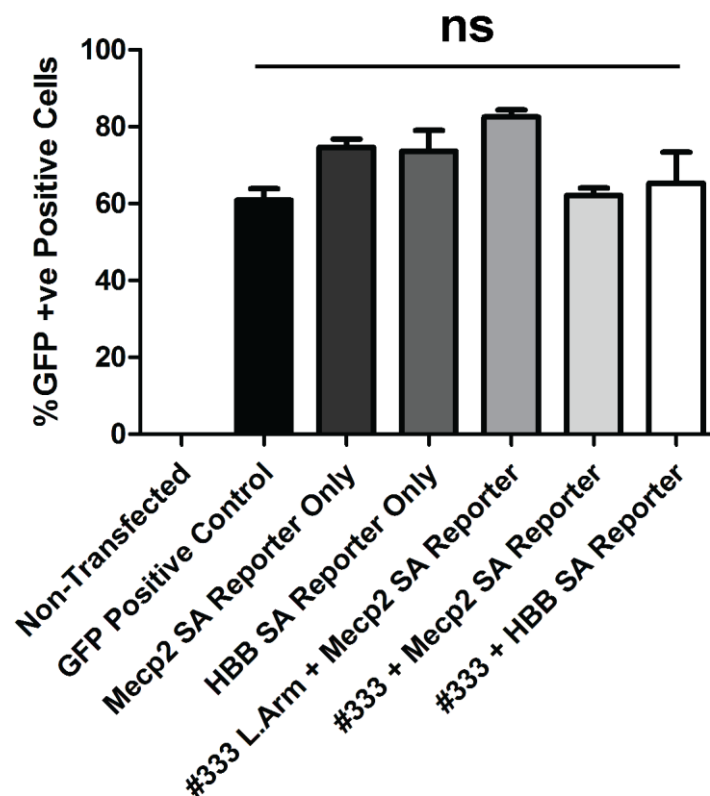
however, transfection with either the *Mecp2* intron 2 or the *HBB* SA reporter construct only (**Fig. 5.21-C&D**) also led to a large proportion of cells expressing GFP protein (76.8% and 75.6% of cells positive respectively) as did transfection of the *Mecp2* intron 2 reporter construct along with a single TALEN left arm (84.4% of cells positive; **Fig. 5.21-E**). Similarly, transfection of either reporter construct along with both TALEN # 333 arms led to a large proportion of GFP positive cells (64% for intron 2 *Mecp2* SA and 73.4% for *HBB* SA; **Fig. 5.21-F&G**).

One-way ANOVA analysis showed that the proportion of GFP positive cells in these reactions did not differ significantly from the reporter plasmid only transfection, or from each other ( $p < 0.05$ ; **Fig. 5.22**). Overall, these results suggest that for unknown reasons, GFP is being expressed from the reporter plasmid, independently of the presence of a known mammalian promoter. For this reason, the results from these experiments could not be used to assess the effect of altering the splice acceptor site on protein expression levels.



**Figure 5.21 – Flow cytometry analysis of GFP reporter construct transfected cells**

Representative histograms of GFP protein expression as determined by flow cytometry. Shows results from viable cells of (A) WT non-transfected cells, (B) GFP positive control, (C) *Mecp2* SA reporter plasmid, (D) *HBB* ex3 SA reporter plasmid, (E) TALEN # 333 left arm + *Mecp2* SA reporter, (F) TALEN # 333 left and right arm + *Mecp2* SA reporter, (G) TALEN # 333 left and right arm + *HBB* SA reporter. x-axis shows GFP fluorescence intensity and the y-axis shows the number of cells. Non-transfected cells (A) were used to identify sub-populations positive and negative for GFP expression (horizontal black lines). The % value of each population as a proportion of the total viable cell population examined is shown in the top left and right of each histogram.



**Figure 5.22 – High background GFP protein expression from reporter construct**

Summary of flow cytometry results from GFP reporter construct transfections. Results show mean ± SEM of GFP-positive cells of three replicates from the total viable cell population for the groups previously described (see Fig 5.21). Groups compared using one-way ANOVA. ns = non-significant.

## 5.11 Discussion

Insertion of a repair construct into non-coding regions of a gene, using TALEN or CRISPR-Cas9, represents a novel strategy for overcoming some of the obstacles involved in developing genome-editing based therapies for nervous system disorders. In this chapter previously validated TALEN constructs (**Chapter 4**) were used to insert a repair construct into intron 2 of the *Mecp2* gene. Results showed that co-transfection of cells with the TALEN arms and a repair construct led to efficient insertion of the repair sequence into the genome. However, further protein expression analysis revealed that successful insertion of the repair construct using TALEN did not lead to subsequent expression of repaired MeCP2 protein, perhaps indicating a failure of the repair construct to splice to endogenous exons.

An unexpected difficulty that arose during the cloning of the repair construct was the extremely low yield obtained when preparing plasmid stocks. Plasmids have a metabolic cost for their host cells, both from the increase in the amount of DNA to be copied during cell division, as well as the production of large amounts of recombinant protein (Corchero and Villaverde, 1998; Summers, 1998). This metabolic cost is increased if the plasmid encodes a protein which is toxic to bacterial cells, leading to a decrease in the growth rate of cells containing the plasmid and a subsequent overtake of the culture by plasmid free cells. Several studies have used *E. coli* cells to express and purify recombinant MeCP2 protein (Nan et al., 1993; Klose and Bird, 2004; Laget et al., 2010), but in each of these cases gene transcription was controlled by an inducible promoter, meaning that the culture could first be allowed to grow to a high density before transcription, thus mitigating the effects of any toxicity. In this study, *Mecp2* transcription was driven by the T7 bacteriophage promoter in bacterial cells. Overnight cultures showed very little growth suggesting high toxicity of the insert. In order to reduce the cells metabolic load the culture temperature was lowered to 30°C, which increased the culture density by a significant amount but still did not lead to sufficient plasmid yields. Since our target cells were mammalian, it was unnecessary to have any bacterial transcription of the repair construct. The repair construct was therefore sub-cloned into alternative plasmid vectors that did not contain a bacterial promoter. Two versions of the repair construct were cloned. One, in which the repair sequence was flanked by TALEN # 333 target sites, and another in which it was flanked by 400 bp of sequence containing target sites for CRISPR B52 and B86. This time high yields were obtained for the TALEN repair construct, supporting the hypothesis that transcription of the repair construct was toxic in bacteria. In contrast, bacterial growth and plasmid yield was still extremely poor from cells transformed

with the CRISPR repair construct, even when multiple cultures were combined in an attempt to increase the yield. It is possible this was due to the presence of inverted terminal repeats of the CRISPR target regions. Since the TALEN construct was producing suitable yields it was decided to focus on this strategy.

By careful design of primers for the PCR assay, the successful insertion of the repair construct could be easily detected. Two flanking genomic primers were chosen as well as two specific to the insert. By having flanking primers at either end, the orientation of the insertion could also be determined. This revealed that, as expected, the obligate heterodimer version of the FokI nuclease lead to a strong bias for insertion in the correct orientation. As previously described, this is because the TALEN target sites are not reconstituted after correct insertion (Maresca et al., 2013) but instead the inserted DNA is flanked at either end by homozygous sites, either two sequential left arm or right arm target sequences. Since the obligate heterodimer FokI requires both a left and right arm sequence in close approximation, the insert can no longer be cut by the enzyme and remains permanently ligated into the genome. In contrast, the WT version of the enzyme did not lead to a greater proportion of correct insertion. This is because the WT version of FokI can also function as a homodimer, meaning that sequential left arm or right arm target sites are still vulnerable to re-cutting. While the results of the PCR assay strongly suggest that the repair construct was successfully inserted, it will be necessary to confirm this by Sanger sequencing of the PCR products. In addition, this will also allow the nature of the 5' and 3'-junctions to be characterised to assess the precision of the insertion. In the WT FokI reactions a variety of smaller than expected PCR products also appeared on the gel. Since these did not appear in the obligate heterodimer reactions, it is possible that repeated rounds of ligation and re-cutting leads to truncation of the insert. However, since primers C1 and C2 bind to overlapping regions, it would be expected that the same additional PCR products would appear in both reactions. However, different products appeared in each of the PCR reactions. In future experiments these products should be characterised by Sanger sequencing in order to determine their nature. Overall, the results suggest that heterodimer FokI nuclease promotes the permanent ligation of the repair construct into the genome in the correct orientation, and that this benefit outweighs the significantly higher cutting activity seen in the WT version.

To examine the effect of altering the relative amounts of TALEN arms and repair construct transfected into cells, a variety of ratios were tested. Surprisingly, even at low TALEN arm concentrations, with a cutting efficiency of only 5%, a significant amount of insertion was

detected by the PCR assay. This was perhaps helped by the large molar excess of repair construct plasmid (3:1) in this reaction. Increasing the amount of each TALEN arm DNA in the reaction by a third led to a 200% increase in cutting efficiency but led to only a modest 38% increase in the frequency of repair construct insertion. In this reaction, the amount of repair construct plasmid was reduced suggesting that the availability of the repair construct is a significant rate-limiting step for insertion. A further increase in the TALEN concentration by a third led only to a 33% increase in cutting efficiency and a modest 12% increase in the insertion rate, suggesting that cutting efficiency does not scale linearly with increasing amounts of TALEN arm plasmid, but reaches a plateau which will likely vary depending on the TALEN sequence. Since there is a limited amount of DNA that can be inserted into a cell during transfection without causing toxicity, it is important to have TALEN arm : repair construct ratios that are optimised for maximum insertion, with the need for large amounts of repair plasmid to be balanced with the need to have sufficient levels of TALEN cutting. This is difficult to achieve with a TALEN construct with only modest activity levels, as large amounts of plasmid are needed to achieve adequate cutting. The development of improved versions of the obligate heterodimer FokI nuclease, with activity levels closer to those of the WT version, would aid future efforts.

To determine if repair construct insertion led to subsequent protein expression, flow cytometry was used to detect the mCherry fluorescent tag. In addition to being extremely sensitive, it allowed the number of cells expressing the repair construct to be precisely quantified. The repair construct itself did not contain a mammalian promoter, so mCherry protein translation would only occur if the plasmid was cut, ligated into the genome, and spliced to upstream exons of the *Mecp2* gene. As expected, transfection with the repair construct only did not lead to cells expressing mCherry. Surprisingly, when the cells were co-transfected with both obligate heterodimer TALEN arms and the repair construct, there was no significant increase in the number of cells expressing mCherry. Since the previous experiments showed that a combination of TALEN and repair construct could lead to insertion of the repair construct into intron 2 of *Mecp2*, it was hypothesised that the absence of protein expression may due to a failure of the repair construct to splice to the upstream *Mecp2* exons. For the splice acceptor it was decided to use the last 50 bp of the intron, which normally contain all the required splice acceptor elements (Reed and Maniatis, 1988; Gao et al., 2008). To test if the failure of protein expression was due to this region not being sufficient for splicing, a GFP reporter construct was made using either this sequence or a

previously validated splice acceptor region the  $\beta$ -globin gene (Hockemeyer et al., 2009). Once again the reporter plasmid did not contain a mammalian promoter, so GFP protein translation should only have occurred if the reporter was inserted into the genome and spliced to upstream exons. Similar to the previous experiments, control cells were transfected with the reporter construct only to ensure there was no background transcription of the reporter construct. Strangely, transfection with this construct led to GFP protein expression in the majority of cells (>75%), even greater than transfection with the positive control in which GFP is in a plasmid containing the PGK promoter. The reasons for this are unclear. The initial reporter construct, containing the *Mecp2* intron 2 splice acceptor sequence, was synthesised commercially by GeneArt and sub-cloned into a standard promoterless vector. The *Mecp2* intron 2 sequence was then subsequently replaced by the  $\beta$ -globin sequence, but the construct remained within the same vector. The plasmid sequence was confirmed by Sanger sequence analysis before the plasmid was shipped, which confirmed that no mammalian promoter was present. The high levels of background protein expression of the reporter construct rendered it unsuitable for determining if modifying the splice acceptor sequence led to subsequent repair reporter construct protein expression. Since no logical explanation could be found for the protein translation of the GFP reporter without genomic insertion, the reporter construct would need to be resynthesized into a different plasmid to test if there was an unknown issue with the original plasmid. Unfortunately, these experiments were carried out at the end of the available time for the project, and so it was not possible to further investigate this issue. Another possible reason for the lack of protein expression could be low levels of transcription driven by the *Mecp2* promoter in our P19 cells. Previous studies have shown that P19 cells express detectable levels of MeCP2, even in their undifferentiated state (Hwang et al., 2007), suggesting these cells would be suitable for detecting repaired protein. However, future experiments would test this directly in our own batch of cells, both at the transcript levels using RT-qPCR, as well as at the protein level using western blot, to confirm that these cells do indeed express MeCP2 protein.

The *Mecp2* intron 2 is an extremely large intron (> 42,000 bp) which presents particular challenges for the splicing machinery due to the distance between splice donor and acceptor sites. One study suggests that in order to reduce this distance, repetitive elements such as SINEs and LINEs can form stems which enable intronic RNA to fold and reduce the distance between the donor and acceptor sites (Shepard et al., 2009). Another study provided evidence that large introns may in fact be spliced out in stages, with intermediate recursive splice sites



across the intron being used (Kelly et al., 2015). These studies make it clear that the precise mechanisms by which large introns are spliced has not been determined. This makes it more difficult to rationally design a construct for insertion into intron 2 of *Mecp2* that will be efficiently spliced to upstream exons. It is possible that even with all the required splicing elements being present in the repair construct, splicing failed because the unique mechanisms used in splicing a large intron meant that this was not sufficient to out-compete the endogenous splice site. For example, if the intron RNA does indeed fold in on itself, in order to bring together distant splice donor and acceptor sites, the repair construct splice sequences may not be optimally located in the loop structure to attract the splicing machinery. Perhaps future studies could focus on smaller introns like intron 1 and 3, where standard splicing mechanisms will apply, meaning that the splice elements closest to the 5'-donor site will be preferentially utilised.

Overall, the results from this chapter suggest that TALEN constructs can be used to ligate an exogenous repair construct into specific non-coding regions of the *Mecp2* gene. They also confirm that, as described previously (Maresca et al., 2013), the obligate heterodimer FokI nuclease can be used to bias the orientation of insertion by preventing correctly orientated DNA from being subsequently re-cut from the genome. The results also show that the current design of the repair construct did not enable the exogenous DNA to be spliced to upstream exons, and thus be subsequently expressed. It could not be determined from these experiments whether this was due to an issue with the splice acceptor sequences chosen for the repair construct, or if it was due to complications caused by the unique splicing mechanisms employed for large introns. Future experiments would focus on, first, troubleshooting the GFP reporter construct issues so that alternative splice site sequences could be rapidly tested. Second, PCR and Sanger sequencing would be used to characterise genomic/repair construct boundaries to check for DNA loss from the repair construct during insertion. Third, RT-qPCR would be employed to determine if the repair construct is being incorporated into the final mRNA transcript. Fourth, target sites in smaller *Mecp2* intron would be identified to test if the failure of splicing was due to problems associated with targeting a large intron.

# Chapter 6

---

## General Discussion

### 6.1 Introduction

RTT is an extremely severe and life-long disorder but studies in mice have given hope that the disease could be reversible in patients, even in adulthood (Guy et al., 2007; Robinson et al., 2012). This has led to a huge effort to try and develop novel therapies, but this has been hampered by a fundamental lack of knowledge about the downstream pathways involved in the gene function, making it extremely difficult to identify drug targets that could help ameliorate the disease. For this reason, recent efforts have focused on targeting the disease at the gene level, mainly using viral based delivery of *Mecp2* gene products (Gadalla et al., 2012; Garg et al., 2013). While this has shown some potential, there remain several outstanding issues, including the dangers of overexpression toxicity, due to a lack of endogenous control mechanisms in the transgene constructs, and the fact that current viral delivery vectors such as AAV do not lead to permanent insertion into the genome. This means that in dividing cells like glia, which have been shown to be relevant for the RTT phenotype (Lioy et al., 2011), the transgene is quickly diluted in the population as it is not copied along with chromosomal DNA during cell division.

Also relevant for the development of RTT therapies is the question of what tissues are responsible for the phenotype. It has been widely assumed in the field that the RTT phenotype can be explained solely by an absence of MeCP2 from cells in the nervous system (Chen et al., 2001; Guy et al., 2001; Giacometti et al., 2007). This was based on the evidence from several mouse models which showed that specific deletion of *Mecp2* from neurons and astrocytes leads to the development of a severe RTT-like phenotype and lethality that appeared indistinguishable from global KO mouse models. On this basis it was argued that RTT was a nervous system specific disorder. However, RTT is a disorder with a wide range of phenotypes in multiple organs in the body. A major weakness of these studies was that the phenotyping carried out was limited to only a few aspects of the disorder such as lifespan, bodyweight, and activity levels and did not assess aspects of the disorder with a plausible peripheral component such as cardiac and respiratory dysfunction, bone and muscle abnormalities, and gait. It is therefore possible that some RTT phenotypes could be treated by

targeting relevant peripheral tissues, which is appealing due to the greater accessibility of these tissues compared to the isolated nervous system.

Based on this there were two major aims of this thesis. First, to comprehensively examine the role of the peripheral tissues in the development of the RTT phenotype, and second, to utilise genome editing tools such as TALEN and CRISPR-Cas9 to develop an alternative, gene-based, RTT therapy.

## **6.2 Major findings and conclusions**

### ***6.2.1 Peripheral contribution to RTT***

#### ***6.2.1.1 Findings***

In order to examine the role of peripheral MeCP2, a novel mouse model was used, in which MeCP2 was selectively expressed in neurons and astrocytes of the nervous system, thus creating a functional peripheral KO mouse. As expected from previous studies of neuron specific KO mice (Chen et al., 2001; Guy et al., 2001), mice expressing MeCP2 exclusively in the nervous system did not show the early lethality seen in global KO mice, and showed only extremely mild RTT-like symptoms when assessed by a well-established scoring system, even after one year of age. This mild score was usually due to the hypoactivity observed in these mice, which was apparent even by casual observation when handling the mice. More detailed phenotyping confirmed these reduced activity levels, with the mice showing moderately reduced spontaneous activity levels in the open field and impaired activity on an accelerating rotarod, which is sensitive to endurance fatigue as well as to balance issues. When the mice were more severely challenged on an elevated accelerating treadmill a markedly reduced exercise capacity was observed, with an almost 50% reduction in performance compared to WT. The reason for this reduced activity and exercise capacity is unclear and these mice did not differ from WT in the other major RTT-like phenotypes including breathing, gait, balance, and cardiac function. To investigate a possible role for muscle dysfunction in the locomotor phenotype, structural analysis of skeletal muscle was carried out and revealed some evidence of muscle fibrosis, however this was variable and only seen in about half the mice, meaning that the results did not quite reach overall significance. The most striking result of this study was that the bone abnormalities previously described for global KO models (Kamal et al., 2015) were also seen in the peripheral KO

mice, and in fact no significant difference was seen between peripheral and global KO mice, suggesting this is an exclusively peripheral phenotype.

### *6.2.1.2 Significance*

The work presented in this thesis is the first time that a comprehensive phenotyping has been carried out to examine the role of MeCP2 outside of the nervous system, and answers several important questions regarding the role of MeCP2 throughout the body and the appropriate target cells for therapies. First, it confirms that most of the major aspects of the disorder are due to the absence of MeCP2 from the nervous system. While, several previous studies have suggested this (Chen et al., 2001; Guy et al., 2001; Giacometti et al., 2007), these assumptions were based on very superficial phenotyping of neuron-specific KO animals. This study is the first to examine the role of the nervous system in breathing, cardiac function, bone abnormalities, gait, exercise capacity, balance and blood biochemistry. In addition, it provides the first comprehensive pathological screening of multiple tissue systems, carried out by a qualified veterinary pathologist, which conclusively demonstrates that a loss of functional MeCP2 does not lead to any major structural or pathological abnormalities in any of the tissues tested. The results strongly suggest that therapies targeting the nervous system will lead to the reversal of most of the major RTT phenotypes and should therefore be the main focus of any future therapeutic strategies. A surprising finding of this study was the marked decrease in exercise capacity seen in the peripheral KO mice. Since RTT patients are usually confined to wheelchairs, this is not an aspect of the disorder that would be considered clinically relevant and has therefore not been investigated before. However, this fatigue could become more relevant when seen in the context of disease reversal. It suggests that even if a treatment led to a full recovery of MeCP2 protein expression in the brain, patients would still display a strongly reduced ability to carry out exercise. Whether this would actually manifest in human patients is not clear from this study, and its importance is therefore difficult to assess. Since the animals in this study were assessed at 15 weeks, the equivalent of a young adult in humans, it is possible that this phenotype could progress over time to become more significantly debilitating. Another novel and significant finding from this study was that the bone abnormalities previously described in RTT mice (Kamal et al., 2015) are likely to be caused exclusively by an absence of MeCP2 from peripheral tissues. RTT patients suffer from a number of skeletal abnormalities, including early osteoporosis, spinal deformities that have to be corrected by surgery, and a preponderance of low-impact fractures (Keret et al., 1988; Guidera et al., 1991; Zysman et al., 2006; Downs et al., 2008; Percy et al., 2010), and

the results of this study suggest that therapies will need to be targeted to peripheral tissues to impact on this phenotype. This is important as peripheral organs are more accessible to therapies than the highly protected environment of the nervous system.

### *6.2.1.3 Caveats and technical considerations*

The cre-based strategy employed to generate a functional peripheral KO mouse achieved very high levels of reactivation in the nervous system and very low levels in the peripheral tissues. However, it cannot be discounted that a lack of MeCP2 in 10% of neurons and astrocytes could have a mild impact on some of the phenotypes seen in the peripheral KO mouse. An argument against this is that presumably brain-specific measures such as balance and nest building behaviour are indistinguishable from WT in these mice. Additionally, previous global reactivation mice studies have achieved only 60-70% reactivation in the brain and yet show an almost complete reversal of phenotypes including motor defects (Robinson et al., 2012). Due to the severity of the stop/y mice phenotype at the 15 week time-point chosen for behavioural studies, it was not possible to assess gait in these mice using the motorised treadmill system. This meant that no global KO data was available to use as a comparison for WT and peripheral KO data. This was unfortunate; however, a previous study has shown that even at 10 weeks these animals show significantly impaired gait across a number of parameters compared to WT (Gadalla et al., 2014). The fact that peripheral KO mice, even at a later time-point, show no evidence of any gait impairment, strongly suggests that gait dysfunction is primarily related to nervous system dysfunction.

### *6.2.1.4 Future experiments*

Following on from the work described in this study, a number of follow-up experiments should be carried out to further build on these results. First, to further assess the validity of the functional peripheral KO model it would be useful to carry out long-term potentiation experiments. As previously described, the evidence suggest that a reactivation level of 90% in the brain is sufficient to restore normal brain function, but this could be more convincingly demonstrated by electrophysiological assessment. Second, while the results clearly demonstrate that the bone abnormalities seen in RTT mice models are due to a peripheral absence of MeCP2, it is not possible to determine if the primary locus of dysfunction is in the bone tissue itself. To investigate this, a bone specific KO mouse should be generated and assessed using the same functional tests used in this study. Third, while there was some evidence of muscle pathology in peripheral KO animals, this phenotype was variable and

consequently the result did not reach statistical significance when these mice were compared with WT. However, it is plausible that this is a progressive phenotype and that peripheral KO mice would show a more severe pathology if assessed at a later time-point. Therefore, this experiment should be repeated over a longer period of time, perhaps 6 months and then one year, to look for evidence of this.

## ***6.2.2 Genome editing as a treatment for RTT***

### ***6.2.2.1 Findings***

The previously described challenges involved in developing a traditional gene therapy treatment for RTT demonstrates the need for an alternative approach. In this thesis, a novel strategy using newly developed genome editing tools was tested in a proof-of-principle study. Genome editing relies on DNA repair pathways to introduce genetic changes, with the HR pathway, which is only active in dividing cells, enabling precise mutation repair, while the NHEJ pathway, which is highly active in all cells, can be used to ligate repair constructs into precise target locations. Since other work in this thesis clearly demonstrated that the nervous system should be the primary target of any RTT therapies, it was decided to develop a strategy suitable for post-mitotic neurons. This involved ligating a promoterless repair construct into intron 2 of the *Mecp2* gene and providing suitable sites to enable splicing of the construct to endogenous upstream exons. Initial work focused on generating TALEN and CRISPR constructs capable of targeting intron 2 and stimulating ligation of the repair construct. Several TALEN pairs and CRISPR-Cas9 constructs were designed, synthesised, and then assayed for cutting efficiency. The cutting efficiency of individual TALEN pairs was highly variable, with two of the four pairs showing very little activity and one showing only moderate activity. One highly efficient TALEN pair was found that induced double stranded breaks in over 40% of cells using the WT version of the FOK1 nuclease, however this cutting activity was significantly reduced when using the obligate heterodimer version of the enzyme. In contrast, all four of the CRISPR-Cas9 constructs used showed significant cutting efficiency, with the best performing guide inducing DNA breaks in over 60% of cells, and the worst performer still cutting in over 20% of cells. The best performing TALEN and CRISPR-Cas9 constructs were then used to try and stimulate the ligation of a repair construct into intron 2 of the *Mecp2* gene, in a mouse cell line. This repair construct contained the WT coding sequence of exons 3 and 4 of the gene, the site of the majority of RTT-causing mutations, as well as a fluorescent reporter, an acceptor splice site, and an appropriate

transcriptional terminator. The hypothesis was that, once ligated into the genome, this repair construct would splice to the upstream endogenous exons 1 and 2 and thus replace the downstream mutated exons in the final mRNA transcript. When cells were transfected with the best performing TALEN pair and a repair construct with suitable target sites, robust insertion of the DNA insert was seen in the correct genomic location, as determined by a PCR based assay. Despite the decreased cutting efficiency in the obligate heterodimer version of the FokI nuclease, this version led to a greater proportion of correctly orientated repair construct insertion, probably because the repair construct is vulnerable to being re-cut from the genome when using the WT version of FokI. However, when cells were then examined for protein expression of the repair construct, no expression was detected. This suggested that the inserted repair construct was not splicing correctly to the upstream endogenous exons and thus wasn't being incorporated into the final mRNA transcript. To determine if this was the case a GFP reported construct was engineered, containing either the last 50 bp of *Mecp2* intron 2, assumed to contain a splice acceptor site, or the *HBB* Ex3 Splice acceptor site, which has previously been shown to facilitate splicing of an exogenous construct ligated into the genome (Hockemeyer et al., 2009). Unfortunately, this reporter plasmid showed extremely high levels of background GFP protein expression, despite containing no mammalian promoter, which prevented its use for further experiments.

#### 6.2.2.2 Significance

The work presented in this thesis is the first attempt to develop a genome editing based strategy for the treatment of a neurological disorder like RTT. Of particular significance is the development of a novel strategy for editing that overcomes the challenges involved in targeting post-mitotic cells like neurons, especially the low levels of HR repair seen in these cell types. The work described here shows at a proof-of-principle level, that genome editing tools such as TALEN and CRISPR can be used to insert a repair construct precisely into a non-coding region of the *Mecp2* gene and could thus provide an alternative therapeutic approach. Crucially, this strategy overcomes several challenges to gene-based treatments for RTT. First, ligation of the repair construct relies on the NHEJ pathway rather than the HR pathway. This is essential for any editing approach whose primary target is post-mitotic neurons. Second, the strategy is applicable to almost any RTT causing mutation. HR based editing approaches rely on targeting and repairing specific disease-causing mutations. However, by replacing both exon 3 and exon 4 of the *Mecp2* gene with WT copies, this approach could be a broad-based treatment for RTT. Third, standard gene therapy approaches

are hampered by the danger of overexpression toxicity. This is a particular problem for RTT patients as *Mecp2* duplication has been shown to lead to significant pathology (Van Esch et al., 2005). However, by ligating a repair construct into the *Mecp2* locus and splicing to upstream exons, gene transcription remains under the control of the endogenous promoter, thus recapitulating normal expression patterns. Last, since roughly half of the cells in RTT patients express the WT allele, and it is currently not possible to specifically target mutant cells only, it is essential that any strategy does not damage the normal function of these cells. By targeting only non-coding regions of the gene, this strategy ensures that failed insertions, or mutagenic changes introduced by double strand break repair, does not lead to disruption of WT protein function. Overall, these features make this approach extremely attractive for the treatment of both RTT and other genetic disorders of the nervous system.

### 6.2.2.3 Caveats and technical considerations

There were several technical challenges that impacted on the development of this strategy. First, there was significant difficulty in the cloning of the *Mecp2* repair construct. The repair construct was synthesised commercially and, unexpectedly, plasmids containing the coding sequence of exon 3 and 4 caused toxicity when transformed into bacterial cells, to the extent that it was not possible to culture bacteria adequately to obtain sufficient quantities of plasmid DNA for cell transfections. Various attempts were made to mitigate this, including culturing at lower temperatures to lower the metabolic burden, but these were not successful. Sequence analysis of the plasmid revealed the presence of an unexpected bacterial promoter. Since bacterial transcription of the construct was not required, the repair construct, flanked by either TALEN or CRISPR target sites, was transferred to a promoterless plasmid. This prevented toxicity in the repair construct flanked by the TALEN target sites, but not for the CRISPR target site flanked version. Due to time constraints it was decided that the next experiments would therefore be carried out using the TALEN repair construct only.

A second cloning issue was encountered when attempting to engineer the GFP reporter construct to test the function of the splice acceptor sites. The reporter was designed such that GFP protein translation would only occur if the reporter was successfully inserted into intron 2 of *Mecp2* and correctly spliced to upstream exons. For this reason the plasmid did not contain any mammalian promoter, to prevent plasmid transcription of the GFP. Despite this, extremely large background levels of GFP protein was seen in all control transfections, including those in which the reporter construct was transfected without the presence of the



TALEN arms. Since this prevented the repair construct from being inserted into the target region, it is not clear what caused protein expression in these cases. It was not possible to further investigate this as it occurred at the end of the time available for the thesis.

Another major issue that occurred was the failure of the repair construct to lead to protein expression. It was suspected that this was due to the inability of the construct to correctly splice to upstream exons. Since the last 50 bp of an intron usually contains the necessary splicing elements (Reed and Maniatis, 1988; Gao et al., 2008), it was decided to use the last 50 bp of *Mecp2* intron 2 as the repair construct splice acceptor. When this failed it was decided to use a simple GFP reporter to test the effectiveness of alternative splice sites, however this was hampered by the issues encountered with the reporter, as previously described. Further reading into the mechanisms involved in the splicing of large introns indicated that it was more complex than initially thought and that alternative strategies may be required.

#### *6.2.2.4 Future experiments*

The work described in this thesis is only the initial step in the development of this editing strategy for RTT and number of future experiments will be required to make further progress. The most obvious next steps would be to characterise and quantify the repair construct insertion, and then to determine the cause of the lack of protein expression. First, PCR and Sanger sequencing should be used to characterise genomic/repair construct boundaries. It is possible that DNA loss at the junction sites could occur during the ligation process, and this could lead to an impairment of splicing or transcription termination. Second, qRT-PCR should be employed to determine if the repair construct is being incorporated into the final mRNA transcript. This would enable the frequency of the incorporation of the repair construct into the mRNA to be quantified and reveal if the repair construct is being successfully spliced to endogenous exons. Third, the GFP reporter construct for testing splice acceptor sites could be sub-cloned into a different plasmid to try and prevent the high background protein levels seen in this plasmid, including the removal of the methionine start codon from the coding sequence. Successful cloning of this plasmid would allow the effect of different splice sites to be quickly and easily assessed. Last, since the splicing of large introns is complex and not well understood, target sites in the much smaller intron 3 of *Mecp2* could be identified and the appropriate TALEN arms generated, in order to determine if successful splicing occurs in this simpler intron.

## 6.3 Summary

The aim of this thesis was to comprehensively determine the crucial regions responsible for the RTT phenotype, and then to use this information to design therapies suitable for these regions. Phenotyping of a novel mouse model, in which *Mecp2* was selectively reactivated in neurons and astrocytes, revealed that the majority of the RTT phenotype is due to a loss of functional MeCP2 from the nervous system, but that some disease aspects such as bone abnormalities and reduced exercise capacity have a large peripheral component. Using this information a novel genome editing based strategy was developed, specifically designed to utilise repair mechanisms known to be active in post-mitotic neurons. A series of genome editing tools capable of targeting non-coding regions of the *Mecp2* gene were then engineered and enabled the ligation of a DNA repair construct into the genome. This suggests that genome editing could be a viable strategy for tackling RTT. Future studies will further develop this strategy, including the reengineering of the construct to enable splicing to endogenous *Mecp2* exons.

# References

---

- Abdala APL, Dutschmann M, Bissonnette JM, Paton JFR (2010) Correction of respiratory disorders in a mouse model of Rett syndrome. *Proc Natl Acad Sci* Available at: <http://www.pnas.org/content/early/2010/09/28/1012104107> [Accessed March 11, 2012].
- Abolhassani N, Iyama T, Tsuchimoto D, Sakumi K, Ohno M, Behmanesh M, Nakabeppu Y (2010) NUDT16 and ITPA play a dual protective role in maintaining chromosome stability and cell growth by eliminating dIDP/IDP and dITP/ITP from nucleotide pools in mammals. *Nucleic Acids Res* 38:2891–2903.
- Alvarez-Erviti L, Seow Y, Yin H, Betts C, Lakhal S, Wood MJA (2011) Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes. *Nat Biotechnol* 29:341–345.
- Alvarez-Saavedra M, Sáez MA, Kang D, Zoghbi HY, Young JI (2007) Cell-specific expression of wild-type MeCP2 in mouse models of Rett syndrome yields insight about pathogenesis. *Hum Mol Genet* 16:2315–2325.
- Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nat Genet* 23:185–188.
- Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154:1370–1379.
- Asaka Y, Jugloff DGM, Zhang L, Eubanks JH, Fitzsimonds RM (2006) Hippocampal synaptic plasticity is impaired in the *Mecp2*-null mouse model of Rett syndrome. *Neurobiol Dis* 21:217–227.
- Baikie G, Ravikumara M, Downs J, Naseem N, Wong K, Percy A, Lane J, Weiss B, Ellaway C, Bathgate K, Leonard H (2014) Gastrointestinal dysmotility in Rett syndrome. *J Pediatr Gastroenterol Nutr* 58:237–244.
- Baker SA, Chen L, Wilkins AD, Yu P, Lichtarge O, Zoghbi HY (2013) An AT-Hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell* 152:984–996.
- Ballas N, Liroy DT, Grunseich C, Mandel G (2009) Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. *Nat Neurosci* 12:311–317.
- Bankiewicz KS, Forsayeth J, Eberling JL, Sanchez-Pernaute R, Pivrotto P, Bringas J, Herscovitch P, Carson RE, Eckelman W, Reutter B, Cunningham J (2006) Long-term clinical improvement in MPTP-lesioned primates after gene therapy with AAV-hAADC. *Mol Ther* 14:564–570.

- Barney CC, Feyma T, Beisang A, Symons FJ (2015) Pain experience and expression in Rett syndrome: Subjective and objective measurement approaches. *J Dev Phys Disabil* 27:417–429.
- Barzel A, Paulk NK, Shi Y, Huang Y, Chu K, Zhang F, Valdmanis PN, Spector LP, Porteus MH, Gaensler KM, Kay MA (2015) Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* 517:360–364.
- Baubec T, Ivánek R, Lienert F, Schübeler D (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* 153:480–492.
- Ben-Shachar S, Chahrour M, Thaller C, Shaw CA, Zoghbi HY (2009) Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum Mol Genet* 18:2431–2442.
- Bertelli E, Regoli M, Fonzi L, Occhini R, Mannucci S, Ermini L, Toti P (2007) Nestin expression in adult and developing human kidney. *J Histochem Cytochem* 55:411–421.
- Bhakta MS, Henry IM, Ousterout DG, Das KT, Lockwood SH, Meckler JF, Wallen MC, Zykovich A, Yu Y, Leo H, Xu L, Gersbach CA, Segal DJ (2013) Highly active zinc-finger nucleases by extended modular assembly. *Genome Res* 23:530–538.
- Bhatnagar S, Zhu X, Ou J, Lin L, Chamberlain L, Zhu LJ, Wajapeyee N, Green MR (2014) Genetic and pharmacological reactivation of the mammalian inactive X chromosome. *Proc Natl Acad Sci* 111:12591–12598.
- Bianciardi L, Fichera M, Failla P, Di Marco C, Grozeva D, Mencarelli MA, Spiga O, Mari F, Meloni I, Raymond L, Renieri A, Romano C, Ariani F (2015) *MECP2* missense mutations outside the canonical MBD and TRD domains in males with intellectual disability. *J Hum Genet* Available at: <http://www.nature.com/jhg/journal/vaop/ncurrent/full/jhg2015118a.html> [Accessed October 26, 2015].
- Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim Y-G, Chandrasegaran S (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* 21:289–297.
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16:6–21.
- Bissonnette JM, Knopp SJ (2006) Separate respiratory phenotypes in methyl-CpG-binding protein 2 (*Mecp2*) Deficient Mice. *Pediatr Res* 59:513–518.
- Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* 95:10570–10575.
- Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291–336.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326:1509–1512.

- Bogdanove AJ, Schornack S, Lahaye T (2010) TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol* 13:394–401.
- Bogdanove AJ, Voytas DF (2011) TAL effectors: Customizable proteins for DNA targeting. *Science* 333:1843–1846.
- Brinkman EK, Chen T, Amendola M, van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*.
- Buchovecky CM, Turley SD, Brown HM, Kyle SM, McDonald JG, Liu B, Pieper AA, Huang W, Katz DM, Russell DW, Shendure J, Justice MJ (2013) A suppressor screen in mouse *Mecp2* implicates cholesterol metabolism in Rett syndrome. *Nat Genet* 45:1013–1020.
- Carroll D (2012) A CRISPR approach to gene targeting. *Mol Ther* 20:1658–1660.
- Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39:e82–e82.
- Chahrour M, Jung SY, Shaw C, Zhou X, Wong STC, Qin J, Zoghbi HY (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320:1224–1229.
- Chang Y-F, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76:51–74.
- Chao H-T, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J, Neul JL, Gong S, Lu H-C, Heintz N, Ekker M, Rubenstein JLR, Noebels JL, Rosenmund C, Zoghbi HY (2010) Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* 468:263–269.
- Chen L, Chen K, Lavery LA, Baker SA, Shaw CA, Li W, Zoghbi HY (2015) MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc Natl Acad Sci* 112:5509–5514.
- Chen RZ, Akbarian S, Tudor M, Jaenisch R (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat Genet* 27:327–331.
- Choi J-H, Yu N-K, Baek G-C, Bakes J, Seo D, Nam HJ, Baek SH, Lim C-S, Lee Y-S, Kaang B-K (2014) Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. *Mol Brain* 7:17.
- Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186:757–761.
- Colvin L, Fyfe S, Leonard S, Schiavello T, Ellaway C, Klerk N de, Christodoulou J, Msall M, Leonard H (2003) Describing the phenotype in Rett syndrome using a population database. *Arch Dis Child* 88:38–43.

- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823.
- Conti V, Gandaglia A, Galli F, Tirone M, Bellini E, Campana L, Kilstrup-Nielsen C, Rovere-Querini P, Brunelli S, Landsberger N (2015) MeCP2 affects skeletal muscle growth and morphology through non cell-autonomous mechanisms. *PloS One* 10:e0130183.
- Corchero JL, Villaverde A (1998) Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol Bioeng* 58:625–632.
- Corti M, Elder M, Falk D, Lawson L, Smith B, Nayak S, Conlon T, Clément N, Erger K, Lavassani E, Green M, Doerfler P, Herzog R, Byrne B (2014) B-cell depletion is protective against anti-AAV capsid immune response: a human subject case study. *Mol Ther — Methods Clin Dev* 1:14033.
- Cradick TJ, Fine EJ, Antico CJ, Bao G (2013) CRISPR/Cas9 systems targeting  $\beta$ -globin and *CCR5* genes have substantial off-target activity. *Nucleic Acids Res* 41:9584–9592.
- Daniels RW, Rossano AJ, Macleod GT, Ganetzky B (2014) Expression of multiple transgenes from a single construct using viral 2A peptides in *Drosophila*. *PLoS ONE* 9:e100637.
- Daniel T, Lioy, Saurabh K. Garg, Caitlin E. Monaghan, Jacob Raber, Kevin D. Foust, Brian K. Kaspar, Petra G. Hirrlinger, Frank Kirchhoff, John M. Bissonnette, Nurit Ballas, Gail Mandel (2011) A role for glia in the progression of Rett syndrome. *Nature* 475:497–500.
- Dani VS, Chang Q, Maffei A, Turrigiano GG, Jaenisch R, Nelson SB (2005) Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett Syndrome. *Proc Natl Acad Sci USA* 102:12560–12565.
- Deacon RMJ (2006) Assessing nest building in mice. *Nat Protoc* 1:1117–1119.
- De Felice C, Guazzi G, Rossi M, Ciccoli L, Signorini C, Leoncini S, Tonni G, Latini G, Valacchi G, Hayek J (2010) Unrecognized lung disease in classic Rett syndrome: a physiologic and high-resolution CT imaging study. *Chest* 138:386–392.
- De Felice C, Rossi M, Leoncini S, Chisci G, Signorini C, Lonetti G, Vannuccini L, Spina D, Ginori A, Iacona I, Cortelazzo A, Pecorelli A, Valacchi G, Ciccoli L, Pizzorusso T, Hayek J (2014) Inflammatory lung disease in Rett syndrome. *Mediators Inflamm* 2014:560120.
- Delacour A, Nepote V, Trumpp A, Herrera PL (2004) Nestin expression in pancreatic exocrine cell lineages. *Mech Dev* 121:3–14.
- Delcuve GP, Khan DH, Davie JR (2012) Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors. *Clin Epigenetics* 4:5.

- Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* 471:602–607.
- Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu J-K, Shi Y, Yan N (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335:720–723.
- Derecki NC, Cronk JC, Lu Z, Xu E, Abbott SBG, Guyenet PG, Kipnis J (2012) Wild-type microglia arrest pathology in a mouse model of Rett syndrome. *Nature* 484:105–109.
- Dow LE, Fisher J, O'Rourke KP, Muley A, Kastenhuber ER, Livshits G, Tschaharganeh DF, Socci ND, Lowe SW (2015) Inducible *in vivo* genome editing with CRISPR-Cas9. *Nat Biotechnol* 33:390–394.
- Downs J, Bebbington A, Woodhead H, Jacoby P, Jian L, Jefferson A, Leonard H (2008) Early determinants of fractures in Rett syndrome. *Pediatrics* 121:540–546.
- Dragatsis I, Zeitlin S (2001) A method for the generation of conditional gene repair mutations in mice. *Nucleic Acids Res* 29:e10.
- Dragich JM, Kim Y-H, Arnold AP, Schanen NC (2007) Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. *J Comp Neurol* 501:526–542.
- Duque S, Joussemet B, Riviere C, Marais T, Dubreil L, Douar A-M, Fyfe J, Moullier P, Colle M-A, Barkats M (2009) *Intravenous* administration of self-complementary AAV9 enables transgene delivery to adult motor neurons. *Mol Ther J Am Soc Gene Ther* 17:1187–1196.
- Elian M, Rudolf N de M (1991) EEG and respiration in Rett syndrome. *Acta Neurol Scand* 83:123–128.
- El-Osta A, Kantharidis P, Zalcberg JR, Wolffe AP (2002) Precipitous release of methyl-CpG binding protein 2 and histone deacetylase 1 from the methylated human multidrug resistance gene (MDR1) on activation. *Mol Cell Biol* 22:1844–1857.
- Engerström IW (1992) Rett syndrome: The late infantile regression period--a retrospective analysis of 91 cases. *Acta Paediatr Oslo Nor* 1992 81:167–172.
- Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS ONE* 4:e5553.
- Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* 3:e3647.
- Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. *Genes Dev* 17:419–437.
- Fine EJ, Appleton CM, White DE, Brown MT, Deshmukh H, Kemp ML, Bao G (2015) *Trans*-spliced Cas9 allows cleavage of *HBB* and *CCR5* genes in human cells using compact expression cassettes. *Sci Rep* 5:10777.

- Forlani G, Giarda E, Ala U, Di Cunto F, Salani M, Tupler R, Kilstrup-Nielsen C, Landsberger N (2010) The MeCP2/YY1 interaction regulates *ANT1* expression at 4q35: novel hints for Rett syndrome pathogenesis. *Hum Mol Genet* 19:3114–3123.
- Foust KD, Nurre E, Montgomery CL, Hernandez A, Chan CM, Kaspar BK (2008) *Intravascular* AAV9 preferentially targets neonatal neurons and adult astrocytes. *Nat Biotechnol* 27:59–65.
- Frock RL, Hu J, Meyers RM, Ho Y-J, Kii E, Alt FW (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* 33:179–186.
- Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31:822–826.
- Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32:279–284.
- Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*.
- Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppl C, Nowrouzi A, Bartholomae CC, Wang J, Friedman G, Holmes MC, Gregory PD, Glimm H, Schmidt M, Naldini L, von Kalle C (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* 29:816–823.
- Gadalla KK, Bailey ME, Spike RC, Ross PD, Woodard KT, Kalburgi SN, Bachaboina L, Deng JV, West AE, Samulski RJ, Gray SJ, Cobb SR (2012) Improved survival and reduced phenotypic severity following AAV9/*MECP2* gene transfer to neonatal and juvenile male *mecp2* knockout mice. *Mol Ther* 21:18–30.
- Gadalla KKE, Bailey MES, Cobb SR (2011) MeCP2 and Rett syndrome: reversibility and potential avenues for therapy. *Biochem J* 439:1–14.
- Gadalla KK, Ross PD, Riddell JS, Bailey ME, Cobb SR (2014) Gait analysis in a *Mecp2* knockout mouse model of Rett syndrome reveals early-onset and progressive motor deficits. *PLoS One* 9:e112889.
- Gao K, Masuda A, Matsuura T, Ohno K (2008) Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* 36:2257–2267.
- Garg SK, Liou DT, Cheval H, McGann JC, Bissonnette JM, Murtha MJ, Foust KD, Kaspar BK, Bird A, Mandel G (2013) Systemic delivery of MeCP2 rescues behavioural and cellular deficits in female mouse models of Rett syndrome. *J Neurosci Off J Soc Neurosci* 33:13612–13620.
- Genovese P, Schirotti G, Escobar G, Di Tomaso T, Firrito C, Calabria A, Moi D, Mazzieri R, Bonini C, Holmes MC, Gregory PD, van der Burg M, Gentner B, Montini E, Lombardo A, Naldini L (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* 510:235–240.



- Georgel PT, Horowitz-Scherer RA, Adkins N, Woodcock CL, Wade PA, Hansen JC (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J Biol Chem* 278:32181–32188.
- Ghosh RP, Horowitz-Scherer RA, Nikitina T, Shlyakhtenko LS, Woodcock CL (2010a) MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. *Mol Cell Biol* 30:4656–4670.
- Ghosh RP, Nikitina T, Horowitz-Scherer RA, Gierasch LM, Uversky VN, Hite K, Hansen JC, Woodcock CL (2010b) Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry (Mosc)* 49:4395–4410.
- Giacometti E, Luikenhuis S, Beard C, Jaenisch R (2007) Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. *Proc Natl Acad Sci* 104:1931–1936.
- Girard M, Couvert P, Carrié A, Tardieu M, Chelly J, Beldjord C, Bienvenu T (2001) Parental origin of de novo *MECP2* mutations in Rett syndrome. *Eur J Hum Genet EJHG* 9:231–236.
- Glaze DG, Frost JD, Zoghbi HY, Percy AK (1987) Rett's syndrome: characterization of respiratory patterns and sleep. *Ann Neurol* 21:377–382.
- Goffin D, Allen M, Zhang L, Amorim M, Wang I-TJ, Reyes A-RS, Mercado-Berton A, Ong C, Cohen S, Hu L, Blendy JA, Carlson GC, Siegel SJ, Greenberg ME, Zhou Z (2012) Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nat Neurosci* 15:274–283.
- Gold WA, Williamson SL, Kaur S, Hargreaves IP, Land JM, Pelka GJ, Tam PPL, Christodoulou J (2014) Mitochondrial dysfunction in the skeletal muscle of a mouse model of Rett syndrome (RTT): implications for the disease phenotype. *Mitochondrion* 15:10–17.
- Green MR (1991) Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu Rev Cell Biol* 7:559–599.
- Guidera KJ, Borrelli J, Raney E, Thompson-Rangel T, Ogden JA (1991) Orthopaedic manifestations of Rett syndrome. *J Pediatr Orthop* 11:204–208.
- Guilinger JP, Pattanayak V, Reyon D, Tsai SQ, Sander JD, Joung JK, Liu DR (2014) Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods* 11:429–435.
- Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming G, Song H (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* 17:215–222.
- Guy J, Gan J, Selfridge J, Cobb S, Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315:1143–1147.
- Guy J, Hendrich B, Holmes M, Martin JE, Bird A (2001) A mouse *Mecp2*-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet* 27:322–326.

- Hagberg B (2002) Clinical manifestations and stages of Rett syndrome. *Ment Retard Dev Disabil Res Rev* 8:61–65.
- Hagberg B, Aicardi J, Dias K, Ramos O (1983) A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann Neurol* 14:471–479.
- Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18:6538–6547.
- Hess SE, Rohr S, Dufour BD, Gaskill BN, Pajor EA, Garner JP (2008) Home improvement: C57BL/6J mice given more naturalistic nesting materials build better nests. *J Am Assoc Lab Anim Sci JAALAS* 47:25–31.
- High K (2002) AAV-mediated gene transfer for hemophilia. *Genet Med* 4:56S – 61S.
- Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKolver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 27:851–857.
- Hoffbuhr KC, Moses LM, Jerdonek MA, Naidu S, Hoffman EP (2002) Associations between *MECP2* mutations, X chromosome inactivation, and phenotype. *Ment Retard Dev Disabil Res Rev* 8:99–105.
- Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD (2008) MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol Cell* 29:525–531.
- Holkers M, Maggio I, Liu J, Janssen JM, Miselli F, Mussolino C, Recchia A, Cathomen T, Gonçalves MAFV (2013) Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res* 41:e63–e63.
- Holt N, Wang J, Kim K, Friedman G, Wang X, Taupin V, Crooks GM, Kohn DB, Gregory PD, Holmes MC, Cannon PM (2010) Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to *CCR5* control HIV-1 *in vivo*. *Nat Biotechnol* 28:839–847.
- Hsu M-F, Meng T-C (2010) Enhancement of insulin responsiveness by nitric oxide-mediated inactivation of protein-tyrosine phosphatases. *J Biol Chem* 285:7919–7928.
- Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157:1262–1278.
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31:827–832.
- Huang H-S, Allen JA, Mabb AM, King IF, Miriyala J, Taylor-Blake B, Sciaky N, Dutton JW, Lee H-M, Chen X, Jin J, Bridges AS, Zylka MJ, Roth BL, Philpot BD (2012) Topoisomerase inhibitors unsilence the dormant allele of *Ube3a* in neurons. *Nature* 481:185–189.

- Hütter G, Nowak D, Mossner M, Ganepola S, Müssig A, Allers K, Schneider T, Hofmann J, Kücherer C, Blau O, Blau IW, Hofmann WK, Thiel E (2009) Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med* 360:692–698.
- Hwang CK, Song KY, Kim CS, Choi HS, Guo X-H, Law P-Y, Wei L-N, Loh HH (2007) Evidence of endogenous mu opioid receptor regulation by epigenetic control of the promoters. *Mol Cell Biol* 27:4720–4736.
- Isaacs JS, Murdock M, Lane J, Percy AK (2003) Eating difficulties in girls with Rett syndrome compared with other developmental disabilities. *J Am Diet Assoc* 103:224–230.
- Jao L-E, Wente SR, Chen W (2013) Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A* 110:13904–13909.
- Jeppesen DK, Bohr VA, Stevnsner T (2011) DNA repair deficiency in neurodegeneration. In: *Prog Neurobiol*, pp 166–200. England: 2011 Elsevier Ltd.
- Jones PL, Jan Veenstra GC, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19:187–191.
- Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 14:49–55.
- Julu PO, Kerr AM, Apartopoulos F, Al-Rawas S, Engerström IW, Engerström L, Jamal GA, Hansen S (2001) Characterisation of breathing and associated central autonomic dysfunction in the Rett disorder. *Arch Dis Child* 85:29–37.
- Kamal B, Russell D, Payne A, Constante D, Tanner KE, Isaksson H, Mathavan N, Cobb SR (2015) Biomechanical properties of bone in a mouse model of Rett syndrome. *Bone* 71:106–114.
- Karra D, Dahm R (2010) Transfection techniques for neuronal cells. *J Neurosci* 30:6171–6177.
- Kathrin Plath, Susanna Mlynarczyk-Evans, Dmitri A. Nusinow, Panning B (2002) Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet* 36:233–278.
- Kaufmann WE, Taylor CV, Hohmann CF, Sanwal IB, Naidu S (1997) Abnormalities in neuronal maturation in Rett syndrome neocortex: preliminary molecular correlates. *Eur Child Adolesc Psychiatry* 6 Suppl 1:75–77.
- Kelly Beumer, Gargi Bhattacharyya, Marina Bibikova, Jonathan K. Trautman, Dana Carroll (2006) Efficient gene targeting in *Drosophila* with zinc-finger nucleases.
- Kelly S, Georgomanolis T, Zirkel A, Diermeier S, O'Reilly D, Murphy S, Längst G, Cook PR, Papantonis A (2015) Splicing of many human genes involves sites embedded within introns. *Nucleic Acids Res*:gkv386.

- Kemi OJ, Haram PM, Wisløff U, Ellingsen Ø (2004) aerobic fitness is associated with cardiomyocyte contractile capacity and endothelial function in exercise training and detraining. *Circulation* 109:2897–2904.
- Keret D, Bassett GS, Bunnell WP, Marks HG (1988) Scoliosis in Rett syndrome. *J Pediatr Orthop* 8:138–142.
- Kerr AM, Webb P, Prescott RJ, Milne Y (2003) Results of surgery for scoliosis in Rett syndrome. *J Child Neurol* 18:703–708.
- Kim HJ, Lee HJ, Kim H, Cho SW, Kim J-S (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* 19:1279–1288.
- Kim J-S, Lee HJ, Carroll D (2010) Genome editing with modularly assembled zinc finger nucleases. *Nat Methods* 7:91; author reply 91–92.
- Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci* 93:1156–1160.
- Kishino T, Lalande M, Wagstaff J (1997) *UBE3A/E6-AP* mutations cause Angelman syndrome. *Nat Genet* 15:70–73.
- Klose RJ, Bird AP (2004) MeCP2 behaves as an elongated monomer that does not stably associate with the sin3a chromatin remodeling complex. *J Biol Chem* 279:46490–46496.
- Kokura K, Kaul SC, Wadhwa R, Nomura T, Khan MM, Shinagawa T, Yasukawa T, Colmenares C, Ishii S (2001) The Ski protein family is required for MeCP2-mediated transcriptional repression. *J Biol Chem* 276:34115–34121.
- Kormann MSD, Hasenpusch G, Aneja MK, Nica G, Flemmer AW, Herber-Jonat S, Huppmann M, Mays LE, Illenyi M, Schams A, Griesse M, Bittmann I, Handgretinger R, Hartl D, Rosenecker J, Rudolph C (2011) Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. *Nat Biotechnol* 29:154–157.
- Kosai K, Kusaga A, Isagai T, Hirata K, Nagano S, Murofushi Y, Takahashi T, Takashima S, Matsuishi T (2005) 58. Rett syndrome is reversible and treatable by MeCP2 gene therapy into the striatum in mice. *Mol Ther* 11:S24–S24.
- Krejci L, Altmannova V, Spirek M, Zhao X (2012) Homologous recombination and its regulation. *Nucleic Acids Res* 40:5795–5818.
- Kriaucionis S, Bird A (2004) The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic Acids Res* 32:1818–1823.
- Kriaucionis S, Paterson A, Curtis J, Guy J, Macleod N, Bird A (2006) Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. *Mol Cell Biol* 26:5033–5042.

- Kudo S (1998) Methyl-CpG-binding protein MeCP2 represses Sp1-activated transcription of the human leukosialin gene when the promoter is methylated. *Mol Cell Biol* 18:5492–5499.
- Laget S, Joulie M, Le Masson F, Sasai N, Christians E, Pradhan S, Roberts RJ, Defossez P-A (2010) The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS ONE* 5:e11982.
- Leonard H, Thomson MR, Glasson EJ, Fyfe S, Leonard S, Bower C, Christodoulou J, Ellaway C (1999) A population-based approach to the investigation of osteopenia in Rett syndrome. *Dev Med Child Neurol* 41:323–328.
- Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69:905–914.
- Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end joining pathway. *Annu Rev Biochem* 79:181–211.
- Li H et al. (2011a) *In vivo* genome editing restores haemostasis in a mouse model of haemophilia. *Nature* 475:217–221.
- Lioy DT, Garg SK, Monaghan CE, Raber J, Foust KD, Kaspar BK, Hirrlinger PG, Kirchhoff F, Bissonnette JM, Ballas N, Mandel G (2011) A role for glia in the progression of Rett syndrome. *Nature* Available at: <http://www.nature.com/nature/journal/vaop/ncurrent/full/nature10214.html> [Accessed March 6, 2012].
- Lister R et al. (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341:1237905.
- Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, Yang B (2011b) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* 39:359–372.
- Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86:367–377.
- Li W, Teng F, Li T, Zhou Q (2013a) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol* 31:684–686.
- Li Y, Wang H, Muffat J, Cheng AW, Orlando DA, Lovén J, Kwok S-M, Feldman DA, Bateup HS, Gao Q, Hockemeyer D, Mitalipova M, Lewis CA, Vander Heiden MG, Sur M, Young RA, Jaenisch R (2013b) Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons. *Cell Stem Cell* 13:446–458.
- Lugtenberg D et al. (2009) Structural variation in Xq28: *MECP2* duplications in 1% of patients with unexplained XLMR and in 2% of male patients with severe encephalopathy. *Eur J Hum Genet EJHG* 17:444–453.

- Luikenhuis S, Giacometti E, Beard CF, Jaenisch R (2004) Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc Natl Acad Sci U S A* 101:6033–6038.
- Lyon MF (1989) X chromosome inactivation as a system of gene dosage compensation to regulate gene expression. *Prog Nucleic Acid Res Mol Biol* 36:119–130.
- Lyst MJ, Bird A (2015) Rett syndrome: a complex disorder with simple roots. *Nat Rev Genet* 16:261–275.
- Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg ME, Bird A (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci* 16:898–902.
- Maeder ML et al. (2008) Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 31:294–301.
- Maezawa I, Swanberg S, Harvey D, LaSalle JM, Jin L-W (2009) Rett syndrome astrocytes are abnormal and spread mecp2 deficiency through gap junctions. *J Neurosci* 29:5051–5061.
- Maier DA, Brennan AL, Jiang S, Binder-Scholl GK, Lee G, Plesa G, Zheng Z, Cotte J, Carpenito C, Wood T, Spratt SK, Ando D, Gregory P, Holmes MC, Perez EE, Riley JL, Carroll RG, June CH, Levine BL (2013) Efficient clinical scale gene modification via zinc finger nuclease-targeted disruption of the HIV co-receptor CCR5. *Hum Gene Ther* 24:245–258.
- Mak AN-S, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335:716–719.
- Mandell JG, Barbas CF (2006) Zinc finger tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* 34:W516–W523.
- Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R (1997) Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev* 11:156–166.
- Marcus CL, Carroll JL, McColley SA, Loughlin GM, Curtis S, Pyzik P, Naidu S (1994) Polysomnographic characteristics of patients with Rett syndrome. *J Pediatr* 125:218–224.
- Maresca M, Lin VG, Guo N, Yang Y (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res* 23:539–546.
- Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, Emmanuelle Charpentier (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.
- McArthur AJ, Budden SS (1998) Sleep dysfunction in Rett syndrome: a trial of exogenous melatonin treatment. *Dev Med Child Neurol* 40:186–192.

- McBurney MW, Rogers BJ (1982) Isolation of male embryonal carcinoma cells and their chromosome replication patterns. *Dev Biol* 89:503–508.
- McCarty DM (2008) Self-complementary AAV vectors; advances and applications. *Mol Ther J Am Soc Gene Ther* 16:1648–1656.
- McCauley MD, Wang T, Mike E, Herrera J, Beavers DL, Huang T-W, Ward CS, Skinner S, Percy AK, Glaze DG, Wehrens XHT, Neul JL (2011) Pathogenesis of lethal cardiac arrhythmias in *Mecp2* mutant mice: implication for therapy in Rett syndrome. *Sci Transl Med* 3:113ra125–113ra125.
- McGraw CM, Samaco RC, Zoghbi HY (2011) Adult neural function requires MeCP2. *Science* 333:186–186.
- Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151:1417–1430.
- Michal Szczepek, Vincent Brondani, Janine Büchel, Luis Serrano, David J Segal, Toni Cathomen (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc finger nucleases. *Nat Biotechnol* 25:786–793.
- Miller JC et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29:143–148.
- Miller JC, Holmes MC, Wang J, Guschin DY, Lee Y-L, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO, Rebar EJ (2007) An improved zinc finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* 25:778–785.
- Mingozzi F, High KA (2013) Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* 122:23–36.
- Mnatzakanian GN, Lohi H, Munteanu I, Alfred SE, Yamada T, MacLeod PJM, Jones JR, Scherer SW, Schanen NC, Friez MJ, Vincent JB, Minassian BA (2004) A previously unidentified *MECP2* open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat Genet* 36:339–341.
- Moehle EA, Moehle EA, Rock JM, Rock JM, Lee Y-L, Lee YL, Jouvenot Y, Jouvenot Y, DeKolver RC, DeKolver RC, Gregory PD, Gregory PD, Urnov FD, Urnov FD, Holmes MC, Holmes MC (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A* 104:3055–3060.
- Moretti P, Bouwknecht JA, Teague R, Paylor R, Zoghbi HY (2005) Abnormalities of social interactions and home-cage behavior in a mouse model of Rett syndrome. *Hum Mol Genet* 14:205–220.
- Moretti P, Zoghbi HY (2006) MeCP2 dysfunction in Rett syndrome and related disorders. *Curr Opin Genet Dev* 16:276–281.

- Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501–1501.
- Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468:443–446.
- Mussolino C, Alzubi J, Fine EJ, Morbitzer R, Cradick TJ, Lahaye T, Bao G, Cathomen T (2014) TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Res*:gku305.
- Mussolino C, Morbitzer R, Lütge F, Dannemann N, Lahaye T, Cathomen T (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 39:9283–9293.
- Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, Daimon T, Sezutsu H, Yamamoto T, Sakuma T, Suzuki K-IT (2014) Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat Commun* 5:5560.
- Nan X, Campoy FJ, Bird A (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 88:471–481.
- Nan X, Meehan RR, Bird A (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* 21:4886–4892.
- Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393:386–389.
- Nan X, Tate P, Li E, Bird A (1996) DNA methylation specifies chromosomal localization of MeCP2. *Mol Cell Biol* 16:414–421.
- Narkar VA, Downes M, Yu RT, Embler E, Wang Y-X, Banayo E, Mihaylova MM, Nelson MC, Zou Y, Juguilon H, Kang H, Shaw RJ, Evans RM (2008) AMPK and PPARdelta agonists are exercise mimetics. *Cell* 134:405–415.
- Nathwani AC et al. (2011) Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med* 365:2357–2365.
- Neul JL, Fang P, Barrish J, Lane J, Caeg EB, Smith EO, Zoghbi H, Percy A, Glaze DG (2008) Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* 70:1313–1321.
- Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, Leonard H, Bailey MES, Schanen NC, Zappella M, Renieri A, Huppke P, Percy AK (2010) Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol* 68:944–950.
- Neul JL, Zoghbi HY (2004) Rett syndrome: a prototypical neurodevelopmental disorder. *The Neuroscientist* 10:118–128.



- Nuber UA, Kriaucionis S, Roloff TC, Guy J, Selfridge J, Steinhoff C, Schulz R, Lipkowitz B, Ropers HH, Holmes MC, Bird A (2005) Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. *Hum Mol Genet* 14:2247–2256.
- O'Connor RD, Zayzafoon M, Farach-Carson MC, Schanen NC (2009) MeCP2 deficiency decreases bone formation and reduces bone volume in a rodent model of Rett syndrome. *Bone* 45:346–356.
- Ogier M, Wang H, Hong E, Wang Q, Greenberg ME, Katz DM (2007) Brain-derived neurotrophic factor expression and respiratory function improve after ampakine treatment in a mouse model of Rett syndrome. *J Neurosci* 27:10912–10917.
- Panda SK, Wefers B, Ortiz O, Floss T, Schmid B, Haass C, Wurst W, Kühn R (2013) Highly efficient targeted mutagenesis in mice using TALENs. *Genetics* 195:703–713.
- Panighini A, Duranti E, Santini F, Maffei M, Pizzorusso T, Funel N, Taddei S, Bernardini N, Ippolito C, Viridis A, Costa M (2013) Vascular dysfunction in a mouse model of Rett syndrome and effects of curcumin treatment. *PLoS One* 8:e64863.
- Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31:839–843.
- Pattanayak V, Ramirez CL, Joung JK, Liu DR (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection. *Nat Methods* 8:765–770.
- Paulk NK, Wursthorn K, Wang Z, Finegold MJ, Kay MA, Grompe M (2010) Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia *in vivo*. *Hepatology* 51:1200–1208.
- Penaud-Budloo M, Le Guiner C, Nowrouzi A, Toromanoff A, Chérel Y, Chenuaud P, Schmidt M, von Kalle C, Rolling F, Moullier P, Snyder RO (2008) Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. *J Virol* 82:7875–7885.
- Percy AK, Lane JB (2005) Rett syndrome: model of neurodevelopmental disorders. *J Child Neurol* 20:718–721.
- Percy AK, Lee H-S, Neul JL, Lane JB, Skinner SA, Geerts SP, Annese F, Graham J, McNair L, Motil KJ, Barrish JO, Glaze DG (2010) Profiling scoliosis in Rett syndrome. *Pediatr Res* 67:435–439.
- Perez EE et al. (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc finger nucleases. *Nat Biotechnol* 26:808–816.
- Perez-Pinera P, Ousterout DG, Gersbach CA (2012) Advances in targeted genome editing. *Curr Opin Chem Biol* 16:268–277.
- Philippe Horvath, Rodolphe Barrangou (2010) CRISPR/Cas, the immune system of bacteria and archaea.

- Prashant Mali, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, George M. Church (2013) RNA-guided human genome engineering via Cas9.
- Ramirez CL, Foley JE, Wright DA, Müller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA, Cathomen T, Voytas DF, Joung JK (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* 5:374–375.
- Ramirez J-M, Ward CS, Neul JL (2013) Breathing challenges in Rett syndrome: lessons learned from humans and animal models. *Respir Physiol Neurobiol* 189:280–287.
- Ramocki MB, Peters SU, Tavyev YJ, Zhang F, Carvalho CMB, Schaaf CP, Richman R, Fang P, Glaze DG, Lupski JR, Zoghbi HY (2009) Autism and other neuropsychiatric symptoms are prevalent in individuals with *MECP2* duplication syndrome. *Ann Neurol* 66:771–782.
- Ramocki MB, Tavyev YJ, Peters SU (2010) The *MECP2* Duplication Syndrome. *Am J Med Genet A* 152A:1079–1088.
- Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F (2015) *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 520:186–191.
- Rastegar M, Hotta A, Pasceri P, Makarem M, Cheung AYL, Elliott S, Park KJ, Adachi M, Jones FS, Clarke ID, Dirks P, Ellis J (2009) *MECP2* isoform-specific vectors with regulated expression for Rett syndrome gene therapy. *PLoS ONE* 4 Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2728539/> [Accessed October 30, 2015].
- Reed R, Maniatis T (1988) The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev* 2:1268–1276.
- Robinson L, Guy J, McKay L, Brockett E, Spike RC, Selfridge J, Sousa DD, Merusi C, Riedel G, Bird A, Cobb SR (2012) Morphological and functional reversal of phenotypes in a mouse model of Rett syndrome. *Brain* 135:2699–2710.
- Roende G, Ravn K, Fuglsang K, Andersen H, Vestergaard A, Brøndum-Nielsen K, Jensen J-EB, Nielsen JB (2011) Patients with Rett syndrome sustain low-energy fractures. *Pediatr Res* 69:359–364.
- Rothkamm K, Krüger I, Thompson LH, Löbrich M (2003) Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol Cell Biol* 23:5706–5715.
- Rougeulle C, Glatt H, Lalande M (1997) The Angelman syndrome candidate gene, *UBE3A/E6-AP*, is imprinted in brain. *Nat Genet* 17:14–15.
- Roze E, Cochen V, Sangla S, Bienvenu T, Roubergue A, Leu-Semenescu S, Vidaihet M (2007) Rett syndrome: an overlooked diagnosis in women with stereotypic hand movements, psychomotor retardation, Parkinsonism, and dystonia? *Mov Disord Off J Mov Disord Soc* 22:387–389.

- Sakuma T, Nishikawa A, Kume S, Chayama K, Yamamoto T (2014) Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep* 4:5400.
- Samulski RJ, Muzyczka N (2014) AAV-mediated gene therapy for research and therapeutic purposes. *Annu Rev Virol* 1:427–451.
- Sander JD et al. (2011) Selection-free zinc finger nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* 8:67–69.
- San Filippo J, Sung P, Klein H (2008) Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* 77:229–257.
- Santos M, Summavielle T, Teixeira-Castro A, Silva-Fernandes A, Duarte-Silva S, Marques F, Martins L, Dierssen M, Oliveira P, Sousa N, Maciel P (2010) Monoamine deficits in the brain of methyl-CpG binding protein 2 null mice suggest the involvement of the cerebral cortex in early stages of Rett syndrome. *Neuroscience* 170:453–467.
- Sastre C, Rubio-Navarro A, Buendía I, Gómez-Guerrero C, Blanco J, Mas S, Egido J, Blanco-Colio LM, Ortiz A, Moreno JA (2013) Hyperlipidemia-Associated Renal Damage Decreases Klotho Expression in Kidneys from ApoE Knockout Mice. *PLoS ONE* 8:e83713.
- Scharenberg AM, Duchateau P, Smith J (2013) Genome engineering with TAL-effector nucleases and alternative modular nuclease technologies. *Curr Gene Ther* 13:291–303.
- Schwank G, Koo BK, Sasselli V, Dekkers JF, Heo I, Demircan T, Sasaki N, Boymans S, Cuppen E, van der Ent CK, Nieuwenhuis EE, Beekman JM, Clevers H (2013) Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* 13:653–658.
- Sclafani AM, Skidmore JM, Ramaprakash H, Trumpp A, Gage PJ, Martin DM (2006) Nestin-Cre mediated deletion of *Pitx2* in the mouse. *Genes N Y N* 2000 44:336–344.
- Segatto M, Trapani L, Di Tunno I, Sticozzi C, Valacchi G, Hayek J, Pallottini V (2014) Cholesterol metabolism is altered in Rett syndrome: a study on plasma and primary cultured fibroblasts derived from patients. *PLoS ONE* 9:e104834.
- Shahbazian MD, Antalffy B, Armstrong DL, Zoghbi HY (2002a) Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Hum Mol Genet* 11:115–124.
- Shahbazian MD, Young JI, Yuva-Paylor LA, Spencer CM, Antalffy BA, Noebels JL, Armstrong DL, Paylor R, Zoghbi HY (2002b) Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* 35:243–254.
- Sharma S (2007) Age-related nonhomologous end joining activity in rat neurons. *Brain Res Bull* 73:48–54.

- Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, Wang L, Hodgkins A, Iyer V, Huang X, Skarnes WC (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* 11:399–402.
- Shepard S, McCreary M, Fedorov A (2009) The peculiarities of large intron splicing in animals. *PLoS ONE* 4:e7853.
- Skene PJ, Illingworth RS, Webb S, Kerr ARW, James KD, Turner DJ, Andrews R, Bird AP (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol Cell* 37:457–468.
- Smith C, Abalde-Atristain L, He C, Brodsky BR, Braunstein EM, Chaudhari P, Jang Y-Y, Cheng L, Ye Z (2015) Efficient and allele-specific genome editing of disease loci in human iPSCs. *Mol Ther* 23:570–577.
- Song C, Feodorova Y, Guy J, Peichl L, Jost KL, Kimura H, Cardoso MC, Bird A, Leonhardt H, Joffe B, Solovei I (2014) DNA methylation reader MeCP2: cell type- and differentiation stage-specific protein distribution. *Epigenetics Chromatin* 7:17.
- Sorek R, Kunin V, Hugenholtz P (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 6:181–186.
- Spruijt CG et al. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152:1146–1159.
- Steffenburg U, Hagberg G, Hagberg B (2001) Epilepsy in a representative series of Rett syndrome. *Acta Paediatr Oslo Nor* 1992 90:34–39.
- Stettner GM, Huppke P, Brendel C, Richter DW, Gärtner J, Dutschmann M (2007) Breathing dysfunctions associated with impaired control of postinspiratory activity in *Mecp2*<sup>-/-</sup> knockout mice. *J Physiol* 579:863–876.
- Sugino K, Hempel CM, Okaty BW, Arnson HA, Kato S, Dani VS, Nelson SB (2014) Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. *J Neurosci Off J Soc Neurosci* 34:12877–12883.
- Summers D (1998) Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Mol Microbiol* 29:1137–1145.
- Sung YH, Baek I-J, Kim DH, Jeon J, Lee J, Lee K, Jeong D, Kim J-S, Lee H-W (2013) Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol* 31:23–24.
- Sun N, Zhao H (2013) Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol Bioeng* 110:1811–1821.
- Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F (2014) *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat Biotechnol*.

- Szcepek M, Brondani V, Büchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc finger nucleases. *Nat Biotechnol* 25:786–793.
- Tarquinio DC, Hou W, Neul JL, Kaufmann WE, Glaze DG, Motil KJ, Skinner SA, Lee H-S, Percy AK (2015) The changing face of survival in Rett Syndrome and *MECP2*-related disorders. *Pediatr Neurol* 53:402–411.
- Tebas P et al. (2014) Gene editing of *CCR5* in Autologous CD4 T cells of persons infected with HIV. *N Engl J Med* 370:901–910.
- Trappe R, Laccone F, Cobilanschi J, Meins M, Huppke P, Hanefeld F, Engel W (2001) *MECP2* mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin. *Am J Hum Genet* 68:1093–1101.
- Tronche F, Kellendonk C, Kretz O, Gass P, Anlag K, Orban PC, Bock R, Klein R, Schütz G (1999) Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. *Nat Genet* 23:99–103.
- Tudor M, Akbarian S, Chen RZ, Jaenisch R (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc Natl Acad Sci* 99:15536–15541.
- Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11:636–646.
- Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K, Lugtenberg D, Bienvenu T, Jensen LR, Gecz J, Moraine C, Marynen P, Fryns J-P, Froyen G (2005) Duplication of the *MECP2* region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am J Hum Genet* 77:442–453.
- Viemari J-C, Roux J-C, Tryba AK, Saywell V, Burnet H, Peña F, Zanella S, Bévangut M, Barthelemy-Requin M, Herzing LBK, Moncla A, Mancini J, Ramirez J-M, Villard L, Hilaire G (2005) *Mecp2* deficiency disrupts norepinephrine and respiratory systems in mice. *J Neurosci Off J Soc Neurosci* 25:11521–11530.
- Villard L, Kpebe A, Cardoso C, Chelly PJ, Tardieu PM, Fontes M (2000) Two affected boys in a Rett syndrome family: clinical and molecular findings. *Neurology* 55:1188–1193.
- Voituron N, Hilaire G (2011) The benzodiazepine Midazolam mitigates the breathing defects of *Mecp2*-deficient mice. *Respir Physiol Neurobiol* 177:56–60.
- Voituron N, Zanella S, Menuet C, Dutschmann M, Hilaire G (2009) Early breathing defects after moderate hypoxia or hypercapnia in a mouse model of Rett syndrome. *Respir Physiol Neurobiol* 168:109–118.
- Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153:910–918.
- Wang J et al. (2015) Wild-type microglia do not reverse pathology in mouse models of Rett syndrome. *Nature* 521:E1–E4.

- Wang W, Ye C, Liu J, Zhang D, Kimata JT, Zhou P (2014a) CCR5 gene disruption via lentiviral vectors expressing Cas9 and single guided rna renders cells resistant to HIV-1 infection. PLoS ONE 9:e115987.
- Wang X, Wang Y, Huang H, Chen B, Chen X, Hu J, Chang T, Lin R-J, Yee J-K (2014b) Precise gene modification mediated by TALEN and single-stranded oligodeoxynucleotides in human cells. PLoS One 9:e93575.
- Wan M, Lee SSJ, Zhang X, Houwink-Manville I, Song H-R, Amir RE, Budden S, Naidu S, Pereira JLP, Lo IFM, Zoghbi HY, Schanen NC, Francke U (1999) Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. Am J Hum Genet 65:1520–1529.
- Weese-Mayer DE, Lieske SP, Boothby CM, Kenny AS, Bennett HL, Silvestri JM, Ramirez J-M (2006) Autonomic nervous system dysregulation: breathing and heart rate perturbation during wakefulness in young girls with Rett syndrome. Pediatr Res 60:443–449.
- Wegener E, Brendel C, Fischer A, Hülsmann S, Gärtner J, Huppke P (2014) Characterization of the MeCP2R168X knockin mouse model for Rett syndrome. PLoS ONE 9:e115444.
- Wolfe SA, Nekludova L, Pabo CO (2000) DNA recognition by Cys2His2 zinc finger proteins. Annu Rev Biophys Biomol Struct 29:183–212.
- Wong K, Leonard H, Jacoby P, Ellaway C, Downs J (2015) The trajectories of sleep disturbances in Rett syndrome. J Sleep Res 24:223–233.
- Wu Z, Yang H, Colosi P (2010) Effect of genome size on AAV vector packaging. Mol Ther 18:80–86.
- Yannick Doyon, Thuy D Vo, Matthew C Mendel, Shon G Greenberg, Jianbin Wang, Danny F Xia, Jeffrey C Miller, Fyodor D Urnov, Philip D Gregory, Michael C Holmes (2010) Enhancing zinc finger nuclease activity with improved obligate heterodimeric architectures. Nat Methods 8:74–79.
- Yin H, Xue W, Chen S, Bogorad RL, Benedetti E, Grompe M, Koteliensky V, Sharp PA, Jacks T, Anderson DG (2014) Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat Biotechnol 32:551–553.
- Young D, Nagarajan L, DE Klerk N, Jacoby P, Ellaway C, Leonard H (2007) Sleep problems in Rett syndrome. Brain Dev 29:609–616.
- Zeev BB, Yaron Y, Schanen NC, Wolf H, Brandt N, Ginot N, Shomrat R, Orr-Urtreger A (2002) Rett syndrome: clinical manifestations in males with *MECP2* mutations. J Child Neurol 17:20–24.
- Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29:149–153.

- Zhou C, Moore L, Yool A, Jaunzems A, Byard RW (2015) Renal tubular epithelial vacuoles—a marker for both hyperlipidemia and ketoacidosis at autopsy. *J Forensic Sci* 60:638–641.
- Zhou Z, Hong EJ, Cohen S, Zhao W, Ho HH, Schmidt L, Chen WG, Lin Y, Savner E, Griffith EC, others (2006) Brain-specific phosphorylation of MeCP2 regulates activity-dependent *Bdnf* transcription, dendritic growth, and spine maturation. *Neuron* 52:255–269.
- Zincarelli C, Soltys S, Rengo G, Rabinowitz JE (2008) Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol Ther J Am Soc Gene Ther* 16:1073–1080.
- Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen Z-Y, Liu DR (2015) Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat Biotechnol* 33:73–80.
- Zysman L, Lotan M, Ben-Zeev B (2006) Osteoporosis in Rett syndrome: A study on normal values. *ScientificWorldJournal* 6:1619–1630.